# Rettevejledning til reeksamen på Økonomistudiet sommer 2013 Økonometri A

## Spørgsmål 1

En virksomheds vandforbrug per dag, $X$, er givet ved en normalfordeling med middelværdi 4 $m^3$ (kubikmeter) og varians 1.

1. Hvad er sandsynligheden for at forbruget er højere end 6 $m^3$ på en dag.

Fordelingen ved X er givet ved: $X\tilde{}N(4,1)$.Sandsynligheden findes ved at standardisere og anvende standard normalfordelt variabel eller ved direkte opslag i fx EXCEL. $\Pr(X > 6) = 1 - \Phi(\frac{6-4}{1}) = 0,023$

En vandspareordning indføres og det nye vandforbrug per dag, $Y$, er også givet ved en normalfordeling med middelværdi $3,8$ $m^3$ og varians 1. Lad $Z = X - Y$ være forskellen i vandforbrug mellem før og efter vandspareordningen.

2. Hvad er $\Pr(Z > 0)$?

Først findes fordelingen for Z. Denne er en lineær kombination af normalfordelte variable og dermed selv normal fordelt. Det er nødvendigt at antage uafhængighed for at kunne finde middelværdi og varians. $Z\tilde{}N(4-3,8;1+1) = N(0,2;2)$.Herefter findes sandsynligheden som ovenfor $\Pr(Z > 0) = 1 - \Phi(\frac{-0,2}{\sqrt{2}}) = 0,5562$, eller ved opslag i EXCEL for

Lad $\overline{X} = \frac{1}{100}\sum_{i=1}^{100} X_i$ være gennemsnitligt forbrug over 100 dage før vandspareordningen er etableret og lad $\overline{Y} = \frac{1}{100}\sum_{i=1}^{100} Y_i$ være gennemsnitligt forbrug over 100 dage efter vandspareordningen er etableret. Lad nu $Z = \overline{X} - \overline{Y}$.

3. Hvad er $\Pr(Z > 0)$?Kommenter i forhold til forrige spørgsmål?

Igen skal fordelingen for Z findes. Først findes fordelingen for $\overline{X}$. Denne er atter en linær kombination af normalfordelte stokastiske variable og nødvendigt at antage uafhængighed. $\overline{X}\tilde{}N(\frac{1}{100}\sum\mu_x, \frac{1}{100^2}\sum\sigma_x^2) = N(\mu_x, \frac{1}{100}\sigma_x^2) = N(4, \frac{1}{100})$ og vice versa for $\overline{Y}$,dvs. $\overline{Y}\tilde{}N(3,8;\frac{1}{100})$. Herefter kan Z bestemmes som ovenfor. $Z\tilde{}N(4-3,8;\frac{1}{100}+\frac{1}{100}) = N(0,2;\frac{2}{100})$. $\Pr(Z > 0) = 1-\Phi(\frac{-0,2}{\sqrt{\frac{2}{100}}}) = 0,9214$.Variansen af $\overline{X}$ meget mindre end af X, da den er gennemsnittet af 100 realiseringer af $X$ og vice versa for $\overline{Y}$.Dermed lettere at fastslå om Z $> 0$.

## Spørgsmål 2

I en lille kommune blev der ved sidste kommunalvalg afgivet 10.000 stemmer. De 6000 stemte på borgmesteren og 4000 stemte på modkandidaten. Vi trækker nu tilfældigt med tilbagelægning 10 vælgere: Lad $X$ være antallet af vælgere, der stemte på borgmesteren.

1. Opstil en sandsynlighedsmodel for $X$. Hvad er sandsynlighederne for, at der er henholdsvis ingen og alle stemte på borgmesteren? Dvs. find $\Pr(X = 0)$ og $\Pr(X = 10)$.

X er binomialfordelt med antalsparameter 10 og sandsynlighedsparameter 0,6. Uafhængighed kommer af tilfældig trækning med tilbagelægning. $\Pr(X = 0) = 0,000105$, $\Pr(X = 10) = 0,00604$. Findes ved EXCEL.

Antag at stikprøven er trukket uden tilbagelægning. Lad $Y$ være antallet af vælgere, der stemte på borgmesteren.

2. Opstil en sandsnlighedsmodel for $Y$. Hvad er sandsynlighederne for, at der er henholdsvis ingen og alle stemte på borgmesteren? Dvs. find $\Pr(Y = 0)$ og $\Pr(Y = 10)$. Kommenter i forhold til forrige spørgsmål.

Y er hypergeometrisk fordelt idet der er tale om tilfældige trlkninger uden tilbagelægning. Antallet af forsøg er 10, antallet af succeser i populationen er 6000 og fiaskoer er 4000. Dermed kan sandsynlighederne beregnes. $\Pr(Y = 0) = 0,000104$ og $\Pr(Y = 10) = 0,006028$. Først bemærkes at sandsynligheder næsten ens, hvilket er tilfældet hvis antallet af forsøg er lille i forhold til populationens størrelse. Dernæst kan det ogås bemærkes at sandsynligheden for ekstreme værdier (0 og 10) er mindre i den hypergeometriske fordeling, da trækning uden tilbagelægning mindsker sandsynligheden for ekstreme observationer.

Vælgerne kan opdeles i højtuddannede og lavtuddannede. Halvdelen af vælgerne er højtuddannede. Vi ved også at sandsynligheden for at en vælger stemte på borgmesteren givet at hun var højtuddanet er $0,7$ og sandsynligheden for at vælgeren stemte på modkandidaten giivet hun er lavtuddannet er $0,5$.

3. Hvad er sandsynligheden for at en vælger er højtuddannet givet vælgeren stemte på modkandidaten. Brug prior og posterior i fortolkning af resultatet..

Lad U være stokastisk variabel, der angiver uddannelse, med to udfald for lav og høj uddannelse. Vi får oplyst at $\Pr(U = high) = 0,5$.Lad S være en stokastisk variable for den kandidat der stemmes på med to udfald enten borgmester eller modkandidat. Vi får oplyst at $\Pr(X = borgmester|U = high) = 0,7$ og $\Pr(X = \text{mod}\, kandidat|U = low) = 0,5$.

Anvend Bayes opdatering. Lad modkandidatione være givet ved $M$ og borgmesteren ved $B$.

$$\begin{aligned}
\Pr(U &= high|X = M) \\
&= \frac{\Pr(X = M|U = high)|\Pr(U = high)}{\Pr(X = M)} \\
&= \frac{(1 - \Pr(X = B|U = high))\Pr(U = high)}{\Pr(X = M)} \\
&= \frac{(1 - \Pr(X = B|U = high))\Pr(U = high)}{\Pr(X = M|U = high)\Pr(U = high) + \Pr(X = M|U = low)\Pr(U = low)} \\
&= \frac{0,3 \cdot 0,5}{0,3 \cdot 0,5 + 0,5 \cdot 0,5} \\
&= 0,375.
\end{aligned}$$

Før vi får at vide at vælgeren stemmer på modkandidaten er sandsynligheden 0,5 for at vedkommende er højtuddannet (prior), men efter vi får dette oplyst finder vi at sandsynligheden er 0,375 (posterior).

## Spørgsmål 3

En butikskæde med 10 forretninger placeret i nærheden af Universitetets Campus er interesseret i at måle effekten af at annoncere for salg af øl i diverse studenteraviser. I marts måles forretningers salg af øl, derefter iværksættes en større annonce kampagne og salget af øl måles i april måned. Resultaterne er gengivet i nedenstående tabel

| Butik | marts<br>før<br>kr. | april<br>efter<br>kr. | differens<br>"effekt"<br>kr. |
|---|---|---|---|
| 1 | 1368,1 | 1360,5 | -6,3 |
| 2 | 1389,0 | 1770,6 | 381,6 |
| 3 | 1660,8 | 1623,3 | -37,5 |
| 4 | 1464,6 | 1388,4 | -76,2 |
| 5 | 1441,8 | 1580,4 | 138,6 |
| 6 | 1359,6 | 1461,9 | 102,3 |
| 7 | 1665,0 | 1486,8 | -178,2 |
| 8 | 1228,2 | 1581,9 | 353,7 |
| 9 | 945,9 | 891,6 | -54,3 |
| 10 | 1450,2 | 1181,7 | -268,5 |
| | | | |
| gns ($\overline{x}$) | 1397,2 | 1432,7 | 35,5 |
| spredning ($s$) | 206,9 | 249,7 | 211,3 |

Du skal ikke nødvendigvis bruge alle informationerne i ovenstående tabel. Det kan antages at differensen("effekten") for en butik kan beskrives ved en normalfordeling med middelværdi $\mu$ og varians $\sigma^2$.

1. Estimer de to parametre $\mu$ og $\sigma$.

$\widehat{\mu} = \sum_{i=1}^{n} X_i = \overline{X} = 35,5$

$\widehat{\sigma}^2 = s^2 = (211,3)^2$ og dermed $\widehat{\sigma} = 211,3$

2. Angiv egenskaberne ved estimatoren for $\mu$.

Middelret (dvs. at gennemsnit af estimator= den parameter der skal estimeres). Endvidere går variansen mod nul når

antallet af observationer bliver stort. Så estimatoren er konsistent

3. Udregn et 90% konfidensinterval for middelværdien (NB 90% interval)

$\bar{X} \pm t_{0,95}(9) * \sqrt{\frac{S^2}{n}} = 35,5 \pm 1,833 * \sqrt{\frac{(211,3)^2}{10}} = [-87,0 : 158,0]$
(Så der er faktisk ingen effekt, da 0 er i konfidensintervallet)

4. Test hypotesen at effekten af kampagnen er større end 20.

kommentar måske ikke det mest oplagte at teste på basis af at ovenstående interval indeholder 0.

$H_0 : \mu = 20 \quad H_A : \mu > 20$

t=$\frac{\bar{X}-\mu_0}{\sqrt{S^2/n}} = \sqrt{10}\frac{35,5-20}{211,3} =$ 0,22 som er t-fordelt med DF=9

Sss=P(t(9)>0,22)=0,41=41% hypotesen opretholdes. Dermed (kun) en effekt på 20.

Ledelsen af kæden ønsker at basere fremtidige tilsvarende undersøgelser på en mere enkel opgørelse. De vil kun tælle antallet af butikker med en forøgelse af salget og dermed estimere sandsynligheden ($p$) for at en butik har et øget salg. I ovenstående undersøgelse er der 4 butikker, der har haft et øget salg

5. Opstil en statistisk model, der beskriver dette, argumenter for modellen og estimer parameteren $p$, hvor $p$ angiver sandsynligheden for et øget salg.

X= antallet af butikker med øget salg. X vil være Bin(10,p).

der er to udfald for den enkelte butik. Der antages at være den samme sandsynlighed p, for at salget bliver øget
og butikkernes salg antages at være uafhængige af hinanden.

$\widehat{p} = \frac{x}{n} = \frac{4}{10} = 0,4$

6. Test hypotesen at $p = \frac{1}{2}$.

Her er det naturligt at vælge at alternativet skal være p $>\frac{1}{2}$.

Der observeres et antal på 4. Sss=P(X=>4) hvor X er bin $(10, \frac{1}{2})$

Sss bliver dermed 82,8% hypotesen opretholdes (dvs. ingen ændring)