NTNU

Kunnskap for ei betre verd

Institutt for matematiske fag

Eksamensoppgåve i **ST1201/ST6201 Statistiske metoder**

Fagleg kontakt under eksamen: Nikolai Ushakov

Tlf: 45128897

Eksamensdato: 04. desember 2015

Eksamenstid (frå–til): 09:00 – 13:00

Hjelpemiddelkode/Tillatne hjelpemiddel: C:
– Tabeller og formler i statistikk, Tapir forlag,
– K.Rottman. Matematisk formelsamling,
– Stempla gult A4-ark med egne håndskrevne formlar og notat,
– Kalkulator: HP30S, Citizen SR-270X, Citizen SR-270X College eller Casio fx-82ES PLUS.

Annan informasjon:

Målform/språk: nynorsk

Sidetal: 4

Sidetal vedlegg: 0

Kontrollert av:

Dato Sign

Oppgåve 1

La X vere χ^2-fordelt med $2n$ frihetsgradar ($n > 2$).

a) Vis at

$$\mathrm{E}(X^{-1}) = \frac{1}{2(n-1)} \quad \text{og} \quad \mathrm{E}(X^{-2}) = \frac{1}{4(n-1)(n-2)}.$$

Oppgåve 2

Litteraturen oppgir at gjennomsnittleg lengd i Norge av halen til ein pattedyrart er 30 cm. Ein biolog mener at gjennomsnittet (det vil seie forventningsverdien for halelengda til eit tilfeldig valt individ) er større, og ho gjer eit forsøk for å undersøkje dette. Ho får målt halelengdene y_i til eit tilfeldig utvalg på 10 individ, og får desse resultata (i cm):

y_i	32.8	36.8	30.9	34.0	38.2	33.4	21.0	33.7	34.6	26.2

Det blir oppgitt at $\bar{y} = 32.16$ og $\sum(y_i - \bar{y})^2 = 233.324$.

a) Utfør ein test for å undersøkje om forventa halelengd er større enn 30 cm. Gå ut frå at halelengda er normalfordelt. Bruk signifikansnivå $\alpha = 0.05$.

b) Finn eit 95%-konfidensintervall for forventa halelengd.

Også breiddegraden x_i der dyret oppheldt seg då halen vart målt vart registrert. Tabellen viser breiddegrad og halelengt for kvart dyr:

x_i	63.9	61.5	64.8	65.5	59.0	58.5	68.5	66.0	66.0	66.8
y_i	32.8	36.8	30.9	34.0	38.2	33.4	21.0	33.7	34.6	26.2

Gå ut frå ein lineær regresjonsmodell, der breiddegrad er forklaringsvariabel og halelengd responsvariabel. Biologen har ein mistanke om at halelengda minkar med aukande breiddegrad.

Det blir oppgitt at $\bar{x} = 64.05$, $\sum(x_i - \bar{x})^2 = 100.465$, $\sum(x_i - \bar{x})(y_i - \bar{y}) = -105.88$.

c) Estimer regresjonslinja (finn $\hat{\beta}_0$ og $\hat{\beta}_1$). Utfør ein test for å undersøke mistanke til biologen. Bruk signifikansnivå 0.05. Bruk at $\sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 121.737$.

Oppgåve 3

Gå ut ifrå at $X_1, ..., X_n$ er eit tilfeldig utval frå ei normalfordeling med forventning μ_x og varians σ^2, og at $Y_1, ..., Y_m$ er eit tilfeldig utval frå ei normalfordeling med forventning μ_y og varians $k\sigma^2$. Gå dessutan ut ifrå at X-ane og Y-ane alle er uavhengige av kverandre. Parametrane μ_x og μ_y er ukjente og vi skal i denne oppgåva se på korleis vi kan estimere og lage konfidensintervall for differansen $\mu_x - \mu_y$.

Som estimatorar for μ_x og μ_y brukar vi henholdsvis $\bar{X}$ og $\bar{Y}$.

a) Grunngi at $\bar{X} - \bar{Y}$ er normalfordelt og vis at

$$\mathrm{E}(\bar{X} - \bar{Y}) = \mu_x - \mu_y, \quad \mathrm{Var}(\bar{\mathrm{X}} - \bar{\mathrm{Y}}) = \sigma^2 \left(\frac{1}{n} + \frac{k}{m} \right).$$

b) Gå i dette punktet ut ifrå at begge parametrane σ^2 og k er kjente. Utlei då eit $(1-\alpha)$-konfidensintervall for differansen $\mu_x - \mu_y$.

I resten av oppgåva skal vi gå ut ifrå at parametaren σ^2 er ukjent, mens parametaren k er kjent.

La S_x^2 og S_y^2 vere empirisk varians for henholdsvis X-ene og Y-ene, dvs.

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2, \quad S_y^2 = \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \bar{Y})^2.$$

Då har vi som kjent at

$$\frac{(n-1)S_x^2}{\sigma^2} \sim \chi^2_{n-1}, \quad \frac{(m-1)S_y^2}{k\sigma^2} \sim \chi^2_{m-1},$$

samt at S_x^2 og $\bar{X}$ er uavhengige og S_y^2 og $\bar{Y}$ er uavhengige.

c) Vis at

$$S_p^2 = \frac{n-1}{n+m-2} S_x^2 + \frac{m-1}{k(n+m-2)} S_y^2$$

er en forventningsrett estimator for σ^2 og at

$$(n+m-2)S_p^2/\sigma^2 \sim \chi^2_{n+m-2}.$$

Vis vidare at

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{S_p^2(\frac{1}{n} + \frac{k}{m})}}$$

er Student t-fordelt med $n + m - 2$ fridomsgrader. når du gir svara for dette punktet, spesifiser spesielt kva for kjente eigenskapar og sammanhengar mellom ulike typar fordelinger du nytter, og grunngi kvifor vilkåra for desse eigenskapane er oppfylte i den aktuelle situasjonen.

d) Utlei eit $(1 - \alpha)$-konfidensintervall for differansen $\mu_x - \mu_y$.

Gå ut ifrå at du får vere med i planlegginga av forsøket. Det totale talet på målinger som kan utføres, $N = n + m$, er gitt ut frå økonomien i prosjektet, men kan du avgjere talet på målinger av kvar type, n og m. Korleis vil du velje n og m for at forventa lengd på konfidensintervallet skal bli minst mogleg?

Oppgåve 4

Som regel er mars kaldare enn april i Noreg. La X vere gjennomsnittstemperaturen i mars og Y gjennomsnittstemperaturen i april ved Værnes eit tilfeldig valt år, begge målt i °C. Gå ut frå at X er normalfordelt med forventningsverdi μ_X og varians σ^2, og at Y er normalfordelt med forventningsverdi μ_Y og varians σ^2. Gjennomsnittstemperaturen i °C ved Værnes for åra 2001-2012 var slik:

	2001	2002	2003	2004	2005	2006
x_i (mars)	-2.5	0.5	3.3	2.6	-0.7	-4.6
y_i (april)	4.1	7.2	5.0	7.9	5.8	4.9

	2007	2008	2009	2010	2011	2012
x_i (mars)	3.3	0.8	1.9	-0.5	1.2	3.8
y_i (april)	5.0	5.9	6.9	4.8	6.7	3.2

Det blir oppgitt at

$$\sum_{i=1}^{12} x_i = 9.10,$$

$$\sum_{i=1}^{12} y_i = 67.40,$$

$$\sum_{i=1}^{12} x_i^2 = 77.07,$$

$$\sum_{i=1}^{12} y_i^2 = 399.30,$$

$$\sum_{i=1}^{12} (x_i - y_i)^2 = 364.53,$$

der $i = 1$ står for år 2001, $i = 2$ for 2002 osv.

a) Gå ut frå at mars-temperaturane frå år til år er uavhengige, og april-temperaturane frå år til år er uavhengige. Vi ønskjer ved hypotesetesting å prøve å påvise at differansen mellom forventa gjennomsnittstemperatur i april og mars er mindre enn 5°C. Vi kan anten bruke ein toutvalstest eller ein partest. Kva vil du gjere? Argumenter for valet ditt, og utfør testen du vel. Bruk signifikansnivå $\alpha = 0.05$. σ^2 er ukjent.

Oppgåve 5

Vi har to tilfeldige (stokastiske) variablar X og Y. La X ha varians $\text{Var}(X) = 5$, og Y ha varians $\text{Var}(Y) = 9$. Vidare er kovariansen mellom X og Y gitt som $\text{Cov}(X, Y) = 1$.

a) Rekn ut $\text{Cov}(2X + Y, X - Y)$.