Løsningsforslag vinter 2012-2013
Økonometri A
2. Årsprøve

Målbeskrivelse:

Kurset har som mål at introducere studerende til sandsynlighedsteori og statistik. Målet er, at de studerende efter at have gennemført faget kan:

- Forstå og benytte de vigtigste sandsynlighedsteoretiske begreber som: sandsynlighed, simultane-, marginale- og betingede- sandsynligheder, fordeling, tæthedsfunktion, uafhængighed, middelværdi, varians og kovarians samt at selvstændigt kunne anvende disse begreber på konkrete problemstillinger
- Kende resultatet fra den centrale grænseværdi sætning
- Kende og genkende de mest anvendte diskrete og kontinuerte fordelinger som: Bernoulli, Binomial, Poisson, multinomial, negative binomial fordeling, hypergeometrisk, geometrisk, lige-, normal-, Chi-i-anden-, eksponential, gamma-, t-, F-fordeling samt at selvstændigt kunne arbejde med disse fordelinger i konkrete problemstillinger
- Forstå de vigtigste statistiske begreber som: tilfældige udvælgelse, likelihood funtionen, sufficiens, stikprøvefunktion, egenskaber ved stikprøvefunktionen, estimation heruden maksimum likelihood og moment estimation, konsistens, konfidensinterval, hypoteseprøvning, teststørrelser, hypoteser, testsandsynlighed, signifikansniveau og type I og II fejl
- Være i stand til selvstændigt at gennemføre en simpel statistisk analyse, som involverer estimation, inferens og hypoteseprøvning, f.eks. sammenligning af middelværdien i to populationer eller uafhængighedstest for diskrete stokastiske variable.
- Indlæse og kombinere datasæt, lave nye variable, udtrække en stikprøve og udføre simple statistiske analyser ved hjælp af statistik-pakken SAS
- Beskrive resultatet af egne analyser og overvejelser i et klart og tydeligt sprog

Opgave 1

Et universitet undersøger sammensætningen af sine studerende. Lad E være en stokastisk variabel, der angiver, om en studerende er færdig på normeret tid. Der findes to typer af studerende A og B. Fra tidligere undersøgelser er det kendt, at 26 pct. af type A studerende består til tiden, og 38 pct. af type B studerende består til tiden.

1. Hvor stor en andel er type A på universitetet, hvor 34 pct. består til tiden? Angiv teoremer, der anvendes.

$P(E) = P(E|A)P(A)+P(E|B)P(B) = P(E|A)P(A)+P(E|B)(1-P(A)) = 0,26P(A) + 0,38(1 - P(A)) = 0,34 \Longrightarrow P(A) = \frac{0,34-0,38}{0,26-0,38} = \frac{1}{3}$. Loven om den totale sandsynlighed anvendes.

Universitetet ønsker at få flere studerende af type B, men ved ikke præcis hvilke studerende der er type A og type B. Men universitetet kender den simultane fordeling af type og studieretning fra andre studier. Den er givet ved følgende tabel i pct.

	Studieretning		
Type	X	Y	Z
A	12	20	1
B	23	10	34

Studieretningerne angives ved X, Y, og Z. Antag at sandsynligheden for at bestå på normeret tid ikke afhænger af studieretning, når der betinges på type A (eller B). Dvs. $P(E|A, j) = P(E|A)$, $j = X, Y, Z$.

2. Hvad er sandsynligheden for at bestå for de forskellige studieretninger?

Igen kan loven om den totale sandsynlighed anvendes. Da betingede sandsynligheder også er sandsynligheder. $P(E|j) = P(E|A, j)P(A|j)+P(E|B, j)P(B|j) = P(E|A)P(A|j) + P(E|B)P(B|j)$.

Vha. tabellen og loven om den totale sandsynlighed får vi: $P(E|X) = 0,26 \cdot \frac{12}{35} + 0,38 \cdot \frac{23}{35} \approx 0,34$. $P(E|Y) = 0,26 \cdot \frac{20}{30} + 0,38 \cdot \frac{10}{30} = 0,3$. $P(E|Z) = 0,26 \cdot \frac{1}{35} + 0,38 \cdot \frac{34}{35} \approx 0,38$.

Lidt drastisk beslutter universitetet at lukke studieretning Y.

3. Hvor mange studerende består til tiden med denne politik?

$P(E) = P(E|X)P(X) + P(E|Z)P(Z) = 0,34 \cdot 0,5 + 0,38 \cdot 0,5 = 0,36$.Dvs. nu består lidt flere, da Y studiet har mange type A studerende, der har den mindste sandsynlighed for at bestå.

Opgave 2

I gamle dage fik Nationalbanken lavet guldmønter hos en møntmager. For en korrekt fremstillet mønt var vægten givet ved $N(120, 1)$, idet det ikke var muligt med den givne teknologi at få vægten helt præcis.

1. Hvilken værdi havde1 pct. percentilen for en mønt, når mønten var korrekt fremstillet?

$Z = \frac{X-120}{1}, Z_{0,01} = -2,33$.Dvs. $X_{0,01} = 120 - 2,33 \cdot 1 = 117,67$

Mange møntmagere var uærlige dengang. For at sikre mod snyd udtog Nationalbanken en tilfældig stikprøve for at teste ved vejning om møntmageren var uærlig. Møntmageren ansås for at være uærlig, hvis vægten af de udtagne mønter var under den forventede værdi fratrukket en værdi, R. Nationalbanken udtog 100 mønter. Antag uafhængighed af de enkelte mønters vægt.

2. Find fordelingen for de 100 mønter. Hvis sandsynligheden for at straffe en ærlig møntmager skulle være mindre end 1 pct., hvad skulle R så være?

$Y = \sum_{i=1,\ldots 100} X_i$, pga. uafhængighed er $Y\tilde{}N(12000, 100)$. $Z = \frac{Y-12000}{10}, Z_{0,01} = -2,33$,Dvs. $Y_{0,01} = 12000 - 2,33 \cdot 10 = 11976,7$. Hvis $R = 12000 - 11976,7 = 23,3$ så vil sandsynligheden være højst 1 pct. for at en ærlig møntmager straffes.

En møntmager tog $0,1$ af hver mønt og stak i lommen.

3. Hvad var sandsynligheden for at den 'uærlige' møntmager blev straffet med den fundnu værdi for R?

Lad Q være 100 mønter fra denne møntmager. $Q\tilde{}N(11990, 100).P(Q \leq 12000 - R) = P(Q \leq 11976,7) = P(Z \leq \frac{11976,7-11990}{10}) = P(Z \leq -1,33) = 9,2\%$.Dvs. lille sandsynlighed for at blive straffet, men dog større end for en ærlig møntmager.

Opgave 3

Fra en fagforening med ca. 10.000 medlemmer er der i august 2012 udvalgt 18 tilfældige medlemmer.

Deres årlige løn er angivet i nedenstående tabel.

nr.	løn i kroner
1	507.210
2	429.908
3	381.885
4	476.345
5	619.946
6	486.789
7	686.245
8	480.168
9	594.396
10	487.908
11	573.509
12	620.133
13	559.988
14	537.888
15	561.573
16	592.065
17	576.306
18	567.075

Det kan antages, at lønnen kan beskrives med en normalfordeling. Dvs. $X_i \sim N(\mu, \sigma^2)$ $i = 1, 2, 3, ..., 18$.

Du kan benytte at gennemsnittet af ovenstående 18 observationer er 541.074 og at spredningen (s) er 74.859.

1. Estimer parametrene μ og σ.Angiv egenskaberne for estimatoren for μ.

   $\widehat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = 541.074$ her er n=18

   $\widehat{\sigma 2} = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = (74.859)^2$

   og dermed er $\widehat{\sigma} = s = 74.859$

   $E(\widehat{\mu}) = \mu$. og $V(\widehat{\mu}) = \frac{\sigma 2}{\sqrt{n}}$ som går mod nul når antallet af observationer øges. Og dermed er estimatoren konsistent.

2. Angiv et 95% konfidensinterval for μ.

   95% konfidensintervallet er givet

   $\bar{X} \pm t_{0,975}(n-1)\frac{S}{\sqrt{n}}$ her er $t_{0,975}(11) = 2,11$

   så intervallet bliver fra 503.8 til 578.3

I den offentlige debat er det blevet hævdet, at denne fagforening har en gennemsnits løn, der er større end kr. 500.000.

3. Test hypotesen at den gennemsnitlige løn for fagforeningen er lig kr. 500.000 og angiv alternativ hypotesen.

   $\text{H}_0 : \mu = 500.000$ mod $\text{H}_A : \mu > 500.000$ (ensidet test) trststørrelse t=$\frac{\bar{X}-500.000}{\sqrt{\frac{S^2}{18}}} = \sqrt{18}\frac{(\bar{X}-500.000)}{74.859} = 2,33$ som er t-fordelt med 17 frihedsgrader. Signifikanssandsynligheden (eller p-værdi) bliver 1,67% som er mindre end 5% og dermed afvises H_0 og konklusionen bliver at gennemsnitslønnen er er større end kr. 500.000.

Blandt de 18 udtrukne medlemmer er der 12 mænd og dermed 6 kvinder.

4. Argumenter for at man kan bruge binomialfordelingen til at beskrive antallet af mænd blandt de 18 udtrukne. Overvej om der kunne være alternative fordelinger.

   Der er ca. 10.000 medlemmer i fagforeningen, som kan opdeles i to grupper mænd og kvinder. Trækkes en simpel tilfældig stikprøve blandt i denne population, så vil stikprøven blive hypergeometrisk fordelt. Denne hypergeometriske fordeling kan approximeres til binomialfordeling (svarende til at vi lægger medlemmer tilbage efter at de er blevet udtrukket).

5. Estimer andelen af mænd i fagforeningen og udregn et tilhørende 95% konfidensinterval.

   $\widehat{p} = \frac{x}{n} = \frac{12}{18} = \frac{2}{3}$

   95% konfidensintervallet (approximativt) er givet

   $\widehat{p} \pm 1,96\sqrt{\frac{\widehat{p}*(1-\widehat{p})}{n}} = \frac{2}{3} \pm 1,96\sqrt{\frac{\frac{2}{3}*\frac{1}{3}}{18}}$ giver et interval fra 0,45 til 0,88

En større stikprøve trækkes i januar 2013. I alt 25 medlemmer udtrækkes og nu er er 13 mænd og 12 kvinder.

6. Test om fordelingen af mænd er uændret fra august 2012 til januar 2013

   Der er tale om at teste sandsynlighedsparametrene i to uafhængige binomialfordelinger.

   $H_0 : p_1 = p_2 \quad H_A : p_1 <> p_2$ (forskellig fra)

   Der kan udføres tre test, (der viser ikke chi^2 testen og ej heller testen baseret på en betingning af summen af de to binomialfordelinger som giver test i den hypergeometriske fordeling)

   U=$\frac{\widehat{p_1}-\widehat{p_2}}{\sqrt{\widehat{p}*(1-\widehat{p})*(\frac{1}{n_1}+\frac{1}{n_2})}}$ her er $\widehat{p} = \frac{12+13}{18+25} = \frac{25}{43}$

   U= $\frac{\frac{12}{18}-\frac{13}{25}}{\sqrt{\frac{25}{43}*(1-\frac{25}{43})*(\frac{1}{18}+\frac{1}{25})}} = 0,96$ som ligger i acceptområdet (er numerisk mindre end de "berømte 1,96) og Signifikanssandsynligheden (p-værdi) bliver 33,6% som er større end grænsen på 5 %. Hypotesen opretholdes, der kan ikke konstateres en ændring.