Eksamen på Økonomistudiet sommeren 2020
Sandsynlighedsteori og Statistik
2. årsprøve
9. juni, 2020
(3-timers prøve med hjælpemidler)

- RETTEUDKAST

Opgaven består af tre delopgaver, som alle skal besvares. De tre opgaver kan regnes uafhængigt af hinanden.

Opgave 1 og 2 indgår tilsammen med samme vægt som opgave 3.

Opgave 1

Vi betragter nu en to-dimensionel fordeling (X,Y).

Om den marginale fordeling for Y gælder:

P(Y=-1)=0,35 P(Y=0)=0,25 og P(Y=1)=0,4.

Derudover er oplyst *nogle* af sandsynligheder for (X,Y) se nedenstående tabel.

	Y=-1	Y=0	Y=1	
X=0	0,10	0,05	0,30	0,45
X=1	0,25	0,20	0,10	0,55
	0,35	0,25	0,40	1

1. Beregn samtlige simultane sandsynligheder for (X,Y).

2. Beregn P(X=1) og E(X).

   - P(X=1)=E(X)=0,55

   Betragt Z=X*Y.

3. Angiv sandsynlighedsfunktionen for Z,

   og udregn middelværdi og varians.

Z=-1	Z=0	Z=1	
0,25	0,65	0,10	1

   - $E(Z)$=-1*0,25+0*0,65+1*0,10=-0,15
   - $E(Z^2)$ =1*0,25+0*0,65+1*0,10=0,35
   - $V(Z) = E(Z^2) - E(Z)$=0,35-(-0,15)$^2 = 0,3275$

4. Beregn P(X=1|Z=0).

   - $P(X = 1|Z = 0) = \frac{P(X=1,Y=0)}{P(Z=0)} = \frac{0,20}{0,65}$

5. Beregn E(X|Z=0) og V(X|Z=0) .

   - $P(X = 0|Z = 0) = \frac{P(X=0)}{P(Z=0)} = \frac{0,45}{0,65}$
   - $E(X|Z = 0) = P(X = 1|Z = 0) = \frac{0,20}{0,65}$
   - husk at X er 0 eller 1 derfor er X$^2 = X$
   - $E(X^2|Z = 0) = E(X|Z = 0) = \frac{0,20}{0,65}$
   - $V(X|Z = 0) = E(X^2|Z = 0) - [E(X|Z = 0)]^2 =$
   - $\frac{0,20}{0,65} - [\frac{0,20}{0,65}]^2 = 0,21$

Opgave 2

Lad X være binomialfordelt med n=39 og p=1/9. $X \sim Bin(39, \frac{1}{9})$

1. Udregn P(X=0) og P(X=1) og P(X=2).

P(X=0)	P(X=1)	P(X=2)
0,0101	0,0493	0,1171

Lad $X_1 \sim Bin(39, \frac{1}{8})$ $X_2 \sim Bin(38, \frac{1}{8})$ og $X_3 \sim Bin(37, \frac{1}{8})$

2. Udregn $P(X_1 => 3)$ $P(X_2 => 2)$ og $P(X_3 => 1)$.

$P(X_1 => 3)$	$P(X_2 => 2)$	$P(X_3 => 1)$
$= 1 - P(X_1 <= 2)$	$= 1 - P(X_2 <= 1)$	$= 1 - P(X_2 = 0)$
0,8812	0,9598	0,9929

Betragt en terning med 9 sider, som er nummereret fra 1 til 9. Alle udfald er lige sandsynlige.

Der foretages 39 uafhængige kast med denne terning. Man er interesseret i hændelsen: at det 3. største tal er 8.

3. Argumenter for at hændelse {det tredie største tal er 8} kan skrives som en foreningsmængde af hændelserne:

   {ingen 9'ere og mindst 3 8'ere} {præcis 1 9'er og mindst 2 8'ere}

   og {præcis 2 9'ere og mindst 1 8'er}.

   DE TRE MÆNGDER ER DISJUNKTE

4. Udregn sandssynligheden for at det tredie største tal er 8. (vink: inddrag ovenstående resultater)

   P{ingen 9'ere og mindst 3 8'ere}=P(mindst 3 8'ere|ingen 9'ere)*P(ingen 9'ere)=

   P{1 9'ere og mindst 2 8'ere}=P(mindst 2 8'ere|1 9'ere)*P(1 9'ere)=

   P{2 9'ere og mindst 1 8'ere}=P(mindst 1 8'ere|2 9'ere)*P(2 9'ere)=

   P(mindst 3 8'ere|ingen 9'ere)=$P(X => 3|X \sim bin(39, 1/8)) = 0,8812$

   P(mindst 2 8'ere|1 9'ere)=$P(X => 2|X \sim bin(38, 1/8)) = 0,9598$

   P(mindst 1 8'ere|2 9'ere)=$P(X => 1|X \sim bin(37, 1/8)) = 0,9929$

0,8812*0,0101+0,9598*0,0493+0,9929*0,1171=**0,17**

Opgave 3

De meget omtalte PISA undersøgelser måler mange emner blandt de deltagende elever. I denne opgave er der fokus på måling af elevernes læsekompetence, som betegnes læse-scoren. Læsescoren er sammensat af en række læsespørgsmål.

Det kan antages, at læsescoren følger en normalfordeling med standardafvigelse på 10 i dette eksempel. Man er interesseret i at undersøge om gennemsnittet er ændret fra undersøgelsen foretaget i 2015 til undersøgelsen foretaget i 2018.

I 2015 er der udtaget 10 elever fra årets PISA målinger.

I 2018 er der udtaget 20 elever fra årets PISA målinger.

Der er opstillet følgende statistiske model:

$X_1, ...X_{10} \sim N(\mu, 10^2)$

$Y_1, ...Y_{20} \sim N(\nu, 10^2)$

alle stokastiske variable er uafhængige.

så der gælder $V(X) = V(Y) = 10^2$. Dermed $s.e.(X) = s.e.(Y) = 10$

I nedenstående tabel er vist resultaterne af elevernes læse-score.

	2015	2018	begge år
antal	10	20	30
sum	4.991,6	9.833,4	14.825,0
gennemsnit	499,2	491,7	494,2
definition af gns.	$\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i$	$\overline{y} = \frac{1}{20}\sum_{i=1}^{20} y_i$	$\overline{z} = \frac{1}{30}\sum_{i=1}^{30} z_i$
SAK	26.490,5	167.523,0	194.378,3
definition af SAK	$\sum_{i=1}^{10}(x_i - \overline{x})^2$	$\sum_{i=1}^{20}(y_i - \overline{y})^2$	$\sum_{i=1}^{30}(z_i - \overline{z})^2$

kilde: Beregninger på et særligt udvalg af de danske PISA data.

note: når alle 30 data betragtes under et, så er de begnet med Z.

1. Opskriv likelihood-funktionen $L(\mu, \nu)$.

   Angiv log-likelihood-funktionen og scorefunktionerne.

   Generel formel når $\mathrm{X}_i \sim N(\mu, \sigma^2)$

$L(\mu,\sigma^2) = \Pi_{i=1}^n \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2}} \exp(-\frac{1}{2}(X_i-\mu)^2/\sigma^2) =$

$(\frac{1}{\sqrt{2\pi}})^n (\frac{1}{\sigma^2})^{n/2} \exp(-\frac{1}{2}\sum_{i=1}^n (X_i-\mu)^2/\sigma^2) =$

$(\frac{1}{\sqrt{2\pi}})^n (\frac{1}{\sigma})^{n} \exp(-\frac{1}{2}\sum_{i=1}^n (X_i-\mu)^2/\sigma^2) =$

Nu er σ^2 kendt og lig med 10^2 endvidere er n=10 eller 20

så

$L(\mu,\nu) =$

$(\frac{1}{\sqrt{2\pi}})^{10} (\frac{1}{10})^{10} \exp(-\frac{1}{2}\sum_{i=1}^{10} (X_i-\mu)^2/10^2) * (\frac{1}{\sqrt{2\pi}})^{20} (\frac{1}{10})^{20} \exp(-\frac{1}{2}\sum_{i=1}^{20} (Y_i-\mu)^2/10^2)$

$l(\mu,\nu) = \ln[L(\mu,\nu)] =$

$10 * ln[\frac{1}{\sqrt{2\pi}}] + 10 * \ln(\frac{1}{10}) - \frac{1}{2}\frac{1}{10^2}\sum_{i=1}^{10}(X_i-\mu)^2 +$

$20 * ln[\frac{1}{\sqrt{2\pi}}] + 20 * \ln(\frac{1}{10}) - \frac{1}{2}\frac{1}{10^2}\sum_{i=1}^{20}(Y_i-\nu)^2.$

score funktion

$\frac{d}{d\mu} l(\mu,\nu) = -\frac{1}{10^2}\sum_{i=1}^{10}(X_i-\mu) =$

$-\frac{1}{10^2}\sum_{i=1}^{10} X_i - 10\frac{1}{10^2}\mu.$

$\frac{d}{d\nu} l(\mu,\nu) = -\frac{1}{10^2}\sum_{i=1}^{20}(Y_i-\nu) =$

-$\frac{1}{10^2}\sum_{i=1}^{20} Y_i - 20\frac{1}{10^2}\nu$

2. Vis at $\widehat{\mu} = \frac{1}{10}\sum_{i=1}^{10} x_i = \overline{x}$

   og $\widehat{\nu} = \frac{1}{20}\sum_{i=1}^{20} y_i = \overline{y}$:

   $-\frac{1}{10^2}\sum_{i=1}^{10} X_i - 10\frac{1}{10^2}\mu = 0$ giver

   $-\sum_{i=1}^{10} X_i - 10\mu = 0$ giver

   $\widehat{\mu} = \frac{1}{10}\sum_{i=1}^{10} X_i = \overline{X}$

3. Angiv Hesse-matricen.:

   $\frac{d}{d\mu} l(\mu,\nu) = -\frac{1}{10^2}\sum_{i=1}^{10} X_i - 10\frac{1}{10^2}\mu$

   $\frac{d^2}{d^2\mu} l(\mu,\nu) = -10\frac{1}{10^2}$

   $\frac{d^2}{d\mu d\nu} l(\mu,\nu) = 0$

   H=$\begin{bmatrix} \frac{-10}{10^2} & \\ & \frac{-20}{10^2} \end{bmatrix}$

4. Angiv et 95% konfidensinterval for μ.

   $V(\widehat{\mu}) = \frac{10^2}{10}$ 499,2$\pm 1,96\sqrt{\frac{100}{10}}$ [493,0 - 505,4]

   I 2015 var OECD's tilsvarende gennemsnit for læsning på 488.

5. Giv en kommentar til dette. Inddrag det beregnede konfidensinterval fra sp.4.

   DK var "bedre"end OECD

   Antag nu at $\mu = \nu$. Den fælles parameter kaldes γ.

   Dermed haves 30 uafhængige identiske fordelte stokastiske variable.

6. Vis at $\widehat{\gamma} = \overline{z}$.

   $\widehat{\gamma} = \frac{\sum_{i=1}^{10} X_i + \frac{1}{20}\sum_{i=1}^{20} Y_i}{10+20}$ =494,2

7. Test $\mathrm{H}_0 : \mu = \nu$ $(=\gamma)$ mod $\mathrm{H}_A : \mu \neq \nu$.

   Brug et LR test. Og kommenter resultatet.

   $l(\widehat{\mu}, \widehat{\nu})$ er prop. med

   bemærk at:

   $10 * ln[\frac{1}{\sqrt{2\pi}}] + 10 * \ln(\frac{1}{10}) + 20 * ln[\frac{1}{\sqrt{2\pi}}] + 20 * \ln(\frac{1}{10}) =$

   $30 * ln[\frac{1}{\sqrt{2\pi}}] + 30 * \ln(\frac{1}{10})$

   Så derfor vil $l(\widehat{\mu}, \widehat{\nu})$ i praksis være

   $-\frac{1}{2}\frac{1}{100}\sum_{i=1}^{10}(X_i - \widehat{\mu})^2 - \frac{1}{2}\frac{1}{100}\sum_{i=1}^{20}(Y_i - \widehat{\nu})^2 =$

   $-\frac{1}{2}\frac{1}{100}\sum_{i=1}^{10}(X_i - \overline{X})^2 - \frac{1}{2}\frac{1}{100}\sum_{i=1}^{20}(Y_i - \overline{Y})^2 =$

   $-\frac{1}{2}\frac{1}{100}26.490,5 - \frac{1}{2}\frac{1}{100}167.523,0$

   $l(\widehat{\gamma})$ er i praksis

   $-\frac{1}{2}\frac{1}{100}\sum_{i=1}^{10}(X_i - \widehat{\gamma})^2 = \frac{1}{2}\frac{1}{100}\sum_{i=1}^{20}(Y_i - \widehat{\gamma})^2 =$

   $-\frac{1}{2}\frac{1}{100}$194.378,3

   $Q = -2[l(\widehat{\gamma}) - l(\widehat{\mu}, \widehat{\nu})] =$

   $\frac{1}{100}$194.378,3$-\frac{1}{100}26.490,5 - \frac{1}{100}167.523,0 =$

   $\frac{194.378{,}3 - 26.490{,}5 - 167.523{,}0}{100} = 3,6$

som er χ^2 med 1 frihedsgrad. det kritiske område er til højre for 3,84

Sss= 5,6%

Antag nu at alle 30 målinger har samme middelværdi

som betegnes γ.

8. Test $\text{H}_0 : \gamma = 488$ mod $\text{H}_A : \gamma \neq 488$. Brug et Wald test.

   Kommenter resultatet.

   $V(\widehat{\gamma}) = \frac{10^2}{30}$

   494,2$\pm 1,96 * \sqrt{\frac{100}{30}} = [490,6 - 497,8]$

   eller

   Z=$\frac{494{,}2-488}{\sqrt{\frac{100}{30}}} = 3,4 > 1,96$

9. Angiv fordelingen af U. Idet du fortsat antager at alle

   30 målinger har samme middelværdi.

   Hvor $U = \frac{\overline{X}-\overline{Y}}{10*\sqrt{\frac{1}{10}+\frac{1}{20}}}$

   $E(\overline{X}) = \gamma \quad V(\overline{X}) = \frac{10^2}{10}$

   $E(\overline{Y}) = \gamma \quad V(\overline{Y}) = \frac{10^2}{20}$

   generelt gælder

   $E(\overline{X} - \overline{Y}) = E(\overline{X}) - E(\overline{Y}) = 0$

   uafhængige

   $V(\overline{X} - \overline{Y}) = V(\overline{X}) + V(\overline{Y}) = 10^2(\frac{1}{10} + \frac{1}{20})$

   $\sqrt{V(\overline{X} - \overline{Y})} = 10\sqrt{\frac{1}{10} + \frac{1}{20}}$

   så $\frac{\overline{X}-\overline{Y}}{10*\sqrt{\frac{1}{10}+\frac{1}{20}}} \sim N(0,1)$