NTNU

Kunnskap for ei betre verd

Institutt for matematiske fag

Eksamensoppgåve i **ST0103 Brukarkurs i statistikk**

Fagleg kontakt under eksamen: Jarle Tufto

Tlf: 99 70 55 19

Eksamensdato: 3. desember 2016

Eksamenstid (frå–til): 09:00-13:00

Hjelpemiddelkode/Tillatne hjelpemiddel: C: Bestemt enkel kalkulator. Tabeller og formler i statistikk (Tapir akademisk forlag). Eitt gult A4-ark med eigne handskrivne notat.

Annan informasjon:
Nokre formlar for bruk i regresjonsanalyse er gitte i vedlegg.

Alle svara skal grunngjevast (t.d. ved at mellomrekning blir tatt med eller ved tilvising til teori eller døme frå pensum).

I vurderinga tel kvart av dei ti bokstavpunkta likt.

Målform/språk: nynorsk

Sidetal: 5

Sidetal vedlegg: 1

Kontrollert av:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☐	**2-sidig** ☒
sort/hvit ☒	**farger** ☐
skal ha fleirvalskjema ☐	

Dato Sign

Oppgåve 1

Ved eit botanisk institutt vart gjort ei undersøking av førekomsten av ein bestemt type sopp i myrjord. Dette vart gjort ved at ein på eit bestemt myrområde tok opp sylindre av jord med ein jordskrue, vaska røtene, og deretter bestemte delen av rotceller med sopp. Delen vart gitt med eining prosent.

La X vere resultatet av ein slik prøve. Frå lang erfaring reknast som kjend at X er normalfordelt med forventing 42.0 og standardavvik 4.0. Resultat frå ulike prøvar reknast å vere stokastisk uavhengige.

a) Kva er sannsynet for at ei måling av X er mindre enn 35?

Finn sannsynet for at ei måling er mellom 35 og 45.

Tekn inn dei to sannsyna som areal på ei enkel skisse av sannsynstettleiken til X.

b) Anta i dette punktet at det vart gjort fire prøver.

Kva er sannsynet for at gjennomsnittet av dei fire prøvene er under 35?

Kva er sannsynet for at minst tre av dei fire prøvene er mellom 35 og 45?

Oppgåve 2

La situasjonen vere som i Oppgåve 1. På ein annan del av myrområdet består jordsmonnet hovudsakleg av bleika sand. Ein ville finne ut om andelen av sopp er ein annan i denne jordtypen enn i myrjorda der prøvene beskrivne i Oppgåve 1 vart tatte.

Det vart tatt 20 prøver i det nye området på same måte som beskrive i begynninga av Oppgåve 1. Dette ga målingane $X_1, X_2, \ldots, X_{20}$ gitt nedanfor. Målingane vart antatte å vere realisasjonar av uavhengige og identisk normalfordelte variablar med ukjent forventing μ og same standardavvik som for dei første prøvene, $\sigma = 4.0$. *(Analysane nedanfor skal altså gjerast med kjent σ.)*

Målingar:

37.99	46.49	36.92	48.10	42.70	49.91	39.10	46.43	39.52	40.40
47.11	41.32	42.61	47.99	44.19	47.40	49.42	45.52	44.84	41.45

Du kan bruke at $\sum_{i=1}^{20} X_i = 879.41$.

a) Finn eit punktestimat for μ basert på målingane, og berekn standardfeilen, dvs. standardavviket for estimatoren.

Finn også eit 95% konfidensintervall for μ.

b) Forklar kort kvifor problemstillinga til botanikarane medfører testing av

$$H_0 : \mu = 42.0 \text{ mot } H_1 : \mu \neq 42.0.$$

Gjennomfør testinga med dei gitte dataene og angi konklusjonen når signifikansnivået vert sett til 0.05. Berekn også den tilhøyrande p-verdi. Kva for fortolking har den?

c) Kva betyr type I-feil og type II-feil i hypotesetesting?

Kva er sannsynet for type I-feil i testen i punktet foran?

Kva er sannsynet for type II-feil ved denne testen dersom μ i røynda er 45.0?

Oppgåve 3

Som ein del av ei større geologisk undersøking har ein studert kjerneprøver av sand med formål å beskrive samanhengen mellom effektiv diffusjon og graden av sementering.

Den effektive diffusjon beskriv kor lett gassar diffunderer gjennom sandkjernen og vert representert ved ei måling d. Graden av sementering for ein sandkjerne er eit tall mellom 0 og 1 og vert betekna med c.

Ein ønsker å beskrive d som funksjon av c og forventar ein samanheng

$$d = \theta(1 - c)^{\beta}, \tag{1}$$

der θ og β er ukjente parametrar.

For å estimere θ og β gjør ein eit eksperiment der ein startar med å måle den effektive diffusjon D for ein sandkjerne med $c = 0$. Deretter aukar ein sementeringa gradvis ved tilsetting av koparsulfat, og ein måler dei korresponderande verdiar av c og D. Dette gir totalt 10 samanhøyrande verdiar av D og c, $(D_1, c_1), (D_2, c_2), \ldots, (D_{10}, c_{10})$.

a) Vis, ved å ta den naturlege logaritma på kvar side av likskapen i (1), at vi kan beskrive den forventa samanhengen (1) ved

$$y = \alpha + \beta x$$

der $y = \ln d$, $\alpha = \ln \theta$ og $x = \ln(1-c)$.

Basert på dette bestemmer ein seg for å analysere dataene med ein enkel lineær regresjonsmodell

$$Y_i = \alpha + \beta x_i + e_i, \quad i = 1, 2, \ldots, 10 \tag{2}$$

der $Y_i = \ln D_i$ og $x_i = \ln(1 - c_i)$.

Kva for antakingar ligg generelt til grunn for bruk av modellen (2)?
Diskuter kort i kva for grad dei kan tenkast oppfylt i den gitte situasjonen.

Tabellen nedanfor gir verdiane av c_i og dei tilhøyrande responsar $D_i = d_i$, samt dei transformerte x_i og y_i som vert brukt i regresjonsanalysen.

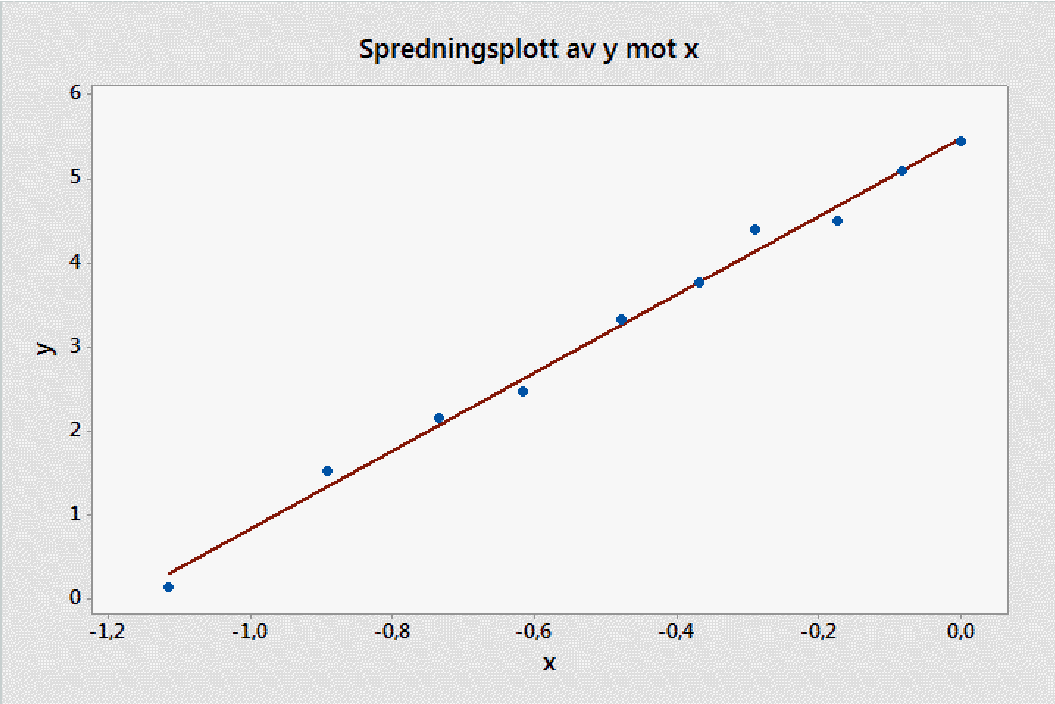
i	c_i	d_i	$x_i = \ln(1-c_i)$	$y_i = \ln d_i$
1	0,000	234.102	0.000	5.456
2	0.080	165.961	-0.083	5.112
3	0.160	90.990	-0.174	4.511
4	0.250	82.496	-0.288	4.413
5	0.307	43.413	-0.367	3.771
6	0.380	28.156	-0.478	3.338
7	0.460	11.903	-0.616	2.477
8	0.520	8.747	-0.734	2.169
9	0.590	4.589	-0.892	1.524
10	0.672	1.165	-1.115	0.153

Det vert oppgitt at $\sum x_i = -4.747$, $\sum x_i^2 = 3.440$, $\sum y_i = 32.924$, $\sum y_i^2 = 134.273$, $\sum x_i y_i = -10.108$.

Du kan også bruke at $SS_E = 0.1898$ og $\sum(x_i - \bar{x})^2 = 1.187$.

I rekninga kan du bruke formlar som er gitt i vedlegget og i tabellane.

Eit spreiingsplott for dei berekna (x_i, y_i) med inntekna regresjonslinje er til orientering gitt på neste side.



b) Berekn punktestimata $\hat{\alpha}$ og $\hat{\beta}$ for α og β basert på minste kvadraters metode ved å bruke dei gitte resultata.

Finn også eit punktestimat for parameteren θ fra ligning (1).

Berekn punktestimatet S for σ og bruk dette til å finne standardfeilen for estimatet $\hat{\beta}$. Finn også eit 95% konfidensintervall for β.

Kor stor del av variasjonen i responsane vert forklart av regresjonsmodellen? Gi ein kommentar i lys av spreiingsplottet.

Oppgåve 4

Ein biolog er interessert i å undersøke førekomsten av ein sjeldan insektsart, i det følgjande kalt art A, i eit bestemt område. Det vert sett opp ei såkalla malaisefelle (ein stor, telt-liknande struktur) og ein antar at insekt av art A vert fanga i fella som ein Poisson-prosess med intensitet (rate) λ pr. time. La X vere talet på insekt av art A som vert fanga i løpet av ei veke (= 168 timer). Biologen reknar før undersøkinga med at λ er 0.01. Denne verdi for λ skal du bruke i punktene a) og b).

a) Gjer greie for at X er Poisson-fordelt med forventing $\mu = 1.68$.

Kva er sannsynet for at biologen ikkje vil finne nokre insekt av art A i fella etter ei veke?

Kva er sannsynet for at det er minst tre insekt av art A i fella?

Anta nå at fella vart ståande i to veker. Gitt at det vart fanga minst tre insekt av art A i løpet av dei to vekene, kva er sannsynet for at ingen vart fanga den første veka?

b) La T vere tida, målt i timar, til det første insektet av art A vert fanga i fella. *(Sjå i dette punktet bort frå at fella er montert bare ei begrensa tid.)*

Kva for fordeling har T?

Kva er sannsynet for at ingen insekt av art A vert fanga i løpet av dei første 48 timane, dvs. $T > 48$?

Gitt at ingen insekt av art A vert fanga i løpet av dei første 48 timane, kva er sannsynet for at det heller ikkje vert fanga nokre insekt av art A i løpet av dei neste 48 timane? Kommenter resultatet.

Biologen er ikkje lenger sikker på sitt tidlegare anslag av λ og vil estimere λ. Han lar fella stå i fem veker og finn Y insekt av art A.

c) Kva for sannsynsfordeling har Y?

Sett opp ein forventingsrett estimator $\hat{\lambda}$ for λ basert på Y.

Kva vert estimatet om det vert observert at $Y = 15$? Kva vert standardfeilen (dvs. estimert standardavvik) for dette estimatet?

Supplement til "Noen resultater fra regresjonsanalysen"
i *Tabeller og formler i statistikk, Tapir akademisk forlag.*

Formlane bygger på summane:

$$\sum_{i=1}^{n} x_i,\ \ \sum_{i=1}^{n} x_i^2,\ \ \sum_{i=1}^{n} y_i,\ \ \sum_{i=1}^{n} y_i^2,\ \ \sum_{i=1}^{n} x_i y_i$$

Minste kvadraters metode gir då:

$$\begin{aligned}\hat{\beta} &= \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2} \\ \hat{\alpha} &= \frac{(\sum_{i=1}^{n} y_i) - \hat{\beta}(\sum_{i=1}^{n} x_i)}{n}\end{aligned}$$

Frå læreboka:

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{SS_T} = \underbrace{\sum_{i=1}^{n}(\hat{\alpha} + \hat{\beta}x_i - \bar{y})^2}_{SS_R} + \underbrace{\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2}_{SS_E}$$

Her er:

$$\begin{aligned}SS_T &= (\sum_{i=1}^{n} y_i^2) - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2 \\ SS_E &= (\sum_{i=1}^{n} y_i^2) - \hat{\alpha}(\sum_{i=1}^{n} y_i) - \hat{\beta}(\sum_{i=1}^{n} x_i y_i)\end{aligned}$$

Forventingsrett estimator for σ^2:

$$S^2 = \frac{SS_E}{n-2}$$

Statistiske eigenskapar ved estimatorane:

$$E(\hat{\alpha}) = \alpha, \qquad E(\hat{\beta}) = \beta$$

$$Var(\hat{\alpha}) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad Var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$