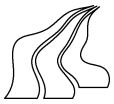


Institut for Matematiske Fag
Aalborg Universitet

Statistisk analyse i et sociologisk perspektiv

- en praktisk tilgang via grafiske modeller

Speciale skrevet af:

Signe O. W. Jensen

Institut for Matematiske Fag
Aalborg Universitet

Titel:

Statistisk analyse i et sociologisk perspektiv
- en praktisk tilgang via grafiske modeller

Tema:

Speciale

Projektperiode:

MAT6, forårssemesteret 2009

Skrevet af:

Signe O. W. Jensen

Vejleder:

Poul Svante Eriksen

Oplagstal: 9

Sidetal: 72

Afsluttet den 5. juni 2009

Synopsis:

Dette speciale beskriver, hvordan en statistisk analyse kan foretages ved at benytte grafiske modeller kombineret med sociologisk teori. Der foretages en analyse af data, og i den forbindelse beskrives metodiske og teoretiske overvejelser, som har betydning for analysens validitet og reliabilitet. Teorien bag grafiske modeller beskrives indledningsvis samt forskellige statistiske test og modelselektionsalgoritmer. Den sociologiske teori, repræsenteret ved Robert Karasek, beskrives efterfølgende. Dernæst præsenteres analysemetoden, der indeholder en beskrivelse af faktoranalyse, hvordan konstruktion af grafiske modeller foregår i programmerne HUGIN, DIGRAM og MIM og endelig beskrives ordinal regression, der anvendes i analysen af de grafiske modeller. De variable, som skal danne grundlag for modellerne, konstrueres indledningsvis i analysen. Med udgangspunkt i disse data konstrueres tre grafiske modeller, som efterfølgende analyseres og fortolkes via statistik og sociologisk teori. Citater fra respondenter inddrages endvidere for at give analysens resultater yderligere styrke. Validiteten og reliabiliteten af både metode, teori og analyse vurderes afslutningsvis for dermed at kunne fremhæve de vigtigste resultater i den endelige konklusion.

Forord

Nærværende speciale er skrevet som den afsluttende opgave i forbindelse med Kandidatuddannelsen ved Institut for Matematiske Fag på Aalborg Universitet.

I dette speciale kombineres to vidt forskellige discipliner ud fra tanken om, at disse discipliner kan supplere hinanden på konstruktiv vis og dermed medvirke til at give specialet en ekstra dimension. De to discipliner som skal vekselvirke i det følgende er Statistik og Sociologi. Kombinationen af disse to grene byder på mange fordele i forhold til analyse af videnskabelige teorier, men på grund af de forskellige traditioner inden for de to grene opstår også problemer. Et traditionelt projekt inden for statistik fokuserer oftest på matematisk teori, eventuel implementering i et program og efterfølgende analyse. Et sociologisk projekt vægter derimod beskrivelse af metodeovervejelser, forklaring af relevansen for den pågældende problemstilling, udvælgelse af teori med tilhørende teoridiskussion, overvejelser omkring validitet og videnskabsteoretisk standpunkt. Kombinationen af de to fagområder er derfor en balancegang mellem henholdsvis at vægte statistikkens og sociologiens principper.
Formålet med dette speciale er, at præsentere hvordan en statistisk analyse kan foretages via grafiske modeller og hvordan sociologi kan inddrages som forklaringsfaktor i forhold til analysens resultater, men også som drivkraft i forbindelse med selve konstruktionen af de grafiske modeller. Hvor statistikken kan afdække sammenhænge mellem variable, kan sociologien hjælpe til at afdække retningen af denne sammenhæng og endvidere give mulige svar på spørgsmålet *hvorfor?*. Ved at kombinere de to discipliner vil analysen således nærme sig kausale forklaringer af resultaterne.
Specialet er dermed fokuseret på at være praktisk anvendeligt i forhold til dataanalyse og vil derfor vægte analytiske metoder frem for matematiske beviser for den anvendte teori. Denne tilgang er valgt ud fra personlig interesse og for at imødekomme en naturlig kombination af statistik og sociologi bedst muligt.
I Kapitel 1 præsenteres den problemstilling som skal undersøges i specialet. Kapitlet fokuserer derfor på at beskrive den aktuelle problemstilling, samt hvordan indsamlingen af data til belysning af denne er foregået. I den forbindelse diskuteres designets fordele og ulemper med hensyn til den valgte metode.
Kapitel 2 fokuserer på den matematiske teori som skal udgøre grundlaget for den statistiske analyse. Heri præsenteres teorien bag grafiske modeller og endvidere beskrives forskellige hypotesetest og korrelationsmål som anvendes i forbindelse med dataanalysen. Derudover beskrives procedurer i forhold til modelselektion, som er relevant for selve modelkonstruktionen. Der henvises iøvrigt til Appendiks for centrale grafteoretiske begreber.
Den sociologiske teori som ligeledes skal inddrages i analysen præsenteres i Kapitel 3, hvor en model til at vurdere arbejdsmiljø introduceres. Det beskrives endvidere, hvordan den sociologiske teori tænkes at blive anvendt i forbindelse med analysen og afslutningsvis gives en kritik af teorien, hvor der fokuseres på svaghederne ved den aktuelle model.
I Kapitel 4 præsenteres analysemetoden, der primært indeholder faktoranalyse og ordinal regression. Disse bliver således kort forklaret. Endvidere beskrives de tre programmer, HUGIN, DIGRAM og MIM, som anvendes i forbindelse med konstruktionen af de grafiske modeller.
Selve analysen foretages i Kapitel 5, hvor de variable som er centrale for problemstillingen konstrueres via faktoranalyse. Derefter analyseres de modeller som er fremkommet via henholdsvis HUGIN, DIGRAM og MIM ved at foretage ordinal regression på udvalgte variable. I analysen inddrages endvidere den sociologiske teori for at give en fortolkning af analysens resultater og i den proces citeres flere af respondenterne for at understrege relevansen og validiteten af disse fortolkninger.
I Kapitel 6 vurderes validiteten og reliabiliteten af analysen, både med hensyn til den metode, der er anvendt i forbindelse med indsamlingen af data, den sociologiske teori og de statistiske metoder. Kapitlet har således til formål at give en overordnet evaluering af undersøgelsen for dermed at blive i stand til at vurdere, hvorvidt analysens resultater er valide og pålidelige, hvilket er afgørende for

anvendelsen.
Selvom specialet forudsætter kendskab til statistik og grafteori for at forstå den bagvedliggende matematiske teori, appellerer det bredt i den forstand, at folk uden særlig matematisk viden på dette område, stadig kan forholde sig til analysen, de overvejelser som går forud for denne og de fortolkninger som følger af analysens resultater.

Det datasæt som anvendes og analyseres i specialet er stillet til rådighed af den danske Diabetesforening. I den forbindelse rettes en tak til Tina Vaarning og Signe Hasseriis for deres samarbejdsvilje og hjælpsomhed.

Endelig takkes vejleder Poul Svante Eriksen for god og konstruktiv vejledning.

Signe O. W. Jensen

Summary

This dissertation is focused on practical data analysis through both statistical and sociological theory. This thesis will show how a statistical analysis can be carried out and is in this respect an attempt to provide the reader with an understanding of the methodological and theoretical considerations that are required to perform a valid and reliable analysis, and also the advantages that lie in combining the two rather different disciplines: Statistics and Sociology.

This dissertation explores the special problems concerned with the work environment of diabetics. This area is being studied on behalf of the Danish Diabetic Union who are interested in Danish studies and literature regarding this field of research. That is why the data that form the basis of the analysis are also provided by the Danish Diabetic Union.
The research design adopted in the study i.e. a cross sectional design is initially described. The different problems that arise with this research design and of course the advantages are presented. The data collection method is in this case a survey and the construction of the survey is therefore of great interest, as it is crucial to the validity and reliability of the answers. Furthermore the distribution of the survey and the method of selection of respondents are described which also has great impact on the overall study. This initial description of the study serves to define the context in which the study takes place which is important to both the sociological theory and the choice of statistical analyses.
The statistical part of the data analysis is primarily concerned with graphical modelling. At first a theoretical approach to the concept of graphical modelling is given but since this thesis is focused on the practical aspects of analysis, the theory is only given and not proven. The most important feature of graphical modelling is that the concept of conditional independence is related to graph theory i.e. from a graphical model it is possible to determine if two variables are conditional independent which is a very useful property in analysis. The analysis of the graphical models contains both hypothesis testing and the use of measures of correlation which therefore also are described. Furthermore several model selection procedures are presented. Although graph theory is essential to understanding and analyzing graphical models the central concepts of graph theory is referred to the appendix.
The sociological theory which is crucial in regard to the interpretations of the analysis is subsequently described. This part of the thesis is primarily focusing on the demand-control model invented by Robert Karasek and the later addition of the social support dimension. This demand-control-support model is used continuously throughout the analysis in order to describe and interpret the results. With this model at hand it is possible to categorize the work environment of the respondents and to evaluate the risk of work related illness which is of key importance to the problem of investigation.
The practical approach as regards the statistical analysis is described before the actual analysis. On the basis of the demand-control-support model there is an a priori expectation in regard to the graphical models of the data. This a priori model is used to compare the initial expectations of data associations with the posterior knowledge of data dependencies. The variables that are necessary to the a priori model are constructed through factor analysis of the data and in that connection factor analysis in general is described. Subsequently, three different programs that are able to construct graphical models are presented. The algorithms and tests that are used in the construction and model selection procedure are described and evaluated with regard to the type of data that are to be analyzed. Finally the theory of ordinal regression is introduced as this is essential to the analysis of the models. Through ordinal regression it is possible to evaluate the risk of illness, fatigue etc. via a series of explanatory variables that are determined by the graphical model.
The open responses from the study population is furthermore included in the analysis of the models as they can give further information and strengthen the interpretations. With the inclusion of sociology, the statistical analysis is hereby given an extra dimension and with this combination, the analysis is approaching causal inference. The statistical analysis determines direction and strength of association

between variables and the sociological theory explain why this association occurred and when it is possible to organize the variables according to a time dimension the analysis is indeed approaching causal inference. Furthermore both the methodology, theory and analysis are evaluated with regard to validity and reliability in order to make the final conclusions.
The study shows that the physical conditions and the amount of control and support is of key importance to the experienced work environment. It is furthermore suggested that a pamphlet be made which can be distributed at the diabetics work place as to enlighten the colleagues of the consequences of diabetes in regard to health and work.

Indhold

Indledning

At foretage undersøgelser af politiske, økonomiske, erhvervsmæssige eller andre problemstillinger er en stor og vigtig del af samfundet. Gennem undersøgelser læres nyt, tidligere overbevisninger og forestillinger viser sig måske at være forældede og må tages op til revision, mens nye, interessante og nogle gange uventede sammenhænge afdækkes. Undersøgelser kan dermed tilføre ny viden, som har betydning for udviklingen af samfundet. Undersøgelser på makroniveau er nødvendige for at kunne målrette en indsats på et givet område, de er nødvendige for bevilling af økonomiske ressourcer og ikke mindst nødvendige for at sikre gennemførelse af diverse ideer og projekter. Ved at foretage indledningsvise undersøgelser søges det eksempelvis at afklare økonomien på et projekt, samt hvorvidt og hvornår der følger et fornuftigt resultat. Ligeledes foretages undersøgelser på mikroniveau, hvor der oftest er fokus på individet, hverdagen og hverdagslivets betydning. Relevansen og anvendeligheden af valide undersøgelser er i høj grad til stede, da det er via disse undersøgelser, at det bliver muligt at beskrive det omgivende samfund på både godt og ondt.
Størstedelen af de undersøgelser der foretages kan med andre ord samles under ét, hvor den gennemgående problemstilling er af samfundsmæssig relevans. Ens for alle kvantitative undersøgelser er endvidere, at der altid skal inddrages statistisk analyse i forbindelse med databehandling, for at kunne drage fornuftige konklusioner. Dataanalyse via statistiske metoder er således en uundgåelig del af enhver kvantitativ undersøgelse og er med til at sikre, at undersøgelsen kan anvendes. Anvendeligheden, af de undersøgelser der foretages, er naturligvis af afgørende betydning, og er næsten udelukkende bestemt af, hvorvidt selve dataindsamlingen er foregået på en fornuftig måde samt validiteten af den efterfølgende analyse. Der er således flere aspekter af en undersøgelse, som gør sig gældende og ikke mindst i forbindelse med fortolkningen af analysens resultater. Disse har selvsagt en central betydning, men kræver dog mere end blot kenskab til design, metoder og statistik.
Den statistiske analyse giver grundlæggende mulighed for at opstille overordnede modeller for data, og dermed kan sammenhænge beskrives via diverse test og korrelationer, og muligvis kan modellen forudsige udfaldet af en given kombination af variable, men der bør endvidere inddrages fortolkninger af de tal, som analysen resulterer i. Det er med andre ord vigtigt ikke blot at kunne analysere data, men også at kunne fortolke og drage konklusioner på baggrund af analysens resultater. I den forbindelse er det nødvendigt at udnytte en a priori viden inden for den aktuelle problemstilling. Med denne viden er det muligt at forholde sig kritisk til de resultater som fremkommer i den statistiske analyse, og samtidig se de sammenhænge, som var uventede a priori. Først med inddragelse af en ekspertviden bliver det interessant at foretage undersøgelser og statistiske analyser. Med en a priori viden bliver det muligt efter analyse at præsentere en konklusion på den pågældende problemstilling og give nye forslag til løsninger på de aktuelle problematikker.
I dette speciale tages der udgangspunkt i en problemstilling, som den danske diabetesforening har ønsket at rette fokus mod, der omhandler arbejdsmiljøet for erhvervsaktive diabetikere. Formålet med dette speciale er at undersøge de problematikker der gør sig gældende i forhold til diabetikeres arbejdsmiljø via en analyse af Diabetesforeningens indsamlede data. I den proces, hvor problemstilingen skal belyses, udnyttes netop både statistiske metoder og sociologisk teori. Formålet med at anvende statistisk analyse i kombination med sociologisk viden er at sikre en mere valid analyse, hvor resultaterne endvidere kan fortolkes og dermed vise sig relevante i en samfundsmæssig kontekst. Det følgende kapitel beskriver således den problemstilling, som den danske Diabetesforening ønsker at belyse, samt det design som Diabetesforeningen har anvendt i forbindelse med undersøgelsen.

Problemstilling

Arbejdsmiljø har stor betydning for ansattes sundhed og på denne baggrund er det et område, som der generelt bruges mange ressourcer på i forbindelse med forskning. Et dårligt arbejdsmiljø kan resultere i både stress og somatiske sygdomme, hvorfor det er vigtigt at undersøge hvilke faktorer, der påvirker arbejdsmiljø og dermed hvordan det fysiske og psykiske arbejdsmiljø kan forbedres. I den forbindelse foretog den danske Diabetesforening i juni 2008 en spørgeskemaundersøgelse af erhvervsaktive diabetikere med henblik på at afdække eventuelle problemer, når diabetes kombineres med karriere. Undersøgelsen havde fokus på det arbejdsmiljø, som diabetikerne færdes i, og hvorvidt denne gruppe af personer oplever et dårligt arbejdsmiljø som følge af deres sygdom.
På trods af mange undersøgelser omkring arbejdsmiljø er der, så vidt vides, ikke nogle danske undersøgelser eller dansk litteratur, der beskriver arbejdsmiljø blandt diabetikere, hvilket er årsag til, at Diabetesforeningen har ønsket et samarbejde. Dette kapitel har til formål at beskrive den problemstilling, som Diabetesforeningen ønsker at belyse samt de metodiske overvejelser, som ligger til grund for spørgeskema, respondentudvælgelse og dataindsamling. Først gives en kort beskrivelse af diabetes og betydningen af at have diabetes i forhold til generelle symptomer og følgesygdommme. Dernæst præsenteres det undersøgelsesdesign som Diabetesforeningen har taget udgangspunkt i, med henblik på at beskrive fordele og ulemper ved dette. Følgende afsnit er inspireret af [Diab] og [dV07, Kapitel 10 og 11].

1.1 Diabetes

Produktion af hormonet insulin er nødvendig, for at kroppens celler kan optage glukose, og det er denne produktion, som diabetikere har problemer med. Hvis produktionen af insulin er for lav, vil blodets indhold af glukose stige, hvilket kan give flere ubehagelige følger. Nogle af de symptomer som gør sig gældende for diabetikere er hovedpine, træthed, kvalme, kløe mm. Behandlingen af diabetes er i høj grad baseret på egenomsorg, i.e. den enkelte skal selv tage ansvar for egen behandling, hvilket indebærer at måle blodsukker flere gange dagligt, tage insulin, være opmærksom på symptomerne for henholdsvis for højt og for lavt blodsukker, spise sundt og være fysisk aktiv.
Det lave blodsukker kan resultere i insulinchok, hvor diabetikeren kan blive bevidstløs eller få kramper. Forstadiet til insulinchok kaldes føling og dette advarselssignal skal diabetikeren således være meget opmærksom på. Hvis diabetikeren ikke passer sin diabetes godt kan der opstå komplikationer blandt andet i form af hjertekarsygdomme, nedsat nyrefunktion og dårligt syn.
Det er således klart, at der kan opstå problemer for nogle diabetikere i forbindelse med at være aktive på arbejdsmarkedet. Diabetesforeningens undersøgelse blev også foretaget, da der modtages mange henvendelser fra medlemmer, der oplever problemer på arbejdspladsen og derfor efterspørger råd og vejledning i forhold til oplysning omkring diabetes.

1.2 Design

Designet er en tværsnitsundersøgelse, hvilket vil sige, at data er indsamlet på ét tidspunkt blandt én gruppe respondenter. Denne type studiedesign er den mest økonomiske, både med hensyn til tid og penge, da der ikke er tale om en opfølgning af respondenter eller undersøgelse af interventionseffekten i en eksperimentelgruppe. Dermed er det muligt at undersøge en stor gruppe og således indhente store datamængder for relativt få midler og kræfter. I tværsnitsundersøgelsen konstrueres gruppen af respondenter på baggrund af eksisterende forskelle, i.e. i dette tilfælde er diabetes den afgørende faktor for allokering til gruppen. I den indledende fase tages der således ikke hensyn til, at respondenterne kan adskille sig fra hinanden på andre variable. Tværsnitsundersøgelsens manglende tidsdimension betyder,

at denne type design ikke kan anvendes i forhold til at måle forandring, men til at måle forskelle mellem grupper. Ved at foretage gentagne tværsnitsundersøgelser er det dog muligt at opnå en tidsdimension, hvor grupper fra forskellige undersøgelser kan sammenlignes, og på denne måde kan en forandring over tid måles. Et sådant prospektivt design vil være en mulighed, hvis en forandring i diabetikeres arbejdsmiljø skal måles. Det er dog vigtigt at bemærke, at dataindsamlingsmetoden skal bibeholdes for at sikre sammenligningsgrundlaget.
Den interne validitet i en tværsnitsundersøgelse kompromitteres netop på grund af den manglende tidsdimension, da dette giver problemer i forhold til at klarlægge kausalitet. Kausale sammenhænge er generelt svære at påvise og i tilfælde, hvor det end ikke er muligt at ordne de pågældende variable tidsmæssigt, således retningen af sammenhængen kan fastslås, er det umuligt. Denne type design er dermed mest velegnet til deskriptiv analyse, hvilket er et vigtigt kritikpunkt når den forklarende analyse tilstræbes. I den aktuelle problemstilling er det imidlertid klart, at diabetes er den uafhængige variabel som eventuelt kan påvirke arbejdsmiljøet, mens den omvendte situation er mere usædvanlig. Fortolkningen af en variabel som årsag forudsætter naturligvis, at der i analysen er kontrolleret for confoundere, ved at sikre, at alle andre forskelle mellem de pågældende grupper er udlignet. I praksis er det ikke muligt at kontrollere for samtlige variable, der kunne tænkes at have betydning, men jo flere der er kontrolleret for i analysen des mere sikkert er det, at sammenhængen også er kausal. Endvidere er det en fordel i forhold til validiteten, at overvejelser omkring mulige kausale sammenhænge er foretaget a priori, da det derved sikres, at argumenterne er uafhængige af data. Den omvendte ressonering, ad hoc, forklarer en sammenhæng i data efter at denne er observeret og er således afhængig af data. Overvejelser a priori er endvidere en fordel, da de kan være med til at fastslå retningen af en kausal sammenhæng. Selvom a priori betragtninger er at foretrække fremfor ad hoc-tilgangen, er sidstnævnte metode alligevel anvendelig, hvis analysen afdækker en uventet sammenhæng. I sådanne tilfælde må der tilføres mening til den observerede association via sociologisk viden og sociologiske ideer.
Ud over den interne validitet er det vigtigt at sikre ekstern validitet, hvilket vil sige, at gruppen af respondenter, der udvælges, udgør et repræsentativt sample i forhold til den population, som undersøgelsens resultater skal generaliseres til. I forhold til den aktuelle problemstilling er det således nødvendigt, at den udvalgte gruppe af diabetikere er repræsentativ for populationen af diabetikere. I praksis sikres den eksterne validitet ved at anvende probabilistiske metoder ved udvælgelse af respondenter. En af fordelene ved tværsnitsdesignet er, at det som nævnt er meget økonomisk, hvorfor det er muligt at undersøge store samples, hvilket kan være med til at sikre den eksterne validitet. Efter et tilfældigt udtræk af respondenter kan dataindsamlingen begynde. Da dataindsamlingsmetoden kan påvirke respondenternes besvarelser er det afgørende, at der er konsistens i forhold til dataindsamlingsmetoden. Herved opnås et struktureret datasæt, hvor det er muligt at sammenligne grupper inden for variable på systematisk vis. I indsamlingen af data er det endvidere vigtigt, at centrale variable er tildelt tilpas variation i svarkategorierne, således de er anvendelige i analysen. Da data i en tværsnitsundersøgelse analyseres ved at undersøge variationen på én variabel i forhold til variationen på en anden variabel er det væsentligt, at svarkategorierne ikke bliver for brede og dermed ikke måler variationen på en tilpas fin skala.
I ethvert design og enhver undersøgelse er der nogle etiske spørgsmål som skal afklares. Som hovedregel skal respondenterne deltage frivilligt under informeret samtykke og skal endvidere sikres anonymitet. I tværsnitsundersøgelser er der sjældent problemer i forhold til etikken. Kravet om anonymitet, som kan være problematisk i andre studiedesigns, er let at tilfredsstille, da der generelt ikke er behov for at gemme personoplysninger, i og med der ikke er tale om et længerevarende studie. Den manglende tidsdimension er endvidere en fordel, da der således ikke er tale om at måle effekten af en intervention, men kun fokuseres på eksisterende forskelle i de pågældende grupper. Den praktiske udvælgelse af respondenter samt konstruktion og distribution af spørgeskema i den aktuelle undersøgelse bliver beskrevet i de følgende afsnit.

1.2.1 Arbejdsspørgsmål

Det der skal afdækkes via undersøgelsen er arbejdsmiljøet for diabetikere, hvilket kan siges at udgøres af rummeligheden på arbejdspladsen, i.e. at der er plads til forskellighed og at der tages indbyrdes hensyn til hinanden. Indledningsvis blev opstillede Diabetesforeningen en række arbejdsspørgsmål, som udgør grundlaget for konstruktionen og indholdet af spørgeskemaet. De spørgsmål, som der overordnet blev arbejdet ud fra var

1. Yder ledelse/kolleger tilstrækkelig støtte?
2. Er der gode vilkår for egenomsorg?
3. Har diabetikere højere sygefravær end andre?
4. Føler diabetikere sig i højere grad udbrændte?

Det første spørgsmål har til hensigt at give et billede af det psykiske arbejdsmiljø. De problemer, der kan opstå i denne forbindelse tænkes at følge af manglende viden omkring diabetes, hvilket kan resultere i manglende forståelse for, at diabetikeren må sænke arbejdstempoet eller pludselig holder pause. Spørgeskemaet skal derfor klarlægge, hvorvidt diabetikeren oplever at modtage tilstrækkelig støtte fra arbejdspladsen. Spørgsmål 2 omhandler primært det fysiske arbejdsmiljø i forhold til, om der er mulighed for at tage medicin uforstyrret, om der er sund mad i kantinen etc. Hvis egenomsorgen vanskeliggøres i forbindelse med arbejdet kan behandlingen blive en ekstra byrde for diabetikeren. Spørgsmål 3 følger af en antagelse om, at diabetikere forsøger at skjule deres sygdom og derfor fortsætter med at arbejde på trods af utilpashed og sygdom indtil de må opgive, hvorefter sygefraværet stiger markant. Spørgeskemaet skal derfor afdække, hvordan sygefraværet fordeler sig blandt diabetikere. Som sammenligningsgrundlag har Diabetesforeningen inddraget resultater fra Det Nationale Forskningscenter for Arbejdsmiljø. Det sidste spørgsmål hænger i høj grad sammen med de krav, der stilles til diabetikeren på arbejdspladsen, hvor høje krav kombineret med træthed på grund af sygdom kan resultere i udbrændthed. Svarene på dette spørgsmål skal ligeledes vurderes i forhold til data fra Det Nationale Forskningscenter for Arbejdsmiljø.
De fire spørgsmål har således til formål at afdække arbejdsmiljøet for diabetikere og det er svar på disse overordnede arbejdsspørgsmål, der skal opnås via en række beslægtede spørgsmål i spørgeskemaet. I dette speciale søges det blandt andet at besvare de fire arbejdsspørgsmål via statistisk og sociologisk analyse, og derudover undersøges egne forventninger til sammenhænge mellem udvalgte variable.

1.2.2 Spørgeskema

Spørgeskemaet er konstrueret på baggrund af de opstillede arbejdsspørgsmål og er derfor opdelt i forskellige emner. Hvert emne indledes med en kort tekst for at sikre et godt flow i spørgeskemaet. Der indledes med spørgsmål om respondentens sociale baggrund i.e. køn, alder, uddannelse o.lign. Disse er forholdsvis simple spørgsmål som hurtigt er overstået, hvilket er medvirkende til at sikre, at respondenten besvarer hele skemaet. Derefter spørges der ind til respondentens helbred generelt og som følge af at have diabetes. Efter denne indledende del følger en række spørgsmål om tilknytning og forhold til arbejdsmarkedet, herunder arbejdstid og ansættelsens længde. Herefter stilles der spørgsmål om sammenhængen mellem arbejde og privatliv, i.e. hvordan arbejdet påvirker privatlivet og omvendt. Det psykiske arbejdsmiljø bliver ligeledes undersøgt via en lang række spørgsmål omkring indflydelse, krav og samarbejde, før der spørges ind til arbejdspladsen som helhed med særligt fokus på rummelighed. I denne del af skemaet spørges der endvidere ind til antallet af sygedage og sygeperioder. Afslutningsvis optræder spørgsmål omkring diabetes og arbejde, blandt andet hvorvidt ledelse og kolleger har kendskab til respondentens diabetes og hvordan sygdommen påvirker arbejdet.
I praksis består spørgeskemaet primært af lukkede spørgsmål, i.e. spørgsmål med afkrydsningsmuligheder og kun få åbne spørgsmål, hvor respondenterne kan give en længere forklaring eller uddybning af svar. Generelt er svarmulighederne til de lukkede spørgsmål opstillet på ordinalskala, hvilket giver

repsondenterne mulighed for at angive hvor ofte de oplever situationerne, som beskrives i de pågældende udsagn, eller i hvilken grad de er enige i disse. Svarkategorien 'Ved ikke' er i langt de fleste tilfælde udeladt ligesom midterkategorien i ordinalskalaerne, der afspejler den neutrale holdning, udelades i en række spørgsmål. Dette valg har både fordele og ulemper. Respondenterne bliver dermed nødt til at tage stilling til de forskellige udsagn frem for at bruge de nævnte kategorier som en nem udvej. Samtidig er det dog tænkeligt, at respondenterne mangler netop disse svarkategorier og dermed afkrydser en svarmulighed som ikke stemmer overens med deres reelle overbevisning i.e., at respondenterne i stedet svarer tilfældigt på spørgsmålene.
Det er ikke muligt at vurdere hvornår et spørgeskema er for langt, men med 133 udsagn som respondenterne skal tage stilling til nærmer det aktuelle spørgeskema sig grænsen. Problemet med de mange spørgsmål opvejes dog af, at de er relevante for respondenterne, hvorfor der er incitament til at besvare hele skemaet, hvilket også afspejles i responsraten på 58%. En del af spørgsmålene omkring træthed, stress og rummelighed er inddraget fra en undersøgelse foretaget af Det Nationale Forskningscenter for Arbejdsmiljø fra 2004 for at blive i stand til at sammenligne besvarelserne fra diabetikerne med normalbefolkningen. I den forbindelse skal det imidlertid overvejes, hvorvidt resultaterne fra denne undersøgelse kan sammenholdes med de nye resultater, da undersøgelserne er foretaget med stor tidsforskel. Et argument, der understøtter sammenligningen er, at tidsforskellen ikke har betydning for respondenternes oplevede niveau af træthed og stress. Omvendt kan den nye situation på arbejdsmarkedet øge frygten for fyringer og dermed risiko for at udvikle stress.

1.2.3 Respondentudvælgelse og distribution

Adgang til relevante respondenter kan ofte være et problem i forbindelse med undersøgelser, men i dette tilfælde havde Diabetesforeningen en naturlig fordel, da de på baggrund af medlemsdatabasen kunne foretage et repræsentativt udtræk af diabetikere, der alle var aktive på arbejdsmarkedet. Ud af de 240.000 personer, der er diagnosticerede med diabetes er knap 30% medlem af Diabetesforeningen [Diaa]. Der blev foretaget to udtræk af diabetikere i alderen 30-60 år. Det første udtræk angik type 1 diabetikere, hvor der blev samplet 500 personer. Blandt gruppen af type 2 diabetikere blev der ligeledes udtrukket 500 personer i.e. et samlet stratificeret udtræk på 1000 diabetikere. Ifølge [Diaa] er andelen af type 1 diabetikere 5-10% af den samlede gruppe af diabetikere. Dermed repræsenterer det stratificerede sample ikke den reelle fordeling af henholdsvis type 1 og type 2 diabetikere i befolkningen. Forudsat, at typen af diabetes ikke har betydning for respondenternes besvarelser vil dette dog ikke påvirke validiteten af undersøgelsen nævneværdigt. I forbindelse med et tema omkring arbejdsliv i Diabetesforeningens juniudgave af medlemsbladet, blev den aktuelle undersøgelse nævnt før udsendelse af spørgeskemaet. Dette har muligvis virket yderligere motiverende i forhold til at besvare spørgeskemaet for de medlemmer, der efterfølgende modtog spørgeskemaet.
Distributionen af spørgeskemaet foregik med almindelig post, hvor respondenterne modtog en papirudgave af spørgeskemaet og en frankeret svarkuvert. Dermed sikres det, at alle i samplerammen modtager spørgeskemaet fremfor eksempelvis distribution via email, hvor internetadgang og emailkonto er nødvendig. Endvidere sikrer inklusionen af svarkuvert, at returnering af besvarelsen er overkommelig for langt de fleste. Respondenterne fik en deadline for returnering af spørgeskemaet og fik efter dennes overskridelse tilsendt en rykker med ny deadline. Da besvarelserne var anonyme, var det ikke muligt at afgøre hvilke personer, der havde returneret spørgeskemaet, hvorfor spørgeskemaet blev udsendt til hele samplerammen igen efter første deadline. Selvom det er usandsynligt, at samme person har besvaret og returneret spørgeskemaet to gange er det alligevel en mulighed, der skal tages højde for i forhold til reliabiliteten. Dette problem kunne være undgået, hvis spørgeskema og/eller konvolut var blevet forsynet med et identifikationsnummer. Før den endelige analyse kan identifikationsnumre let fjernes, hvormed respondenternes anonymitet sikres. Af det samlede udtræk valgte 580 personer at besvare og returnere spørgeskemaet. Dermed er der opnået en responsrate på 58%. Af respondentgruppen udgør type 1 diabetikerne 279, svarende til 48.1%, mens type 2 diabetikerne udgør 51.9% svarende til 301 personer.
Med udgangspunkt i besvarelserne fra disse respondenter vil der i dette speciale blive foretaget en

analyse, der skal klarlægge diabetikernes oplevede arbejdsmiljø. Det skal blandt andet afdækkes om der er forskel i besvarelserne som følge af typen af diabetes eller andre variable og selvfølgelig, om der er signifikant forskel mellem diabetikere og normalbefolkningen. Da der ikke er spurgt direkte til arbejdsmiljøet vil analysen baseres på en fortolkning af spørgsmålenes betydning for arbejdsmiljøet ud fra en subjektiv vurdering. Analysen vil, som beskrevet i indledningen, både fokusere på statistik og sociologi, da begge elementer er nødvendige i forhold til at gennemføre en solid analyse med mulighed for at identificere kausale sammenhænge. Den teori, der skal være generende for den statistiske analyse bliver således præsenteret i næste kapitel, hvorefter den sociologiske teoretiker Robert Karasek introduceres i kraft af krav-kontrol modellen, der skal være medvirkende til at beskrive arbejdsmiljø.

Grafiske modeller

Dette kapitel fokuserer på den grundlæggende teori om diskrete grafiske modeller. Herunder en række centrale definitioner og konstruktion af modeller. Endvidere beskrives hypotesetest både i forhold til sammenligning af modeller og i forhold til at teste signifikans af kanter i en grafisk model. Endelig introduceres forskellige metoder til modelselektion og teorien udvides til også at indeholde orienterede grafer og kædegrafer. Først skal nogle grundlæggende egenskaber i forbindelse med grafiske modeller dog præsenteres. Kapitlet er inspireret af [Edw00, Kapitel 2 og 5-7].

2.1 Relation mellem modeller og grafer

Grafiske modeller indeholder, som navnet antyder, elementer af både grafteori og statistiske modeller. Indledningsvis beskrives det således, hvordan disse relaterer til hinanden. Der henvises iøvrigt til Appendiks i forhold til notation og betydning af centrale grafteoretiske begreber.
Analyse af data tager udgangspunkt i at konstruere en model, der kan beskrive data. En model, M, er en familie af mulige tæthedsfunktioner, f_θ, hvor $\theta \in \Theta$ er en ukendt parameter. Data betragtes således som et uafhængigt sample fra f_θ. I tilfælde med diskrete data vil der være tale om en sandsynlighedsfunktion. En uorienteret graf, G, består af en endelig mængde knuder, V, og kanter, E, hvor hver knude repræsenterer en variabel fra de pågældende data. Via en række Markovegenskaber, som defineres i nedenstående, er det muligt at relatere den statistiske egenskab betinget uafhængighed til grafteorien.

Definition 2.1.1. Parvis Markovegenskab for uorienterede grafer
En sandsynlighedsfunktion opfylder den parvise Markovegenskab med hensyn til grafen G, hvis

$$X \perp Y | V \setminus \{X, Y\}$$

for ethvert par af knuder, hvor $X \nsim Y$.

Konstruktionen af en model sker i overensstemmelse med Markovegenskaberne, hvilket vil sige, at efter konstruktion fortolkes en manglende kant mellem to variable som betinget uafhængighed givet alle øvrige variable.

Definition 2.1.2. Lokal Markovegenskab for uorienterede grafer
En sandsynlighedsfunktion opfylder den lokale Markovegenskab med hensyn til grafen G, hvis

$$X \perp V \setminus (X \cup \mathrm{bd}(X)) | \mathrm{bd}(X)$$

for ethvert $X \in V$.

Ud fra Definition 2.1.2 fortolkes en manglende kant til X som betinget uafhængighed givet alle de variable, der er tilstødende.

Definition 2.1.3. Global Markovegenskab for uorienterede grafer
En sandsynlighedsfunktion opfylder den globale Markovegenskab med hensyn til grafen G, hvis

$$X \perp Y | S$$

for disjunkte $X, Y, S \subseteq V$ således S separerer X og Y i G.

Disse definitioner er således nøglen i forbindelse med at relatere grafteori og betinget uafhængighed. Definitionerne danner med andre ord grundlaget for fortolkning af sammenhænge mellem variable i grafiske modeller.
Når to variable er betinget uafhængige faktoriserer sandsynlighedsfunktionen som

$$P(A = a, B = b|C = c) = P(A = a|C = c)P(B = b|C = c). \tag{2.1}$$

Denne egenskab kan ligeledes relateres til grafteorien ved følgende definition.

Definition 2.1.4. Faktoriseringsegenskaben
En sandsynlighedsfunktion faktoriserer med hensyn til grafen G, hvis sandsynlighedsfunktionen kan skrives som

$$p(i) = \prod_{c \in C} \phi_c(i),$$

hvor C er mængden af kliker i G og ϕ_c er en funktion som kun afhænger af i gennem i_c.

Definition 2.1.1, 2.1.2, 2.1.3 og 2.1.4 er alle ækvivalente under forholdsvis generelle betingelser og på baggrund af disse definitioner kan alle marginale og betingede uafhængigheder, der gælder for alle tætheder i modellen, tolkes direkte af grafen. I de følgende afsnit præsenteres teorien bag den gren af grafiske modeller, der betegnes *diskrete* i.e. modeller, der kun indeholder diskrete variable.

2.2 Diskrete modeller

Konstruktionen af en grafisk model tager udgangspunkt i en d-dimensional datatabel, hvor $d = |\Delta|$ i.e. antallet af variable, der udgør mængden Δ. Hver variabel $X \in \Delta$ har et antal kategorier a_X. Antallet af celler i en given tabel er således $I = \prod_{\mathcal{X} \in \Delta} a_{\mathcal{X}}$ og mængden af celler noteres $\mathcal{I}$. Sandsynligheden for, at en observation falder i celle $i = (i_1, \ldots, i_d) \in \mathcal{I}$ betegnes p_i. Antallet af observationer i celle i betegnes n_i og det forventede antal observationer er givet ved $m_i = Np_i$, hvor $N = \sum_{i \in \mathcal{I}} n_i$ i.e. det totale antal observationer.
Fordelingen af observationer kan betragtes som multinomial, da der er $\binom{N}{n_i}$ måder at placere n_i observationer, det vil sige, likelihoodfunktionen for tabellen $\{n_i\}_{i \in \mathcal{I}}$ er givet ved

$$L(\{p_i\}_{i \in \mathcal{I}} | \{n_i\}_{i \in \mathcal{I}}) = \frac{N!}{\prod_{i \in \mathcal{I}} n_i!} \prod_{i \in \mathcal{I}} p_i^{n_i}. \tag{2.2}$$

Til at modellere de diskrete data anvendes loglineære modeller af formen

$$\ln(p_i) = \sum_{\delta \subseteq \Delta} u_i^{\delta},$$

hvor $\{u_i\}$ er interaktionsled. Det bemærkes, at u_i^{δ} kun afhænger af i gennem marginalcellerne i_δ. Hvert led $u_i^X, X \in \Delta$ tolkes således som hovedeffekten af at inkludere variablen X i modellen. Enhver model indeholder et konstantled, svarende til $\delta = \emptyset$, der angiver referenceniveauet i.e. modellen uden effekt af variable. Interaktion mellem flere variable modelleres ved at medtage to-faktor (eller fler-faktor) interaktionsled af formen u_i^{XY}, hvor $X, Y \in \Delta$. Hver model er defineret ved det/de maksimale interaktionsled svarende til det/de u_i^{δ}, der beskriver effekten af det største antal variable. De variable, der indgår i de maksimale interaktionsled kaldes også modellens *generatorer*.

Definition 2.2.1. Hierarkisk model
En hierarkisk model defineres som

$$\ln(p_i) = \sum_{\delta \in g} u_i^{\delta}$$

hvor g betegner mængden af generatorer.

Listen af generatorer udgør endvidere *modelformlen* for den pågældende model. For de hierarkiske modeller gælder, at hvis $t > s$ og s-faktor interaktionsleddet er sat lig nul, så bliver alle t-faktor interaktionsled, der indeholder samme variable, sat lig nul. De grafiske modeller er en underklasse af hierarkiske modeller og er defineret som følger.

Definition 2.2.2. Grafisk model
En grafisk model er en hierarkisk model, som afviger fra den fulde model ved, at en mængde af to-faktor interaktionsleddene sættes lig nul.

For en grafisk model svarer generatorerne til klikerne i grafen. Givet en graf er det således muligt at bestemme modelformlen for den tilhørende grafiske model ved at identificere klikerne i grafen.

Eksempel 2.2.3. Loglineær model
En model for $\Delta = \{X, Y, Z\}$ med maksimale interaktionsled u_i^{XY} og u_i^{XZ} er givet ved modelformlen (XY, XZ). Den loglineære model er således af formen

$$\ln(p_i) = \emptyset + u_i^X + u_i^Y + u_i^Z + u_i^{XY} + u_i^{XZ}.$$

Modellen kan også udtrykkes ved cellesandsynligheder. Lad henholdsvis X, Y og Z have $a_X = x, a_Y = y$ og $a_Z = z$, så er sandsynligheden for en observation i celle $i = (i_X, i_Y, i_Z) = (j, k, l)$ givet ved p_{jkl}, hvor j kan antage værdierne $1, \ldots, x$, k kan antage værdierne $1, \ldots, y$ og l kan antage værdierne $1, \ldots, z$. I ovenstående model er Y og Z betinget uafhængige givet X. Der gælder således, at

$$P(Y = k, Z = l | X = j) = P(Y = k | X = j) P(Z = l | X = j).$$

Da $P(A|B) = P(A \cap B)/P(B)$ fås

$$\frac{p_{jkl}}{p_{j++}} = \frac{p_{jk+}}{p_{j++}} \frac{p_{j+l}}{p_{j++}},$$

hvor $+$ indikerer summation over det pågældende indeks. Cellesandsynligheden p_{jkl} er således givet ved

$$p_{jkl} = \frac{p_{jk+} p_{j+l}}{p_{j++}}. \tag{2.3}$$

□

Ud fra Eksempel 2.2.3 ses det, at disse modeller er overparametriserede, da de maksimale interaktionsled alene beskriver effekten af de øvrige interaktionsled i og med, at disse er indeholdt i det lineære underrum, som udspændes af de maksimale interaktionsled. Dermed ses det, at modellerne netop kan skrives som i Definition 2.2.1.
I forlængelse af begrebet marginale celler introduceres begrebet marginale tabeller. En marginal tabel på δ for antallet af observationer noteres $N_\delta = \{n_{i_\delta}\}_{i_\delta \in \mathcal{I}_\delta}$. Da p_i kan beregnes ud fra de marginale cellesandsynligheder for de maksimale interaktionsled, som vist i (2.3), er de marginale tabeller $\{N_\delta\}_{\delta \in g}$ sufficiente for den fulde tabel. Det vil sige, den fulde datatabel kan reduceres til de marginale tabeller uden at miste information.
Når der opstilles en model for data er det naturligvis interessant at undersøge, hvor godt den pågældende

model beskriver data. Deviansen mellem den opstillede model, M_0, og den mættede model, M_f, beregnes således via likelihood ratio testet

$$G^2 = 2(\hat{l}_f - \hat{l}_0), \tag{2.4}$$

hvor $\hat{l}$ betegner den maksimerede loglikelihood. Ved at tage logaritmen til (2.2), idet cellerne indiceres $1, \ldots, I$, opnås loglikelihooden

$$\begin{aligned} l(\{p_i\}|\{n_i\}) &= \ln\left(\frac{N!}{\prod_{i\in\mathcal{I}} n_i!}\right) + \sum_{i\in\mathcal{I}} n_i \ln(p_i) \\ &= \ln\left(\frac{N!}{\prod_{i=1}^{I} n_i!}\right) + \sum_{i=1}^{I} n_i \ln(p_i) \\ &= \ln\left(\frac{N!}{\prod_{i=1}^{I} n_i!}\right) + \sum_{i=1}^{I-1} n_i \ln(p_i) + n_I \ln(p_I), \end{aligned}$$

hvor $p_I = 1 - \sum_{i=1}^{I-1} p_i$, da $\sum_{i=1}^{I} p_i = 1$. Maksimumlikelihoodestimaterne (MLE'erne) er de værdier af p_i, der maksimerer ovenstående og beregnes som følger:

$$\begin{aligned} \frac{\partial l}{\partial p_i} &= \frac{n_i}{p_i} + \frac{n_I}{p_I}\frac{\partial p_I}{\partial p_i} \\ 0 &= \frac{n_i}{\hat{p}_i} - \frac{n_I}{p_I} \\ \hat{p}_i &= n_i \frac{p_I}{n_I}. \end{aligned} \tag{2.5}$$

Ved at summere over i fås

$$\begin{aligned} \sum_{i\in\mathcal{I}} \hat{p}_i &= \sum_{i\in\mathcal{I}} n_i \frac{p_I}{n_I} \\ 1 &= N\frac{p_I}{n_I} \\ \frac{1}{N} &= \frac{p_I}{n_I}. \end{aligned} \tag{2.6}$$

Ved at indsætte (2.6) i (2.5) opnås MLE'erne $\hat{p}_i = \frac{n_i}{N}$ under M_f. Deviansen mellem en given model M_0 og M_f er dermed

$$\begin{aligned} G^2 &= 2\left(\sum_{i=1}^{I} n_i \ln(\hat{p}_i^f) - \sum_{i=1}^{I} \ln(\hat{p}_i^0)\right) \\ &= 2\left(\sum_{i=1}^{I} n_i \ln\left(\frac{n_i}{N}\right) - \sum_{i=1}^{I} \ln(\hat{p}_i^0)\right) \\ &= 2\sum_{i=1}^{I} n_i \ln\left(\frac{n_i}{\hat{m}_i^0}\right), \end{aligned}$$

hvor $\hat{m}_i^0$ er estimatet af det forventede antal observationer under M_0. For indlejrede modeller, $M_0 \subseteq M_1$, beregnes deviansdifferensen som følger

$$\begin{aligned} D &= G_0^2 - G_1^2 \\ &= 2\sum_{i=1}^{I} n_i \ln\left(\frac{n_i}{\hat{m}_i^0}\right) - 2\sum_{i=1}^{I} n_i \ln\left(\frac{n_i}{m_i^1}\right) \\ &= 2\sum_{i=1}^{I} n_i \ln\left(\frac{\hat{m}_i^1}{\hat{m}_i^0}\right). \end{aligned} \tag{2.7}$$

Fordelen ved (2.4) og (2.7) er, at begge er asymptotisk χ^2_v-fordelt under M_0, hvor antallet af frihedsgrader, v, er givet ved differensen i antallet af parametre mellem de to modeller. Det er dermed muligt at vurdere ekstremiteten af enten deviansen, G^2, eller deviansdifferensen, D, ved at sammenholde disse med en passende χ^2-fordeling. Ved et fastlagt signifikansniveau, α, kan en given model således forkastes hvis $P(D > d_{\text{obs}}) < \alpha$. I forlængelse af dette vil det følgende afsnit beskrive anvendelsen af hypotesetest i forbindelse med grafiske modeller.

2.3 Hypotesetest

De test som foretages skal afgøre hvilke variable, der er betinget uafhængige, da det er denne egenskab der har betydning for inklusion eller eksklusion af kanter mellem variable. I udgangspunktet er de følgende hypotesetest således test for betinget uafhængighed, hvor

$$H_0 : X \perp Y | Z.$$

Et test for betinget uafhængighed er med andre ord et test af

$$p_{jkl} = \frac{p_{j+l}p_{+kl}}{p_{++l}}$$

jf. Eksempel 2.2.3. I praksis testes en kant ved at foretage et test af modellen uden den pågældende kant mod modellen med den pågældende kant, og på baggrund af p-værdien vælges en af de to modeller.
I dette afsnit fokuseres på tre forskellige typer af test, der hver især er anvendelige inden for et særligt område. Først præsenteres χ^2-testet, der anvendes i forbindelse med store samples, mens eksakte test er mere hensigtsmæssige når samplestørrelsen er lille. Endelig beskrives et rangtest, der egner sig specielt til ordinalskalerede variable.

2.3.1 χ^2-test

Som beskrevet i det forrige anvendes χ^2-fordelingen blandt andet når to modeller sammenlignes med henblik på at afdække hvilken model, der har det bedste fit i forhold til data. Teststatistikken for χ^2-testet er givet ved

$$\chi^2 = \sum_{i=1}^{I} \frac{(n_i - \hat{m}_i)^2}{\hat{m}_i},$$

hvor $\hat{m}_i$ er det forventede antal observationer i celle i under nulhypotesen. Det ses, at χ^2-testet ligeledes er deviansbaseret, hvorfor store værdier af χ^2 er kritiske for nulhypotesen. Teststatistikken χ^2 giver anledning til en p-værdi og på baggrund af denne vil nulhypotesen enten accepteres eller forkastes. Teststatistikken kan anvendes i forbindelse med modelselektion og herunder *decomposable edge deletion*-proceduren. Denne procedure tager udgangspunkt i dekomposable modeller, det vil sige modeller hvis grafer er triangulerede og dermed ikke indeholder cykler uden korde af længde større end 3. I denne procedure er M_0 og M_1 dekomposable modeller, hvor M_0 opstår ved at fjerne en kant fra M_1. Da M_0 skal være dekomposabel er det således ikke alle kanter, der kandiderer til at blive fjernet. Det kræves eksempelvis, at den pågældende kant kun må være indeholdt i én klike. Hvis en kant er indeholdt i flere, vil fjernelse medføre minimum en 4-cykel uden korde. Anvendelsen af χ^2-testet i forbindelse med modelselektion beskrives i afsnit 2.5.

2.3.2 Eksakt test

Fordelen ved at bruge eksakte test er, at disse test ikke er asymptotiske og derfor bør anvendes i forbindelse med små samples. Fishers eksakte test tager udgangspunkt i sandsynligheden for de tabeller, N, som det er muligt at konstruere, mens marginalerne, N_0, fra den observerede tabel, N_{obs},

fastholdes. Det er muligt at beregne disse sandsynligheder, da de er givet ved successive hypergeometriske fordelinger. Teststatistikken er i dette tilfælde

$$T = P(N|H_0, N_0)$$

og p-værdien beregnes som $\sum_t P(T \leq t)$, hvor $t = P(N_{\text{obs}}|H_0, N_0)$. Problemet med Fishers eksakte test er, at den anvender metoden *exhaustive enumeration*, det vil sige for hver mulig tabel beregnes t, hvilket er meget beregningstungt i og med store samples medfører et stort antal mulige tabeller. Et alternativ til denne metode er at bruge Monte Carlo sampling, hvor et stort tal K vælges og derefter samples K tilfældige tabeller med tilhørende T-værdier t_i, hvor $i = 1, \ldots K$. Der defineres en indikatorfunktion $\mathbb{1}[t \leq T]$ og p-værdien estimeres således ved

$$\frac{1}{K} \sum_{i=1}^{K} \mathbb{1}[t_i \leq t],$$

hvorefter et konfidensinterval for en proportion kan beregnes. En variant af denne metode er den sekventielle Monte Carlo sampling, hvor samplingen stoppes, hvis der er fundet h tabeller, hvor $\mathbb{1}[t_i \leq t] = 1$.

2.3.3 Rangtest

Den sidste type test adskiller sig fra de øvrige, da disse test tager udgangspunkt i rangen af de forskellige observationer. Rangtestet, som beskrives her, er et ikke-parametrisk test, i.e. det er ikke baseret på antagelser om en bestemt fordeling af data. Observationerne ordnes således i ikke aftagende rækkefølge efter deres numeriske størrelse. Rangen af en observation svarer til observationens position i rækken af observationer. I tilfælde, hvor to eller flere observationer er lige store beregnes rangen som gennemsnittet af observationernes position.

Eksempel 2.3.1. Rang af observationer
Lad $y = (6, 2, 3, 5, 6, 9)$. Efter ordning fås $\tilde{y} = (2, 3, 5, 6, 6, 9)$. Rangen af $\tilde{y}_1 = 2$ er således 1, rangen af $\tilde{y}_2 = 3$ er 2 og rangen af $\tilde{y}_3 = 5$ er 3. Da $\tilde{y}_4 = \tilde{y}_5 = 6$, så er rangen af disse observationer 4.5 og rangen af $\tilde{y}_6 = 9$ er 6. □

Wilcoxon test er et rangtest, der bruges til at sammenligne fordelingsfunktioner, F_1 og F_2, mellem to populationer. Testet forudsætter således en dikotom variabel, der angiver de to populationer. Fordelingen af populationerne på en ordinal variabel udgør testet, hvor der testes for homogenitet, $F_1 = F_2$, mod den alternative hypotese; at F_1 er stokastisk større (eller mindre) end F_2. Det er endvidere muligt at foretage testet på forskellige strata. Teststatistikken er givet ved

$$T = |W - E[W|H_0]|,$$

hvor W angiver summen af rangene for den første population over (eventuelle) strata s i.e.

$$W = \sum_s R_{1s}.$$

Den tilhørende p-værdi beregnes som

$$P(T \geq T_{\text{obs}}|H_0).$$

Da Wilcoxon test tager udgangspunkt i stokastisk ordning er testen særlig anvendelig i forhold til data med ordinalskalerede variable, dog ikke i tilfælde, hvor F_1 hverken er konsekvent større eller mindre end F_2. I disse tilfælde er der ikke entydig afvigelse fra nulhypotesen og Wilcoxon mister sin styrke.
I tilfælde hvor flere populationer skal sammenlignes kan Kruskal-Wallis testet anvendes. I denne opsætning tages der udgangspunkt i en nominal variabel, der angiver k forskellige populationer på en ordinal variabel. Der testes således igen for homogenitet,

$$F_1 = F_2 = \cdots = F_k,$$

mod den alternative hypotese, hvor mindst én af fordelingsfunktionerne er stokastisk større end de øvrige.
Der er nu introduceret flere typer af test, der kan afgøre, hvorvidt der er statistisk sammenhæng mellem forskellige variable. Som nævnt tidligere anvendes hypotesetest i forbindelse med sammenligning af modeller, men endvidere anvendes disse test når det undersøges, om der skal tilføjes eller fjernes en kant mellem to variable i.e. i forbindelse med konstruktionen af en grafisk model. Særligt rangtestet er anvendeligt i forbindelse med ordinalskalerede variable, som udgør størstedelen af variablene i forbindelse med den aktuelle problemstilling. Rangtestene, som beskrevet i ovenstående, tester dog kun sammenhæng mellem én ordinalskaleret og en dikotom eller nominal variabel. Det kan endvidere være nødvendigt at undersøge sammenhænge mellem to ordinale variable, hvilket er muligt med et Jonckheere-Terpstra test, der ligesom Kruskal-Wallis tester homogenitet mellem en række fordelingsfunktioner. I stedet for at foretage et test kan to ordinalskalerede variable dog også undersøges ved at bruge et korrelationsmål, som det beskrives i følgende afsnit.

2.4 Korrelationsmål

De data, der er indsamlet udgøres i høj grad af ordinalskalerede variable. Dermed er det muligt at foretage en naturlig rangering af respondenternes svar, hvilket skal udnyttes i den senere analyse for at styrke validiteten. Fordelen ved at bruge et associationsmål som udnytter ordinalitet er, at det bliver muligt at vurdere både styrken og retningen af en sammenhæng. I nedenstående præsenteres således to forskellige mål som beskriver association mellem ordinalskalerede variable. Afsnittene er inspireret af [Agr06].

2.4.1 γ og τ

I samfundvidenskabelige problemstillinger anvendes ofte Goodman og Kruskals γ. Dette mål beregnes som et forhold mellem antallet af *konkordante* og *diskordante* par for en aktuel kontingenstabel. Tabellen konstrueres ved at udnytte ordinaliteten af de pågældende variable i.e. første celle, øverste venstre, indeholder de respondenter, der har svaret i den 'laveste' kategori på begge variable, mens tabellens sidste celle, nederste højre, indeholder de respondenter, der har svaret i den 'højeste' kategori på begge variable. Cellerne i tabellen svarer således til forskellige kombinationer af respondenternes svar på de to variable, og udgør dermed grundlaget for de konkordante og diskordante par.
Et konkordant par er, som navnet antyder, et par af observationer, som er i indbyrdes overensstemmelse. Det vil eksempelvis sige et observationspar, hvor respondenten har angivet svar i den lave kategori på både X og Y; alternativt i den høje kategori på både X og Y. Antallet af konkordante par i en kontingenstabel $N = \{n_{ij}\}$ er givet ved

$$C = \sum_{i'>i} \sum_{j'>j} n_{ij} n_{i'j'}.$$

De diskordante par er til gengæld observationspar, hvor der er uoverensstemmelse mellem respondentens svar på de to variable således vil et svar i den lave kategori på X eksempelvis korrespondere med et svar i den høje kategori på Y. Antallet af diskordante par er dermed

$$D = \sum_{i'<i} \sum_{j'<j} n_{ij} n_{i'j'}.$$

I og med der kun summeres over de celler, hvor begge kategorier er skarpt større eller mindre udgør de konkordante og diskordante par ikke det totale antal observationspar i tabellen. Ligesom det var tilfældet i rangtestet, hvor nogle observationer må tildeles samme rang, opstår der nogle par af observationer, som ikke er skarpt større eller mindre på begge variable. Antallet af observationspar som er lige store

på henholdsvis X og Y er givet ved

$$T_X = \sum_i \frac{n_{i+}(n_{i+}-1)}{2} \qquad \text{og} \qquad T_Y = \sum_j \frac{n_{+j}(n_{+j}-1)}{2}.$$

De par der er lige store på både X og Y, i.e. cellerne i tabellens diagonal, hvis X og Y har samme antal svarkategorier, tælles ved

$$T_{XY} = \sum \frac{n_{ij}(n_{ij}-1)}{2},$$

hvormed det totale antal observationspar i en tabel er givet ved $C+D+T_X+T_Y-T_{XY}$. Det er imidlertid kun de konkordante og diskordante par som har betydning for γ. Dette mål er givet ved

$$\gamma = \frac{C-D}{C+D},$$

hvilket vil sige, at $C > D$ angiver en positiv association mellem X og Y. Associationsmålet påvirkes ikke af, hvorvidt X eller Y betragtes som henholdsvis uafhængig og afhængig variabel, i.e. γ er et såkaldt symmetrisk mål.
Et relateret mål er Kendalls τ, som til forskel fra γ inddrager de observationspar som netop hverken er konkordante eller diskordante i.e. de par som er indeholdt i T_X, T_Y og T_{XY}. Dette mål er givet ved

$$\tau = \frac{C-D}{\sqrt{(C+D+T_Y-T_{XY})(C+D+T_X-T_{XY})}}.$$

Det ses, at for en kontingenstabel, hvor $T_X = T_Y = T_{XY} = 0$, så er $\tau = \gamma$; ellers vil $\tau < \gamma$. Både τ og γ antager værdier i intervallet $[-1, 1]$, hvor de to yderpunkter indikerer henholdsvis fuldstændig negativ og fuldstændig positiv association.
De to mål er imidlertid blot estimationer af den sande korrelation, hvorfor et hypotesetest efterfølgende er nødvendigt. Det testes således om den aktuelle γ- eller τ-værdi er signifikant forskellig fra 0. Da γ og τ er asymptotisk normalfordelte er det i praksis muligt at konstruere et konfidensinterval og på baggrund af dette afgøre, hvorvidt den alternative hypotese accepteres. Associationsmålet kan dermed afgøre, hvorvidt der skal tilføjes eller fjernes en kant mellem to variable. Både korrelationer og hypotesetest er således af betydning når en grafisk model skal konstrueres ud fra data. I det følgende bliver flere fremgangsmåder for denne konstruktion gennemgået.

2.5 Modelselektion

I praksis vil konstruktionen af modeller foregå via computerprogrammer som eksempelvis HUGIN eller DIGRAM, hvorfor dette afsnit vil være en overordnet introduktion til nogle af de selektionsprocedurer, som benyttes, når der skal opstilles en grafisk model for en mængde variable. Der er to centrale ingredienser i selektionsprocedurerne; selektionskriteriet og søgestrategien, hvilke bliver fremhævet i det følgende.

2.5.1 Stepwise selektion

Der er to selektionsprocedurer inden for stepwise selektion, som skal beskrives i dette afsnit. Søgestrategien for den første kaldes *backward* og initialiserer normalt med den fulde model, som derefter simplificeres ved fjernelse af kanter. Selektionskriteriet for at beholde eller fjerne en kant er i dette tilfælde normalt et hypotesetest. Proceduren fjerner således den kant, der har den største p-værdi og terminerer når alle modellens resterende kanter har p$< \alpha$. Det er muligt at fiksere kanter, i.e. at fastholde en kant selvom den ikke er statistisk signifikant, hvis det vurderes, at kanten har betydning i forhold til problemstilling eller på baggrund af a priori viden. Backward selektion kan effektiviseres via *kohærensprincippet*, det vil sige hvis en kant én gang er testet signifikant, så kan den ikke efterfølgende fjernes.

Den anden procedure bruger en *forward* søgestrategi og initialiserer normalt med den tomme model, hvorefter kanter tilføjes i henhold til selektionskriteriet; igen typisk et hypotesetest. Den kant, der har den mindste p-værdi tilføjes først og proceduren fortsætter indtil test af de resterende kanter alle giver p> α. Forskellen mellem de to typer af stepwise selektion er, at backward selektion begynder med en model, der er konsistent med data, hvilken simplificeres ved fjernelse af kanter. Denne fremgangsmåde sikrer, at alle modeller til og med den endelige model ligeledes er konsistente med data. Forward begynder i stedet med en simpel model, der er inkonsistent med data og som efterfølgende udvides til en model, der muligvis er acceptabel, ved at tilføje kanter. Når der anvendes hypotesetest i udvælgelsen af kanter, foretager backward selektion et test mellem modeller, hvor i hvert fald den store model er acceptabel, mens forward selektion foretager et test mellem modeller, hvor hverken nulhypotesen eller den alternative hypotese er valid. Frem for at vælge én af ovenstående søgestrategier er det muligt at kombinere disse således der både søges forlæns og baglæns.
Et alternativ til hypotesetest i forbindelse med modelselektion er at anvende et *informationskriterie*, der ligeledes afgør, hvilken model, der har det bedste fit. Stepwise selektion kan eksempelvis baseres på Akaikes informationskriterie (AIC), der er givet ved

$$-2\ln(Q) + 2p,$$

hvor Q angiver den maksimerede værdi af likelihooden for den aktuelle model, mens p betegner antallet af parametre i modellen. Med dette mål vælges den model, der har den mindste AIC-værdi. Et lignende mål er det bayesianske informationskriterie (BIC), der er givet ved

$$-2\ln(Q) + p\ln(N),$$

hvor N angiver antallet af observationer. Disse mål kan således være med til at afgøre, hvilken model, der er mest konsistent med data og dermed er informationskriterierne også selektionskriterier.

2.5.2 EH-proceduren

Modelselektion via denne procedure initierer med den fulde model og tester en følge af modeller i forhold til denne via χ^2-testet af deviansen G^2. Hvis p> α accepteres den pågældende model således, da deviansen dermed ikke er kritisk for nulhypotesen. Ligesom det var tilfældet med backward selektion gælder kohærensprincippet, i.e. hvis en model, M, accepteres, så vil alle modeller, der indeholder M som submodel ligeledes accepteres. Modellerne tilføjes en liste over accepterede modeller og tilsvarende laves en liste over forkastede modeller. EH-proceduren søger efter de mest simple modeller, det vil sige modeller, med det færreste antal kanter. Når proceduren er færdig vælges således en minimalt accepteret model fra listen over accepterede modeller; hvis der er flere minimalt accepterede modeller beror valget af model på et skøn. Hvor stepwise selektion således kun returnerer én endelig model, returnerer EH-proceduren flere mulige modeller, hvilket kan gøre valget af model tidskrævende.
Valg af model, via modelselektion, er, som det også har vist sig i ovenstående, primært eksplorativ, i.e. der tages udgangspunkt i data og på baggrund af signifikante datasammenhænge konstrueres en fornuftig model. Dette afspejler netop det største validitetsproblem i forbindelse med grafiske modeller, da validiteten af en model blandt andet hviler på, at modellen ikke er valgt på baggrund af data. Ved at inddrage sociologisk viden er det til gengæld muligt at konstruere en model, der også opbygges på baggrund af en a priori viden om sammenhænge mellem variable. Med hensyn til modelselektion er det således en fordel at anvende sociologiske teorier, da disse kan være med til at afgøre, om visse kanter bør fikseres eller om en kant eventuelt kan erstattes af en pil i.e. en indikation af retningen for en sammenhæng. I forlængelse af dette introduceres nu forskellige typer af orienterede grafer.

2.6 Orienterede grafer

Indtil videre har der kun været fokus på de uorienterede grafer, i.e. grafer der udelukkende består af kanter. I dette afsnit behandles først de såkaldte DAGs, der er acykliske grafer kun bestående af orienterede kanter, i.e. grafer uden orienterede cykler. Dernæst behandles kædegrafer, som er en kombination

af orienterede og uorienterede grafer, hvilket vil sige, at de indeholder både pile og kanter. I tilfældet med de uorienterede grafer blev en manglende kant fortolket som betinget uafhængighed givet alle øvrige variable. Med indførelsen af orienterede kanter, pile, ændres denne fortolkning, da der nu kan siges at blive inddraget en tidsdimension.

2.6.1 DAG

Gruppen af grafer, der betegnes DAGs, indeholder som nævnt ingen orienterede cykler. Med andre ord eksisterer der en ordning, ikke nødvendigvis entydig, af variablene $\{X_1, X_2, \ldots, X_n\}$ således det kun for $i < j$ gælder, at $X_i \to X_j$ er mulig. Når en DAG skal konstrueres er det dermed oplagt at inddrage a priori viden om, hvordan variablene forholder sig til hinanden med hensyn til en tidsdimension således de kan ordnes på en meningsfuld måde. Ikke overraskende vil det derfor være hensigtsmæssigt at inddrage sociologisk viden i forbindelse med konstruktionen af en DAG.
Eftersom variablene kan ordnes som ovenstående er det muligt at faktorisere den simultane tæthed af $\{X_1, X_2, \ldots, X_n\}$ som

$$f(X_1)f(X_2|X_1)\cdots f(X_n|X_{n-1}, X_{n-2}, \ldots, X_1). \tag{2.8}$$

Hvis $f(X_j|X_{j-1}, X_{j-2}, \ldots, X_1)$ ikke afhænger af X_i, i.e. hvis

$$X_i \perp X_j | \{X_1, X_2, \ldots, X_j\} \setminus \{X_i, X_j\}, \quad \text{hvor } i < j$$

så forbindes X_i og X_j ikke med en pil. Fortolkningen af en manglende kant i en DAG svarer altså til den parvise Markovegenskab for uorienterede grafer; at de pågældende variable er betinget uafhængige givet alle øvrige variable, men med den forskel, at variablene nu tolkes som betinget uafhængige givet alle *tidligere* variable. Den simultane tæthed som præsenteret i (2.8) kan således også skrives som

$$\prod_{X \in \Delta} f(X|\text{pa}(X)),$$

hvor $\text{pa}(X)$ betegner mængden af forældre til X i.e. de variable, der er orienterede mod X, og hvis

$$X_i \perp X_j | \text{an}(\{X_i, X_j\}),$$

hvor $\text{an}(\{X_i, X_j\})$ angiver mængden af forfædre til X_i og X_j i.e. de variable, der er på en orienteret sti til X_i og X_j, så skal der ikke tilføjes en pil mellem X_i og X_j.
Det er imidlertid ikke kun den parvise Markovegenskab der gør sig gældende for DAGs. Der er ligeledes en pendant til den globale Markovegenskab som kan udledes ved at *moralisere* den pågældende DAG. Moralisering vil sige, at alle forældre forbindes med en uorienteret kant og alle pile erstattes af uorienterede kanter. Herved opnås en moraliseret graf G^m. For at undersøge hvorvidt $X_i \perp X_j | S$, hvor $S \subseteq V$, betragtes først den forfædrene mængde, A, af $\{X_i, X_j\} \cup S$. Da $\text{an}(X) \subseteq A \Rightarrow \text{pa}(X) \subseteq A$, så vil der for enhver variabel i A gælde, at forældrene til variablen er indeholdt i A og dermed kan den simultane tæthed for A skrives

$$\prod_{X \in A} f(X|\text{pa}(X)),$$

svarende til tætheden for subgrafen G_A. Da dette er et produkt af faktorer, der kun er afhængig af variablene X og $\text{pa}(X)$, så faktoriserer det i overensstemmelse med G_A^m, hvor den globale Markovegenskab for uorienterede grafer naturligvis er gældende. Dermed er $X_i \perp X_j | S$ hvis S separerer X_i og X_j i G_A^m.
Dette kan udvides til mængder af variable i.e. for disjunkte mængder S_1, S_2 og S_3 er $S_1 \perp S_2 | S_3$ hvis S_3 separerer S_1 og S_2 i G_A^m, hvor

$$A = (S_1 \cup S_2 \cup S_3) \cup \text{an}(S_1 \cup S_2 \cup S_3) = \text{an}^+(S_1 \cup S_2 \cup S_3).$$

Med hensyn til modelkonstruktion har en DAG en fordel i forhold til de uorienterede grafer, hvor inklusion eller eksklusion af en kant afhænger af tilstedeværelsen eller fraværet af alle de øvrige kanter.

For DAGs beror afgørelsen om at tilføje kanten $X \rightarrow Y$ kun på de andre pile, der peger på Y. Dermed er valget af model for en variabel X_j uafhængig af valget af model for X_{j-1}. I hvert trin af processen kan afhængigheden mellem X_j og de tidligere variable altså modelleres ved en hvilken som helst univariat model, hvor $X_1, X_2, \ldots, X_{j-1}$ er kovariater. Tilsvarende kan et hypotesetest i hvert trin af proceduren tilpasses efter eksempelvis variabeltype, når det undersøges hvilke kanter, der skal tilføjes eller fjernes.
Konstruktionen af en DAG forudsætter som nævnt, at variablene kan ordnes i forhold til hinanden. I tilfælde, hvor det imidlertid ikke er muligt at ordne alle variable, er det hensigstmæssigt at modellere data som en kædegraf, hvor både orienterede og uorienterede kanter inkluderes.

2.6.2 Kædegrafer

Ligesom det blev præsenteret i foregående afsnit tages der i forbindelse med kædegrafer udgangspunkt i en ordning af variable på baggrund af en a priori viden. Til gengæld betragtes nu tilfældet, hvor det kun er muligt at opdele variablene i en ordnet liste af *blokke*. Indenfor en blok er de pågældende variable således 'samtidige', mens blok B_i indeholder variable, der er tidligere end variable i blok B_{i+1}. Heraf følger, at variable indenfor en blok er forbundet via uorienterede kanter, mens variable i forskellige blokke er forbundet med pile. Den simultane tæthed for denne blokstruktur kan skrives

$$f(B_1)f(B_2|B_1)\cdots f(B_n|B_1 \cup B_2 \cup \cdots \cup B_{n-1}).$$

En manglende kant mellem to variable i samme blok, B_i, eller en manglende pil fra $X \in B_j$ til $Y \in B_i$, $j < i$, fortolkes som

$$X \perp Y | B_1 \cup B_2 \cup \cdots \cup B_i$$

i.e. X og Y er betinget uafhængige givet alle tidligere og samtidige variable. Bemærk iøvrigt, at en kædegraf ikke indeholder orienterede cykler.
En finere inddeling af en kædegraf er mulig ved at anvende kædegrafens *komponenter*. En komponent er en mængde af knuder i G, der kun indeholder uorienterede kanter. To komponenter kan være forbundet via pile med samme retning. På baggrund af kædegrafens komponenter kan en *komponent-DAG* konstrueres ved at lade komponenterne fra G udgøre knuderne i komponent-DAG'en og lade disse være forbundet med en pil, hvis det var tilfældet for komponenterne i G.
Ved moralisering af en kædegraf er det, ligesom det var tilfældet for DAGs, muligt at udlede den globale Markovegenskab. Den moralske graf, G_m, for kædegrafen G konstrueres ved at forbinde grænsen af hver komponent i G med kanter og erstatte alle pile med kanter. For at undersøge, hvorvidt $X \perp Y | S$, hvor $S \subseteq V$, betragtes igen den forfædrene mængde, $A = \text{an}^+(S)$, og subgrafen G_A moraliseres derefter. Hvis S separerer X og Y i G_A^m, så er $X \perp Y | S$, hvilket igen kan udvides til disjunkte mængder af knuder.
Modellering af kædegrafer adskiller sig til gengæld fra DAGs, da der nu kræves multivariate modeller. For hver blok B_i skal der findes en multivariat model givet $B_1 \cup B_2 \cup \cdots \cup B_{i-1}$. Valget af model for B_i er dog fortsat uafhængig af valg af model for B_{i-1}. Beslutningen om at tilføje en pil fra $X \in B_j$ til $Y \in B_i$ afhænger således kun af de øvrige pile, der peger på en variabel i B_i og de kanter der er mellem variablene i B_i.
Som beskrevet tidligere skal sociologisk viden inddrages i forbindelse med konstruktionen af de grafiske modeller og generelt i forbindelse med dataanalysen. Den problemstilling, som skal belyses, fokuserer på arbejdsmiljø, hvilket således vil være omdrejningspunktet for næste kapitel. Viden om den eksisterende teori og forskning på dette område vil som tidligere nævnt være afgørende for inklusion af orienterede kanter i grafiske modeller og vil øge validiteten af analysen.

Sociologisk teori

Som oftest opdeles arbejdsmiljø i to forskellige felter; det fysiske og det psykiske, hvilke begge vil blive beskrevet, da begge er aktuelle i forhold til problemstillingen. I dette kapitel gennemgås blandt andet den såkaldte krav-kontrol model, som blev introduceret af Robert Karasek, hvilken fokuserer på generelle faktorer, der har indflydelse på arbejdsmiljø. Kapitlet er skrevet på baggrund af [The90] og [eaR08, Kapitel 11 og 12].

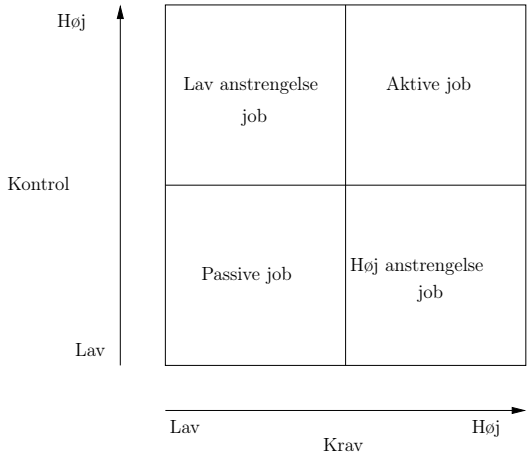
3.1 Krav-kontrol modellen

Der er flere faktorer, der har afgørende betydning for arbejdsmiljø. I det følgende beskrives Karaseks todimensionale model, der giver et mål for arbejdsmiljø ud fra forholdet mellem psykologiske *jobkrav* og den ansattes *kontrol* i.e. beslutningsfrihed. Modellen bygger på antagelsen om, at der er en sammenhæng mellem omgivelser og individ, således arbejdets vilkår medfører en række stressorer, der belaster individet. De psykologiske jobkrav er mentale krav, som stilles til den ansatte eksempelvis i form af arbejdsmængde og tidspres i forbindelse med overholdelse af deadlines. De mentale belastninger, som den ansatte udsættes for, kan imidlertid mindskes ved at ændre på graden af beslutningsfrihed, hvilket blandt andet vil sige, at den ansatte bliver givet mere frihed til at tilrettelægge sit arbejde. Her tænkes på, at den ansatte selv kan bestemme arbejdstempo, kan holde pauser efter behov etc. Indflydelse på jobbet har med andre ord positiv indvirkning på de oplevede krav, og dermed er den ansattes beslutningsfrihed et vigtigt element i processen med at få kontrol over arbejdet. Endvidere skriver Karasek, at: *"[...] if the worker's skill is being utilized and developed, the worker is more likely to feel in control of the many different situations that may arise"* [The90, side 5]. I forlængelse af dette fokuserer Karasek på *læring* i den forstand, at høje jobkrav kan forhindre læring, mens læring via arbejde udvikler selvtillid og selvsikkerhed hos den ansatte, hvilket dermed kan reducere stress o.lign. Formålet med krav-kontrol modellen er at vise, at det er muligt at reducere risikoen for arbejdsrelaterede sygdomme og samtidig sikre produktiviteten på en arbejdsplads. Fra Karaseks side kritiseres de eksisterende 'løsninger' på eksempelvis stressrelaterede sygdomme, der i højere grad fokuserer på at behandle symptomerne end på de underliggende årsager til sygdommen.
Et af de begreber som Karasek arbejder med, og som kan have betydning for arbejdsmiljøet, er *kontrakter* mellem arbejdsgiver og ansat. Udover den konkrete ansættelseskontrakt fokuseres her på den psykologiske kontrakt som består af usagte forventninger mellem de to parter. Ansættelseskontrakten er et eksempel på en transaktionel kontrakt, i.e. en kontrakt der indeholder aftale om løn, arbejdstid o.lign., mens den psykologiske kontrakt er en relationel kontrakt, der fokuserer på mindre håndgribelige faktorer som tillid, anerkendelse, følelse af at være værdsat etc. Den psykologiske kontrakt er primært en social proces, hvor disharmoni mellem parternes overbevisninger kan medvirke til at skabe et dårligt arbejdsmiljø eventuelt i form af ekstra kontrol af den ansatte, der ikke lever op til de usagte forventninger. Den omvendte situation, hvor den ansatte oplever et kontraktbrud, gør sig dog også gældende og vil igen resultere i et dårligt arbejdsmiljø. Problemet med den psykologiske kontrakt er, at hvert individ har en forforståelse af arbejdsrelationen, hvorfor ændringer fra eksempelvis ledelsens side, kan tolkes som et brud, og dermed føre til en følelse af at blive forrådt.

3.1.1 De fire kategorier

Overordnet inddeler Karasek krav-kontrol modellen i fire kategorier svarende til de forskellige kombinationer af henholdsvis lave/høje jobkrav og lav/høj kontrol. Modellen er konstrueret på baggrund af de to dimensioner; krav og kontrol. Dimensionen for de psykologiske krav konstrueres ud fra svar på spørgsmål som: "Er der for meget arbejde?"og "Bliver du nødt til at arbejde hurtigt/hårdt?". På samme måde

konstrueres kontroldimensionen ud fra svar på spørgsmål af typen "Har du meget at skulle have sagt på dit arbejde?", "Har du frihed til at tage beslutninger?", "Er dine opgaver varierede?"etc. Variablene måles således via den ansattes subjektive oplevelse og på denne baggrund kan individets placering i modellen, angivet i Figur 3.1, afgøres. Karaseks hypotese er, at der vil være en høj forekomst af sygdom,



Figur 3.1: Krav kontrol model

stress o.lign. i den gruppe af ansatte, der har job med høj grad af anstrengelse. Der skelnes mellem den gode form for anstrengelse, hvor den ansattes energi kan omsættes til handling og efterlader den ansatte i en ligevægtstilstand umiddelbart efter. Den skadelige form for anstrengelse opstår derimod når energien ikke kan omsættes til handling på grund af forhindringer i form af krav. Pulsen og adrenalinen stiger, men der kommer ingen forløsning i form af en tilfredsstillende løsning på opgaven. Det er dette psykofysiologiske pres, der i sidste ende er skyld i sygdomme som stress ifølge Karasek. Karasek bruger også udtrykket *residual anstrengelse* om den uforløste anstrengelse, som den ansatte ophober i kroppen og som resulterer i sygdom. Det er dog ikke kun handlingsfrihed i forhold til opgaven, der afhjælper anstrengelse *"[...] it may also be the freedom to engage in the informal rituals - the coffee break, the smoke break, or even fridgeting - that serve as supplementary tension release mechanisms during the work day."* [The90, side 34]. Uden disse pauser, hvor det er muligt at slappe af *"[...] no additional work can be effectively accomplished and disorganized activity results."* [The90, side 34].
De aktive job, hvor der stilles høje krav, men samtidig er høj kontrol er til gengæld den kategori, hvor der bør være den laveste forekomst af sygdom. På trods af de høje krav har den ansatte den fornødne kontrol i.e. frihed til at bruge sine evner og dermed løse de opgaver, der bliver stillet. Derudover vil personer i denne kategori opleve at lære mere, hvilket giver dem yderligere kompetencer til den næste opgave. I og med, at den ansatte har mulighed for at handle i overensstemmelse med egen overbevisning, er der kun lidt residual anstrengelse og derfor ikke det psykologiske pres som gjorde sig gældende i førnævnte kategori.
Ansatte i job med lav grad af anstrengelse vil ligeledes have et lavt niveau af residual anstrengelse og dermed lille risiko for sygdomme. Denne kategori indeholder de job, der er de mest afslappede, fordi den ansatte kan *"[...] respond to each challenge optimally, and because there are relatively few challenges to begin with."* [The90, side 36]. Ansatte i sådanne job vurderes endda at blive sundere via arbejdet.
Hvor ovenstående kategori repræsenterede den positive form for afslappede job indeholder den sidste kategori de såkaldte passive job, der har negativ effekt på læring og endvidere resulterer i *"[...] gradual loss of previously acquired skills [...]"* [The90, side 37]. Der er med andre ord tale om job uden udfordringer, hvor der ikke er mulighed for udvikling for den ansatte. Det passive job er således både umotiverende og demotiverende for den ansatte, hvorfor job i denne kategori kan føre til apati. På baggrund af de fire kategorier er det muligt at fordele en række forskellige beskæftigelsesområder i krav-kontrol modellen, hvilket beskrives i det følgende.

3.1.2 Fordeling af job

De aktive job betragtes af Karasek som gruppen af højprestige job som blandt andet tæller advokat, ingeniør og diverse lederstillinger. I denne gruppe rapporteres der, trods aktivitetsniveauet, mindre træthed, hvilket forklares med, at *"[t]hey receive higher incomes and enjoy the highest psychic rewards from work [...]"* [The90, side 42]. Fra et arbejdsmiljømæssigt synspunkt er denne kategori den bedste, da denne gruppe af ansatte ifølge Karasek har den største jobtilfredshed. Det skal endvidere bemærkes, at der er en overvægt af mænd i denne kategori. Dette afsnit vil i den forbindelse afsluttes med en beskrivelse af, hvordan mænd og kvinder fordeler sig i krav-kontrol modellen.
Lav anstrengelse job indeholder generelt jobfunktioner, hvor det er muligt for den ansatte selv at koordinere arbejdet og dermed også fastlægge arbejdstempoet. På grund af det lave niveau af psykologisk anstrengelse betegner Karasek også job i denne kategori som *"[...] a relative psychosocial paradise [...]"* [The90, side 42], hvor træthed nærmest ikke er eksisterende. Som eksempel på disse job fremhæves vedligeholdelses- og reparationsarbejde, hvilket igen har en højere forekomst af mandlige ansatte.
De passive job indeholder operatører af offentlige transportmidler, pedeller og diverse ekspedienter. I den forbindelse fremhæves et resultat fra en svensk undersøgelse, hvor det blev dokumenteret, at personer i passive job havde det laveste niveau af søvnproblemer, hvilket Karasek giver følgende fortolkning: *"The person in the passive job may inhabit a stimulation-deprived world that allows the development of one capability almost to perfection: the ability to sleep."* [The90, side 42].
Job med høj grad af anstrengelse, beskæftigelsesområdet med den højeste forekomst af arbejdsrelaterede sygdomme, er eksempelvis givet ved arbejde som tjener og generelt maskinstyrede job som arbejde ved samlebånd. Denne kategori af job besiddes primært af kvinder, hvilket understreger kønsforskellen i fordelingen af job. På baggrund af krav-kontrol modellen har mænd således generelt mere kontrol med deres arbejde end kvinder og besidder i højere grad de aktive job, mens kvinderne overvejende har passive job. Karasek præsenterer endvidere en negativ korrelation mellem kontrol og krav for kvinder, mens der er fundet en lille, positiv korrelation i mændenes tilfælde. Krav-kontrol modellen blev senere videreudviklet således der blev inddraget en ny dimension, som ligeledes har stor betydning for arbejdsmiljø. Denne nye variabel kaldes *social støtte* og kan fungere som en regulerende faktor i forhold til de sociale relationer på arbejdspladsen.

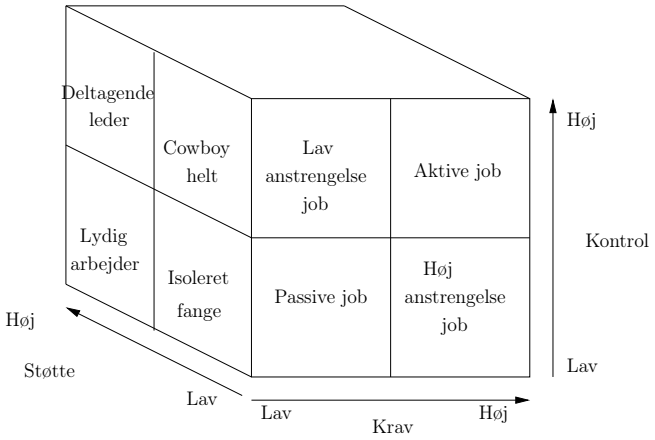
3.1.3 Social støtte

De sociale forhold er en stor del af arbejdsmiljøet, da ansatte bliver underlagt de sociale normer og værdier der eksisterer på arbejdspladsen. Gennem sociale relationer dannes normer for henholdsvis den maksimalt og minimalt accepterede indsats, i.e. der dannes, via forhandling mellem de ansatte, en fælles forståelse af organiseringen og udførelsen af en dags arbejde; af Karasek betegnet *norms of fair performance*. Denne overenskomst er et eksempel på tavs viden; altså viden som kun kan tillæres ved at indgå i det sociale liv på arbejdspladsen. For at opnå social støtte er det således nødvendigt først at blive integreret i arbejdsfællesskabet. Social støtte er således en socialiseringsfaktor, hvilken, ifølge Karasek, beskytter mod stigende krav og derved kan kompensere for en negativ belastning. Det er dog vigtigt at bemærke, at den sociale støtte ikke kan ophæve den negative effekt. Derudover har den sociale kontakt stor betydning for helbredet og i forbindelse med tilegnelse af ny viden. Endelig mener Karasek, at social støtte giver *"[...] a positive sense of identity, based on the socially confirmed value of the individual's contribution to the collective goals and well-being."* [The90, side 70]. Der er opstillet fire handlingstyper af social støtte, der kort præsenteres:

- Følelsesmæssig støtte: At der vises indføling, omsorg og tillid i forhold til den ansatte.
- Instrumentel støtte: At den ansatte får direkte hjælp til at løse opgaver.
- Informationsmæssig støtte: At den ansatte får rådgivning og dermed bedre mulighed for at håndtere belastende situationer.

- Vurderingsmæssig støtte: At den ansatte får feedback, hvilket har betydning for dennes selvvurdering.

Den sociale støtte er således afgørende for det psykiske arbejdsmiljø. På trods af, at det sociale fællesskab bidrager til at håndtere den arbejdsmæssige belastning, er de sociale bånd blevet svagere, da arbejdspladserne generelt oplever stor udskiftning. Dermed bærer de ansattes relationer på arbejdet præg af at være mere omskiftelige og skrøbelige, end det tidligere har været tilfældet.
Med inddragelse af den sociale støtte på arbejdspladsen får krav-kontrol modellen en ekstra dimension og dermed opnås nu fire nye kategorier af job som kan ses i Figur 3.2. Kategorien Deltagende leder



Figur 3.2: Krav-kontrol-støtte modellen

refererer ikke til ledere i almindelighed, men til ansatte, der har en form for indflydelse på kollektive beslutninger. Ansatte i denne kategori, hvor der både er høj støtte og høj kontrol, har mulighed for at ændre arbejdsmiljøet ved netop at være deltagende.
Begrebet cowboyhelt beskriver den type af job, hvor der er høj kontrol, men lav social støtte. Denne person er kompetent, men arbejder hovedsageligt alene. Som eksempel på denne kategori fremhæver Karasek blandt andet kunstneren.
Den isolerede fange refererer generelt til de automatiserede job, hvor der er lav kontrol og lav social støtte. I dette tilfælde er arbejde ved et samlebånd et ideelt eksempel, da dette arbejde er forudbestemt og ofte isoleret fra de øvrige ansatte. Karasek bruger udtrykket fange, da *"[...] such jobs represent a clear sociobiological misfit with human physiological capabilities."* [The90, side 74].
Endelig kategoriseres lav kontrol og høj støtte som den lydige arbejder. Denne kategori indeholder således job med lav handlingsfrihed og repræsenteres ved servicearbejdere, postarbejdere o.lign. Det bemærkes dog, at der er nationale forskelle mellem fordelingen af jobkategorier i modellen, da der er forskel på den sociale støtte afhængig af land. Som eksempel fremhæver Karasek, at lande og beskæftigelsesområder, hvor arbejdere i højere grad er medlem af en fagforening oplever en større social støtte.
Gennem beskrivelsen af denne sociologiske teori bekræftes det således, at den sociale støtte på arbejdspladsen og generelt forholdet mellem krav og kontrol er aktuel i forhold til problemstillingen. I det følgende afsnit bliver det fremhævet, hvorledes Karaseks model og begreber kan inddrages i forbindelse med den sociologiske analyse.

3.2 Anvendelse af teori

Den aktuelle problemstilling tager udgangspunkt i diabetikere og deres opfattelse af arbejdsmiljø. Det skal dermed undersøges, om diabetikere føler en begrænsning på arbejdspladsen på grund af sygdommen. Ved at knytte Karaseks krav-kontrol-støtte model til problemstillingen, er det muligt at begrænse mængden af variable til de, der vil være relevante fra et sociologisk synspunkt. I det følgende bliver der argumenteret for relevansen og anvendeligheden af denne model i forhold til problemstillingen.

Det skal undersøges, hvilke problemer der gør sig gældende blandt diabetikere på arbejdsmarkedet og i den forbindelse er det aktuelt at fokusere på misforhold mellem arbejdskravene til diabetikeren og dennes mulighed for at udføre arbejdet. Det er af afgørende betydning, at der er overensstemmelse mellem krav og kontrol, hvilket i praksis betyder, at diabetikeren skal have den fornødne kontrol over sit eget arbejde til at være i stand til at opfylde de arbejdskrav, der stilles. Dette indebærer, at diabetikeren kan planlægge arbejdsdagen efter dagsformen, i.e. det er muligt at tage en pause, når der er behov for det, arbejdstempoet kan sættes ned o.lign. Et arbejde med fleksible arbejdstider kan i den forbindelse være en fordel. Krav-kontrol modellen kan være med til at kategorisere de forskellige respondenters arbejde og dermed give et mål for belastningen via arbejdet. Med udgangspunkt i Karaseks model er det således muligt at opnå et mål, som kan tolkes som et mål for arbejdsmiljøet. Derudover er der flere faktorer i forhold til det fysiske arbejdsmiljø, som på sigt kan påvirke det psykiske arbejdsmiljø. De følgende eksempler kan umiddelbart virke som bagateller, men de er alle medvirkende til at marginalisere diabetikeren på arbejdspladsen og dermed skabe et skel mellem diabetikeren og de øvrige ansatte. For en diabetiker har sund mad naturligvis en afgørende rolle i hverdagen, hvorfor det er vigtigt, at der findes sunde alternativer når der serveres kage i forbindelse med fødselsdage el.lign. Ligeledes kan usund kantinemad tvinge diabetikeren til at medbringe egen mad. Endvidere har det fysiske arbejdsmiljø betydning i forhold til, om der findes faciliteter til at foretage blodsukkermålinger og insulinbehandlinger. Der er selvsagt stor forskel på om der stilles særlige faciliteter til rådighed eller om de eneste muligheder er et åbent kontorlandskab og toilettet. De her nævnte forhold kan alle medvirke til at få diabetikeren til at føle sig anderledes og uden for det normale, sociale samvær.
Den sociale støtte fra kolleger og ledelse har dermed stor betydning, især hvis de fysiske forhold ikke er optimale. Overordnet er det således vigtigt, at der er et generelt kendskab til og endvidere, at der er forståelse for sygdommen, i.e. at diabetikeren modtager følelsesmæssig støtte. Hvis der imidlertid ikke er forståelse blandt de nærmeste kolleger vil der følgelig ikke være social støtte og diabetikeren vil have svært ved at blive integreret på den pågældende arbejdsplads. Som følge af manglende social støtte vil diabetikeren ikke være beskyttet mod krav i samme grad som de øvrige ansatte. Derudover kan det være aktuelt i nogle situationer, at kolleger yder instrumentel støtte. Dette kan være i tilfælde, hvor diabetikeren er nødsaget til at holde pause, men hvor der ikke er tilstrækkelig kontrol med arbejdet til at udskyde opgaven. Som følge af dette kan det tænkes, at flere diabetikere har dårlig samvittighed eller skyldfølelse over for kolleger, hvilket også vil påvirke arbejdsmiljøet. Det er således muligt at inddrage Karaseks model i forhold til den pågældende problemstilling og især med hensyn til at beskrive den sociale støtte. Der er dog også et par kritikpunkter, som skal præsenteres i forbindelse med modellen.

3.3 Kritik

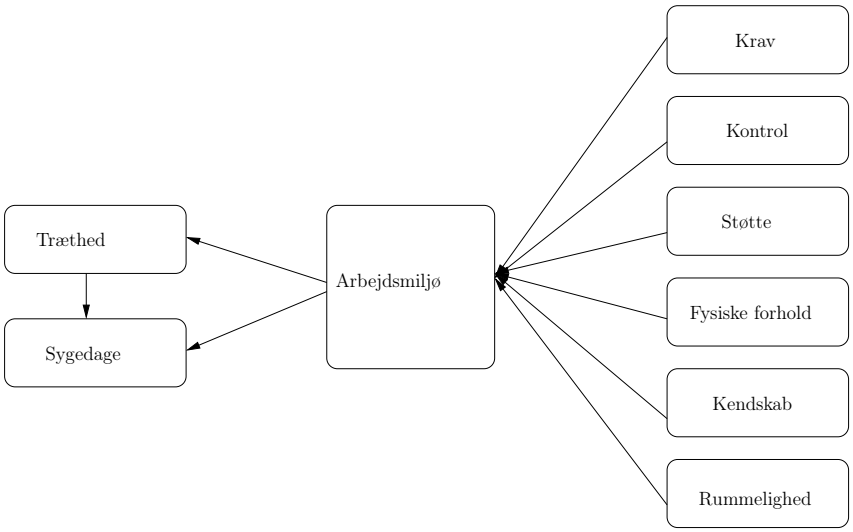
Krav-kontrol-støtte modellen bliver i dette speciale anvendt i forbindelse med at måle arbejdsmiljø, hvilket reelt er et subjektivt begreb. Det er med andre ord ikke muligt at vurdere arbejdsmiljøet i en given virksomhed, da hver ansat har forskellige ønsker og behov. I forhold til Karaseks model er det således en svaghed, at der ikke tages højde for, at de forskellige faktorer påvirker individet i forskellig grad. Nogle miljøfaktorer og stressorer har mindre betydning for nogle individer end andre. Det er dermed ikke muligt at opstille en model, der kan måle graden af arbejdsmæssig belastning uden at inddrage en forståelse af individet. Krav-kontrol-støtte modellen er således en forsimplet model, hvor individdimensionen ikke medtages. Endvidere tildeles individet ikke handlingsrum, i.e. individet har ikke mulighed for at ændre på miljøbelastningerne, hvilket er langt fra virkeligheden på de fleste arbejdspladser, hvor der er tillidsmænd, afholdes TUS-samtaler, nedsættes diverse udvalg blandt de ansatte etc. Den passivitet som Karasek tildeler individet gør sig således ikke gældende på hovedparten af de danske arbejdspladser.
Efter de to sidste kapitler, der har beskrevet henholdsvis det matematiske og det sociologiske grundlag, er det således på tide at præsentere analysemetoderne. Det følgende kapitel vil dermed beskrive de analytiske overvejelser i forhold til data og endvidere vil de programmer, der skal konstruere de grafiske modeller over data, blive præsenteret før den egentlige analyse påbegyndes.

Analysemetode

For at sikre en valid analyse er der en del overvejelser, som skal gøres, før analysen kan indledes. Ved at gøre brug af sociologisk viden skal centrale variable udvælges og heraf følger en indledende databehandling, hvor variable skal rekodes eller reduceres i dimension ved faktoranalyse. Da en stor del af analysen vil foregå via computerprogrammer vil disse kort blive præsenteret for at klarlægge hvilke algoritmer og tests, der anvendes i forbindelse med konstruktionen af grafiske modeller. Den efterfølgende del af analysen fokuserer på at fortolke modellerne, hvilket gøres via diverse test og korrelationskoefficienter som er beskrevet i Kapitel 2, samt ordinal regression, der kort præsenteres til slut i dette kapitel.

4.1 A priori model

Analysen skal opbygges på baggrund af Karaseks model, hvilket vil sige, at der skal konstrueres variable, der beskriver graden af henholdsvis krav, kontrol og støtte på arbejdspladsen. I spørgeskemaet bliver der spurgt ind til flere forskellige aspekter af de tre elementer, hvorfor svarkategorierne til de aktuelle spørgsmål vil blive omregnet til en scorevariabel. Dermed kan respondentens score på de forskellige elementer fungere som mål for graden af henholdsvis krav, kontrol og støtte. Da de fysiske forhold på arbejdspladsen ligeledes har stor betydning for arbejdsmiljøet kræver analysen også, at der konstrueres en variabel for disse. På tilsvarende vis skal der laves en variabel, der beskriver rummeligheden samt kendskabet til diabetes på arbejdspladsen. Ud fra respondentens score på de nævnte variable forventes det, at denne kan oversættes til et mål for arbejdsmiljøet. Som følge af dette forventes en sammenhæng mellem en negativ opfattelse af arbejdets betydning og graden af træthed hos respondenten. Endvidere forventes et dårligt arbejdsmiljø at give et højere sygefravær. A priori forventes således en model som angivet i Figur 4.1. Der forventes at være forskel mellem denne model for henholdsvis mænd og kvinder



Figur 4.1: A priori model

på baggrund af Karaseks studier af kønsforskelle og selvfølgelig for forskellige erhvervsmæssige stillinger jf. afsnit 3.1.2.
Der er ikke på noget tidspunkt i spørgeskemaet spurgt direkte ind til krav, kontrol, støtte eller det oplevede arbejdsmiljø. I stedet forsøges der med spørgsmålene i spørgeskemaet at afdække disse som latente faktorer. Indledningsvis foretages derfor faktoranalyse af en række spørgsmål, som menes at beskrive de latente faktorer, for derved at understøtte antagelsen. I det følgende præsenteres således ideen bag faktoranalyse; inspireret af [Max06].

4.2 Faktoranalyse

Faktoranalysens problemstilling tager udgangspunkt i, at nogle målinger er korrelerede, i.e. det forsøges med forskelligt formulerede spørgsmål at afdække samme bagvedliggende faktor. Årsagen til, at forsøge at indkredse interessefaktoren fremfor at stille et direkte spørgsmål, kan være at den bagvedliggende faktor er udefinerbar eller uhåndgribelig, og derfor bedst afdækkes ved flere relaterede spørgsmål. I forbindelse med holdninger eller andre personlige spørgsmål, kan det ligeledes være hensigtsmæssigt at stille indirekte spørgsmål for at sikre sandfærdige svar. Hvis antagelsen er rigtig, i.e. at spørgsmålene giver svar på én bagvedliggende faktor, så udgør de samlede svar naturligvis en mængde redundant information og det er i den forbindelse, at faktoranalysen skal anvendes. Via faktoranalyse bliver mængden af data reduceret i.e. antallet af målinger reduceres til et mindre antal faktorer, som mindsker mængden af redundant information. Faktoranalysen kan med andre ord betragtes som en effektivisering, hvor overflødig information fjernes og hvormed dimensionen af data nedbringes.
De variable som er præsenteret i Figur 4.1 er ikke målt direkte og skal derfor konstrueres på baggrund af spørgsmål fra spørgeskemaet via faktoranalyse. Ud fra spørgeskemaet udvælges således de spørgsmål som forventes at beskrive de latente variable i Figur 4.1, og det er disse udvalgte spørgsmål, der skal anvendes i faktoranalysen. Med denne fremgangsmåde er der tale om en faktoranalyse, hvor en a priori viden om de latente faktorer skal bekræftes, da modellen er konstrueret på baggrund af faktorer, som ifølge Karaseks model har betydning, og som er kendte. Der er således ikke tale om en decideret eksplorativ tilgang, hvor der ikke på forhånd er nogen ide om, at latente faktorer eksisterer eller hvilke faktorer der er tale om, men nærmere den omvendte situation, hvor de latente faktorer er 'kendte', og i stedet skal de rigtige spørgsmål findes i spørgeskemaet.
Faktoranalysen kan opstilles som en matrixligning

$$\mathbf{Y} = \mathbf{\Lambda f} + \varepsilon,$$

hvor $\mathbf{Y}$ er en søjlevektor, der angiver de p forskellige variable som danner grundlag for faktoranalysen. Hver variabel har en loadingværdi, der beskriver, hvor godt de k udtrukne faktorer, som er givet ved søjlevektoren $\mathbf{f}$, forklarer den pågældende variabel. Disse loadingværdier er givet ved $\mathbf{\Lambda}$, der er en $p \times k$ matrix. Det antages, at $\varepsilon \sim \mathrm{N}_p(0, \sigma^2\mathbf{I})$ og at $\mathbf{f} \sim \mathrm{N}_k(0, \mathbf{I})$.
Efter udvælgelse af de spørgsmål, også kaldet items, som forventes at beskrive en latent faktor foretages faktoranalyserne i praksis i programmet SPSS. Indledningsvis måles Kaiser-Meier-Olkin (KMO), der er et mål for korrelationen mellem de udvalgte variable og dermed et mål for, om det kan forventes, at de dækker over samme grundlæggende faktor, i.e. om det er fornuftigt at foretage en faktoranalyse. Derudover foretages Bartletts test, der tester hvorvidt kovariansmatricen, $\mathbf{\Sigma}$, er givet ved $\sigma^2\mathbf{I}$. Dette testes da det gælder, at

$$\mathrm{Cov}[\mathbf{Y}, \mathbf{Y}] = \mathrm{Var}[\mathbf{Y}]$$

og da

$$\begin{aligned}\mathrm{Var}[\mathbf{Y}] &= \mathrm{Var}[\mathbf{\Lambda f}] + \mathrm{Var}[\varepsilon] \\ &= \mathbf{\Lambda}\mathrm{Var}[\mathbf{f}]\mathbf{\Lambda}^{\mathrm{T}} + \sigma^2\mathbf{I} \\ &= \mathbf{\Lambda\Lambda}^{\mathrm{T}} + \sigma^2\mathbf{I}.\end{aligned}$$

Hvis $\mathbf{\Sigma} = \sigma^2\mathbf{I}$, i.e. hvis nulhypotesen ikke forkastes, så vil loadingværdierne være 0, og dermed vil en faktoranalyse naturligvis ikke give mening. For at gennemføre en faktoranalyse kræves det således, at Bartletts test er signifikant.
Til at udtrække de latente faktorer, $\mathbf{f}$, kan principal komponent metoden anvendes i.e. den faktor, der indeholder mest information, og dermed giver den største datareduktion, udvælges først. Dette sker på baggrund af den pågældende komponents egenværdi. Alternativt kan maximum likelihood-metoden anvendes, hvor $\mathbf{f}$ betragtes som en linearkombination af items, $\mathbf{Y}$, hvori de mest sandsynlige parametre estimeres. Forhåbentlig vil hver faktoranalyse kun returnere én komponent, da spørgsmålene netop udvælges med henblik på at afdække præcis én faktor.

Loadingværdierne, som er givet i $\mathbf{\Lambda}$, er netop de parameterestimater som ML-metoden returnerer. Fortolkningen af loadingværdierne er, at jo større loadingværdi des bedre forklarer det pågældende item den bagvedliggende faktor. I tilfælde, hvor variablene loader på flere faktorer betragtes loadingværdierne således for at afgøre, hvilken fortolkning faktorerne skal tillægges. Når faktoranalysen udtrækker flere faktorer, er det muligt at foretage en rotation, hvilket giver variablene nye loadingværdier på de udtrukne faktorer. Denne situation opstår med andre ord når $k > 1$. I disse tilfælde er $\mathbf{\Lambda}$ ikke entydig, og kan erstattes af $\mathbf{\Lambda M}$, hvor M er en ortogonal matrix. Overordnet er der to forskellige former for rotation, hvor en *ortogonal* rotation svarer til, at de pågældende faktorer er ukorrelerede. Alternativt kan der foretages en *oblique*, i.e. skævvinklet, rotation som også accepterer, at faktorerne kan være korrelerede. Begge metoder tilpasser således loadingværdierne for variablene, men hvor en ortogonal rotation er begrænset til ukorrelerede faktorer, har en oblique rotation mere frihed til at sikre henholdsvis meget høje og meget lave loadingværdier. Denne metode præsenterer netop formålet med at foretage rotation, da det er muligt at lette tolkningen af de bagvedliggende faktorer, når et item eksempelvis har en høj loading på første faktor, men lav loading på en anden. Hvis der imidlertid kun udtrækkes én faktor er det ikke muligt at foretage en rotation.
Respondenternes oprindelige svar på de udvalgte spørgsmål skal afspejles i den nye samlede faktor, hvorfor der konstrueres en faktorscore. Denne nye variabel opnås ved at foretage en lineær regression, hvor de forklarende variable er givet ved loadingværdierne for de forskellige items multipliceret med respondenternes svar på items. Med faktorscoren opnås således et mål for den latente faktor, som kan inddrages i analysen på lige fod med de øvrige variable.
Efter konstruktion af de variable som forventes at have betydning i forhold til problemstillingen foretages den egentlige analyse i.e. konstruktionen af grafiske modeller, der afspejler afhængighedsstrukturen i data, i særlige programmer. De følgende afsnit vil således beskrive de aktuelle programmer, som kan anvendes i denne forbindelse.

4.3 HUGIN

På baggrund af et givet datasæt kan programmet HUGIN generere grafiske modeller. Til dette anvender HUGIN den såkaldte PC-algoritme, som via hypotesetest afdækker betingede uafhængigheder. Algoritmen initierer med en komplet, uorienteret graf og tester derefter alle de forskellige variable. Hvis X og Y viser sig at være betinget uafhængige fjernes den pågældende kant. Søgestrategien, der er implementeret i denne algoritme, er således et eksempel på backward selektion.
Efter PC-algoritmen er en mængde kanter i den oprindelige graf blevet fjernet, således der kun er et *skelet* tilbage, som udelukkende består af uorienterede kanter. Via algoritmen opnås dermed både viden om, hvilke variable der er betinget afhængige, repræsenteret ved en kant, og viden om hvilke variable, der er betinget uafhængige givet en mængde af de øvrige variable. Der indføres herefter orienterede kanter i overensstemmelse med følgende regelsæt fra [Büh07].

1. For alle par af variable $X \nsim Y$, der har en fælles nabo Z indføres v-strukturen $X \to Z \leftarrow Y$.
2. Kanten mellem X og Y orienteres som $X \to Y$ når der findes en orienteret kant $Z \to X$ således $Z \nsim Y$.
3. Kanten mellem X og Y orienteres som $X \to Y$ når der findes en kæde $X \to Z \to Y$.
4. Kanten mellem X og Y orienteres som $X \to Y$ når der findes to kæder $X \sim Z \to Y$ og $X \sim W \to Y$ således $Z \nsim W$.
5. Kanten mellem X og Y orienteres som $X \to Y$ når der findes to kæder $X \sim Z \to W$ og $Z \to W \to Y$ således $X \nsim W$.

Reglerne er hierarkiske i den forstand, at Regel 2 kun bliver aktuel, hvis Regel 1 ikke kan opfyldes og tilsvarende gør sig gældende for de øvrige regler. Hvis ingen af ovenstående regler kan anvendes orienteres

de resterende kanter tilfældigt, dog uden at bryde de eksisterende regler. Ved denne fremgangsmåde konstrueres en DAG, som netop er en grafisk model for data. Denne DAG er dog ikke nødvendigvis entydigt bestemt, hvorfor det er hensigtsmæssigt at inddrage sociologisk viden og teori for dermed at konstruere og vælge den model, der ud fra et sociologisk synspunkt bedst beskriver data.
En videreudvikling af PC-algoritmen, kaldet NPC-algoritmen, kan ligeledes anvendes i HUGIN. Denne algoritme udnytter i testsekvensen, at det kun er nødvendigt at teste betinget uafhængighed mellem X og Y givet de variable, der er indeholdt i stien mellem X og Y. Dermed kan NPC-algoritmen betragtes som en effektiviseret udgave af den oprindelige PC-algoritme.
Det er endvidere muligt at tilføje en viden til nettet af variable, som HUGIN genererer via NPC-algoritmen, a priori. En forhåndsviden om relationer mellem variable kan med andre ord indføres ved at fiksere de pågældende kanter; endvidere kan relationens retning defineres ved en orienteret kant. Hvis det vurderes, at en del af kanterne fra Figur 4.1 er gældende på baggrund af Karaseks model, uagtet resultaterne af de statistiske test i algoritmen, kan disse således fikseres.
Ved at anvende HUGIN kan a priori viden udnyttes i analysen, hvilket er en klar fordel, men HUGIN har imidlertid en anden ulempe som er væsentlig i forhold til de aktuelle data. HUGIN er ikke i stand til at udnytte ordinaliteten i variablene. Dette følger af, at HUGIN anvender χ-test og ikke et rangtest som eksempelvis Wilcoxon eller associationsmål som γ eller τ.
I stedet for programmet HUGIN kan DIGRAM eller MIM anvendes, der ligesom HUGIN konstruerer grafiske modeller, men som netop udnytter ordinalitet i variablene. Følgende afsnit er inspireret af [Kre03].

4.4 DIGRAM og MIM

DIGRAM anvender γ i forbindelse med analyse af ordinale data, hvor der estimeres en p-værdi for den pågældende γ og på denne måde undersøges, om der skal være en kant mellem to variable. Ligesom det var tilfældet med HUGIN er det muligt at fiksere kanter mellem udvalgte variable. Endvidere kan det forhindres, at der tilføjes en kant mellem to variable, hvilket ligeledes er en fordel hvis en a priori viden kan afvise denne sammenhæng.
Modelselektionen i DIGRAM er imidlertid mere krævende end i HUGIN, da søgestrategien skal fastlægges før analysen påbegyndes. Blandt selektionsprocedurerne er forskellige udgaver af stepwise selektion, hvor backward er default søgestrategi. Ideen bag denne tilgang til modelselektion er, at en model skal konstrueres på baggrund af velovervejede argumenter, da det antages, at der altid vil være en a priori viden som kan udnyttes. Efter hvert trin i søgeproceduren kræves det således, at der træffes en beslutning om inklusion, eksklusion, fiksering el.lign, før søgeproceduren kan fortsætte. På denne måde reduceres antallet af tilfældige kanter, hvilket er en fordel i sammenligning med HUGIN, men til gengæld bliver modelselektionen, især for et større antal variable, en længerevarende proces med DIGRAM. Endvidere arbejder DIGRAM med kædegrafer, jf. afsnit 2.6.2, og forudsætter derfor, at variablene inddeles i forskellige blokke i overensstemmelse med en tidsdimension. Før analysen skal samtlige variable således rangeres i henhold til denne tidsdimension, hvilket ligeledes kan være en stor opgave.
Endelig kan programmet MIM anvendes. Fordelen ved dette program er, at der både er mulighed for at lade MIM konstruere en grafisk model på baggrund af hypotesetest, ligesom det var tilfældet med HUGIN, og mulighed for at påvirke modelkonstruktionen ved at tilføje eller slette kanter uanset signifikans. MIM kan ligeledes udnytte ordinalitet i variablene hvis det specificeres, og kræver ikke, men kan også behandle, kædegrafer. Ved at anvende MIM er det således muligt at påvirke modelkonstruktionen i det omfang det ønskes.
Efter konstruktion af de grafiske modeller skakl udvalgte kanter undersøges nærmere for at afgøre eksempelvis styrken af den pågældende sammenhæng. Dette gøres generelt ved ordinal regression som beskrives i følgende.

4.5 Ordinal regression

I tilfælde hvor mange variable er associerede er det ikke hensigtsmæssigt at beregne en korrelationskoefficient både på grund af besværligheder i forbindelse med fortolkningen men også, da der vil være ganske få observationer når flere variable skal samles i en krydstabel, hvilket gør eksempelvis γ-koefficienten usikker. I disse situationer er det således en fordel at anvende regression, da de enkelte variables påvirkning og betydning for den afhængige variabel dermed estimeres. Da størstedelen af variablene i denne analyse er ordinalskalerede, og da dette ønskes udnyttet, anvendes ordinal regression.
Regressionsligningen er i dette tilfælde givet ved

$$\text{link}(p_j) = \alpha_j - (\beta_1 X_1 + \cdots + \beta_n X_n), \tag{4.1}$$

hvor $p_j = P(Y \leq j)$ for $j = 1, \ldots, J-1$. I ovenstående angiver link$(\cdot)$ linkfunktionen og α_j er threshold for det pågældende niveau, j, af den afhængige, ordinalskalerede variabel. Hvert niveau af responsvariablen, Y, har således en threshold-værdi svarende til intercept i en almindelig logistisk regression.
I udgangspunktet anvendes logit-transformationen som linkfunktion, hvilken er givet ved

$$\ln\left(\frac{p_j}{1-p_j}\right).$$

Parameterestimaterne, β_i, er således logit-transformerede og for at opnå den prædikterede sandsynlighed for det pågældende niveau af den afhængige variabel skal link(p_j) efterfølgende transformeres via den inverse logit

$$\frac{\exp(\eta_j)}{1+\exp(\eta_j)}, \text{ hvor } \eta_j = \alpha_j - (\beta_1 X_1 + \cdots + \beta_n X_n).$$

Den sandsynlighed som herved beregnes er den kumulerede sandsynlighed for niveau $1, \ldots, j$ af den afhængige variabel. Sandsynligheden for et specifikt niveau kan efterfølgende beregnes som

$$\begin{aligned} P(Y = j) &= P(Y \leq j) - P(Y \leq j-1) \\ &= \frac{\exp(\eta_j)}{1+\exp(\eta_j)} - \frac{\exp(\eta_{j-1})}{1+\exp(\eta_{j-1})} \end{aligned}$$

Der gælder naturligvis, at $P(Y \leq J-1) = 1$, da $J-1$ er det sidste niveau af Y.
Ordinal regression forudsætter, som (4.1) også viser, at effekten af de forklarende variable er uafhængig af niveauet af den afhængige variabel i.e., at effekten er den samme for de forskellige niveauer af Y. Dette er imidlertid ikke altid tilfældet, hvorfor der skal foretages et test af parallelitet i forbindelse med regressionen. Hvis effekterne er de samme, vil dette resultere i parallelle regressionslinjer/-planer. Nulhypotesen, at $\beta_1 = \cdots = \beta_n$, testes således mod den alternative hypotese, at mindst én parameter er forskellig fra de øvrige. Det er således dette test, der refereres til, når begrebet parallelitetstest nævnes i analysen.
Hvis nulhypotesen forkastes er det i stedet muligt, hvis $J-1 = 3$, at foretage to logistiske regressioner, hvor de forklarende variable sammenlignes med hensyn til signifikans og retning. Herved kan det afdækkes hvilke variable, der ikke har samme effekt for de tre niveauer.
Regressionen foretages i SPSS, hvor både parallelitetsantagelsen og regressionsmodellens fit i forhold til data testes. Derudover beskrives andelen af varians, som modellen kan forklare, ved Nagelkerkes R^2, der tilhører intervallet $[0, 1]$. Det er på baggrund af disse tre output, at validiteten af regressionen vurderes og dermed validiteten af de tilhørende fortolkninger.

I næste kapitel foretages analysen, hvor de grafiske modeller konstrueres via de forskellige programmer og efterfølgende sammenlignes. Analysen indledes, som beskrevet, med faktoranalyser for derved at opnå de aktuelle variable som skal udgøre grundlaget for modellerne. Derefter undersøges modellerne nærmere ved både ordinal regression og beregning af korrelationskoefficienter for derved at blive i stand til at klarlægge styrke og retning af de forskellige sammenhænge.

Analyse

Analysen opdeles i to dele som tilsammen udgør en fuld analyse af problemstillingen. Faktoranalysen, som blev nævnt i foregående kapitel, udgør grundlaget for at konstruere de forskellige variable, som er nødvendige i forhold til a priori modellen. Faktoranalysen foretages således med en god portion forhåndsviden og generelle antagelser om, hvilke spørgsmål, der har forklaringskraft i forhold til de variable som skal afdækkes. I tilfælde af, at en faktoranalyse ikke kan gennemføres tages der i hvert tilfælde stilling til, hvordan dette problem bedst løses.
Efter de indledende faktoranalyser er datagrundlaget for de efterfølgende grafiske modeller skabt, og anden del af analysen vil således fokusere på konstruktionen og analysen af grafiske modeller. Analysen fokuserer på fortolkning af de udvalgte grafiske modeller, hvor interessante sammenhænge undersøges nærmere via regression, beregning af korrelationskoefficienter og sociologisk viden. I forbindelse med analysen vil der løbende blive inddraget citater fra respondenterne, da disse kan være med til at give en dybere forståelse og dermed bedre fortolkning af analysens resultater.

5.1 Indledende faktoranalyser

De variable som skal anvendes i forbindelse med faktoranalyse udvælges på baggrund af spørgeskemaet. De aktuelle variable rekodes før faktoranalysen for at sikre, at svarkategorierne er i overensstemmelse. Endvidere foretages indledningsvis en missing value analyse, hvor eventuelle missings erstattes på baggrund af EM-algoritmen. Denne algoritme er en iterativ procedure, hvor det komplette data, $\mathbf{Y}$, i.e. de observerede data, $\mathbf{Y}_{\text{obs}}$, inklusiv missings, $\mathbf{Y}_{\text{mis}}$, estimeres via maksimering af loglikelihoodfunktionen for det komplette data. Indledningsvis gættes således på værdierne af de manglende data og derefter beregnes $\mathrm{E}[l(\theta|\mathbf{Y})]$ og denne maksimeres efterfølgende. Den værdi $\tilde{\theta}$ af θ, der maksimerer $\mathrm{E}[l(\theta|\mathbf{Y})]$ anvendes derefter som ny værdi og sådan fortsætter algoritmen indtil der er opnået konvergens [Lee04, Afsnit 9.2]. Efter erstatningen af missings foretages faktoranalyserne, som beskrives i det følgende.

5.1.1 Krav-variablen

På baggrund af svar på nedenstående spørgsmål foretages en faktoranalyse, der skal afdække et bagvedliggende mål for graden af krav på arbejdspladsen.

- Er dit arbejde ujævnt fordelt, så det hober sig op?
- Skal du overskue mange ting på én gang i dit arbejde?
- Er det nødvendigt at arbejde meget hurtigt?
- Kræver dit arbejde en høj grad af kunnen eller færdigheder?
- Kræver dit arbejde, at du tager hurtige beslutninger?
- Kræver dit arbejde, at du husker meget?

Disse spørgsmål er udvalgt, da de menes at repræsentere en række problematikker i forhold til de krav, der stilles på arbejdspladsen, som kan have indflydelse på det oplevede arbejdsmiljø. De ovenstående spørgsmål afdækker både krav i form af krav til evner for udførelse af arbejdet og i form af krav til mængden af arbejde. I spørgeskemaet er der angivet fem svarmuligheder, der rekodes til tre kategorier. Således samles 'Altid' og 'Ofte' i én kategori, 'Sommetider' forbliver uændret, mens 'Sjældent' og 'Aldrig' ligeledes samles. I forbindelse med fortolkningen af resultaterne fra faktoranalysen er det vigtigt at bemærke, at udsagnene i dette tilfælde er negative.
Det første output fra SPSS viser, at det er fornuftigt at foretage en faktoranalyse af de nævnte items,

da der er korrelation mellem disse, hvorfor der er en mængde redundant, eller overlappende, information som kan udnyttes. Korrelationen måles ved KMO som generelt skal være minimum 0.7 for at opnå en passende datareduktion. Endvidere er Bartletts test signifikant.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,800
Bartlett's Test of Sphericity	Approx. Chi-Square	641,440
	df	15,000
	Sig.	,000

Figur 5.1: KMO for bagvedliggende faktor

Det næste output viser den faktor, som er udtrukket på baggrund af ML-metoden. Det ses således, at der er én bagvedliggende faktor, hvor variablene har en forholdsvis høj loading. Dermed skal alle variablene danne grundlag for fortolkningen af den latente faktor, hvor spørgsmålet om, hvorvidt der er mange ting, der skal overskues på én gang skal tillægges den største forklaringskraft. I og med ovenstående spørgsmål

Factor Matrix[a]

	Factor
	1
Er dit arbejde ujævnt fordelt, så det hober sig op?	,404
Skal du overskue mange ting på én gang i dit arbejde?	,726
Er det nødvendigt at arbejde meget hurtigt?	,479
Kræver dit arbejde en høj grad af kunnen eller færdigheder?	,527
Kræver dit arbejde, at du tager hurtige beslutninger?	,672
Kræver dit arbejde, at du husker meget?	,574

Extraction Method: Maximum Likelihood.

a. 1 factors extracted. 3 iterations required.

Figur 5.2: Loading på den bagvedliggende faktor

fra spørgeskemaet netop er udvalgt med henblik på at afdække et mål for krav på arbejdspladsen, giver det mening, med udgangspunkt i faktoranalysen, at fortolke den udtrukne faktor som et mål for de krav der stilles.
For at kunne vurdere graden af krav for hver enkelt respondent estimeres, som nævnt i Kapitel 4, en faktorscore. For hvert item i analysen estimeres en scorekoefficient som multipliceres med værdien på den aktuelle svarkategori, hvormed faktorscoren opnås.

Factor Score Coefficient Matrix

	Factor
	1
Er dit arbejde ujævnt fordelt, så det hober sig op?	,128
Skal du overskue mange ting på én gang i dit arbejde?	,407
Er det nødvendigt at arbejde meget hurtigt?	,165
Kræver dit arbejde en høj grad af kunnen eller færdigheder?	,193
Kræver dit arbejde, at du tager hurtige beslutninger?	,324
Kræver dit arbejde, at du husker meget?	,226

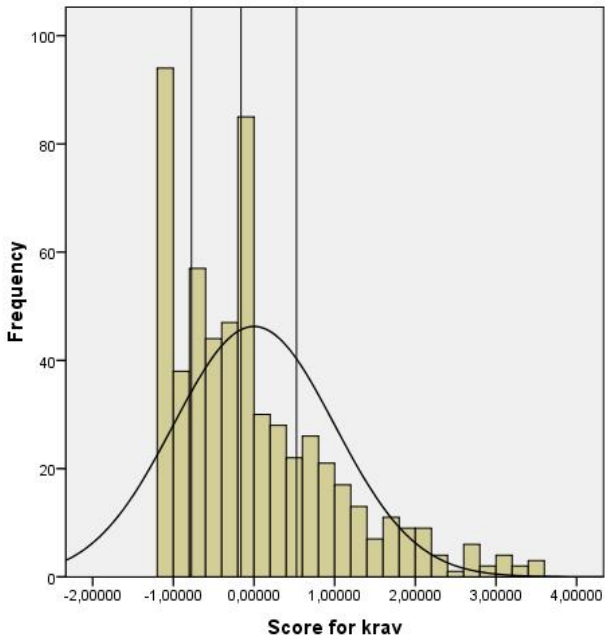
Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.
Factor Scores Method: Anderson-Rubin.

Figur 5.3: Scorekoefficienter for items

I denne analyse, hvor spørgsmålene er negativt formulerede, vil en respondent have en lav faktorscore, hvis respondenten hovedsagligt har svaret 'Ofte'. Med andre ord er en lav faktorscore udtryk for generelt høje krav på arbejdspladsen. For respondenterne kan de høje krav være et problem i forhold til deres egenomsorg, hvilket nedenstående citat viser:

> *"Får næsten aldrig målt blodsukker på arbejdet - bl.a. pga tid & forhold. Nogen gange kan jeg heller ikke nå at tage insulin [...]"* Respondent 110.

Med dette udsagn står det klart, at det kan være problematisk at være erhvervsaktiv diabetiker og at det særligt for denne gruppe, kan gå ud over helbredet, hvis der stilles for store krav til eksempelvis arbejdsmængde og -tempo.
Faktorscoren som SPSS konstruerer fordeler sig, som det fremgår af Figur 5.4, med mere end 50% af observationerne under scoreværdien 0. Fordelingen er tydeligvis højreskæv, hvilket indikerer, at der er forholdsvis mange respondenter, der får en lav score, mens antallet af respondenter med en højere score falder støt.



Figur 5.4: Fordeling af krav-variablen med angivelse af kvartiler og normalkurve

Det er nødvendigt at inddele denne nye variabel i passende kategorier, der angiver graden af krav.

Krav-variablen, som den er givet ved faktorscoren, er på en fininddelt skala, som ikke er hensigtsmæssig at anvende i praksis. I stedet skal variablen have en grovere inddeling for dermed at lette fortolkningen. En mulighed er at lade middelværdien være cutpoint, hvorved der opnås en dikotom variabel, som beskriver henholdsvis høje og lave krav, men det vil imidlertid være mere hensigtsmæssigt at inddele i tre kategorier, da det dermed bliver muligt at udnytte ordinalitet også i krav-variablen. Derudover vil den midterste kategori være en neutral kategori, hvor mængden af krav kan tolkes som værende tilpas. De tre nye intervaller konstrueres ved at benytte øvre og nedre kvartil for fordelingen af scoren som cutpoints, i.e. første interval svarer til de første 25%, det midterste interval indeholder 50% og det sidste indeholder de resterende 25% af observationerne svarende til et cutpoint ved 75%.

5.1.2 Kontrol-variablen

Graden af kontrol over arbejdet er den anden af Karaseks hoveddimensioner. En række spørgsmål, som menes at afdække denne faktor, er udvalgt fra spørgeskemaet og præsenteres her.

- Har du stor indflydelse på beslutninger om dit arbejde?
- Er dit arbejde varieret?
- Har du indflydelse på, hvordan du gør dit arbejde?
- Har du indflydelse på, hvad du laver på dit arbejde?
- Har du indflydelse på mængden af dit arbejde?
- Har du indflydelse på dit arbejdsmiljø?
- Kan du holde pause, når du har behov for det?
- Kan du spise mellemmåltider, når du har behov for det?

Disse spørgsmål omhandler respondenternes indflydelse på flere aspekter af deres arbejde herunder blandt andet udførelse og mængde. Endvidere er der medtaget to spørgsmål, som ikke direkte henvender sig til respondenternes arbejdsopgaver, men som tænkes at afdække et andet aspekt af respondenternes kontrol med deres arbejde; nemlig friheden til at planlægge arbejdet efter deres fysiske behov og begrænsninger. Ovenstående items er ikke målt på samme skala, og rekodes derfor således de alle har tre kategorier. Modsat forrige faktoranalyse er udsagnene i dette tilfælde positive, hvilket betyder, at den eventuelle faktorscore vil være høj, hvis en respondent generelt har svaret 'Sjældent' på de forskellige items, og dermed vil en høj faktorscore indikere, at respondenten ikke har kontrol.
Det første output fra SPSS viser igen, at det er fornuftigt at foretage en faktoranalyse på de udvalgte variable.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,788
Bartlett's Test of Sphericity	Approx. Chi-Square	1438,671
	df	28,000
	Sig.	,000

Figur 5.5: KMO for bagvedliggende faktor

Analysen returnerer to faktorer som hver især kan forklare nogle af de pågældende items. Da der udtrækkes flere faktorer, er det muligt at rotere den første umiddelbare 'løsning'. Der foretages derfor en oblique rotation, som giver flere muligheder for forskellige loadingværdier på de bagvedliggende faktorer. Analysen viser, at der er to forskellige faktorer, som muligvis er korrelerede som følge af

rotationsmetoden, som kan siges at forklare respondenternes besvarelser på de udvalgte items. Som nævnt indledningsvis adskiller de to sidste items sig fra de øvrige, hvilket kan være forklaringen på, hvorfor disse loader på en anden faktor. I forhold til at give en fortolkning af de to faktorer fokuseres på de højeste loadingværdier for hver faktor. På denne baggrund vurderes den første faktor at være et mål for indflydelsen på udførelsen af arbejdet i.e kontrol med de reelle arbejdsopgaver. Den anden faktor skal udelukkende fortolkes med udgangspunkt i de to sidste spørgsmål, og vurderes dermed at være et mål for graden af kontrol til at opfylde fysiske behov når de opstår.

Pattern Matrix[a]

	Factor	
	1	2
Har du stor indflydelse på beslutninger om dit arbejde?	,720	
Er dit arbejde varieret?	,510	
Har du indflydelse på hvordan du gør dit arbejde?	,782	
Har du indflydelse på hvad du laver på dit arbejde?	,816	
Har du indflydelse på mængden af dit arbejde?	,416	
Har du indflydelse på dit arbejdsmiljø?	,426	
Kan du holde pauser, når du har behov for det?		,901
Kan du spise mellemmåltider, når du har behov for det?		,755

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Figur 5.6: Loading på de bagvedliggende faktorer

Med to faktorer konstrueres ligeledes to faktorscores, som således beskriver henholdsvis kontrol med arbejdsopgaver og kontrol med opfyldelse af behov. Scorekoefficienterne er vist i Figur 5.7 og på baggrund af disse konstrueres to faktorscores for hver respondent. De to scores inddeles i passende mindre

Factor Score Coefficient Matrix

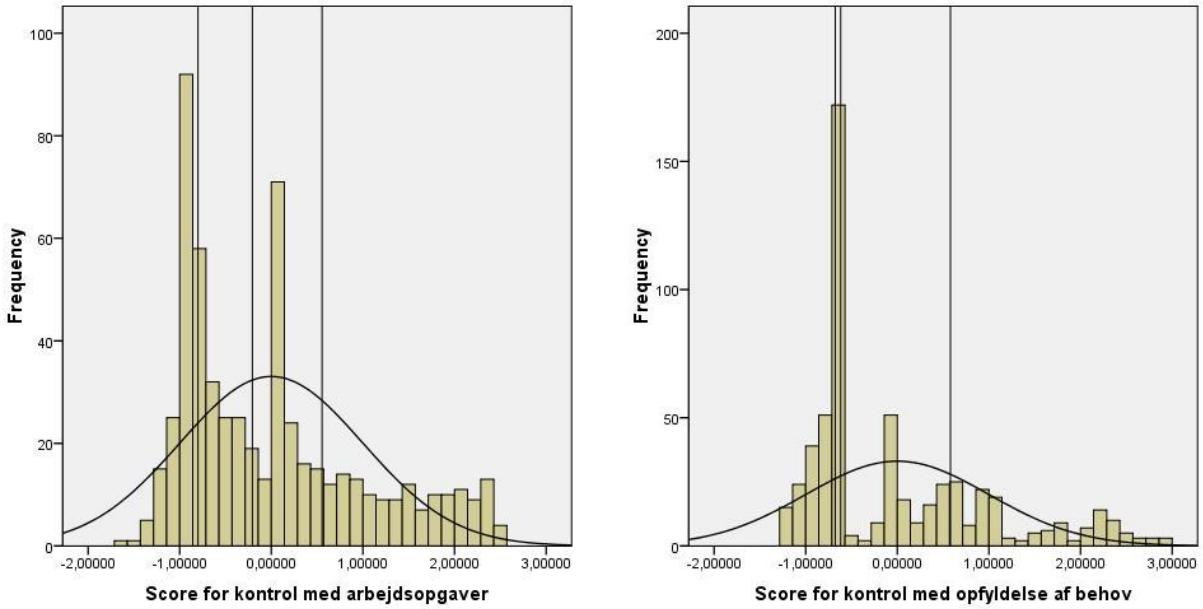
	Factor	
	1	2
Har du stor indflydelse på beslutninger om dit arbejde?	,270	-,051
Er dit arbejde varieret?	,125	-,023
Har du indflydelse på hvordan du gør dit arbejde?	,392	-,053
Har du indflydelse på hvad du laver på dit arbejde?	,396	-,105
Har du indflydelse på mængden af dit arbejde?	,092	-,012
Har du indflydelse på dit arbejdsmiljø?	,097	,006
Kan du holde pauser, når du har behov for det?	-,156	,818
Kan du spise mellemmåltider, når du har behov for det?	-,052	,298

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.
Factor Scores Method: Anderson-Rubin.

Figur 5.7: Scorekoefficienter for items

intervaller, der beskriver henholdsvis høj grad af kontrol, lav grad af kontrol og en midterkategori med

tilpas kontrol. Inddelingen i disse intervaller følger samme procedure som beskrevet i forrige afsnit og resulterer dermed i tre intervaller for hver score, hvor det midterste interval udgør halvdelen af observationerne. Dermed repræsenterer det første og sidste interval igen de mest ekstreme besvarelser. De to scorevariable har begge en højreskæv fordeling, hvor langt de fleste observationer ligger under middelscoreværdien 0, hvilket indikerer, at der generelt er flest respondenter, der vurderes at have høj eller tilpas kontrol. Med kendskab til de udfordringer, der kan opstå i forbindelse med diabetes er det



Figur 5.8: Fordeling af kontrol-variablene med angivelse af kvartiler og normalkurve

positivt, at størstedelen af respondenterne har kontrol med deres arbejde, og dermed ikke er tvunget til at tilsidesætte deres helbred. Et eksempel på en af de respondenter, der selv mener at have god kontrol med arbejdet, beskrives ved følgende citat:

> *"Jeg er næsten 100% selvstyrende og lever på en arbejdsplads med tillid og plads til min(e) sygdomm(e)"*
> Respondent 175.

Med resultatet af denne faktoranalyse kan kontrol-variablen nu antage to forskellige former, hvilket skal medtages i a priori modellen. I og med analysen har givet en ekstra kontrolfaktor opstår der fire ekstra typer af job i forhold til Karaseks oprindelige model, Figur 3.1. Dette illustrerer netop en af svaghederne ved Karaseks model; at der ikke er mulighed for at tage højde for forskellige typer af krav eller kontrol. En metode hvorpå dette problem kan løses er ved, at kombinere de to mål for graden af kontrol til ét samlet mål, der således angiver en 'total' kontrol. For at kunne udnytte Karaseks model vælges denne løsning selvom det kunne være interessant at undersøge, hvorvidt der er forskel på betydningen af de to typer af kontrol for det oplevede arbejdsmiljø. Det samlede mål for kontrol defineres til at betegne lav grad af kontrol, hvis dette var tilfældet for én eller begge kontrolvariable og ligeledes defineres det samlede mål til at betegne høj grad af kontrol, hvis det var tilfældet for én eller begge kontrolvariable. De resterende observationer placeres i midterkategorien, hvor graden af kontrol vurderes at være tilpas.

5.1.3 Støtte-variablen

I Karaseks udvidede model, Figur 3.2, inddrages en variabel, der beskriver den støtte som en medarbejder modtager fra arbejdspladsen. Denne variabel tænkes at blive afdækket ved en faktoranalyse på følgende items.

- Får du på din arbejdsplads information om vigtige beslutninger, ændringer og fremtidsplaner i god tid?

- Bliver dit arbejde anerkendt og påskønnet af ledelsen?
- Bliver du respekteret af ledelsen på din arbejdsplads?
- Får du al den information du behøver, for at klare dit arbejde godt?
- Bliver du behandlet retfærdigt på din arbejdsplads?
- Bliver ansatte med helbredsproblemer godt behandlet af virksomheden?

Disse spørgsmål omhandler alle den støtte som den enkelte respondent oplever både på det faglige og det personlige plan. I disse items genkendes endvidere både den følelsesmæssige og den informationsmæssige støtte som blev beskrevet tidligere, hvorfor det forstærker forventningen om, at en faktoranalyse er fornuftig, hvilket bekræftes ved KMO og Bartletts test. Faktoranalysen viser, at der netop er én faktor

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,843
Bartlett's Test of Sphericity	Approx. Chi-Square	1091,177
	df	15,000
	Sig.	,000

Figur 5.9: KMO for bagvedliggende faktor

som alle de aktuelle items loader forholdsvis højt på. De items der skal tillægges mest værdi i forhold til fortolkning af den udtrukne faktor, er spørgmålene om henholdsvis anerkendelse og respekt, hvilket afspejler, at den følelsesmæssige støtte udgør det vigtigste aspekt i forhold til den bagvedliggende variabel. Den nye variabel der dannes på baggrund af faktoranalysen, kan dermed tolkes som et mål for den følelsesmæssige støtte, som respondenterne får fra arbejdspladsen.
Udsagnene i denne analyse er alle positive. Dermed vil en lav score betyde, at respondenten generelt har svaret 'Ofte', hvilket vil sige, at der modtages støtte fra arbejdspladsen. Graden af støtte vurderes

Factor Matrix[a]

	Factor
	1
Får du på din arbejdsplads information om vigtige beslutninger?	,662
Bliver dit arbejde anerkendt og påskønnet af ledelsen?	,733
Bliver du respekteret af ledelsen på din arbejdsplads?	,753
Får du al den information du behøver for at klare dit arbejde godt?	,666
Bliver du behandlet retfærdigt på din arbejdsplads?	,571
Bliver ansatte med helbredsproblemer godt behandlet af virksomheden?	,583

Extraction Method: Maximum Likelihood.

a. 1 factors extracted. 3 iterations required.

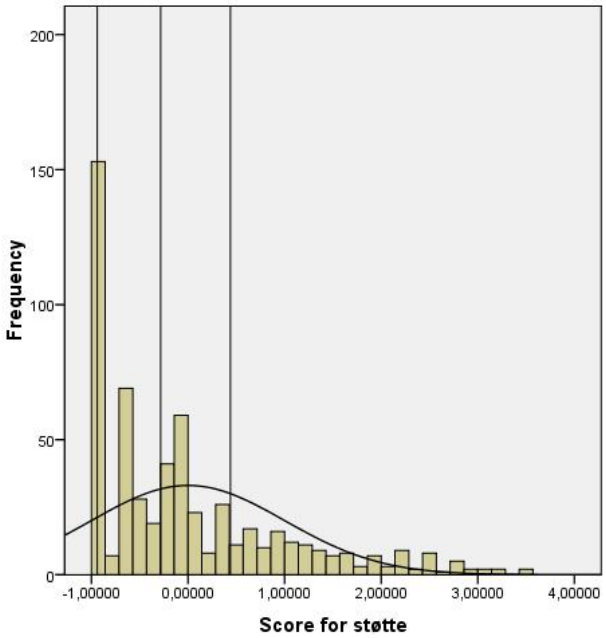
Factor Score Coefficient Matrix

	Factor
	1
Får du på din arbejdsplads information om vigtige beslutninger?	,213
Bliver dit arbejde anerkendt og påskønnet af ledelsen?	,287
Bliver du respekteret af ledelsen på din arbejdsplads?	,315
Får du al den information du behøver for at klare dit arbejde godt?	,216
Bliver du behandlet retfærdigt på din arbejdsplads?	,153
Bliver ansatte med helbredsproblemer godt behandlet af virksomheden?	,160

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.
Factor Scores Method: Anderson-Rubin.

Figur 5.10: Loadingværdier og scorekoefficienter for items

nu ud fra faktorscoren som opdeles i tre intervaller, der beskriver henholdsvis høj og lav grad af støtte samt den neutrale midterkategori 'Nogenlunde støtte'.



Figur 5.11: Fordeling af støtte-variablen med angivelse af kvartiler og normalkurve

Scoren fordeler sig med mere end halvdelen af observationerne under middelværdien, hvilket giver endnu en skæv fordeling, hvor de fleste respondenter således har angivet, at modtage høj eller nogenlunde støtte. Ud af de 580 respondenter er der omkring 150, der bliver vurderet til at få en lav grad af støtte fra arbejdspladsen. De øvrige respondenter modtager nogenlunde eller høj støtte. En af respondenterne udtrykker dette ønske:

> *"At folk kender mere til diabetes. Folk snakker bare og har ikke en skid forstand på det. De tror det er lidt bare man spiser sundt og tager sin insulin [...]"* Respondent 14.

I dette tilfælde knyttes støtte på arbejdspladsen sammen med øget kendskab til diabetes og et overordnet ønske om mere forståelse fra de øvrige ansatte. Den frustration der kommer til udtryk i dette citat, viser tydeligt, at støtte på arbejdspladsen har stor betydning for trivslen, og understreger dermed vigtigheden af at inkludere en variabel, der beskriver graden af støtte i a priori modellen.
I forhold til at besvare Arbejdsspørgsmål 1 fra afsnit 1.2.1 betragtes histogrammet, hvor de tre intervaller for kategorierne er afbilledet. På baggrund af disse ses det, at intervallet for lav støtte er det bredeste, hvilket indikerer, at der er størst spredning i denne kategori, i.e. der er flere ekstreme besvarelser, end det eksempelvis er tilfældet i første og andet interval, men til gengæld er det et fåtal af respondenterne, som oplever meget lav grad af støtte. Hvorvidt denne fordeling giver be- eller afkræftende svar på spørgsmålet beror på en subjektiv vurdering fra Diabetesforeningens side, og derfor svares der ikke endeligt på Arbejdsspørgsmål 1 her.
Med konstruktionen af støtte-variablen er det nu også muligt at anvende Karaseks udvidede model i analysen. Kombinationen af kontrol og støtte er beskrevet i modellen i Figur 3.2, og det er på baggrund af de forskellige typer af job, som denne model opstiller, at arbejdsmiljøet blandt andet skal belyses. Variablene skal således danne grundlag for de typer af job, som vurderes enten at have en positiv eller negativ virkning på arbejdsmiljø og specielt forventes det, at disse variable har indflydelse på, hvorvidt respondenterne mener, at arbejdet har en positiv eller negativ indvirkning på deres diabetes.

5.1.4 Variabel for de fysiske forhold

Som det blev præsenteret i a priori modellen, tænkes de fysiske forhold ligeledes at have betydning for arbejdsmiljøet, og derfor forsøges det via faktoranalyse, at afdække en variabel, der kan beskrive dette aspekt. I spørgeskemaet er der generelt ikke blevet stillet mange spørgsmål omkring de fysiske

omgivelser, da der har været et større fokus på det psykiske arbejdsmiljø. Forhold som temperatur og støj er således ikke medtaget, hvorfor faktoranalysen i dette tilfælde kun er baseret på fire spørgsmål, der lyder som følger:

- Er du tilfreds med forholdene, hvorunder du udfører dine blodsukkermålinger?
- Er du tilfreds med forholdene, hvorunder du tager din insulin?
- Synes du, at forholdene på din arbejdsplads gør, at du måler blodsukker mindre end du burde?
- Tager udvalget af mad i kantinen hensyn til, at der er medarbejdere med diabetes?

Fremfor at foretage en faktoranalyse er det muligt at konstruere en variabel udelukkende på baggrund af de tre første spørgsmål og dermed lade disse være målet for de fysiske forhold. Dette kan forsvares i og med ovenstående spørgsmål er tilpas konkrete til, at der nærmere er tale om et direkte mål end et mål af et indirekte, latent fænomen. Før faktoranalysen er det iøvrigt vigtigt at bemærke, at det tredje spørgsmål er negativt, mens de øvrige er positive. For at faktoranalysen kan give mening skal udsagnene ensrettes i.e. svarkategorierne på det negative udsagn kodes omvendt af de øvrige. Ved at foretage denne rekodning sikres det, at et svar i 'Ofte' får samme fortolkning på alle udsagn.
Med de udvalgte items viser det sig imidlertid, at en faktoranalyse ikke er fornuftig, da der returneres følgende output.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,593
Bartlett's Test of Sphericity	Approx. Chi-Square	893,717
	df	6,000
	Sig.	,000

Figur 5.12: KMO for bagvedliggende faktor

Med så lav en KMO-værdi kan der praktisk talt ikke opnås en datareduktion. Dermed er faktoranalysen ikke hensigtsmæssig i dette tilfælde, og det vælges i stedet at konstruere en variabel for de fysiske forhold ved direkte sammenlægning af respondenternes svar på de to første variable. Årsagen til, at det kun er de to første variable, der skal belyse de fysiske forhold er, at disse er de mest konkrete og relevante for gruppen af respondenter, og da det endvidere ikke er sandsynliggjort, at der findes en bagvedliggende faktor, som kan forklare respondenternes besvarelser på alle de aktuelle udsagn, fravælges de resterende items.
Denne nye variabel inddeles i tre kategorier, der defineres som henholdsvis gode, acceptable og dårlige forhold. I kategorien 'Acceptable forhold' gemmer sig alle de respondenter, der har svaret, at 'Det er okay, men vil gerne have bedre forhold'. Som det ses af nedenstående er der meget få respondenter, der har decideret dårlige forhold. Langt størstedelen af respondenterne placerer sig i kategorien, hvor de nuværende forhold er gode. De tal som viser sig i frekvenstabellen stemmer iøvrigt også godt overens

Fysiske forhold

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Gode forhold	513	88,4	92,9	92,9
	2 Acceptable forhold	31	5,3	5,6	98,6
	3 Dårlige forhold	8	1,4	1,4	100,0
	Total	552	95,2	100,0	
Missing	System	28	4,8		
Total		580	100,0		

Figur 5.13: Frekvenstabel

med, at kun 10% af respondenterne ikke har informeret deres chef om, at de har diabetes. Det vil sige, ud af de respondenter som har oplyst om deres diabetes, har de tilsyneladende alle gode forhold på arbejdspladsen. Med hensyn til at besvare Arbejdsspørgsmål 2, "Er der gode vilkår for egenomsorg?", vurderes det således, at der generelt er gode forhold og dermed gode vilkår for egenomsorgen, og der kan dermed svares bekræftende på dette spørgsmål.

5.1.5 Variabel for rummelighed

Den problematik som blev beskrevet i forrige afsnit gør sig ligeledes gældende i forhold til variablen for rummeligheden på arbejdspladsen; nemlig at der er meget få spørgsmål, som kan belyse respondenternes opfattelse af rummelighed. De udvalgte spørgsmål er:

- Er der plads til ældre medarbejdere?
- Er der plads til ansatte med forskellige skavanker og handicaps?
- Føler din chef, at din diabetes tager så meget tid, at det går ud over arbejdet?
- Er der sunde alternativer, når der serveres kage eller slik?

Disse spørgsmål søger alle at belyse, hvorvidt den pågældende virksomhed er i stand til at tilbyde et godt arbejdsmiljø til de ansatte, som i en eller flere henseender er anderledes. Selvom det er muligt at gennemføre analysen, viser det første output desværre, at faktoranalysen er uhensigtsmæssig. Der er iøvrigt taget højde for, at tredje item er negativt, mens de øvrige items er positive. Med dette resultat

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,553
Bartlett's Test of Sphericity	Approx. Chi-Square	122,713
	df	6,000
	Sig.	,000

Figur 5.14: KMO for bagvedliggende faktor

af analysen vælges det i stedet at konstruere en variabel for rummelighed på baggrund af de første to items. Dette valg af items foretages for det første, da disse spørgsmål er mere direkte formulerede i forhold til at afdække rummeligheden og for det andet, da disse spørgsmål er de eneste som, ifølge faktoranalysen, har forklaringskraft med hensyn til en latent faktor. Efter sammenlægning og rekodning af de oprindelige variable er der dannet en variabel for rummeligheden, hvilken fordeler sig som følger på de forskellige kategorier.

Rummelighed

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Stor rummelighed	331	57,1	62,7	62,7
	2 Nogenlunde rummelighed	136	23,4	25,8	88,4
	3 Lille rummelighed	61	10,5	11,6	100,0
	Total	528	91,0	100,0	
Missing	System	52	9,0		
Total		580	100,0		

Figur 5.15: Frekvenstabel

Størstedelen af respondenterne angiver således, at der er stor rummelighed på deres arbejdsplads, mens kun 10% føler, at der er dårlig plads til medarbejdere, som eksempelvis er fysisk udfordrede.

5.1.6 Variabel for kendskabet til diabetes

Blandt respondenterne i denne undersøgelse antages det, at kendskabet til diabetes er afgørende for arbejdsmiljøet. Hvis medarbejderen mødes med forståelse og hvis det ikke er nødvendigt at give en længere forklaring omkring påvirkningen og betydningen af diabetes i forhold til arbejdsliv, så forventes respondenten at have et bedre arbejdsmiljø end en diabetiker, som befinder sig på en arbejdsplads med kun lidt kendskab til diabetes. Dette kommer også til udtryk i følgende:

> *"Jeg tror det vil hjælpe rigtig mange både med og uden diabetes at folk ved hvad det er!"* Respondent 189.

På trods af vigtigheden af en sådan variabel, er der kun stillet tre spørgsmål, der forsøger at afdække det generelle kendskab til diabetes. Disse spørgsmål er:

- Synes du, at din chef og kolleger ved tilstrækkeligt om diabetes?
- Ved dine kolleger eller chef, hvad de skal gøre, hvis du får føling?
- Ved dine kolleger eller chef, hvad de skal gøre, hvis du får insulinchok?

På baggrund af kun tre spørgsmål er det ikke fornuftigt at foretage en faktoranalyse, selvom KMO-værdien på 0.696 er acceptabel og Bartletts test endvidere er signifikant. Dette vurderes ud fra nedenstående output.

Warnings

The number of degrees of freedom (0) is not positive. Factor analysis may not be appropriate.

Figur 5.16: Advarsel fra SPSS

Antallet af frihedsgrader er i udgangspunktet seks, da de tre items hver har to frihedsgrader i henholdsvis middelværdien og variansen af deres fordeling. Faktoranalysen bruger imidlertid tre frihedsgrader på at estimere loadingværdierne og én frihedsgrad på variansen af fejlleddet, ε. Hvis der blot udtrækkes én faktor, vil der således ikke være noget problem, men i dette tilfælde udtrækkes tilsyneladende to faktorer, hvorfor der bruges yderligere to frihedsgrader på at foretage rotationen af loadingværdierne.
Som tidligere vælges det derfor at konstruere en variabel for kendskabet til diabetes manuelt. Den løsning der vælges i dette tilfælde er at lade det første item alene beskrive kendskabet til diabetes, hvilket er oplagt, da det første spørgsmål netop er formuleret direkte, og dermed er et direkte mål for kendskabet til diabetes på arbejdspladsen. De to efterfølgende spørgsmål er således uddybninger af det første og det kan derfor forsvares, at der ses bort fra disse.

5.1.7 Variabel for træthed

Den sidste variabel som skal konstrueres via faktoranalyse er en variabel, der måler træthed og udkørthed blandt respondenterne. Denne variabel forventes at blive aktuel som resultatet af, og dermed et mål for, et dårligt arbejdsmiljø. Faktoranalysen tager udgangspunkt i følgende items.

- Hvor tit har du sovet dårligt og uroligt?
- Hvor tit har du følt dig udkørt?
- Hvor tit har du været fysisk udmattet?
- Hvor tit har du været træt?
- Hvor tit har du tænkt: "Nu kan jeg ikke klare mere"?

- Hvor tit har du manglet energi og kræfter?

Med disse seks spørgsmål forventes det at opnå et mål for, hvor meget arbejdet påvirker respondenterne både mentalt og fysisk. Med valget af ovenstående items er en faktoranalyse hensigtsmæssig og der er dermed mulighed for, at de aktuelle items er udtryk for samme bagvedliggende faktor. Dette vurderes på baggrund af KMO og Bartletts test i det første output. Endvidere viser analysen, at samtlige items

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,872
Bartlett's Test of Sphericity	Approx. Chi-Square	1876,489
	df	15,000
	Sig.	,000

Figur 5.17: KMO for bagvedliggende faktor

kan føres tilbage til én faktor. Særligt tre items har høje loadingværdier på den latente faktor. Disse tre items er henholdsvis spørgsmålene om, hvor tit respondenten har følt sig udkørt, træt og manglet energi og kræfter. Dermed skal de tillægges den største forklaringskraft i forhold til en fortolkning af

Factor Matrix[a]

	Factor
	1
Hvor tit har du sovet dårligt og uroligt?	,552
Hvor tit har du følt dig udkørt?	,867
Hvor tit har du været fysisk udmattet?	,744
Hvor tit har du været træt?	,853
Hvor tit har du tænkt: "Nu kan jeg ikke klare mere"?	,599
Hvor tit har du manglet energi og kræfter?	,840

Extraction Method: Maximum Likelihood.

a. 1 factors extracted. 4 iterations required.

Figur 5.18: Loading på den bagvedliggende faktor

den udtrukne faktor og a priori antagelsen om, at der ville blive afdækket en variabel for træthed, bekræftes således. I og med der kun udtrækkes én faktor, er det ikke muligt at få andre loadingværdier på items ved at foretage en rotation. Ligesom det var tilfældet i de øvrige faktoranalyser konstrueres

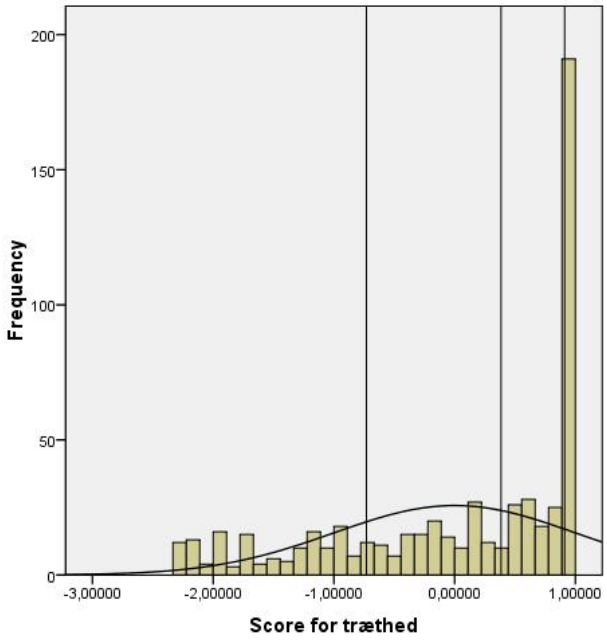
Factor Score Coefficient Matrix

	Factor
	1
Hvor tit har du sovet dårligt og uroligt?	,073
Hvor tit har du følt dig udkørt?	,323
Hvor tit har du været fysisk udmattet?	,154
Hvor tit har du været træt?	,290
Hvor tit har du tænkt: "Nu kan jeg ikke klare mere"?	,086
Hvor tit har du manglet energi og kræfter?	,263

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.
Factor Scores Method: Anderson-Rubin.

Figur 5.19: Scorekoefficienter for items

en faktorscore, der kan beskrive graden af træthed blandt respondenterne. Spørgsmålene som udgør grundlaget i analysen er alle negativt formulerede og dermed vil en lav score være udtryk for, at den pågældende respondent er meget træt.
Den faktorscore som analysen estimerer rekodes til en ny variabel med tre kategorier i overensstemmelse med den tidligere beskrevne metode. Den endelige træthedsvariabel beskriver således om respondenterne ofte, sommetider eller sjældent er trætte. Nedenstående output viser histogram for den oprindelige faktorscore. Som det ses er fordelingen af faktorscoren nogenlunde ens med undtagelse af den sidste



Figur 5.20: Fordeling af træthedsvariablen med angivelse af kvartiler og normalkurve

søjle, der er markant højere end de øvrige.
For at svare på Diabetesforeningens Arbejdsspørgsmål 4, som blev opstillet i afsnit 1.2.1, skal træthedsvariablen sammenholdes med resultater fra Det Nationale Forskningscenter for Arbejdsmiljø for at afgøre, hvorvidt der er flere blandt diabetikerne, som er udkørte end blandt den øvrige befolkning. Ved at foretage et χ^2-test er det muligt at afgøre, om der er signifikant forskel mellem de to grupper, eller om de svarer ens på de forskellige spørgsmål, i.e. at der er uafhængighed. Fra Diabetesforeningen er der således stillet data til rådighed, hvor personer, der aldrig har haft diabetes har svaret på enslydende spørgsmål. På baggrund af disse data foretages således en sammenligning af de to gruppers svar på de seks spørgsmål. I alle seks tilfælde beregnes en p-værdi, der er signifikant og dermed konkluderes det,

Udkørt		Ofte	Sommetider	Sjældent	
Diabetes	n	117	165	298	580
	%	20.2	28.4	51.4	100
Forventet værdi		88.4	158.6	333	
Ikke diabetes	n	486	917	1973	3376
	%	14.4	27.2	58.4	100
Forventet værdi		514.6	923.4	1938	
Total		603	1082	2271	3956

Figur 5.21: Krydstabel for "Hvor tit har du følt dig udkørt?"

at der er forskel i besvarelserne. Ved at betragte de aktuelle krydstabeller viser det sig, at der er flere diabetikere end forventet under nulhypotesen, som ofte føler sig udkørte, og færre, der sjældent har

dette problem. I dette tilfælde beregnes $\chi = \sum \frac{(\text{Observeret-Forventet})^2}{\text{Forventet}} = 15.54$ og med to frihedsgrader bliver p< 0.001. Følgende citat afspejler ovenstående:

"Jeg føler mit arbejde 'tapper' mig for de kræfter jeg har [...]" Respondent 292.

På trods af dette resultat er der færre diabetikere end forventet, som angiver, at de ofte har været fysisk udmattede. Til gengæld er der generelt flere, der ofte har været trætte. Tilsvarende er der flere

Træt		Ofte	Sommetider	Sjældent	
Diabetes	n	159	148	273	580
	%	27.4	25.5	47.1	100
Forventet værdi		114.9	181.5	283.7	
Ikke diabetes	n	624	1089	1661	3374
	%	18.5	32.3	49.2	100
Forventet værdi		668.1	1055.5	1650.3	
Total		783	1237	1934	3954

$\chi = 27.67 \Rightarrow p < 0.001$

Figur 5.22: Krydstabel for "Hvor tit har du været træt?"

diabetikere end forventet, som ofte har tænkt "Nu kan jeg ikke klare mere", eller har manglet energi og kræfter, som det kommer til udtryk her:

"Jeg kan ikke forstå hvorfor, der er flest dage, hvor der ingen energi er." Respondent 220.

Med disse resultater tyder det på, at diabetikere i nogle tilfælde oplever at være mere trætte og kræftesløse end den øvrige befolkning og dermed selvfølgelig også de øvrige ansatte på arbejdspladsen. Med hensyn til det første spørgsmål omkring dårlig søvn er diabetikerne tilsyneladende bedre stillet, da de generelt sjældent oplever problemer på dette område. Et svar på Arbejdsspørgsmål 4 vil således lyde, at der er forskel på besvarelserne fra henholdsvis diabetikere og den øvrige befolkning, men at der ikke konsekvent er tale om en negativ forskel for gruppen af diabetikere.

5.1.8 Øvrige variable

I forhold til a priori modellen er der stadig tre variable som skal inkluderes, men da disse er indeholdt i spørgeskemaet med direkte spørgsmål, er det ikke nødvendigt at foretage yderligere faktoranalyser. De spørgsmål som skal anvendes i a priori modellen er følgende.

- Synes du, at dit arbejde har en positiv indvirkning på din diabetes/egenomsorg?
- Synes du, at dit arbejde har en negativ indvirkning på din diabetes/egenomsorg?
- Hvor mange sygedage har du haft på dit arbejde inden for de sidste 12 måneder?

De to første variable kræver ikke en rekodning, da svarkategorierne er 'Ja', 'Delvist' og 'Nej'. Den sidste variabel inddeles i intervaller, der angiver antallet af sygedage inden for det sidste år. Ifølge [Lyn08] er der gennemsnitligt 9 sygedage pr. person i den danske befolkning. På baggrund af denne oplysning konstrueres intervallet $[6, 12]$, som tænkes at beskrive det antal sygedage som kan forventes. Frekvensen af intervallerne er præsenteret i Figur 5.23, hvor det ses, at størstedelen af respondenterne har færre sygedage end det umiddelbart forventes, mens omkring 20% befinder sig indenfor normalen. De øvrige, godt 20%, har haft flere sygedage inden for de sidste 12 måneder end det forventes.
I gennemsnit har respondenterne 13.1 sygedage hvilket umiddelbart tænkes at være et signifikant større antal end det generelle gennemsnit på 9. Ved at lægge et 95% konfidensinterval omkring respondenternes gennemsnit bekræftes dette, da intervallet $[10.3, 15.9]$ returneres. Dog er det muligt, at et tilsvarende

Antallet af sygedage på et år

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 0-5 sygedage	300	51,7	57,8	57,8
	2 6-12 sygedage	109	18,8	21,0	78,8
	3 13-19 sygedage	32	5,5	6,2	85,0
	4 20-49 sygedage	47	8,1	9,1	94,0
	5 50-99 sygedage	16	2,8	3,1	97,1
	6 100 sygedage eller mere	15	2,6	2,9	100,0
	Total	519	89,5	100,0	
Missing	System	61	10,5		
Total		580	100,0		

Figur 5.23: Frekvens af sygedage

konfidensinterval omkring det generelle gennemsnit vil overlappe, og dermed kan det ikke afvises, at de sande gennemsnit kan tænkes at være de samme. Dette er der netop forsøgt at tage hensyn til, ved at lade normalen være et interval, $[6, 12]$, omkring den generelle middelværdi. Hvis der skulle tages udgangspunkt i dette interval ville der ikke være dokumenteret en forskel mellem henholdsvis diabetikerne og den øvrige befolkning. Med dette in mente er det ikke muligt at give et bekræftende svar på Arbejdsspørgsmål 3.
Med de tre sidste variable er der redegjort for alle de variable, der skal bruges i forbindelse med a priori modellen og med udgangspunkt i disse er det muligt at konstruere en grafisk model, der således kan være med til enten at be- eller afkræfte a priori antagelserne. Før konstruktionen af de grafiske modeller undersøges dog, hvorvidt respondenternes besvarelser er uafhængige af henholdsvis køn, alder og typen af diabetes. Specielt denne sidste variabel er interessant at undersøge, da dataudtrækket som nævnt i afsnit 1.2.3 er skævt sammensat med hensyn til typen af diabetes i forhold til den normale andel. Det testes således om der er signifikant forskel i besvarelserne fra de forskellige respondenter, hvilket viser sig ikke at være tilfældet. Dermed er det ikke nødvendigt at kontrollere for eksempelvis typen af diabetes i analysen.
Efter de indledende faktoranalyser og konstruktion af variable er der nu opnået et fornuftigt datagrundlag for konstruktion af de grafiske modeller. I det følgende præsenteres således de modeller, som de tre programmer giver, med udgangspunkt i de samme data.

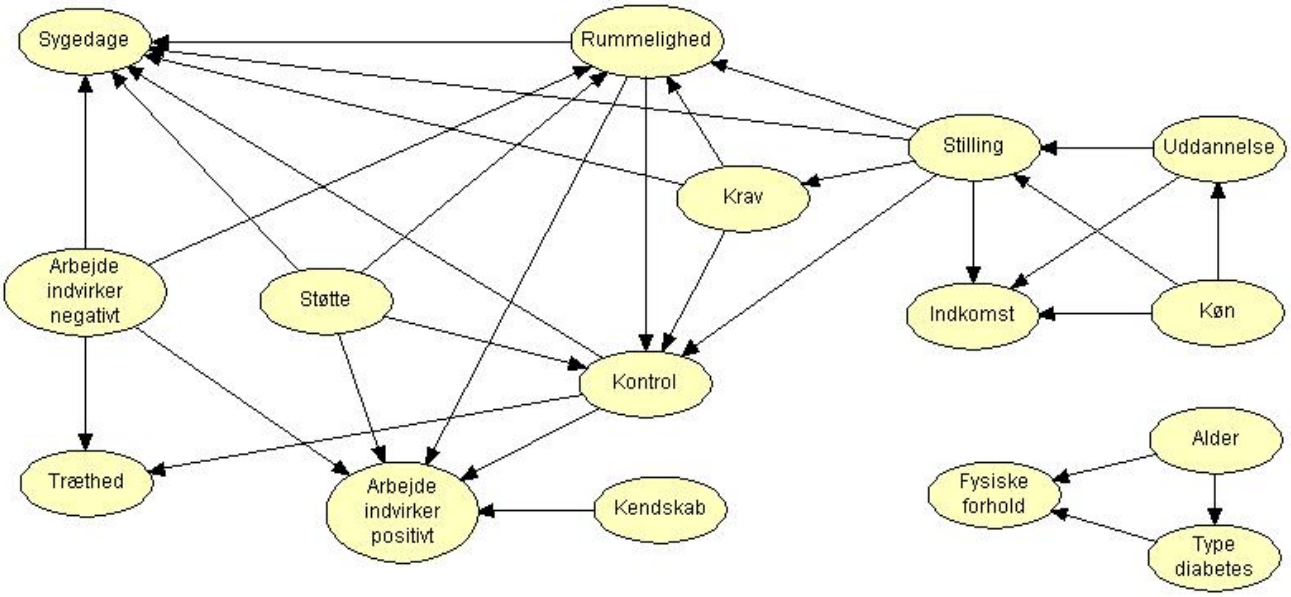
5.2 Konstruktion og analyse af grafiske modeller

I hvert af de følgende afsnit beskrives kort hvilken metode der er anvendt til konstruktion af de grafiske modeller, hvordan modellen stemmer overens med det forventede og endelig undersøges udvalgte kanter og pile med henblik på at give en fortolkning af modellen. I denne proces inddrages endvidere citater fra respondenterne.

5.2.1 HUGIN-modellen

Når data er åbnet i HUGIN, er det muligt at lægge restriktioner på den grafiske model, i forhold til hvilke kanter og pile der tillades, hvis der er en forhåndsviden om afhængigheder eller uafhængigheder i data. Derefter vælges signifikansniveauet, $\alpha = 0.01$, og den algoritme som skal anvendes i forbindelse med at teste betinget uafhængighed mellem variablene. Både PC- og NPC-algoritmen er mulige og i dette tilfælde vælges NPC jf. afsnit 4.3. HUGIN konstruerer derefter en grafisk model. Den første model som returneres er dog ikke den færdige model, da den også indeholder kanter, hvor HUGIN ikke kan afgøre, hvorvidt kanterne skal inkluderes eller ej. Det er derefter muligt at hjælpe HUGIN ved at fjerne de kanter, som ikke vurderes at have betydning. Herefter returneres den endelige model. Modellen kan efterfølgende tilpasses ved at vende 'forkert' orienterede kanter, hvis HUGIN tillader det. Når dette gøres inkluderer HUGIN automatisk de kanter som er nødvendige for at respektere regelsættet fra afsnit 4.3.

På baggrund af de aktuelle data opnås følgende model, hvor en orienteret kant fra X til Y betyder, at Y er betinget afhængig af X givet alle tidligere variable. Det er imidlertid ikke muligt ud fra figuren



Figur 5.24: Grafisk model med NPC algoritmen i HUGIN

at bestemme retning eller styrke af de forskellige associationer og det er derfor interessant at undersøge modellen nærmere for at afklare, hvilken betydning de forskellige variable skal tillægges. Til dette anvendes ordinal regression eller, hvis det er muligt, korrelationskoefficienter.
Ud fra modellen ses det, at der er en sammenhæng mellem arbejdets negative indvirkning og variablene Træthed og Sygedage, ligesom det blev forventet i a priori modellen. Til gengæld er Træthed og Sygedage betinget uafhængige givet Arbejde indvirker negativt og Kontrol. Ved at opstille en krydstabel for Træthed og Sygedage i SPSS, hvor der elaboreres for de to øvrige variable viser det sig, at der som forventet er en generelt positiv korrelation, målt ved γ, således stor træthed er associeret med flere sygedage. Ikke overraskende findes de stærkeste korrelationer blandt de respondenter, der føler, at arbejdet indvirker negativt på deres egenomsorg.
Dette viser tydeligt en af svaghederne ved HUGIN. På et tidspunkt i NPC-algoritmen er kanten mellem Sygedage og Træthed fjernet, fordi test viser, at variablene er uafhængige. Alligevel er der en association mellem de to variable, som nu ikke kan aflæses direkte af modellen. Sammenhængen opdages således kun, da netop de to variable har sociologisk interesse og endvidere kan undersøges nærmere via SPSS. Årsagen til, at HUGIN overser denne sammenhæng er, at algoritmen som tidligere nævnt ikke udnytter ordinaliteten i variablene, hvilket er et væsentligt kritikpunkt i denne kontekst.
I dette tilfælde har et dårligt arbejdsmiljø dermed betydning for, hvor stor træthed respondenterne oplever samt antallet af sygedage. Dette resultat understreger, at det er nødvendigt at afdække, hvilke faktorer der påvirker arbejdsmiljøet, således der kan organiseres en målrettet indsats på området. Det erkendes, at en del af den træthed som respondenterne oplever, ikke udelukkende er relateret til arbejdspladsen, men også i høj grad deres diabetes, hvorfor det ikke er muligt at fjerne al trætheden, men hvis det kan afdækkes hvilke variable, der har særlig betydning for denne gruppe af respondenter, så giver det mulighed for oplysning. Som en af respondenterne foreslår kunne Diabetesforeningen fremstille

> *"[...] en pjece til såvel arbejdsgiver som kolleger om, hvordan det er at have diabetes samt hvorledes en diabetiker bør behandles/taget hensyn til på arbejdspladsen."* Respondent 341.

Ud over arbejdets negative indvirkning er der en række andre variable, der er orienterede mod Sygedage. Derfor foretages en ordinal regression, hvor hovedeffekterne af de enkelte variable estimeres, når der er korrigeret for de øvrige. Blandt parameterestimaterne er det næsten udelukkende de forskellige erhvervs-

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[sygedage = 1]	1,547	,539	8,244	1	,004	,491	2,603
	[sygedage = 2]	2,615	,547	22,858	1	,000	1,543	3,686
	[sygedage = 3]	3,107	,552	31,631	1	,000	2,024	4,189
	[sygedage = 4]	4,253	,577	54,255	1	,000	3,122	5,385
	[sygedage = 5]	4,894	,607	64,899	1	,000	3,703	6,084
Location	[rummelig=1]	-,093	,310	,090	1	,764	-,701	,515
	[rummelig=2]	-,459	,334	1,893	1	,169	-1,113	,195
	[rummelig=3]	0[a]	.	.	0	.	.	.
	[krav=1]	,475	,266	3,195	1	,074	-,046	,996
	[krav=2]	,241	,241	,995	1	,319	-,232	,714
	[krav=3]	0[a]	.	.	0	.	.	.
	[kontrol=1]	-,311	,267	1,353	1	,245	-,835	,213
	[kontrol=2]	-,124	,262	,224	1	,636	-,638	,389
	[kontrol=3]	0[a]	.	.	0	.	.	.
	[støtte=2]	-,400	,225	3,160	1	,075	-,842	,041
	[støtte=3]	0[a]	.	.	0	.	.	.
	[stilling=1]	1,455	,523	7,738	1	,005	,430	2,480
	[stilling=2]	1,830	,473	14,944	1	,000	,902	2,757
	[stilling=3]	1,604	,530	9,139	1	,003	,564	2,643
	[stilling=4]	1,413	,416	11,543	1	,001	,598	2,229
	[stilling=5]	1,394	,456	9,363	1	,002	,501	2,287
	[stilling=6]	0[a]	.	.	0	.	.	.
	[negativt_arbejde=1]	,519	,386	1,809	1	,179	-,238	1,276
	[negativt_arbejde=2]	,806	,243	10,966	1	,001	,329	1,283
	[negativt_arbejde=3]	0[a]	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Figur 5.25: Ordinal regression af Sygedage

mæssige stillinger, der er signifikante og som dermed har en effekt i forhold til den afhængige variabel. Sandsynligheden for at have $0-5$ sygedage på et år, [sygedage = 1], er eksempelvis højere blandt funktionærerne, [stilling = 4], end det er tilfældet for de ufaglærte, [stilling = 2]. De grupper der har størst sandsynlighed for kun at have $0-5$ sygedage er henholdsvis selvstændige og ledende funktionærer. Dette kan forklares med udgangspunkt i Karaseks teori, hvor lederstillinger generelt kategoriseres som aktive job, jf. afsnit 3.1.2, hvor der på trods af det høje aktivitetsniveau er mindre træthed på grund af den øgede økonomiske og psykiske belønning. Det er dette, der antages at afspejle sig i den aktuelle model, blot eksemplificeret ved antallet af sygedage i stedet. Det antages med andre ord, at der er en interaktionseffekt af krav og kontrol, der har betydning for antallet af sygedage.
I forlængelse af ovenstående er det interessant at undersøge pilen fra Stilling til Krav for dermed at finde ud af hvilke stillinger, der medfører henholdsvis de højeste og de laveste krav. Denne sammenhæng kan undersøges ved blot at beregne γ for den pågældende krydstabel, da Krav $\perp$ Stilling | (Køn, Uddannelse, Indkomst). Korrelationskoefficienten bliver $0.191(p = 0.001)$, hvilket indikerer en svag, men signifikant sammenhæng, hvor respondenter i højere stillinger også skal opfylde højere krav. Dermed må respondenterne siges at opføre sig i overensstemmelse med det, på baggrund af teorien, forventede. De høje krav som generelt gør sig gældende for diverse lederstillinger medfører altså ikke flere sygedage, hvilket kan forklares via den tilsvarende høje kontrol. Analysen af de aktuelle data bekræfter med andre ord, at det i høj grad er forholdet mellem krav og kontrol, der er væsentligt, og at høje krav generelt ikke kan forbindes med et dårligt arbejdsmiljø.
Ud af de variable som det blev forventet ville have betydning for arbejdsmiljøet viser det sig, at både Kontrol, Støtte, Kendskab og Rummelighed er orienterede mod variablen Arbejde indvirker positivt, mens eksempelvis de fysiske forhold ikke har betydning i forhold til det oplevede arbejdsmiljø. Ved at foretage en ordinal regression estimeres betydningen af de forskellige variable for sandsynligheden for om arbejdet indvirker positivt eller ej. Herved viser det sig, at sandsynligheden for, at arbejdet indvirker positivt eller delvist positivt på egenomsorgen estimeres til 58%, hvis arbejdet ikke indvirker negativt, [negativt_arbejde = 3], støtten er nogenlunde, [støtte = 2], der er stor rummelighed, [rummelig = 1],

Parameter Estimates

							95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Lower Bound	Upper Bound
Threshold	[positivt_arbejde = 1]	-2,169	,401	29,183	1	,000	-2,956	-1,382
	[positivt_arbejde = 2]	,535	,388	1,902	1	,168	-,225	1,295
Location	[støtte=2]	-,154	,234	,431	1	,512	-,613	,305
	[støtte=3]	0[a]	.	.	0	.	.	.
	[rummelig=1]	-,105	,326	,104	1	,747	-,744	,534
	[rummelig=2]	,101	,348	,084	1	,772	-,582	,784
	[rummelig=3]	0[a]	.	.	0	.	.	.
	[kontrol=1]	-,306	,265	1,339	1	,247	-,825	,213
	[kontrol=2]	-,030	,271	,012	1	,911	-,562	,502
	[kontrol=3]	0[a]	.	.	0	.	.	.
	[diabetes_kendskab=1]	-,543	,278	3,814	1	,051	-1,088	,002
	[diabetes_kendskab=2]	-,312	,241	1,680	1	,195	-,785	,160
	[diabetes_kendskab=3]	0[a]	.	.	0	.	.	.
	[negativt_arbejde=1]	4,458	1,028	18,795	1	,000	2,443	6,473
	[negativt_arbejde=2]	2,041	,267	58,293	1	,000	1,517	2,565
	[negativt_arbejde=3]	0[a]	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Figur 5.26: Ordinal regression af Arbejde indvirker positivt

høj kontrol, [kontrol = 1], og godt kendskab, [diabetes_kendskab = 1]. Derimod er sandsynligheden under 2%, hvis arbejdet indvirker negativt, [negativt_arbejde = 1], støtten er lav, [støtte = 3], der er lille rummelighed, [rummelig = 3], lav kontrol, [kontrol = 3], og dårligt kendskab, [diabetes_kendskab = 3].
Dermed kan støtte, rummelighed, kontrol og kendskabet til diabetes påvirke, hvordan arbejdet indvirker på egenomsorgen, både i positiv og negativ retning, og som følge heraf også arbejdsmiljøet for respondenterne. Igen bekræftes således Karaseks teori med hensyn til variablene Støtte og Kontrol samt a priori antagelserne om betydningen af arbejdspladsens rummelighed og det generelle kendskab til diabetes. Disse fortolkninger er dog kun forsigtige bud på sammenhænge, da meget få af variablene er signifikante, hvilket tabellen i Figur 5.26 viser. Endvidere viser parallelitetstestet, at faktorerne ikke har samme effekt for de forskellige niveauer af den afhængige variabel, da p= 0.001. Ved at rekode variablen for arbejdets positive indvirkning til to dikotome variable, hvor første variabel har svarkategorierne 'Ja' og 'Delvist', mens den anden variabel har svarkategorierne 'Delvist' og 'Nej', er det muligt at foretage to almindelige logistiske regressioner, hvorved det kan undersøges hvilke variable, der har forskellig effekt. Parameterestimaterne fra nedenstående output viser, at eksempelvis negativt_arbejde har forskellig effekt i de to regressioner. I det første output er denne variabel ikke signifikant, men til gengæld viser den

Ja — **Variables in the Equation**

		B	S.E.	Wald	df	Sig.
Step 1	støtte(1)	,592	,433	1,868	1	,172
	rummelig			,420	2	,811
	rummelig(1)	,272	,605	,201	1	,654
	rummelig(2)	,086	,658	,017	1	,896
	diabetes_kendskab			8,231	2	,016
	diabetes_kendskab(1)	,539	,430	1,569	1	,210
	diabetes_kendskab(2)	-,352	,407	,751	1	,386
	kontrol			7,904	2	,019
	kontrol(1)	1,948	,756	6,633	1	,010
	kontrol(2)	2,147	,764	7,901	1	,005
	negativt_arbejde			4,079	2	,130
	negativt_arbejde(1)	-18,066	40192,970	,000	1	1,000
	negativt_arbejde(2)	-2,114	1,047	4,079	1	,043
	Constant	-3,648	,972	14,095	1	,000

Nej — **Variables in the Equation**

		B	S.E.	Wald	df	Sig.
Step 1	støtte(1)	,017	,277	,004	1	,950
	rummelig			,231	2	,891
	rummelig(1)	-,018	,376	,002	1	,962
	rummelig(2)	,104	,396	,069	1	,792
	diabetes_kendskab			5,925	2	,052
	diabetes_kendskab(1)	-,267	,330	,656	1	,418
	diabetes_kendskab(2)	-,639	,276	5,369	1	,020
	kontrol			5,053	2	,080
	kontrol(1)	,225	,311	,522	1	,470
	kontrol(2)	,647	,315	4,216	1	,040
	negativt_arbejde			59,442	2	,000
	negativt_arbejde(1)	4,476	1,035	18,706	1	,000
	negativt_arbejde(2)	1,892	,281	45,417	1	,000
	Constant	-,802	,433	3,432	1	,064

Figur 5.27: Logistisk regression af dikotome variable for Arbejde indvirker positivt

anden regression stærk signifikans. Arbejdets negative indvirkning har altså forskellig effekt på de forskellige niveauer af den afhængige variabel. Det er dog ikke overraskende, at Arbejde indvirker negativt har signifikant betydning for den sidste regression, da det netop er denne, hvor den afhængige variabel

beskriver, at arbejdet ikke indvirker positivt. Tilsvarende er kontrol-variablen signifikant i første output, men ikke signifikant i andet output, hvilket ligeledes viser, at Kontrol primært har betydning i forhold til den dikotome variabel, der beskriver, at arbejdet indvirker positivt. Ved logistisk regression viser det sig således, at flere af de forklarende variable kun har betydning i den ene ende af skalaen, mens de er uden effekt på andre niveauer af den afhængige variabel.
På baggrund af HUGIN-modellen foretages der ligeledes regression med graden af rummelighed som afhængig variabel og henholdsvis Stilling, Krav, Støtte og arbejdets negative indvirkning som faktorer. Ved at betragte forskellige erhvervsmæssige stillinger tegner der sig samme billede som tidligere; jo højere stilling des større sandsynlighed for stor rummelighed. I forhold til betydningen af krav, så viser regressionen, at midterkategorien på krav-variablen, betegnet 'Tilpasse krav', giver den største sandsynlighed for stor rummelighed, mens sandsynligheden bliver mindre for henholdsvis lave og høje krav. Dette resultat bekræfter, at både for høje og for lave krav kan have en negativ betydning, hvilket også præsenteres i Karaseks krav-kontrol model, hvor for lave krav kan føre til de såkaldte passive job, mens for høje krav kan resultere i høj anstrengelse job.
Dette er et vigtigt resultat, da virksomheder og arbejdspladser kunne forledes til at tro, at den bedste måde at tage hensyn til diabetikere er ved ikke at stille særlige krav til dem. Med denne analyse og med reference til Karaseks teori er det imidlertid vist, at det afgørende er, at mængden af krav hverken er for høje eller for lave. Dette er naturligvis en udfordring for virksomhederne, da det kræver, at arbejdet tilpasses den enkelte medarbejder, men hvis det til gengæld kunne resultere i lavere sygefravær vil det klart være en fornuftig investering.
På baggrund af HUGIN-modellen er der afdækket flere interessante sammenhænge som kort opsummeres før analyse af de modeller, som konstrueres via DIGRAM og MIM. Flere af a priori antagelserne blev bekræftet, da der eksempelvis viste sig en sammenhæng mellem træthed og antallet af sygedage med stærkest korrelation for de respondenter som følte, at arbejdet havde en negativ indvirkning på deres egenomsorg. Antallet af sygedage var endvidere interessant i forhold til respondenternes erhvervsmæssige stilling, da analysen viste, at der generelt var færre sygedage blandt respondenter i lederstillinger. Dette resultat gjorde sig gældende selvom respondenterne i de høje stillinger også var den gruppe, hvor der blev stillet de største krav. Med udgangspunkt i Karaseks teori om forholdet mellem krav og kontrol blev denne sammenhæng forklaret. Fordelen ved en høj stilling blev igen præsenteret, da regression viste, at en høj stilling havde betydning for den grad af rummelighed, som respondenten oplever på arbejdspladsen.
De variable som havde betydning for arbejdsmiljøet blev ligeledes afdækket via regression, hvor Støtte, Kontrol, Rummelighed og Kendskab til diabetes gjorde sig gældende. Analysen skulle dog suppleres med to logistiske regressioner, der viste, at effekten af disse variable var forskellig, hvilket besværliggjorde fortolkning.
Selvom modellen således danner grundlag for nogle fornuftige fortolkninger, er der også nogle svagheder ved HUGIN-modellen. Det største problem, i forhold til den efterfølgende fortolkning, er, at det ikke er muligt at vende alle de pile, som reelt burde vendes. Eksempelvis burde pilen fra Arbejde indvirker negativt til Rummelighed være orienteret modsat, da rummeligheden rent logisk er tidligere, i.e. det er rummeligheden, der afgør om arbejdet har en negativ indvirkning og ikke omvendt. Det er med andre ord svært at modellere virkeligheden via HUGIN og dette afspejler sig også i regressionsanalysen med få signifikante variable og problemer med parallelitet. Problemet med orienteringen af kanterne kan imidlertid afhjælpes i de andre programmer, da DIGRAM kræver og MIM giver mulighed for at anvende kædegrafer.

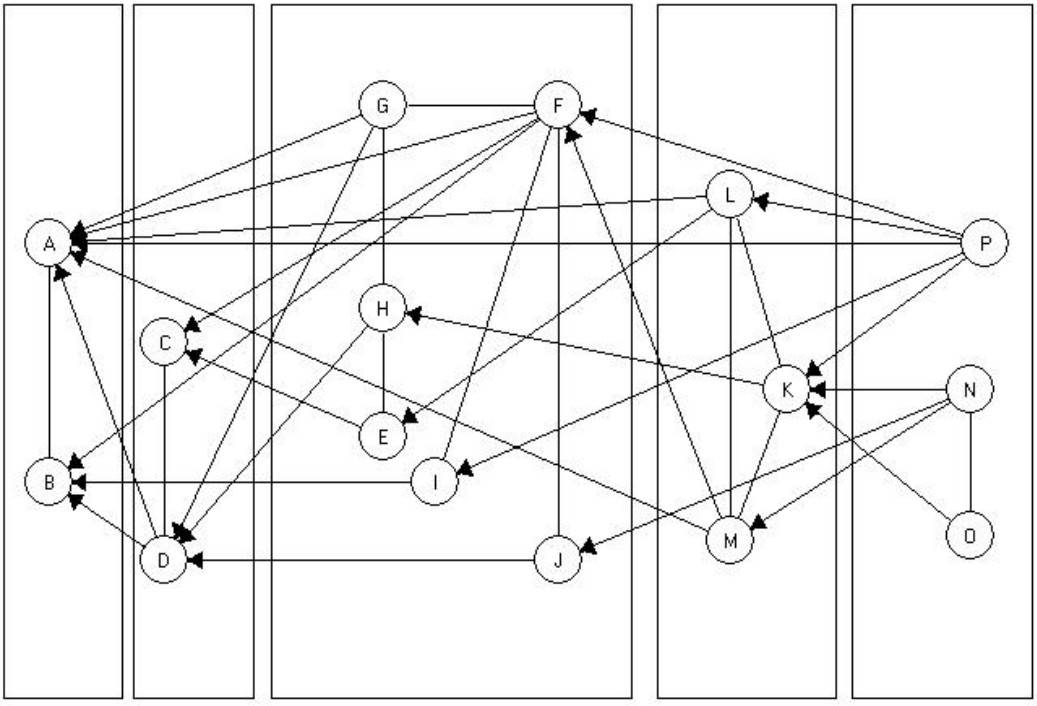
5.2.2 DIGRAM-modellen

Før konstruktion og analyse skal data naturligvis importeres til DIGRAM, hvilket kræver et særligt program. I den forbindelse specificeres den blokstruktur som DIGRAM kræver for at kunne fungere. For store datasæt vil dette selvfølgelig tage tid, men til gengæld vindes meget tid tilbage i forbindelse med analysen af modellen, da de ulogiske kanter, som eksempelvis forekommer i HUGIN, ikke opstår. Med denne model undgås således mange af de pile som var problematiske i forhold til fortolkning af

HUGIN-modellen, da blokstrukturen sikrer, at alle pile orienteres i overensstemmelse med en given tidsdimension. Med et beskedent datasæt som dette, hvor 16 variable skal organiseres, er denne indledende blokstrukturering ikke krævende. Det vælges på forhånd, at

Blok 1 = Diabetestype (N), Alder (O) og Køn (P)
Blok 2 = Erhvervsuddannelse (K), Stilling (L) og Indkomst (M)
Blok 3 = Krav (E), Kontrol (F), Støtte (G), Rummelighed (H), Kendskab til diabetes (I) og Fysiske forhold (J)
Blok 4 = Arbejde indvirker positivt (C)/negativt (D) på diabetes
Blok 5 = Sygedage (A) og Træthed (B)

Blokstrukturen gemmes som et nyt DIGRAM projekt og derefter kan DIGRAM anvendes til selve konstruktionen af grafiske modeller. Der indledes i dette tilfælde med SCREEN-kommandoen, som genererer en initialmodel på baggrund af analyse af henholdsvis to- og tre-vejstabeller, hvor det undersøges, om der er association mellem variablene. Med udgangspunkt i denne model foretages derefter modelselektion via stepwiseproceduren. Som default anvendes backwards selektion, men denne metode kan naturligvis ændres til forward, hvis det ønskes. Signifikansniveauet vælges til 0.01 og med backwards som søgestrategi vil det sige, at ikke-signifikante kanter fjernes fra initialmodellen, hvorved modellen simplificeres.
Ved at inkludere samtlige variable fra a priori modellen og baggrundsvariablene i Blok 1 og 2, opnås en grafisk model som er vist i Figur 5.28 via ovenstående fremgangsmåde i DIGRAM. I nedenstående model er Blok 1 længst til højre og Blok 5 længst til venstre.



Figur 5.28: Grafisk model efter backwards selektion i DIGRAM

Modellen kan simplificeres yderligere, hvis det ønskes, ved at fjerne kanter og pile. Dette gøres nemt ved at bruge DELETE-kommandoen og deri specificere den pågældende kant. Problemet med denne metode er imidlertid, at der ikke tages højde for, hvorvidt afhængighedsstrukturen ændres ved at fjerne de forskellige kanter. Med HUGIN opstod disse problemer ikke, hvilket derfor er en af de overvejelser, som skal gøres i forhold til valg af program. I dette tilfælde er DELETE-kommandoen ikke anvendt.
I forhold til det forventede ses det, at flere af variablene fra Blok 3 er orienterede mod enten Arbejde

indvirker positivt (C) eller Arbejde indvirker negativt (D) i.e. målene for arbejdsmiljø. Derudover er der tilsyneladende sammenhæng mellem Sygedage (A) og Træthed (B). Til forskel fra HUGIN-modellen har de fysiske forhold (J) nu betydning for arbejdets negative indvirkning ligesom kendskabet til diabetes (I) påvirker graden af træthed. Med denne nye model er der således andre sammenhænge som skal undersøges, hvilket igen gøres via ordinal regression i SPSS. Det skal dog bemærkes, at DIGRAM ligeledes giver mulighed for at undersøge de enkelte kanter via test og beregning af korrelationskoefficienter i selve programmet. Dermed kan DIGRAM i højere grad anvendes alene, end det eksempelvis er tilfældet med HUGIN.
I analysen af DIGRAM-modellen undersøges først, hvilken betydning Støtte (G), Rummelighed (H) og Fysiske forhold (J) har for Arbejde indvirker negativt (D). Andelen af varians, som regressionsmodellen kan forklare, er lav, 0.144, men test viser, at analysen trods alt er fornuftig, da effekten af de forklarende variable er den samme. Ud fra parameterestimaterne viser det sig, at hvis der er nogenlunde støtte,

Parameter Estimates

							95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Lower Bound	Upper Bound
Threshold	[negativt_arbejde = 1]	-,372	,690	,291	1	,590	-1,726	,981
	[negativt_arbejde = 2]	1,295	,693	3,490	1	,062	-,064	2,653
Location	[støtte=2]	,951	,234	16,589	1	,000	,494	1,409
	[støtte=3]	0[a]	.	.	0	.	.	.
	[rummelig=1]	,816	,317	6,615	1	,010	,194	1,438
	[rummelig=2]	,149	,330	,205	1	,650	-,497	,796
	[rummelig=3]	0[a]	.	.	0	.	.	.
	[fysiske_forhold=1]	1,373	,682	4,052	1	,044	,036	2,710
	[fysiske_forhold=2]	,313	,755	,172	1	,679	-1,167	1,792
	[fysiske_forhold=3]	0[a]	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Figur 5.29: Ordinal regression af Arbejde indvirker negativt

[støtte = 2], stor rummelighed, [rummelig = 1], og gode fysiske forhold, [fysiske_forhold = 1], så er der under 3% risiko for, at arbejdet har negativ indvirkning på egenomsorgen, [negativt_arbejde = 1]. På en arbejdsplads med lav støtte, [støtte = 3], lille rummelighed, [rummelig = 3], og dårlige fysiske forhold, [fysiske_forhold = 3], er der til gengæld mere end 40% risiko for negativ indvirkning af arbejdet. Det bliver således bekræftet via DIGRAM-modellen, at de pågældende variable kan påvirke arbejdsmiljøet. Dog skal der igen henledes opmærksomhed på, at flere af parameterestimaterne ikke er signifikante, hvorfor det er muligt, at der ikke er effekt af disse.
Ud fra tabellen i Figur 5.29 er det særligt gode fysiske forhold som har stor, og signifikant, betydning, dernæst følger graden af støtte og endelig rummeligheden på arbejdspladsen. Hvis arbejdet ikke skal have en negativ indvirkning på respondenternes egenomsorg er det således en fordel, at der er gode fysiske forhold, god støtte fra kolleger og stor rummelighed. Dette kommer også til udtryk i følgende:

> *"Jeg synes trivsel har meget stor betydning for min diabetes. Glad og positiv=Lavt blodsukker. Stress og negativ=Høj blodsukker."* Respondent 52.

Et godt arbejdsmiljø er altså medvirkende til et bedre blodsukker for denne respondent, hvilket formentlig også gør sig gældende for mange af de øvrige respondenter. Arbejdsmiljøet er med andre ord vigtigt for respondenternes kontrol med deres blodsukker og med et velreguleret blodsukker undgås mange af de ubehagelige følger, som ellers kunne opstå.
I forlængelse af dette undersøges, hvordan Træthed påvirkes af henholdsvis Kontrol, Kendskab og arbejdets negative indvirkning. Regressionen viser både et godt datafit, samme effekter og en højere Nagelkerke, 0.183, end det var tilfældet i de øvrige regressioner. Med regressionens parameterestimater beregnes sandsynligheden for stor træthed, hvis der er høj kontrol, kendskab og ingen negativ indvirkning. Herved opnås en sandsynlighed på ca. 10%, mens tallet er 80%, hvis der er lav kontrol, dårligt kendskab og generelt negativ indvirkning. Graden af træthed påvirkes således i negativ retning, som det

også umiddelbart forventes, hvis arbejdsmiljøet, repræsenteret ved eksempelvis kontrol og kendskab, er dårligt. En kvinde skriver eksempelvis:

"Jeg ville ønske at jeg var bedre til at fortælle mine kolleger at jeg har diabetes, men jeg orker ikke opmærksomheden og spørgsmålene [...]" Respondent 564.

Her gives der udtryk for, at det ville være nemmere for respondenten, hvis der i forvejen var et kendskab til diabetes blandt kolleger, hvilket som tidligere nævnt kunne udbredes via en pjece. Derved skal respondenten ikke bruge tid og energi på at forklare sig overfor de øvrige ansatte og kan derfor bruge energien på at løse sine arbejdsopgaver. Efterspørgslen er i hvert fald til stede:

"Findes der en pjece, der enkelt beskriver diabetes, forholdsregler, følgesygdomme. Den kunne jeg lægge på personalestuen. Pjecen kunne hedde: Du har en kollega/medarbejder, der har diabetes." Respondent 520.

A priori blev det forventet, at træthed havde betydning for antallet af sygedage blandt respondenterne og ud fra modellen tyder det på, at der også er en association mellem de to variable. Grundet blokstrukturen er der dog ikke en orienteret kant fra Træthed til Sygedage, men ved at konstruere en ny model kunne dette vise sig at være tilfældet.

Parameter Estimates

							95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Lower Bound	Upper Bound
Threshold	[sygedage = 1]	1,997	,485	16,982	1	,000	1,047	2,947
	[sygedage = 2]	3,109	,496	39,224	1	,000	2,136	4,082
	[sygedage = 3]	3,607	,504	51,331	1	,000	2,621	4,594
	[sygedage = 4]	4,778	,533	80,358	1	,000	3,734	5,823
	[sygedage = 5]	5,418	,566	91,687	1	,000	4,309	6,527
Location	[køn=1]	,228	,196	1,352	1	,245	-,156	,612
	[køn=2]	0[a]	.	.	0	.	.	.
	[stilling=1]	1,278	,506	6,372	1	,012	,286	2,270
	[stilling=2]	1,025	,475	4,657	1	,031	,094	1,956
	[stilling=3]	,929	,533	3,039	1	,081	-,115	1,973
	[stilling=4]	,991	,416	5,669	1	,017	,175	1,807
	[stilling=5]	,898	,452	3,943	1	,047	,012	1,784
	[stilling=6]	0[a]	.	.	0	.	.	.
	[indkomstgruppe=1]	,068	,477	,020	1	,887	-,867	1,003
	[indkomstgruppe=2]	,614	,211	8,463	1	,004	,200	1,027
	[indkomstgruppe=3]	0[a]	.	.	0	.	.	.
	[kontrol=1]	-,039	,271	,021	1	,886	-,571	,493
	[kontrol=2]	,071	,265	,071	1	,790	-,448	,590
	[kontrol=3]	0[a]	.	.	0	.	.	.
	[støtte=2]	-,341	,221	2,371	1	,124	-,774	,093
	[støtte=3]	0[a]	.	.	0	.	.	.
	[negativt_arbejde=1]	,257	,391	,434	1	,510	-,508	1,023
	[negativt_arbejde=2]	,627	,246	6,475	1	,011	,144	1,110
	[negativt_arbejde=3]	0[a]	.	.	0	.	.	.
	[træthed=1]	,933	,269	12,013	1	,001	,405	1,460
	[træthed=2]	,446	,231	3,738	1	,053	-,006	,899
	[træthed=3]	0[a]	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Figur 5.30: Ordinal regression af Sygedage

Både for mænd og kvinder gælder det, at sandsynligheden for at have $0-5$ sygedage er størst for respondenter i den højeste indkomstgruppe, [indkomstgruppe = 3]. Mod forventning har respondenter i den laveste indkomstgruppe næststørst sandsynlighed for $0-5$ sygedage og derefter følger indkomstgruppen 200.000 − 399.000 kr. Dette resultat kan eventuelt forklares på baggrund af forholdet mellem krav og kontrol. Her tænkes det, at respondenterne i den høje indkomstgruppe generelt besidder de højeste stillinger og derfor følger argumentationen resultatet fra HUGIN-modellen, hvor de aktive job netop giver den største tilfredshed, fordi der er balance mellem krav og kontrol. Det forventes således, at

der er en vekselvirkning mellem Krav og Kontrol, som påvirker antallet af sygedage. Årsagen til, at det er den laveste indkomstgruppe, der følger umiddelbart efter er muligvis, at respondenter i denne gruppe besidder forholdsvis passive job, hvor både mængden af krav og mængden af kontrol er lille. Dermed tænkes det, at disse respondenter generelt sjældent er syge, fordi interaktionen mellem Krav og Kontrol resulterer i, at de ikke oplever et anstrengende arbejdsmiljø. Dette fører dog ikke til den tolkning, at de har et godt arbejdsmiljø. Karasek beskriver blandt andet, at denne gruppe af ansatte kan udvikle en formidabel evne til at sove, fordi de ikke bliver udfordret i deres arbejdsliv, hvilket næppe kan siges at være positivt. I værste tilfælde kan det tænkes, at denne gruppe ligefrem udvikler en apatisk adfærd.
I forhold til antallet af sygedage er det altså i den midterste kategori, at de største problemer findes, muligvis fordi det er i denne gruppe, at der forekommer de største misforhold mellem de krav der stilles, og den kontrol som respondenterne har. I og med regressionen estimerer hovedeffekterne af de enkelte variable, når der er korrigeret for de øvrige, er det dog også muligt, at der findes en alternativ forklaring, der ikke har sammenhæng med Krav og Kontrol. Alternativt kan det tænkes, at resultatet skyldes respondenternes arbejdstider, i.e. hvis der er en overvægt af respondenter med aften- eller natarbejde i denne indkomstgruppe, kan dette være årsag til et højere antal sygedage:

> *"Skiftehold, fast natarbejde og døgnvagt med udkald. Burde ikke være tilladt for personer med diabetes."* Respondent 24.

> *"Jeg arbejder aften og nat [...] Min diabetes er svær at regulere med disse arb tider. Det betyder jeg har mange sygedage pga diabetes [...]"* Respondent 428.

Endvidere er det muligt, at de respondenter, der har en mellemindkomst generelt også arbejder i teams eller i hvert fald har et tættere samarbejde med deres kolleger end det er tilfældet for henholdsvis lav- og højindkomstgrupperne. I en sådan situation føler respondenterne formentlig en større forpligtelse over for kollegerne og vælger dermed at fortsætte arbejdet trods kroppens signaler om behov for pause:

> *"Jeg ville få et bedre arbejdsliv, hvis jeg ikke tænkte på mine kolleger som skal løbe hurtigere, hvis jeg må melde pas på en dårlig dag. Resultatet bliver ofte, at man ta'r en dag man burde blive hjemme og få styr på sygdommen."* Respondent 80.

Ud fra parameterestimaterne i Figur 5.30 er højere grad af kontrol og støtte endvidere med til at øge sandsynligheden for et færre antal sygedage, hvilket ikke er overraskende. På baggrund af regressionen viser det sig også, at mænd generelt har større sandsynlighed for færre sygedage end kvinder. Dette resultat er dog tvivlsomt, da de pågældende parameterestimater ikke er signifikante. I Karaseks teori bliver det beskrevet, at det ofte er kvinder, der har job med høj grad af anstrengelse, hvor andelen af arbejdsrelaterede sygdomme er højest. Med ovenstående tyder det dog på, at Karaseks teori på dette punkt ikke længere er tidssvarende og at tidligere kønsforskelle er blevet udjævnet. Relevansen af kønsforskelle i forhold til den aktuelle problemstilling er endvidere lille, da kønsforskelle, hvis de eksisterer i denne undersøgelse, næppe er noget, som Diabetesforeningen har mulighed for at ændre.
I forbindelse med analysen af sygedage er det hensigtsmæssigt at foretage logistiske regressioner, da testet af parallelitet er signifikant. Det er naturligvis muligt at foretage alle de logistiske regressioner som det kræves for en variabel med seks niveauer, men af hensyn til den efterfølgende fortolkning vælges det, at samle nogle af kategorierne således Sygedage rekodes til tre niveauer, hvorved kun to logistiske regressioner er nødvendige.
Rekodningen af Sygedage foregår således $0-5$ sygedage fortsat er første niveau, $6-12$ angiver fortsat det antal sygedage som forventes almindeligvis, mens de øvrige kategorier samles i én i.e 13+. Med udgangspunkt i denne nye variabel kan to logistiske regressioner nu foretages. Regressionen for det lave niveau af Sygedage viser, at flere af de forklarende variable er signifikante, blandt andre nogle af de erhvervsmæssige stillinger og negativt_arbejde. Regressionen for det høje niveau viser til gengæld, at ingen variable har signifikans og dermed ingen effekt i forhold til Sygedage.
Ud fra Figur 5.31 har variablene således kun betydning i forhold til det lave niveau af Sygedage, hvor eksempelvis en stilling som ledende funktionær, stilling(5), i stedet for en stilling som funktionær,

Lav

Variables in the Equation

		B	S.E.	Wald	df	Sig.
Step 1	køn(1)	-,007	,264	,001	1	,979
	stilling			8,916	5	,112
	stilling(1)	-1,775	,735	5,834	1	,016
	stilling(2)	-1,602	,717	4,987	1	,026
	stilling(3)	-1,138	,802	2,014	1	,156
	stilling(4)	-1,576	,647	5,922	1	,015
	stilling(5)	-1,044	,705	2,192	1	,139
	indkomstgruppe			5,922	2	,052
	indkomstgruppe(1)	-,645	,583	1,224	1	,269
	indkomstgruppe(2)	-,701	,292	5,769	1	,016
	kontrol			1,070	2	,586
	kontrol(1)	-,379	,383	,981	1	,322
	kontrol(2)	-,343	,377	,825	1	,364
	støtte(1)	,231	,300	,593	1	,441
	negativt_arbejde			7,528	2	,023
	negativt_arbejde(1)	-,844	,531	2,527	1	,112
	negativt_arbejde(2)	-,853	,334	6,543	1	,011
	træthed			5,824	2	,054
	træthed(1)	-,874	,367	5,661	1	,017
	træthed(2)	-,480	,301	2,534	1	,111
	Constant	3,504	,760	21,247	1	,000

Høj

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	køn(1)	,411	,300	1,874	1	,171	1,508
	stilling			3,383	5	,641	
	stilling(1)	-,453	,873	,269	1	,604	,636
	stilling(2)	-,869	,817	1,131	1	,287	,419
	stilling(3)	-,246	,933	,070	1	,792	,782
	stilling(4)	-,819	,757	1,170	1	,279	,441
	stilling(5)	-,312	,819	,145	1	,704	,732
	indkomstgruppe			1,472	2	,479	
	indkomstgruppe(1)	-,742	,719	1,064	1	,302	,476
	indkomstgruppe(2)	,101	,323	,098	1	,754	1,106
	kontrol			1,200	2	,549	
	kontrol(1)	-,411	,418	,963	1	,326	,663
	kontrol(2)	-,114	,406	,078	1	,780	,893
	støtte(1)	-,017	,346	,002	1	,961	,983
	negativt_arbejde			2,615	2	,271	
	negativt_arbejde(1)	-,940	,584	2,589	1	,108	,391
	negativt_arbejde(2)	-,223	,361	,384	1	,536	,800
	træthed			2,949	2	,229	
	træthed(1)	,725	,424	2,923	1	,087	2,064
	træthed(2)	,383	,391	,960	1	,327	1,467
	Constant	,506	,837	,366	1	,545	1,659

Figur 5.31: Logistisk regression af dikotome variable for Sygedage

stilling(4), vil give større sandsynlighed for et lavt antal sygedage. I den anden ende af skalaen, hvor antallet af sygedage er højt, har de forklarende variable tilsyneladende ingen effekt. Dermed kan disse ikke drages til ansvar for et stort antal sygedage. For at forklare det høje antal sygedage må der således inddrages andre variable, i.e. variable som ikke er til stede i den nuværende model.
Hvis der imidlertid vendes tilbage til den ordinale regression med den nye variabel for sygedage viser det sig, at effekterne af de uafhængige variable nu er de samme for de tre niveauer af Sygedage. Kodningen af Sygedage er med andre ord afgørende for, hvorvidt testet af parallelitet er signifikant eller ej. Hvis effekterne var de samme skulle en rekodning af kategorierne, hvor 'nabo'-niveauer samles, ikke ændre noget, hvorfor dette er argumentation for, at effekterne netop er forskellige.
Med DIGRAM-modellen er der tilsyneladende både sammenhænge, som ikke gjorde sig gældende i HUGIN-modellen, og sammenhænge som genkendes fra den første model. En af de nye sammenhænge som viser sig i DIGRAM-modellen, er betydningen af de fysiske forhold for det oplevede arbejdsmiljø. Ordinal regression viste, at gode fysiske forhold havde betydning for et godt arbejdsmiljø. Det samme var tilfældet for den støtte og rummelighed, som var til stede på arbejdspladsen. De to modeller som indtil videre er præsenteret, er således enige om, at variablene Støtte og Rummelighed påvirker, hvorvidt arbejdet har en positiv eller negativ indvirkning på respondenterne.
I dette afsnit blev det ligeledes fremlagt, at træthed påvirkes af respondenternes kontrol, det generelle kendskab til diabetes på arbejdspladsen og arbejdets negative indvirkning. Derudover påvirkes antallet af sygedage af henholdsvis den erhvervsmæssige stilling, indkomstgruppe, køn og graden af kontrol. Her viste det sig, at en middelindkomst var den mest kritiske i forhold til antallet af sygedage. Dette resultat blev forklaret med udgangspunkt i Karaseks teori, hvor forholdet mellem krav og kontrol blev anvendt som argumentation, ligesom det var tilfældet i HUGIN-modellen. Analysen af sygedagene krævede dog en logistisk regression, der viste, at de forklarende variable ikke havde betydning i begge ender af skalaen. Dermed er det fortsat uklart hvilke variable, der kan siges at have en generel indvirkning på antallet af sygedage.
Den sidste model som skal anvendes i forbindelse med analysen er en model konstrueret via MIM, hvor det, ligesom det var tilfældet med DIGRAM-modellen, er udnyttet, at variablene kan opdeles i blokke i forhold til en tidsdimension.

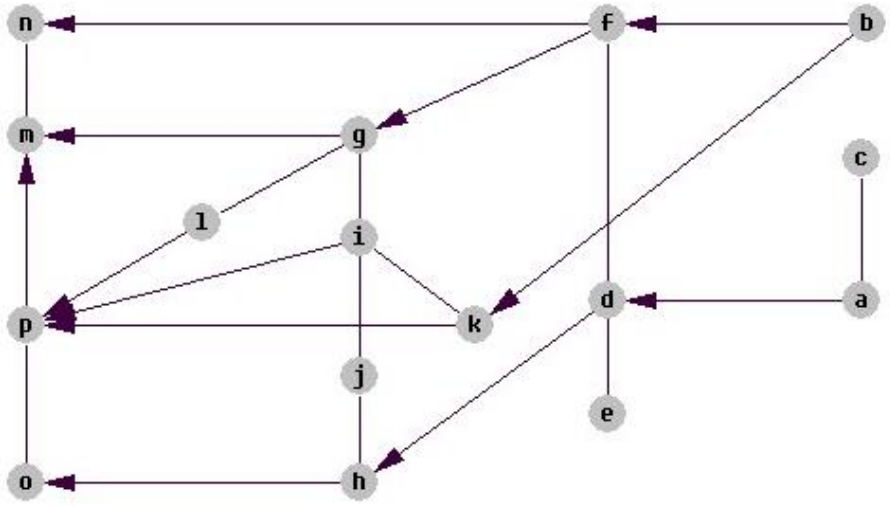
5.2.3 MIM-modellen

I forbindelse med indlæsningen af data specificeres det, at data er ordinalskalerede via kommandoen Ordinal og blokkene defineres via SetBlock kommandoen. Endvidere skal MIM forberedes på, at der arbejdes med kædegrafer, hvilket gøres ved at angive BlockMode +. Hukommelsen i MIM er desværre

ikke tilstrækkelig til at benytte alle de forskellige funktioner; det er blandt andet ikke muligt at foretage modelselektion via stepwiseproceduren, så i stedet vælges en model ved at minimere værdien af informationskriteriet BIC. Med dette selektionskriterie sammenlignes BIC for initialmodellen med de forskellige submodeller og modellen med den midste værdi af BIC vælges.
I MIM-modellen defineres følgende blokstruktur:

Blok 1 = Diabetestype (a), Køn (b) og Alder (c)
Blok 2 = Erhvervsuddannelse (d), Stilling (e) og Indkomst (f)
Blok 3 = Kontrol (g), Krav (h), Støtte (i), Rummelighed (j), Kendskab til diabetes (k) og Fysiske forhold (l)
Blok 4 = Arbejde indvirker positivt (o)/negativt (p) på diabetes
Blok 5 = Træthed (m) og Sygedage (n)

Blokstrukturen er således den samme som gjorde sig gældende i DIGRAM-modellen, men resultatet bliver imidlertid ikke det samme, da selektionskriteriet denne gang er et andet. Modellen som MIM konstruerer ses i nedenstående. Som forventet ud fra a priori modellen er der sammenhæng mellem ar-



Figur 5.32: Grafisk model efter selektion via BIC i MIM

bejdets negative indvirkning (p) og Træthed (m). Endvidere er de fysiske forhold (l), graden af støtte (i) og kendskabet til diabetes (k) alle associerede med Arbejde indvirker negativt. Til gengæld har hverken Kontrol (g) eller Rummelighed (j) betydning for det oplevede arbejdsmiljø. Rummelighed som i både HUGIN- og DIGRAM-modellen spillede en rolle, er altså ikke central i MIM-modellen. Ligeledes er Stilling (e), som i de tidligere modeller var af afgørende betydning, kun associeret med uddannelsesvariablen (d), hvilket ikke er videre interessant. Tilsvarende har baggrundsvariablen Køn (b) ikke samme forklaringskraft som i de to tidligere modeller.
I MIM-modellen er det interessant at undersøge korrelationen mellem Indkomst (f), Træthed (m) og Sygedage (n). Dette gøres ved at opstille en krydstabel, hvor der elaboreres for variablen Indkomst. Da variablene er ordinalskalerede vælges det at beregne γ for hver krydstabel. For den lave indkomstgruppe bliver $\gamma = 0.635(p = 0.004)$, hvilket indikerer, at der er en stærk sammenhæng mellem træthed og sygedage i denne indkomstgruppe, hvor stor træthed er stærkt korreleret med et højt antal sygedage. For mellemindkomstgruppen beregnes $\gamma = 0.324(p < 0.0005)$, hvilket viser, at der i denne gruppe er moderat sammenhæng mellem træthed og sygedage, mens det for den høje indkomstgruppe viser sig, at $\gamma = 0.307(p = 0.001)$. Dermed har respondenter i den høje indkomstgruppe generelt færre sygedage end respondenter med lav indkomst som følge af træthed. En lignende sammenhæng blev beskrevet i forbindelse med HUGIN- og DIGRAM-modellen, hvor respondenter i højere stillinger, og dermed som

regel også højere indkomst, havde lavere sygefravær. Grundet den logiske sammenhæng mellem stilling og indkomst er argumentationen for denne sammenhæng tilsvarende den argumentation, der blev givet tidligere omkring betydningen af aktive vs. passive job.
Det undersøges derefter, hvordan variablene Sygedage, Kontrol og arbejdets negative indvirkning påvirker den træthed som respondenterne føler. Der foretages således en ordinal regression, som viser sig at være fornuftig, da der både er et godt datafit, Nagelkerke er 0.251 og der er samme effekt af de forklarende variable, for de forskellige niveauer af Træthed. Regressionen resulterer således i følgende output, der beskriver, hvordan høj kontrol, [kontrol = 1], medfører, at sandsynligheden for stor træthed, [træthed = 1], bliver mindre end hvis kontrollen var lav. Kontrol på og med arbejdet er med andre ord

Parameter Estimates

							95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Lower Bound	Upper Bound
Threshold	[træthed = 1]	-1,356	,532	6,494	1	,011	-2,400	-,313
	[træthed = 2]	,875	,529	2,733	1	,098	-,162	1,912
Location	[sygedage=1]	,020	,514	,001	1	,969	-,988	1,027
	[sygedage=2]	-,506	,535	,894	1	,344	-1,555	,543
	[sygedage=3]	-,974	,608	2,566	1	,109	-2,165	,218
	[sygedage=4]	-1,237	,584	4,484	1	,034	-2,382	-,092
	[sygedage=5]	-,493	,702	,492	1	,483	-1,869	,884
	[sygedage=6]	0[a]	.	.	0	.	.	.
	[kontrol=1]	1,066	,239	19,905	1	,000	,597	1,534
	[kontrol=2]	,802	,247	10,582	1	,001	,319	1,285
	[kontrol=3]	0[a]	.	.	0	.	.	.
	[negativt_arbejde=1]	-2,129	,388	30,133	1	,000	-2,889	-1,369
	[negativt_arbejde=2]	-1,222	,236	26,691	1	,000	-1,685	-,758
	[negativt_arbejde=3]	0[a]	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Figur 5.33: Ordinal regression af Træthed

af signifikant betydning for den træthed som respondenterne føler, hvor høj kontrol er at foretrække, da det giver respondenterne større frihed til at planlægge arbejdsdagen:

> *"Mit arbejde [...] gør at jeg selv kan bestemme arbejdstider og gå hjem når der er behov for det."* Respondent 368.

Et arbejdsmiljø, hvor medarbejderen ikke har kontrol kan eventuelt resultere i en situation som den, der beskrives her:

> *"Må nødvendigvis sove i flere timer når jeg kommer hjem fra job hver dag. Vil gerne have flexjob."* Respondent 355.

Med dette citat er det tydeligt, at træthed og antallet af sygedage må være associerede, hvilket også tidligere er vist. Ud fra regressionen giver en negativ indvirkning på respondenternes egenomsorg, [negativt_arbejde = 1], større sandsynlighed for at opleve stor træthed, hvilket også var forventningen a priori.
Træthed er ofte en del af respondenternes diabetes, men ved at give mere kontrol og tilpasse arbejdspladsen, både med hensyn til de fysiske forhold, støtten fra kolleger og det generelle kendskab til diabetes, kan den arbejdsrelaterede træthed formentlig reduceres.
Det vigtigste i denne analyse er fortsat at afdække hvilke variable, der har betydning for arbejdsmiljøet, hvorfor det endnu engang undersøges, hvilke variable, der er associerede med en negativ indvirkning fra arbejdet. I denne model er der således foretaget en regression med henholdsvis de fysiske forhold, Støtte, Kendskab til diabetes og arbejdets positive indvirkning som forklarende variable. Regressionen er fornuftig med Nagelkerke på 0.401, et godt datafit og et ikke-signifikant test af parallelitet. Parameterestimaterne for denne regression ses i Figur 5.34, hvor gode fysiske forhold, [fysiske_forhold = 1], støtte, [støtte = 2] og en positiv indvirkning på egenomsorgen, [positivt_arbejde = 1], reducerer

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[negativt_arbejde = 1]	1,045	,713	2,150	1	,143	-,352	2,441
	[negativt_arbejde = 2]	3,021	,724	17,398	1	,000	1,601	4,440
Location	[fysiske_forhold=1]	2,123	,715	8,807	1	,003	,721	3,525
	[fysiske_forhold=2]	,827	,795	1,082	1	,298	-,731	2,385
	[fysiske_forhold=3]	0[a]	.	.	0	.	.	.
	[støtte=2]	1,053	,250	17,762	1	,000	,564	1,543
	[støtte=3]	0[a]	.	.	0	.	.	.
	[diabetes_kendskab=1]	,350	,358	,956	1	,328	-,351	1,051
	[diabetes_kendskab=2]	,387	,280	1,917	1	,166	-,161	,936
	[diabetes_kendskab=3]	0[a]	.	.	0	.	.	.
	[positivt_arbejde=1]	4,010	1,036	14,976	1	,000	1,979	6,041
	[positivt_arbejde=2]	2,456	,273	80,796	1	,000	1,921	2,992
	[positivt_arbejde=3]	0[a]	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Figur 5.34: Ordinal regression af arbejdets negative indvirkning

risikoen for at opleve et dårligt arbejdsmiljø.
MIM-modellen tilfører således ikke nogen afgørende ny information om sammenhængen mellem variablene, men afspejler primært de resultater, som HUGIN- og DIGRAM-modellen gav.

Gennem analyse af de tre modeller står det klart, at Kontrol, Støtte og Kendskab er de gennemgående variable, der har størst betydning for og påvirkning af arbejdsmiljøet for respondenterne. Det er således disse tre variable, som der skal fokuseres på, hvis arbejdsmiljøet skal forbedres. Analysen viser endvidere, at Krav er sekundære i forhold til kontrol-variablen. Høje krav har ifølge modellerne ingen betydning i.e. ingen negativ indflydelse, hvis blot der samtidig er høj kontrol, hvilket er overraskende på baggrund af Karaseks teori og modeller, hvor Krav netop har samme vægt som Kontrol og Støtte.
De fysiske forhold forventedes a priori at have stor betydning for det oplevede arbejdsmiljø. Dette var også tilfældet i DIGRAM- og MIM-modellen, hvor gode fysiske forhold havde signifikant effekt på arbejdets negative indvirkning. I de to regressioner, der blev foretaget og som er vist i Figur 5.29 og 5.34 ses det, at de to modeller er enige om betydningen af de fysiske forhold samt graden af støtte. Til gengæld er rummeligheden fra DIGRAM-modellen erstattet af kendskab til diabetes i MIM-modellen, men parameterestimater og signifikans er stort set de samme for de to variable. I HUGIN-modellen spillede variablen for de fysiske forhold ikke samme rolle, hvilket formentlig kan forklares på baggrund af den blokstruktur som gør sig gældende for både DIGRAM og MIM, men som ikke kan udnyttes i HUGIN.
En del af baggrundsvariablene for respondenterne har vist sig at have betydning; blandt andet var Stilling afgørende i HUGIN-modellen og Indkomst havde betydning for antallet af sygedage i MIM-modellen. Selvom baggrundsvariablene har betydning, er disse ikke analyseret nærmere, da de, på trods af deres samfundsmæssige relevans, ikke er relevante for den aktuelle problemstilling, hvor Diabetesforeningen søger svar på en række spørgsmål, som for det første skal give mere viden på arbejdsmiljøområdet, men som særligt skal gøre Diabetesforeningen i stand til at hjælpe erhvervsaktive diabetikere med de problemer, der opstår på en arbejdsplads.

Validitet og reliabilitet

I dette kapitel vurderes hvor valid den foregående analyse er, da en sådan vurdering naturligvis er afgørende for anvendelsen af analysens resultater. I det følgende beskrives validiteten af data som følge af indsamlingsmetoden og spørgeskemakonstruktionen samt den anvendte teori. Endvidere vurderes validitet og reliabilitet af den statistiske analyse, hvilket overføres til validiteten og reliabiliteten af fortolkningerne.
Vurderingen af validitet har endvidere afgørende betydning for den endelige konklusion, hvor resultaterne af undersøgelsen præsenteres.

6.1 Metoden

Den metode som Diabetesforeningen har valgt i forbindelse med at foretage undersøgelsen af de erhvervsaktive diabetikere, byder på forskellige udfordringer i forhold til validitet og reliabilitet. I dette afsnit beskrives således de problemer, der kan opstå i en tværsnitsundersøgelse, som den Diabetesforeningen har foretaget.
Med hensyn til udvælgelsen af respondenter er der ikke nogle validitetsproblemer, da der netop er anvendt en probabilistisk metode med udgangspunkt i Diabetesforeningens egen database. Der kan dog være et problem, hvis databasen i udgangspunktet ikke er repræsentativ for den samlede gruppe af diabetikere. Som nævnt i afsnit 1.2.3 udgør medlemsdatabasen ca. 30% af alle diagnosticerede diabetikere, hvilket svarer til mere end 70.000 personer. Da der ikke umiddelbart er grund til at tro, at der er signifikant forskel mellem disse 70.000 og de øvrige diabetikere, anses databasen for repræsentativ og dermed vurderes den eksterne validitet at være tilfredsstillende.
Et af de vigtigste punkter i forbindelse med at vurdere undersøgelsen er vurderingen af dataindsamlingen, som i dette tilfælde er foregået via et spørgeskema. Derfor skal validiteten vurderes ud fra selve spørgsmålene, hvor der skal tages stilling til, hvorvidt spørgsmålene måler det, de har til hensigt at måle, i.e. kan spørgsmålene misforstås eller fortolkes forskelligt af respondenterne.
Eftersom respondenterne i udgangspunktet har de nødvendige forudsætninger for at svare på spørgeskemaet og da spørgsmålene generelt er fokuserede på emner, der er aktuelle i forhold til at have diabetes, så er der ikke problemer med selve spørgsmålsformuleringen. Spørgeskemaet tager udgangspunkt i respondenternes egne oplevelser og egen situation, i.e. der skal ikke tages stilling til hypotetiske spørgsmål eller vurderes politiske problemstillinger, som kan være svære at gennemskue. Det vurderes således, at der ikke er problemer med hensyn til at forstå spørgsmålene.
Spørgsmålsformuleringen har ligeledes betydning for reliabiliteten, da spørgsmålene endvidere skal have samme fortolkning uafhængigt af, hvornår spørgsmålene stilles. I og med spørgeskemaet indeholder en sektion omkring helbred og diverse følgesygdomme er det muligt, at nogle besvarelser er påvirket af respondenternes aktuelle helbredsmæssige tilstand. Besvarelserne kan med andre ord være farvet af, hvorvidt respondenterne i situationen, hvor de udfylder spørgeskemaet, har det fysisk dårligt. Tilsvarende gør sig gældende i forhold til spørgsmålene omkring træthed.
Det største problem i forbindelse med spørgeskemaet er imidlertid, at der ikke er taget højde for de respondenter som eksempelvis er selvstændige, når der stilles spørgsmål omkring chefer og kolleger. I disse tilfælde burde der enten være konstrueret spring i spørgeskemaet således respondenterne blev ført videre til andre relevante spørgsmål, eller der burde være en 'Ikke relevant' svarkategori til de aktuelle spørgsmål. Flere af respondenterne beskriver netop dette problem i nogle af de åbne spørgsmål:

> *"Jeg har forsøgt at udfylde så godt som muligt, men skemaet er unyanceret, der er ikke plads til at selvstændige kan udfylde det fyldestgørende, der mangler afkrydsningspunkter [x] Ikke relevant for undertegnede."* Respondent 143.

"Som selvstændig var det svært at svare på en del af spørgsmålene: 'Hvad siger din chef og dine kolleger?'. Har derfor forsøgt at svare efter bedste evne og indimellem med 'ved ikke'." Respondent 580.

De mangelfulde svarkategorier er således et problem som bør tages til efterretning, hvis en lignende undersøgelse skal udføres på et senere tidspunkt, da de ikke kun udgør et problem for respondenterne, men også for undersøgelsens validitet. I dette tilfælde vil svarene fra de selvstændige ikke være pålidelige, da de ikke er relevante. Gruppen af selvstændige udgør i denne undersøgelse 7%, hvilket heldigvis ikke er et stort antal, men der kan altså være validitetsproblemer med eksempelvis variablen Støtte, da denne blandt andet er konstrueret på baggrund af spørgsmålene:

- Bliver dit arbejde anerkendt og påskønnet af ledelsen?
- Bliver du respekteret af ledelsen på din arbejdsplads?

Dette er spørgsmål som en selvstændig erhvervsdrivende kan have svært ved at svare på. På samme måde er variablen for kendskabet til diabetes givet ved spørgsmålet "Synes du, at din chef og kolleger ved tilstrækkeligt om diabetes?", hvilket ligeledes er uhensigtsmæssigt formuleret i forhold til selvstændige. En anden problematik ved spørgeskemaet er længden, hvilket også blev fremhævet i afsnit 1.2.2. På trods af spørgeskemaets længde fuldfører de fleste; formentlig fordi spørgsmålene er aktuelle for diabetikerne og fordi de kan se en fordel i at besvare dem:

"Godt initiativ! Grundigt, vedkommende skema." Respondent 416.

Endelig vurderes det, at målene er konstruktionsgyldige i.e. *"[...] the results we obtain when using the measure fit with theoretical expectations."* [dV07, side 30]. Baggrunden for at foretage denne vurdering er tæt forbundet med den pågældende teori, der naturligvis skal være veldokumenteret for at kunne fungere som målestok for validiteten af de nye mål. Det følgende afsnit vil således fokusere på at vurdere den anvendte teori.

6.2 Teorien

Den a priori viden som primært er udnyttet skyldes den teori, som er udviklet af Robert Karasek gennem 1980'erne omkring arbejdsmiljø. I 1990 blev bogen *Healthy Work* udgivet, som beskriver krav-kontrol modellen samt betydningen af social støtte. Modellen er således ny i sociologisk forstand, men der kræves argumentation for dens relevans og validitet, da arbejdsmiljøet har udviklet sig over de sidste 20 år.
Årsagen til, at det fortsat er fornuftigt at anvende Karaseks model er, at der fokuseres på psykologiske og ikke samfundsmæssige faktorer, i.e. begreberne har stadig betydning fordi den menneskelige psyke ikke ændrer sig i samme grad som eksempelvis samfundet og arbejdsmarkedet. Begreberne krav, kontrol og støtte er altså tilpas generelle til at bevare relevans. Definitionen af job som henholdsvis aktive og passive har muligvis ændret sig, ligesom kønsfordelingen inden for de forskellige jobkategorier formentlig er en anden end det var tilfældet i 1990, men det ændrer ikke på, at krav, kontrol og støtte er gode mål for arbejdsmiljø og risikoen for eventuelle følgesygdomme som stress.
Det største problem i forhold til at anvende Karaseks teori er, at modellen ikke tager højde for individets handlingsmuligheder. Dette kritikpunkt blev ligeledes fremhævet i afsnit 3.3. Karasek 'glemmer', at arbejdsmiljøet ikke er en rigid størrelse, men kan påvirkes og ændres af individet. I løbet af de 20 år er der kommet et større fokus på netop arbejdsmiljø eksempelvis i form af arbejdspladsvurderinger og stiftelsen af Det Nationale Forskningscenter for Arbejdsmiljø, der udsprang i 2004 af det tidligere Arbejdsmiljøinstitut. Dermed er selve arbejdsmiljøet sandsynligvis et andet end det var tilfældet, da Karasek udviklede modellen. At den danske forskning på arbejdsmiljøområdet fortsat tager udgangspunkt i Karaseks model [eaR08, side 247] viser dog med al tydelighed, at modellen stadig er aktuel og anvendelig. Krav-kontrol-støtte modellen er med andre ord ikke ændret fordi den er universel og det bedste, eksisterende mål.

Validiteten og reliabiliteten af krav-kontrol modellen er endvidere beskrevet i Karaseks bog. I forbindelse med konstruktionen af de dimensioner som udgør modellen, blev der foretaget faktoranalyser på en række spørgsmål. Ved at gennemføre samme undersøgelse på forskellige tidspunkter blev en korrelationskoefficient beregnet, hvorved reliabiliteten af målene blev vurderet. I disse studier var korrelationen > 0.9, hvilket indikerer, at de konstruerede skalaer er pålidelige. Den interne skalareliabilitet blev vurderet via Cronbachs α, der er et mål for korrelationen mellem de enkelte spørgsmål, som udgør den samlede skala. Igen blev det bekræftet, at skalaerne var pålidelige med $\alpha > 0.7$ i langt de fleste tilfælde [The90, side 339]. Validiteten af de forskellige mål blev vurderet ved at sammenholde respondenternes besvarelser med ekspertobservationer, hvor korrelationer igen blev beregnet. Med hensyn til kontrol-dimensionen viste der sig: *"[...] high correlation between objective and self-report measures of decision latitude in a variety of studies."* [The90, side 343-344]. Karasek vurderer endvidere, at netop kontrol-variablen er *"[...] the best understood and most statistically reliable [...]"* [The90, side 344]. Med hensyn til krav-variablen siger Karasek: *"Self-reports of psychological job demands correspond well with expert assessments in several other studies."* [The90, side 344], hvormed også denne dimension vurderes at være valid. Målet for social støtte har imidlertid ikke samme høje korrelation, men *"[...] we have little reason to suspect that the problem lies in the substantive validity of the social support scales"* [The90, side 346].
Generelt vurderes det således, at både teorien og de aktuelle mål er valide og pålidelige. Dermed er anvendelsen af krav-kontrol modellen som målestok for validiteten af de nye mål fornuftig. Til gengæld er der stadig tilbage at vurdere, hvorvidt den statistiske analyse er valid, hvilket er fokus i følgende afsnit.

6.3 Analysen

Indledningsvis blev der foretaget en missing value analyse, hvor eventuelle missings blev erstattet via EM-algoritmen. Denne analyse havde til formål at substituere missing værdier med valide værdier og dermed også øge styrken af analysen. Validiteten af faktoranalyserne er beskrevet i hver faktoranalyse, hvor KMO angiver, hvorvidt de udvalgte items er tilpas korrelerede til at være udtryk for samme bagvedliggende faktor, og Bartletts test beskriver, hvorvidt de udvalgte items har loadingværdier $\neq 0$. Ved at tage udgangspunkt i faktoranalysen er det således kun de variable, der statistisk set beskriver samme fænomen, som anvendes som mål.
Efter faktoranalysen blev de nye faktorer opdelt i kategorier ved at definere cutpoints i øvre og nedre kvartil. Ved at inddele faktorerne på denne måde sikres det, at der ikke opstår kategorier, som ikke indeholder observationer, og det er samtidig muligt at beskrive ekstremerne ved at betragte kategoriernes intervaller. Brede intervaller indikerer stor spredning i observationerne og dermed flere ekstreme besvarelser, hvorimod et snævert interval indikerer, at besvarelserne er forholdsvis ens.
I de tilfælde, hvor faktoranalyse ikke var hensigtsmæssig blev nye variable konstrueret ved at samle de items, som blev betragtet som de vigtigste for fortolkningen. Dette vurderes ikke at påvirke validiteten, mens denne beslutning muligvis kan påvirke reliabiliteten, eftersom der er foretaget et valg, som er baseret på en subjektiv vurdering.
Konstruktionen af de tre modeller, som udgør grundlaget for analysen, indeholder ligeledes forskellige problemer i forhold til validitet og reliabilitet. Den første model fremkom via HUGIN, og som nævnt tidligere er en af problematikkerne i den forbindelse, at HUGIN ikke kan udnytte ordinaliteten i variablene og derved overses vigtige sammenhænge, som eksempelvis et rangtest eller en γ-koefficient ville opfange. I forhold til at modellere ordinalskalerede data er HUGIN således ikke hensigtsmæssig. Endvidere er det problematisk, at HUGIN ikke kan arbejde med kædegrafer. I langt de fleste situationer vil det nemlig være tilfældet, at der findes en række baggrundsvariable, eller i hvert fald variable, som er tidligere i forhold til en given tidsdimension, som i en vis udstrækning påvirker en eller flere responsvariable. Til at modellere 'virkeligheden' byder HUGIN således på flere udfordringer. Reliabiliteten af HUGIN-modellen er endvidere problematisk, i og med flere af kanterne orienteres tilfældigt. På baggrund af samme data er det med andre ord muligt at opnå forskellige modeller.
Ved at anvende DIGRAM opstår der ikke alle disse problemer. Dette program er særligt anvendeligt i

forhold til databehandling af den type variable, som er aktuelle i denne undersøgelse. Når holdninger, opfattelser og overbevisninger skal undersøges, er det praktisk talt givet, at svarkategorierne skal være på Likertskala el.lign., hvor det er muligt at afdække retning og styrke af respondenternes besvarelser. DIGRAM er således designet til at udnytte ordinalitet i variablene, og da der endvidere kræves en prædefineret blokstruktur af de pågældende variable lettes den efterfølgende fortolkning af modellen. Den model, som konstrueres via DIGRAM, har dermed større validitet, end det er tilfældet for HUGIN-modellen. Endvidere tilføjer DIGRAM ikke kanter tilfældigt, hvilket sikrer reliabilitet af modellen.
Endelig blev MIM anvendt, hvilket ligeledes kan udnytte ordinalitet i variablene samt en given blokstruktur. Det største problem i forbindelse med anvendelsen af MIM var, at hukommelsen ikke var tilstrækkelig, hvorfor det var meget begrænset, hvilken modelselektion, der kunne foretages. Det var eksempelvis ikke muligt at foretage en stepwise selektion eller at anvende EH-proceduren. MIM har således begrænset anvendelse, hvorfor det vurderes, at DIGRAM-modellen er den mest fornuftige at drage konklusioner på baggrund af.
I forbindelse med analysen af de tre modeller blev ordinal regression og beregning af korrelationskoefficienter udnyttet. Validiteten og reliabiliteten af de forskellige regressioner er beskrevet via parallelitetstest og signifikans af parameterestimaterne, mens korrelationskoefficienterne er forsynet med tilhørende p-værdier. I tilfælde, hvor resultaterne af regressionen ikke var pålidelige, blev der suppleret med logistisk regression. Resultaterne fra regressioner og krydstabeller blev endvidere sammenholdt med teori og udvalgte citater fra respondenterne for dermed at sikre validitet og reliabilitet af fortolkningerne. Alle fortolkninger er således forankret i både den statistiske dataanalyse samt den sociologiske teori.
Validitet og reliabilitet af både metode, teori og analyse har naturligvis betydning for validitet og reliabilitet af de fortolkninger, som tillægges analysens resultater. Der er med andre ord nogle resultater, som er mere troværdige end andre, og det er således disse resultater, som fremhæves i den endelige konklusion.

Konklusion

Analysen af data har afdækket flere interessante sammenhænge mellem de udvalgte variable, hvoraf de mest interessante for problemstillingen og de mest pålidelige præsenteres i det følgende.
Ved de forskellige ordinale regressioner, som er foretaget, har det flere gange vist sig, at de to variable Krav og Kontrol ikke bør behandles separat. Selvom de to mål giver mening hver især er det kombinationen af de to, der har størst betydning når arbejdsmiljøet skal vurderes. Det er med andre ord ikke muligt at beskrive arbejdsmiljøet udelukkende på baggrund af de krav der stilles, da høje krav både kan være positivt og negativt for det oplevede arbejdsmiljø. Det er i højere grad afgørende, hvilken kontrol som følger med kravene, eftersom høj kontrol har vist sig at kunne opveje en eventuel negativ effekt af krav. Med hensyn til at sikre et bedre arbejdsmiljø skal der dermed fokuseres på den kontrol, som diabetikerne har på deres arbejdsplads både i forhold til mængden af arbejde og i forhold til udførelsen af arbejdet. I praksis er det således afgørende, at der er mulighed for at planlægge arbejdsdagen efter dagsformen. Der skal med andre ord være frihed til at veksle mellem forskellige arbejdsopgaver, frihed til at holde pauser, hvis behovet opstår etc.
Hvis diabetikerne ikke har passende kontrol på arbejdspladsen, vil konsekvensen være øget træthed ifølge både DIGRAM- og MIM-modellen. Dette er en meget uheldig konsekvens, da mange af diabetikerne i forvejen mangler energi, hvorfor der bestemt ikke er behov for, at arbejdet ligeledes påvirker i negativ retning. Gennem analysen af modellerne blev det bekræftet, at træthed og sygedage er associerede således øget træthed resulterer i et højere antal sygedage blandt diabetikerne. Den træthed som fører til sygemeldelse kan imidlertid afhjælpes ved netop kontrol. Derudover viste analysen, at kendskabet til diabetes, ligeledes havde stor betydning for diabetikerne. Jo større kendskabet er, des mindre er sandsynligheden for at opleve stor træthed. Som det tidligere er beskrevet, vil udbredelse af viden om diabetes være en stor hjælp for mange af diabetikerne, da de således ikke skal bruge tid og energi på at forklare deres behov og mulige begrænsninger.
Et umiddelbart løsningsforslag, som flere har udtrykt ønske om, er en pjece som kort fortæller om diabetes, således kolleger og ledere eksempelvis kan få information om, hvad de skal gøre, hvis en medarbejder får føling eller insulinchok. Der bliver ikke udtrykt ønske om en landsdækkende kampagne, som ellers kunne give mere fokus på de problemer, som flere diabetikere oplever i forbindelse med at være erhvervsaktive, men blot en pjece, som kan deles ud på arbejdspladsen. Dermed vil diabetikeren kunne koncentrere sig om arbejdet og ikke om spørgsmål eller fordomme fra kolleger.
Udover at øget træthed kan resultere i flere sygedage, har denne association ligeledes betydning for det psykiske arbejdsmiljø. Den konstante træthed og følelsen af at være bagefter kan føre til udbrændthed og stress, blandt andet som følge af dårlig samvittighed over for kolleger og fordi der altid er arbejdsopgaver, som skal indhentes. Arbejdsmiljøet for diabetikerne har vist sig at være påvirket af både støtte, rummelighed, kontrol, de fysiske forhold og kendskabet til diabetes. Krav er således den eneste af a priori variablene som ikke tillægges særlig forklaringskraft i modellerne.
Støtte, rummelighed og kendskabet til diabetes er endvidere tæt forbundet, da de alle omhandler socialpsykologiske apekter omkring trivsel på arbejdspladsen. Overordnet handler det om, at den ansatte føler sig accepteret af de øvrige ansatte. Dette forudsætter, at den enkelte føler sig støttet og føler at der er plads, i.e. rummelighed, på arbejdspladsen. Opsummerende er der således tre faktorer som har afgørende betydning for arbejdsmiljøet: Social accept, kontrol og gode fysiske forhold.
Den sociale accept kan i høj grad opnås via udbredelse af viden eksempelvis gennem en målrettet pjece. Kontrol er imidlertid i højere grad styret af erhvervsmæssig stilling, der har stor betydning for den enkeltes muligheder og frihed. I forhold til Karaseks teori spiller fordelingen af job således fortsat en vigtig rolle, hvor eksempelvis service- og omsorgsorienterede job som tjener, sygeplejerske o.lign. kan være svære for en diabetiker, da mængden af arbejde og pauser er bestemt af udefrakommende faktorer. Endelig er de fysiske forhold til dels bestemt af erhvervsmæssig stilling, men primært af ledelsen på den pågældende arbejdsplads og kan dermed påvirkes af den enkelte ansatte.

I forhold til besvarelse af de opstillede arbejdsspørgsmål er det ikke muligt at give et endeligt svar på, hvorvidt ledelse og kolleger yder tilstrækkelig støtte, da dette er en subjektiv vurdering. Som beskrevet i analysen er der stor spredning i kategorien 'Lav støtte', men til gengæld er der få respondenter i ekstremen af denne kategori. Med hensyn til om der er gode vilkår for egenomsorgen blev dette bekræftet i analysen, da mere end 90% har gode fysiske forhold. Det er således ikke på dette område, at der er behov for en ekstra indsats. Det tredje arbejdsspørgsmål fokuserede på antallet af sygedage. Her viste det sig, at diabetikere ikke kunne siges at have et generelt højere sygefravær end den øvrige befolkning. Til gengæld viste test, at diabetikere oplever større træthed, udkørthed og oftere mangler energi og kræfter end det er tilfældet for andre. Dog er der færre diabetikere end forventet, som ofte føler sig fysisk udmattede. Dermed blev en forskel på disse områder for henholdsvis diabetikere og ikke-diabetikere præsenteret, men uden at der viste sig en generel negativ forskel.
Gennem analysen er der dokumenteret flere interessante sammenhænge mellem forskellige variable og i denne sidste del er det præsenteret hvilke variable, der har størst betydning for arbejdsmiljøet blandt de adspurgte. Med dette speciale er det dermed muligt at målrette en indsats på området og sikre erhvervsaktive diabetikere et bedre arbejdsmiljø, hvor der er større forståelse for og accept af de udfordringer, som diabetikerne skal overvinde i deres arbejdsliv.

Appendiks

6.4 Uorienterede grafer

En uorienteret graf $G = (V, E)$ består af en endelig mængde, V, af knuder og en endelig mængde, $E \subseteq V \times V$, af kanter.

Hvis der er en kant mellem to knuder, X og Y, i.e. $[XY] \in E$, så siges X og Y at være *tilstødende*, hvilket noteres $X \sim Y$. Bemærk, at $X \sim Y$ er det samme som $Y \sim X$.

En graf G er *komplet* hvis der er en kant mellem hvert par af knuder.

Enhver delmængde $v \subseteq V$ inducerer en *subgraf* $G_v = (v, F)$ af G, hvor F består af de kanter i E, hvor begge endepunkter er i v.

En delmængde $v \subseteq V$ er en *klike*, hvis den er komplet og hvis $v \subset w$, så er w ikke komplet.

En følge af knuder $X_0, X_1, \ldots, X_n$, hvor $X_{i-1} \sim X_i$ for $i = 1, 2, \ldots, n$ kaldes en *sti* mellem X_0 og X_n af længde n.

En graf G er *sammenhængende*, hvis der er en sti mellem hvert par af knuder.

En sti $X_1, X_2, \ldots, X_n, X_1$ kaldes en *n-cykel*.

Hvis de n knuder i en n-cykel er forskellige og hvis X_j og X_k kun er tilstødende for $|j - k| \in \{1, n - 1\}$, så er cyklen uden *korder*.

En graf G er *trianguleret* hvis den ikke har nogle n-cykler uden korder, hvor $n \geq 4$.

Hvis alle stier fra $v \subseteq V$ til $w \subseteq V$ skærer $s \subseteq V$, så siges s at *separere* v og w.

Mængden af de knuder i $V \setminus v$, hvor $v \subseteq V$, der er tilstødende til en knude i v kaldes *grænsen* for v, hvilket noteres $\mathrm{bd}(v)$.

6.5 Orienterede grafer

En graf $G = (V, E)$ er orienteret, hvis E er en mængde af ordnede par af knuder.

Hvis $(XY) \in E$, noteres dette $X \to Y$. Bemærk, at $X \to Y$ ikke er det samme som $Y \to X$.

Hvis $X \to Y$ eller $Y \to X$, så siges X og Y at være *tilstødende*, hvilket noteres $X \sim Y$.

En følge af knuder $\{X_1, X_2, \ldots, X_n\}$, hvor $X_i \sim X_{i+1}$ for $i = 1, 2, \ldots, n - 1$ kaldes en *sti*. Hvis $X_i \to X_{i+1}$ for $i = 1, 2, \ldots, n - 1$ så kaldes følgen af knuder en *orienteret sti*.

Hvis første og sidste knude i en orienteret sti er sammenfaldende, i.e. $X_1 = X_n$, så er den orienterede sti en *orienteret cykel*.

Hvis $X \to Y$, så er X *forælder* til Y og Y kaldes tilsvarende et *barn* af X. Mængden af forældre til X

noteres pa(X) og mængden af børn noteres ch(X).

Hvis der er en orienteret sti fra X til Y, så kaldes X en *forfader* til Y og Y kaldes tilsvarende en *efterkommer* af X. Mængden af forfædre til X noteres an(X) og mængden af efterkommere noteres de(X).

For en mængde $S \subseteq V$ defineres $pa(S)$ som mængden af knuder, der ikke tilhører S, men som er forældre til en knude i S.

For en mængde $S \subseteq V$ defineres $ch(S)$ som mængden af knuder, der ikke tilhører S, men som er børn af en knude i S.

For en mængde $S \subseteq V$ defineres $an(S)$ som mængden af knuder, der ikke tilhører S, men som er forfædre til en knude i S. Endvidere defineres *den forfædrene mængde* som foreningsmængden af S og an(S) og noteres $\text{an}^+(S)$.

For en mængde $S \subseteq V$ defineres $de(S)$ som mængden af knuder, der ikke tilhører S, men som er efterkommere af en knude i S.

6.6 Kædegrafer

En graf $G = (V, E)$ kaldes en kædegraf hvis den ikke indeholder orienterede cykler.

Litteratur

[Agr06] Alan Agresti. Ordinal data. Encyclopedia of Statistical Sciences, 2006.

[Büh07] Markus Kalisch & Peter Bühlmann. Estimating high-dimensional directed acyclic graphs with the pc-algorithm. Journal of Machine Learning Research 8, 2007. http://jmlr.csail.mit.edu/papers/volume8/kalisch07a/kalisch07a.pdf.

[Diaa] Diabetesforeningen. Diabetes i danmark. http://www.diabetes.dk/Rundt_om_diabetes/Diabetes_i_tal/Diabetes_i_Danmark.aspx.

[Diab] Diabetesforeningen. Hvad er diabetes? http://www.diabetes.dk/Rundt_om_diabetes/Hvad_er_diabetes.aspx.

[dV07] David de Vaus. *Research Design in Social Research.* SAGE Publications, 2007.

[eaR08] Kjeld Nielsen et al. (Red.). *Virksomhedens personalearbejde - ledelse og administration.* Aalborg Universitetsforlag, 2008.

[Edw00] David Edwards. *Introduction to Graphical Modelling.* Springer, 2. udgave, 2000.

[Kre03] Svend Kreiner. Introduction to digram. http://staff.pubhealth.ku.dk/~skm/skm/, 2003.

[Lee04] Peter M. Lee. *Bayesian Statistics - an introduction.* Hodder Arnold, 2004.

[Lyn08] Kristina Johansen & Elsebeth Lynge. Sygefravær og sygedagpenge i danmark gennem de sidste 30 år. http://www.asusi.dk/publikation/Asusi_DelprojektSygefravaeretsOmfang.pdf, 2008.

[Max06] A. E. Maxwell. Factor analysis. Encyclopedia of Statistical Sciences, 2006.

[The90] Robert Karasek & Töres Theorell. *Healthy Work - Stress, Productivity and The Reconstruction of Working Life.* Basic Books, 1990.