Eksamen på Økonomistudiet sommer 2020

Sandsynlighedsteori og statistik

9. juni 2020

(3½ times eksamen med hjælpemidler)

Besvarelsen uploades på Digital Eksamen som én pdf.fil (inkl. bilag) navngivet udelukkende med eksamensnummeret, f.eks. 12.pdf eller 127.pdf

Dette eksamenssæt består af 5 sider incl denne forside.

Denne eksamen er ændret fra at foregå på Peter Bangsvej til at foregå som en hjemmeeksamen med hjælpemidler.
Læs grundigt teksten i boksen nedenfor, så du undgår at komme i problemer med mistanke om eksamenssnyd.

Pas på at du ikke begår eksamenssnyd!

Det er fx eksamenssnyd, hvis du …

- Kopierer andres tekster uden at sætte citationstegn eller kildehenvise, så det ser ud som om det er din egen tekst. Det gælder også tekst fra gamle rettevejledninger
- Stiller din opgave til rådighed for andre under eksamen
- Kommunikerer med andre om opgaven under eksamen
- Bruger andres idéer eller tanker uden at kildehenvise, så det ser ud om det er din egen idé eller dine tanker
- Genbruger dele af en opgave, som du tidligere har indleveret og fået en bestå karakter for uden at sætte citationstegn eller kildehenvise (selvplagiering)

Du kan læse mere om reglerne for eksamenssnyd på Din Uddannelsesside og i Rammestudieordningens afs. 4.12.

Eksamenssnyd sanktioneres altid med en advarsel og bortvisning fra prøven. I de fleste tilfælde bliver den studerende også bortvist fra universitetet i et semester.

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		0,05	0,3
X=1	0,25		

1. Beregn samtlige simultane sandsynligheder for (X,Y).
2. Beregn P(X=1) og E(X).

   Betragt Z=X*Y.
3. Angiv sandsynlighedsfunktionen for Z,

   og udregn middelværdi og varians.
4. Beregn P(X=1|Z=0).
5. Beregn E(X|Z=0) og V(X|Z=0) .

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).

   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 $P(X_3 => 1)$.

   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}.

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

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.

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

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

3. Angiv Hesse-matricen.

4. Angiv et 95% konfidensinterval for μ.

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

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

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

   Dermed haves 30 uafhængige identiske fordelte stokastiske variable.

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

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

   Brug et LR test. Og kommenter resultatet.

   Antag nu at alle 30 målinger har samme middelværdi

   som betegnes γ.

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

   Kommenter resultatet.

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

   30 målinger har samme middelværdi.

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