DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

# EKSAMENSOPGAVE

| | | | | | | |
|---|---|---|---|---|---|---|
| **Uddannelse og niveau** | **HA, 5. semester, valgfag** | | | | | |
| Termin | V21-22r | | | | | |
| Kursusnavn og eksamenskode(r) | Modellering inden for prescriptive analytics | | | | | 460181E018 |
| Eksamensform og varighed | WOAI | | | | | 3 timer |
| Dato og tidspunkt | | | | | | |
| Hjælpemidler | Alle | X | Anviste | | Ingen | |
| Anden relevant information | **Undgå mistanke om eksamenssnyd!**<br><br>Husk kildehenvisninger og citationstegn, hvis du kopierer andres tekster eller hvis du genbruger dele af en tidligere afleveret opgave (plagiering og selvplagiering). Eksamensbesvarelsen skal udarbejdes **individuelt**. Der udføres plagieringskontrol på alle eksamensopgaver, så snyd og samarbejde mellem studerende vil kunne spores. | | | | | |
| Det er tilladt at aflev. håndskrevet mat. | Ja | | Nej | X | | |
| Anonym eksamen? | Ja | x | Nej | | **Undlad** venligst at anføre navn og studienummer i besvarelsen. Brug i stedet flow-løbenr., som du finder på dit omslag. | |
| Antal sider (inkl. forside) | 3 | | | | | |

## Preamble

- Indeværende eksamensopgave indeholder 10 opgaver. Ud for hver opgave er der anført en *vejledende vægtning* af opgaven, som udelukkende giver information om den relative vægtning af opgaven baseret på opgavens omfang og kompleksitet. Den endelige bedømmelse af besvarelsen vil blive baseret på besvarelsens helhed.
- Du bedes besvare opgaverne enkeltvis – det vil sige, lad være med at slå opgaver samme.
- Såfremt du mener, at der i en opgave mangler informationer for, at du kan løse den, kan du gøre dig relevante antagelser og klart og tydeligt specificere disse.
- Denne eksamen er med alle hjælpemidler (WOAI). Dog er den *individuel* - dermed er *samarbejde med andre ikke tilladt*.
- Hvis du reproducerer modeller, beregninger, forklaringer, osv. fra litteraturen, opgaver eller internettet bedes du klart og tydeligt angive kilder på disse for at undgå beskyldninger om plagiat.

## Eksamensopgave

Et datasæt med 38 kunder er givet i den vedlagte Excel-fil. Excel-filens ark kaldet ”Matrix” indeholder en afstandsmatrix, $\{d_{ij}\}_{i,j=1}^{38}$, som angiver afstanden mellem de 38 kunder. Det vil sige, at $d_{ij}$ angiver afstanden mellem kunderne $i$ og $j$.

### Opgave 1 (Ca. 20%)

Formuler en lineær blandet heltalsmodel, som grupperer de 38 kunder i 5 clustre, således at den største *cluster-diameter* bliver minimeret. Beskriv dine parametre, variabler, objektfunktion og begrænsninger grundigt.

**Svar opgave 1:**

Der er her behov for at benytte beskrevet i afsnittet ”Clustering med minimering af cluster-diameteren” i Kapitel 2 i forelæsningsnoterne. Vi vil her benytte afstanden mellem kunderne *i* og *j* givet ved $d_{ij}$ som omkostningen ved at gruppere kunde *i* og kunde *j* sammen. Det vil sige, at vi bruger $c_{ij} = d_{ij}$. Dermed haves følgende blandede heltalsmodel

$$\min D^{max} \qquad (1)$$

$$s.t. D^{max} \geq D_l, \quad \forall l = 1, \dots, 5 \qquad (2)$$

$$D_l \geq d_{ij}(x_{il} + x_{jl} - 1), \quad \forall l = 1, \dots, 5, i, j = 1, \dots, 38: i \neq j \qquad (3)$$

$$\sum_{l=1}^{5} x_{il} = 1, \quad \forall i = 1, \dots, 38 \qquad (4)$$

$$x_{il} \in \{0,1\}, \quad \forall l = 1, \dots, 5, i = 1, \dots, 38 \qquad (5)$$

$$D_l \geq 0, \quad \forall l = 1, \dots, 5 \qquad (6)$$

Her er variablerne defineret som følger

| | |
|---|---|
| $x_{il}$ | Binær variabel som er lig med 1 hvis og kun hvis kunde *i* er i cluster nummer *l*. Ellers er denne variabel lig med 0 |
| $D_l$ | En øvre grænse for diameteren icluster *l* |
| $D^{max}$ | I en optimal løsning er denne variabel lig med diameteren af det cluster, som har den største diameter |

Objektfunktionen (1) minimerer den største diameter over alle clustre. Begrænsningerne (2) sørger for, at $D^{max}$ antager sin korrekte værdi, nemlig diameteren i den bredeste cluster. Begrænsningerne (2) sørger for, at $D_l$ altid er større end eller lig med afstanden mellem to kunder *i* og *j* som begge er i cluster *l.* Begrænsningerne (4) sørger for, at hver kunde kommer i præcis ét cluster, mens begræsningerne (5) og (6) pålæggerne variablerne de korrekte domæner.

## Opgave 2 (Ca. 15%)

Implementer din model fra Opgave 1 i OPL og løs den ved hjælp af CPLEX. Dokumenter din implementering (fx ved hjælp af screen shots) og fortolk objektfunktionsværdien og beskriv løsningen.

**Svar Opgave 2:**

Nedenfor ses en implementering af modellen fra Opgave i OPL. Paramenteren K angiver hvor mange clustre man ønsker og er altså her specificeret til 5.

Under OPL-implementeringen er ligeledes angivet en oversigt over løsningen jeg har fået. Der er formodentlig alternative optimale løsninger. Vi ser at kunderne 1,…,10 er i cluster 1 mens kunderne16, 19 og 30 er i cluster 3. Clusteren med den største diameter har en diameter på 39 afstandsenheder

```
int n = ...;
int K = ...;
range punkter = 1..n;
range clustre = 1..K;
int d[punkter][punkter] = ...;

dvar boolean x[punkter][clustre];
dvar float+ D[clustre];
dvar float+ Dmax;

minimize Dmax;
subject to
{
  forall ( k in clustre )  D[k] <= Dmax;
  forall ( k in clustre, i,j in punkter: i!=j )
    {
      D[k] >= d[i][j] * (x[i][k] + x[j][k] - 1 );
    }
  forall ( i in punkter )
    {
      sum ( k in clustre ) x[i][k] == 1;
    }
}
```

```
Største cluster diameter : 39
Cluster 1
1, 2, 3, 4, 5, 6, 7, 8, 9, 10,
===================
Cluster 2
12, 14, 21, 23, 24, 29, 32, 33, 34, 35,
===================
Cluster 3
16, 19, 30,
===================
Cluster 4
13, 15, 22, 26, 31, 37, 38,
===================
Cluster 5
11, 17, 18, 20, 25, 27, 28, 36,
===================
```

## Opgave 3 (Ca. 5%)

Kan man, baseret på modellen fra Opgave 1, direkte aflæse af en optimal løsning hvad diameteren er *for hvert cluster*? Argumenter for dit svar?

**Svar Opgave 3:**

Her er svaret: Ikke nødvendigvis. Antag, at den cluster som har den største diameter, har en diameter på $D^* > 1$ afstandsenheder. Antag nu, at den cluster der har den næststørste diameter har en diameter på $D^{**} \leq D^* - 1$. Lad WLOG den største cluster være cluster 1 og den næststørste være cluster 2. Vi kan nu se, at det er en brugbar løsning at sætte $D_2 = d$ for alle $d \in \{D^{**}, \dots, D^*\}$. Det vil sige, at variablen $D_l$, som nævnt i definitionen, blot er en øvre grænse for cluster *l*'s diameter.

I den resterende del af opgaven fokuseres udelukkende på kunderne 1-10.

## Opgave 4 (Ca. 15%)

Formuler en blandet heltalsmodel, som beregner en optimal rute gennem kunderne 1-10 baseret på (dele af) afstandsmatricen fra Excelfilen. Du kan antage, at ruten starter og slutter ved kunde 1. Relater problemet til modeller fra kursets litteratur. Beskriv dine parametre, variabler, objektfunktion og begrænsninger grundigt.

**Svar Opgave 4:**

Dette er en klassisk TSP for de 10 første kunder. Man kan se mere om dette problem i kapitel 7.2 i forelæsningsnoterne. Jeg benytter her en MTZ-baseret model til at opstille denne TSP:

$$\min \sum_{i=1}^{10} \sum_{j=1}^{10} d_{ij} x_{ij}$$

$$s.t.: \sum_{\substack{i=1 \\ i \neq j}}^{10} x_{ij} = 1, \quad \forall j = 1, \dots, 10$$

$$\sum_{\substack{j=1 \\ i \neq j}}^{10} x_{ij} = 1, \quad \forall i = 1, \dots, 10$$

$$1 \leq u_i \leq 10, \quad \forall i = 2, \dots, 9$$

$$u_i - u_j + 10x_{ij} \leq 9 = (10 - 1), \quad \forall i, j = 2, \dots, 10$$

$$x_{ij} \in \{0,1\}, \quad \forall i, j = 1, \dots, 10$$

En generel gennemgang af begrænsningerne i modellen ovenfor kan findes i kapitel 7.2 fra forelæsningsnoterne (Var dette en eksamensbesvarelse, burde man beskrive betydningen af begrænsningerne her).

## Opgave 5 (Ca. 5%)

Kursets litteratur foreslår mere end én model for ovenstående problem. Argumenter for dit valg af model. Nævn mindst én fordel og en ulempe ved den model, du har valgt sammenlignet med den/de modeller, du har fravalgt.

**Svar Opgave 5:**

Jeg har valgt en MTZ-baseret model. Jeg kunne også have valgt en Dantzig-Fulkerson-Johnson (DFJ) eller en Gavish-Graves (GG) formulering. Jeg har valgt en MTZ formulering da den er hurtig og nem at implementere og fordi problemet der skal løses ikke er særligt stort. Derfor vil modellen, forventeligt, producere en optimal løsning hurtigt. En ulempe ved denne model er, at den har en temmelig ringe LP-relaksering, da modellen ikke er særligt stærk. Dette ville meget hurtigt give problemer efterhånden som man øger størrelsen på problemet.

Har man et større problem, vil en GG-baseret model ofte være vejen frem, da denne model er signifikant stærkere mens antallet af variabler og begrænsninger kun stiger moderat sammenlignet med MTZ-formuleringen.

DFJ-formuleringen fungerer ikke som en ”stand-alone” model. Her skal man tage en algoritmisk tilgang hvor man *separerer* subtour-eliminationsbegrænsninger løbenden efterhånden som man opdager, at de er brudte. Dette skyldes, at der er eksponentielt mange begrænsninger. For dette lille eksempel med 10 knuder der skal besøges, vil der være $2^{10} = 1024$ SE-begrænsninger.

## Opgave 6 (Ca. 10%)

Implementer din model fra Opgave 4 i OPL og løs den ved hjælp af CPLEX. Dokumenter din implementering (fx ved hjælp af screen shots) og fortolk objektfunktionsværdien og beskriv løsningen.

**Svar Opgave 6:**

Nedenfor er angivet en OPL implementering af modellen fra Opgave 4:

Ligeledes er en optimal løsning angivet som starter i knude 1, kører til knude 3, så til 2 og så fremdeles. Turens længde er 45 afstandesenheder.

```
int n = ...;
range punkter = 1..n;
int d[punkter][punkter] = ...;
dvar boolean x[punkter][punkter];
dvar float+ u[1..n] in 1..n-1;

minimize sum ( i,j in punkter : i!=j ) d[i][j] * x[i][j];
subject to
{
  forall ( i in punkter )
    {
      sum ( j in punkter : j!=i ) x[i][j] == 1;
      sum ( j in punkter : j!=i ) x[j][i] == 1;
    }

  forall ( i,j in punkter : i != 1 && j!=1 )
    {
        u[i] - u[j] + n*x[i][j] <= n-1;
    }
}
```

```
Optimal turs længde : 45
1->3->2->4->8->5->9->6->10->7->1
```

## Opgave 7 (Ca. 5%)

Argumenter for, at der findes mere end én optimal løsning til problemet fra Opgave 4 og 5.

**Svar Opgave 7:**

Man kan se, at afstandsmatricen er symmetrisk. Det vil sige, at der er lige langt fra kunde i til kunde j, som der er fra kunde j til kunde i. Derfor kan man blot køre turen man har fundet i Opgave 6 baglæns. Dette giver en tour som har samme lægde som den førstfundne hvormed den også er optimal.

Betragt nu tidsvinduerne og servicetiderne givet i Excel-filens ark kaldet ”Tid”. Antag, at tiden det tager at køre fra kunde i til kunde j, er lig med $d_{ij}$. Det vil sige, at afstand og køretid her er det samme.

## Opgave 8 (Ca. 10 %)

Udvid din model fra Opgave 4 til at tage højde for, at service ved kunde $i$ skal *starte* i tidsvinduet angivet ved ”Tidligste start”-”Seneste start”. Der skal også tages højde for servicetider ved kunderne. Beskriv eventuelle nye parametre, variabler og begrænsninger grundigt.

**Svar Opgave 8:**

Tidsvinduer for TSP kan ses beskrevet i Kapitel 7.5 i forelæsningsnoterne. Vi indfører nu følgende notation

| | |
|---|---|
| $\tau_i$ | Angiver tidspunktet hvor service starter ved knude i |
| $a_i$ | Tidligste tidspunkt hvor service kan starte ved kunde i |
| $b_i$ | Seneste tidspunkt hvor service kan starte ved kunde i |
| $s_i$ | Servicetiden ved kunde i |

Følgende begræsninger skal indføres:

$$a_i \leq \tau_i \leq b_i, \quad \forall i = 1, \dots, 10$$

$$\tau_i + (s_i + d_{i1})x_{i1} \leq b_1, \quad \forall i = 2, \dots, 10$$

$$\tau_i \geq d_{1i}x_{1i}, \quad \forall i = 2, \dots, 10$$

$$\tau_i - \tau_j + \left(b_i - a_j + s_{i+}d_{ij}\right)x_{ij} \leq b_i - a_j, \quad \forall i, j = 2, \dots, 10 : i \neq j$$

Den resterende del af modelle, formuleret i Opgave 4, forbliver som den var der.

## Opgave 9 (Ca. 10%)

Implementer modellen fra Opgave 8 i OPL og løs den ved hjælp af CPLEX. Dokumenter din implementering (fx ved hjælp af screen shots) og beskriv den løsning, du fandt. Sammenlign med løsningen fra Opgave 5.

**Svar Opgave 9:**

Nedenfor er angivet en implementering af modellen fra Opgave 8 i OPL. De første 6 linjer er identiske med modellen angivet under Opgave 6.

Under modellen er angivet en optimal tur. I parenteserne angives først kundens nummer, og derefter tidpunktet hvor service starter ved kunden. Fx angiver (5,T=259) at service starter ved kunde 5 til tid T=259.

Det kan tjekkes, at alle tidsvinduer er overholdt.

Vi ser at løsningen har ændret sig, idet kunde 10 nu skal serviceres som den første kunde hvilket ikke var tilfældet for den "rene" TSP løst i Opgave 6 (heller ikke selvom man ændrede retningen på turen). Desuden ses det, at en optimal turs længde er steget markant: fra 45 afstandsenheder til 76 afstandsenheder. Det vil sige, at tidsvinduerne har resulteret i en stigning i turens længde på $\frac{76-45}{45} \times 100 = 68\%$.

```
int a[punkter] = ...;
int b[punkter] = ...;
int s[punkter] = ...;
int bigM[i in punkter][j in punkter] = b[i] - a[j] + s[i] + 2*d[i][j];
dvar float+ tau[i in punkter] in a[i]..b[i];

minimize sum ( i,j in punkter : i!=j ) d[i][j] * x[i][j];
subject to
{
  forall ( i in punkter ){
      sum ( j in punkter : j!=i ) x[i][j] == 1;
      sum ( j in punkter : j!=i ) x[j][i] == 1;
    }

  forall ( i,j in punkter : i != 1 && j!=1 ){
        u[i] - u[j] + n*x[i][j] <= n-1;
    }

  forall ( i in punkter : i!=1 ){
      tau[i] >= d[1][i]*x[1][i];
      tau[i] + ( s[i] + d[i][1] )*x[i][1] <= b[1];
    }
  forall ( i,j in punkter : i != 1 && j != 1){
      tau[i] + s[i] + d[i][j] <= tau[j] + bigM[i][j]* ( 1-x[i][j] );
    }
}
```

```
Optimal turs længde : 76
1->(10,T=60)->(3,T=180)->(2,T=210)->(5,T=259)->(6,T=309)->(9,T=330)->(7,T=386)->(8,T=470)->(4,T=499)->1
```

## Opgave 10 (Ca. 5%)

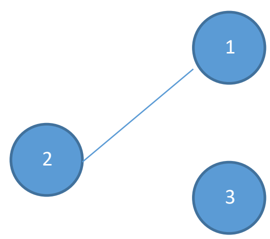
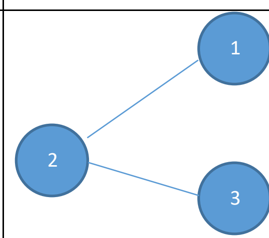
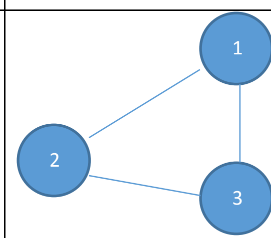
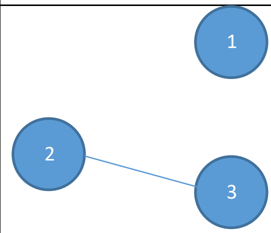
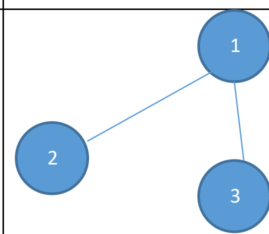
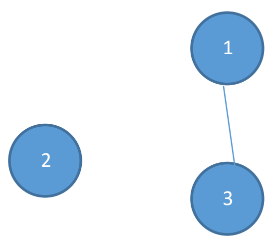
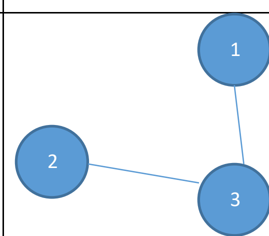
Lad $x_{ij}$ være en binær variabel, som er lig med 1 hvis og kun hvis en optimal rute besøger kunde $i$ lige før kunde $j$. Beskriv med ord, hvad følgende ulighed beskriver:

$$x_{12} + x_{21} + x_{13} + x_{31} + x_{23} + x_{32} \leq 2$$

Relater denne ulighed til ulighederne (7.4) fra kursusnoterne.

**Svar Opgave 10:**

Denne ulighed er en klassisk subtour-elimineringsbegrænsning. Den siger, at for en mængde af tre knuder, kan vi maksimalt have 2 kanter som forbinder disse. Hvis der er 3, vil vi have en subtour, som er noget skidt. For at se dette, betragt følgende tabel

| 0 kanter mellem knuderne 1,2 og 3 | 1 kant mellem de tre knuder 1, 2 og 3 | 2 kanter mellem de tre knuder 1, 2 og 3 | 3 kanter mellem de tre knuder 1, 2 og 3 |
|---|---|---|---|
| Dette er altid OK. Det kunne være, at ruten hen 1->4->2->5->3->6->1 hvorved der ikke er nogle kanter mellem knuderne 1,2 og 3 |  |  |  |
| |  |  | |
| |  |  | |
| | Alt er godt | Alt er godt | Ej godt! |

Uligheden fjerner subture ligesom (7.4) gør. I stedet for at kræve at der skal være mindst én kant der peger ud af gruppen af knuder, siger denne ulighed, at der maksimalt kan være 2 kanter indenfor gruppen, hvilket implicerer at der også skal være mindst én der peger ud, idet hver kunde skal forlades præcis én gang. Dermed er denne ulighed ækvivalent til

$$\sum_{i \in \{1,2,3\}} \sum_{j \in \{4,\ldots,10\}} x_{ij} \geq 1$$

Denne ulighed er netop en subtur elimineringsbegrænsning på samme form som (7.4) med $S = \{1,2,3\}$ og dermed $V \setminus S = \{4, \ldots, 10\}$