ECON 2130 EKSAMEN 2010 VÅR

TALLSVAR

Det anbefales at de 9 deloppgavene merket med A, B,… teller likt – uansett variasjon i vanskelighetsgrad. Svarene er gitt i << … >>.

Oppgave 1

A. I en vanlig kortstokk på 52 kort er det fire ess – ett av hver farge: hjerter, ruter, kløver og spar. Vi trekker et kort rent tilfeldig fra kortstokken. La *A* være begivenheten at det uttrukne kortet er et ess. La *R* være begivenheten at det uttrukne kortet er et "rødt kort" – det vil si et hjerter-kort eller et ruter-kort. Det er i alt 26 "røde kort" i kortstokken.

(**i**) Finn sannsynlighetene, $P(A)$ og $P(A \cup R)$.

(**ii**) Er *A* og *R* disjunkte? Begrunn svaret.

(**iii**) Er *A* og *R* stokastisk uavhengige? Begrunn svaret.

<< **Svar: (i):** $P(A) = \frac{4}{52} = \frac{1}{13}$,

$$P(A \cup R) = P(A) + P(R) - P(A \cap R) = \frac{1}{13} + \frac{1}{2} - \frac{1}{26} = \frac{14}{26} = \frac{7}{13}$$

Man kan også resonnere at $A \cup R$ omfatter 28 muligheter av 52.

(**ii**) Nei (**iii**) Ja, $P(A \cap R) = \frac{1}{26} = \frac{1}{13} \cdot \frac{1}{2} = P(A)P(R).$

>>

B. Vi lager en mini-kortstokk bestående av fem kort, spar-ess, spar-to, spar-tre, spar-fire og spar-fem. Fra mini-kortstokken trekkes tilfeldig ett og ett kort uten tilbakelegging inntil esset dukker opp. La *X* være antall trekninger som må til før esset kommer. Hvis esset trekkes ut i første trekning, får *X* verdien 1, hvis esset kommer i annen trekning, får *X* verdien 2, osv.

(**i**) Vis at *X* er uniformt fordelt over tallene, 1,2,…,5. Med andre ord, vis at

$$P(X = x) = \frac{1}{5}, \quad x = 1, 2, \ldots, 5.$$

(**ii**) Finn $E(X)$ og $\text{var}(X)$.

<< **Svar:** (**i**) La E_j være "ess i *j*-te trekning". $P(X = 1) = P(E_1) = \frac{1}{5}$,

$$P(X = 2) = P(\overline{E}_1 \cap E_2) = \frac{4}{5} \cdot \frac{1}{4} = \frac{1}{5},$$

$$P(X=3) = P(\bar{E}_1 \cap \bar{E}_2 \cap E_3) = \frac{4}{5}\cdot\frac{3}{4}\cdot\frac{1}{3} = \frac{1}{5} \quad \text{osv}$$

Man kan også resonnere: Om vi stokker kortene og legger dem ved siden av hverandre med baksiden opp, er det like sannsynlig om esset ligger på første, andre, tredje…osv. plass. Dette gjelder selv om vi stokker kortene hver gang etter å ha plukket opp et kort pga symmetrien.

(ii) $E(X) = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3, \quad E(X^2) = \frac{1+4+9+16+25}{5} = \frac{55}{5} = 11$

$\operatorname{var}(X) = 11 - 3^2 = 2$

>>

C. Å observere X en gang som beskrevet i punkt **B** kaller vi et *spill*. Hvis X får verdien 4 eller 5, sier vi at vi har fått et *langt spill*.

(i) Hva er sannsynligheten for at et vilkårlig spill blir et langt spill?

(ii) Beregn sannsynligheten tilnærmet for at minst 25 av 50 spill blir lange spill. Bruk heltallskorreksjon ved beregningen.

<< **Svar:** **(i)** $P(X \geq 4) = \frac{2}{5} = 0.4$

(ii) La Y være antall lange spill blant $n = 50$ spill. $Y \sim \text{bin}(n = 50, p = 0.4)$.

$E(Y) = 20, \quad \operatorname{var}(Y) = 50\frac{2}{5}\cdot\frac{3}{5} = 12$, som er stor nok til at normaltilnærmelsen skulle være akseptabel. Med heltallskorreksjon (av regel 5.20 i Løvås):

$$P(Y \geq 25) = 1 - P(Y \leq 24) \approx 1 - G\left(\frac{24,5-20}{\sqrt{12}}\right) = 1 - G(1,30) = 1 - 0,9032 = 0,0968$$

(Uten heltallskorreksjon: $P(Y \geq 25) \approx 1 - G(1,15) = 1 - 0,8749 = 0,1251$)

(Eksakt verdi etter Excel: 0,098)

>>

Oppgave 2

Vi er interessert i å sammenligne prisnivået på hårklipp i Oslo og noen andre større byer i Norge basert på data fra 2005 (publisert i Forbrukerrapporten for oktober 2005). De større byene i undersøkelsen er representert ved Oslo, Trondheim, Bergen, Kristiansand og Tromsø. Materialet omfatter også tall fra noen mindre steder som vi ikke skal se på her.

Prisene gjelder "dameklipp" og "herreklipp" som her betyr:

Dameklipp - Vask, klipp (to-tre centimeter) og føn av en kvinne med langt hår og normal tykkelse.

Herreklipp - Vask, klipp (to-tre centimeter) og tørk av en mann med kort hår og normal tykkelse.

Rent tilfeldige utvalg på 10 frisørsalonger ble trukket fra hver av de fem byene, og intervjueren, som utga seg for å være en vanlig kunde, fikk prisene oppgitt ved henvendelse på telefon. Istedenfor ett rent tilfeldig utvalg på 50 fra alle byene slått sammen, har vi altså her fem rent tilfeldige utvalg på 10 – ett fra hver av de fem byene. Resultatene i kroner er gitt i tabell 1.

Tabell 1 Priser i kroner på dameklipp og herreklipp august 2005

		Andre byer							
OSLO		TRONDHEIM		BERGEN		KRISTIANSAND		TROMSØ	
Dame	Herre	Dame	Herre	Dame	Herre	Dame	Herre	Dame	Herre
480	360	505	335	430	295	490	390	450	270
670	670	507	280	460	255	399	299	450	295
430	330	500	280	490	300	465	320	450	324
325	250	500	350	485	340	385	220	400	260
325	305	630	360	465	320	470	275	419	309
575	575	530	280	390	290	430	230	480	370
489	389	530	310	410	290	400	230	300	200
500	350	557	314	450	200	325	165	430	330
499	369	450	350	340	215	469	339	505	350
370	370	570	330	455	345	450	320	450	260

I denne oppgaven skal vi utelukkende se på prisene for herreklipp i de fire byene under kategorien "andre byer". Vi har da $n = 40$ observasjoner. Disse antas å være observasjoner av stokastiske variable, $Y_1, Y_2, \ldots, Y_n$, som vi antar er uavhengige og identisk fordelte (*uid*), med forventning, $E(Y_i) = \mu$, og varians, $\operatorname{var}(Y_i) = \sigma^2$, for $i = 1, 2, \ldots, n$, der μ og σ^2 oppfattes som ukjente populasjonsstørrelser. μ tolkes som det gjennomsnittlige prisnivået generelt på herreklipp i "andre byer". Merk at, på grunn av måten utvalget er trukket på, så innebærer modellen en implisitt forutsetning om at prisnivået for herreklipp er like stort i Trondheim, Bergen, Kristiansand og Tromsø.

For å lette beregningene i denne oppgaven oppgis (der $y_1, y_2, \ldots, y_{40}$ betegner de observerte herre-prisene fra "Andre byer"):

$$\sum_{i=1}^{40} y_i = 11\,795 \quad \text{og} \quad \sum_{i=1}^{40} (y_i - \bar{y})^2 = 103\,224$$

A. (**i**) Forklar hvorfor $\hat{\mu} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ er en forventningsrett estimator for μ.

(**ii**) Sett opp en forventningsrett estimator for $\text{var}(\hat{\mu})$ og begrunn forventningsrettheten.

[**Hint:** Du kan benytte resultatet at hvis *uid*-modellen gjelder for $Y_1, Y_2, \ldots, Y_n$ der den felles populasjonsvariansen er σ^2, gjelder $E\left(S^2\right) = \sigma^2$ der $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$. Du trenger ikke å vise dette.]

(**iii**) Beregn estimatet basert på $\hat{\mu}$, samt estimert standardfeil for $\hat{\mu}$.

<< **Svar:** (**i**): Av regel 4.12 i Løvås følger

$$E(\hat{\mu}) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}n\mu = \mu$$

(**ii**): Regel 4.17 gir $\text{var}(\hat{\mu}) = \frac{\sigma^2}{n}$ med $\frac{S^2}{n}$ som forventningsrett estimator. Forventningsrettheten følger av regel 4.7 eller avregel 4.12.

(**iii**) Estimat: $\hat{\mu}_{obs} = \frac{11795}{40} = 294.88$, $SE(\hat{\mu}) = \sqrt{\hat{\text{var}}(\hat{\mu})_{\text{obs}}} = \sqrt{\frac{103224}{39 \cdot 40}} = 8.1344$

\>>

B. (**i**) Sett opp en formel for et tilnærmet 99% konfidensintervall for μ.

(**ii**) Beregn intervallet.

<< **Svar:** 0.005-kvantilen i N(0, 1) er 2.576. Et tilnærmet 99% KI er dermed

$$\hat{\mu} \pm (2.576)SE(\hat{\mu}) = \hat{\mu} \pm (2.576)\frac{S}{\sqrt{n}} \quad \to 294.88 \pm 20.95 = [273.93,\ 315.83],$$

eller $[274, 316]$. >>

C. Nå er det ikke sikkert at prisnivået for herreklipp er likt i de fire byene. La $\mu_1, \mu_2, \mu_3, \mu_4$ betegne prisnivået i de fire byene, der indeksen 1 står for Trondheim, 2 for Bergen, 3 for Kristiansand og 4 for Tromsø. For å ta hensyn til muligheten for forskjellig prisnivå, kan vi endre litt på modellen som følger:

Vi antar fortsatt at $Y_1, Y_2, \ldots, Y_{40}$ er uavhengige og med samme varians, $\text{var}(Y_i) = \sigma^2$, men vi lar forventningen variere med byen Y_i stammer fra. Vi antar at de 10 første Y_i-ene stammer fra Trondheim, de 10 neste fra Bergen osv. som framgår av tabell 2:

Tabell 2 Ny modell for priser på herreklipp fra "andre byer".

By	Variable	$E(Y_i)$	$\text{var}(Y_i)$
Trondheim	$Y_1,\dots,Y_{10}$	μ_1	σ^2
Bergen	$Y_{11},\dots,Y_{20}$	μ_2	σ^2
Kristiansand	$Y_{21},\dots,Y_{30}$	μ_3	σ^2
Tromsø	$Y_{31},\dots,Y_{40}$	μ_4	σ^2

(i) Vis at estimatoren fra punkt **A**, $\hat{\mu} = \overline{Y} = \frac{1}{40}\sum_{i=1}^{40} Y_i$, er en forventningsrett estimator for $\mu = \frac{1}{4}(\mu_1 + \mu_2 + \mu_3 + \mu_4)$, der μ nå er en ny parameter definert som gjennomsnittet av $\mu_1,\dots,\mu_4$.

(ii) Vis at $\text{var}(\hat{\mu}) = \frac{\sigma^2}{40}$.

(iii) Foreslå en estimator for σ^2 som er forventningsrett under denne modellen.

<< **Svar:** **(i)** Av regel 4.12,

$$E(\hat{\mu}) = \frac{1}{40}(10\mu_1 + 10\mu_2 + 10\mu_3 + 10\mu_4) = \frac{1}{4}(\mu_1 + \mu_2 + \mu_3 + \mu_4) = \mu$$

(ii) Av regel 4.17, $\text{var}(\hat{\mu}) = \frac{1}{n^2}\text{var}\left(\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{var}(Y_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$

(iii) Modellen er ikke lenger uid siden forventningene varierer. Men de 10 Y_i-ene fra hver enkelt by er uid. Hvis vi betegner Y_i-ene fra by *j* som $Y_{ij}, \quad i = 1,\dots,10, \quad j = 1,\dots,4$, og $\overline{Y}_j$ gjennomsnittet i by *j*, blir $S_j^2 = \frac{1}{9}\sum_{j=1}^{4}(Y_{ij} - \overline{Y}_j)^2$ forventningsrett for σ^2 for $j = 1,\dots,4$. Den kombinerte estimatoren $\hat{\sigma}^2 = \frac{1}{4}\left(S_1^2 + S_2^2 + S_3^2 + S_4^2\right)$ blir dermed forventningsrett.

>>

Oppgave 3

Vi ønsker å sammenligne prisnivået på hårklipp i Oslo med nivået i de andre byene, der situasjonen er som beskrevet i oppgave 2, og vi vil benytte regresjonsmodellen som formulert i Løvås til dette.

Vi starter med prisene for herreklipp. For en vilkårlig frisørsalong som trekkes ut observeres to variable, x og Y, der Y er prisen på herreklipp, og x en såkalt indkator-variabel for Oslo som får verdien 1 hvis salongen kommer fra Oslo og verdien 0 hvis salongen er fra en av de fire andre byene.

Vi har i alt $n = 50$ observasjoner hvorav 10 fra Oslo og 40 fra "andre byer". La Y_i være prisen på herreklipp i frisørsalong i ($i = 1, 2, \ldots, 50$), og x_i er lik 1 hvis salong nr. i er fra Oslo og lik 0 ellers. Vi antar at $Y_1, Y_2, \ldots, Y_n$ er uavhengige og normalfordelte med samme ukjente varians, $\text{var}(Y_i) = \sigma^2$, og forventning, $E(Y_i) = \mu(x_i) = \alpha + \beta x_i$, $i = 1, 2, \ldots, n$.

Regresjonsfunksjonen, $E(Y) = \mu(x) = \alpha + \beta x$, er altså her bare definert for to verdier av x ($x = 1$ og $x = 0$). $\mu(1)$ tolkes som prisnivået på herreklipp i Oslo generelt (i august 2005), og $\mu(0)$ som prisnivået i de fire andre byene (idet vi antar at de fire byene har likt prisnivå). α og β anses som ukjente populasjonsstørrelser.

For å lette beregningene nedenfor, oppgis følgende mellomresultater i tabell 3:

Tabell 3 Mellomresultater for prisene på herreklipp fra tabell 1

	Formel	*Observert verdi*
n	antall observasjoner	50
$\bar{x}$	$\frac{1}{n}\sum_{i=1}^{n} x_i$	0,2
$\bar{Y}$	$\frac{1}{n}\sum_{i=1}^{n} Y_i$	315,26
s_x^2	$\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$	0,1633
S_y^2	$\frac{1}{n-1}\sum_{i=1}^{n} (Y_i - \bar{Y})^2$	6780,89
S_{xy}	$\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})$	16,6408
SS_E	kvadratsum av residualer	249172,-

A.

(**i**) Sett opp formler for minste kvadraters estimatorene, $\hat{\alpha}$ og $\hat{\beta}$, og beregn estimatene. Forklar kort hva som er forskjellen på en estimator og et estimat.

(**ii**) Estimer standardfeilen til $\hat{\beta}$.

(**iii**) Hvor mange % av variasjonen av Y_i-ene i data blir "forklart" av variabelen *x*?

<< **Svar:** (**i**) $\hat{\beta} = \dfrac{S_{xy}}{s_x^2} \quad \to \hat{\beta}_{obs} = \dfrac{16.6408}{0.1633} = 101.9$.

$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{x} \quad \to \hat{\alpha}_{obs} = 315.26 - (101.90)(0.2) = 294.88$

Forskjellen på estimator og estimat er ….

(**ii**) $\hat{\sigma} = \sqrt{\dfrac{SS_E}{48}} \quad \to \sqrt{\dfrac{149172}{48}} = 72.0492$

$$SE(\hat{\beta}) = \frac{\hat{\sigma}}{\sqrt{49 s_x^2}} \quad \to \frac{72.0492}{\sqrt{49(0.1633)}} = 25.4705$$

(**iii**) $r = \dfrac{S_{xy}}{\sqrt{S_y^2 s_x^2}} \quad \to 0.500 \quad \Rightarrow 100r^2 = 25\%$

>>

B. (**i**) Forklar hvorfor β kan tolkes som forskjellen mellom prisnivået på herreklipp i Oslo og prisnivået på herreklipp i de fire andre byene.

(**ii**) Tyder data på at prisnivået på herreklipp er høyere i Oslo enn i de fire andre byene?. Dvs. sett opp og gjennomfør en test for $H_0: \ \beta \le 0$ mot alternativet $H_1: \ \beta > 0$. Bruk signifikansnivå 1% og formuler en konklusjon.

Merk: Hvis du ikke finner akkurat den kvantilen du trenger i tabellene dine, foreslå en verdi på øyemål basert på de nærmeste verdiene i tabellen.

<<**Svar:** (**i**) Forskjell i prisnivå er $\mu(1) - \mu(0) = \alpha + \beta - \alpha = \beta$

(**ii**) Testobservatoren er av formen $T = \dfrac{\hat{\theta} - \theta_0}{SE(\hat{\theta})} = \dfrac{\hat{\beta}}{SE(\hat{\beta})}$ som er t-fordelt med 48 frihetsgrader hvis $\beta = 0$. Med 1% nivå og ensidig test forkaster vi H_0 hvis $T \ge t_{48,\,0.01}$. Tabell D5 i Løvås gir $t_{45,\,0.01} = 2.412$ og $t_{50,\,0.01} = 2.403$, så et naturlig forslag på $t_{48,\,0.01}$ kunne være 2.41.

Observert, $T = \dfrac{\hat{\beta}}{SE(\hat{\beta})} \quad \to \dfrac{101.90}{25.4705} = 4.00..$

Konklusjon: Forkast H_0 - dvs det er sterk evidens i data for at prisnivået er høyere i Oslo.

>>

C. Vi ønsker å gjennomføre tilsvarende analyse som i punkt **B** for prisene på dameklipp. Problemet er altså om data tyder på at prisnivået på dameklipp er høyere i Oslo enn i de fire andre byene. Vi benytter samme modell som for herreprisene – den eneste forskjellen er at Y_i nå står for prisen for dameklipp i frisørsalong *i* i utvalget.

Fra utskriften til en regresjonskjøring i Excel henter vi minste kvadraters estimater på regresjonsparametrene og deres standardfeil som referert i tabell 4.

Tabell 4 Regresjonsestimater for priser på dameklipp fra Oslo og fire andre byer

Regresjons-parametre	Estimat	Standardfeil
α	456,775	11,7776
β	9,525	26,3354

(i) Beregn verdien av testobservatoren du ville bruke for å teste $H_0: \ \beta \le 0$ mot $H_1: \ \beta > 0$.

(ii) Beregn p-verdien tilnærmet for testen ved å benytte resultatet at

$\dfrac{\hat{\beta}-\beta}{SE(\hat{\beta})} \overset{\text{tilnærmet}}{\sim} N(0,1)$ uansett hva den sanne verdien av β er. (Dette gjelder

når antall observasjoner er så stor som i dette tilfellet).

Kommenter p-verdien i lys av problemstillingen.

(iii) Hva ville p-verdien blitt dersom problemet hadde vært å teste $H_0: \ \beta = 0$ mot $H_1: \ \beta \ne 0$?

<<**Svar:** **(i)** $T = \dfrac{\hat{\beta}}{SE(\hat{\beta})} \quad \to T_{obs} = \dfrac{9.525}{26.3354} = 0.36...$

(ii) Hvis $\beta = 0$ er ifølge det generelle resultatet, $T \overset{\text{tilnærmet}}{\sim} N(0,1)$, og p-verdien blir p-verdi $= P_{\beta=0}(T \ge 0.36) \approx 1 - G\left(0.36\right) = 1 - 0.6406 = 0.3594$ (dvs ca 36%), som indikerer liten evidens for at prisnivået for dameklipp er høyere i Oslo enn andre byer.

(iii) Vi bruker samme testobservator med observert verdi $T_{obs} = 0.36.$. P-verdien for det tosidige problemet blir da

$$2 \cdot P_{\beta=0}\left(T \geq | T_{obs} |\right) = 2 \cdot P_{\beta=0}(T \geq 0.36) \approx 2 \cdot (0.36) = 0.72 \text{ (eller 72\%).}$$

>>