Kungl
Tekniska
Högskolan

Matematiska institutionen
avd matematisk statistik

TENTAMEN I 5B1503 STATISTIK MED FÖRSÖKSPLANERING
FÖR B OCH K FREDAGEN DEN 11 JANUARI 2002 KL 14.00–19.00.

Examinator: Gunnar Englund, tel. 7907416

Tillåtna hjälpmedel: Formel- och tabellsamling i Statistik med försöksplanering. Kalkylator.

Införda beteckningar skall förklaras och definieras. Resonemang och räkningar skall vara så utförliga och motiverade att de är lätta att följa. 20 poäng innebär godkänt resultat.

Resultatet anslås senast fredagen den 1 februari 2002 på matematisk statistiks anslagstavla i entréplanet, Lindstedtsvägen 25. Tentamen kommer att finnas tillgänglig på elevexpeditionen sju veckor efter skrivningstillfället.

Uppgift 1

a) Den stokastiska variabeln X har väntevärdet 3 och variansen 2. Beräkna med lämplig approximation standardavvikelsen $D(Xe^X)$. (5 p)

b) Låt X och Y vara oberoende normalfördelade stokastiska variabler. X är $N(3,\ 2^2)$ (väntevärde 3 och standardavvikelse 2) och Y är $N(2.5,\ 1.5^2)$. Beräkna $P(X > 2 \text{ och } Y > 3)$ samt $P(X + Y > 5)$. (5 p)

Uppgift 2

Man har undersökt sambandet mellan halten av en viss fibertyp (x) och kvalitén av producerat papper (y). Följande 13 par av (lämpligt transformerade) observationer (x_i, y_i) (=(halt fibrer, kvalité)) har erhållits:

x_i	5	10	70	75	70	90	95	140	170	190	192	210	230
y_i	0	2	9	6	4	8	6	3	13	12	14	17	18

Vi antar att $(x_1, y_1), \dots, (x_{13}, y_{13})$ kan beskrivas av en enkel linjär regressionsmodell dvs. som oberoende utfall av $(x_1, Y_1), \dots, (x_{13}, Y_{13})$, där varje Y_i är normalfördelad $N(\alpha + \beta x_i, \sigma^2)$

Räknehjälp (grafisk lösning godtages ej):

$$\sum_{i=1}^{13} x_i = 1547 \quad \sum_{i=1}^{13} y_i = 112$$

$$\sum_{i=1}^{13} y_i^2 = 1368 \quad \sum_{i=1}^{13} x_i^2 = 251139$$

$$\sum_{i=1}^{13} x_i y_i = 17978.$$

Skatta förväntad kvalité vid fiberhalten x=100 samt medelfelet för denna skattning. (10 p)

Uppgift 3

En grisuppfödare vill jämföra tillväxten hos grisar för två olika typer av vitaminiserat foder under en viss tidsperiod. 15 grisar får foder A och 15 grisar får foder B. Mätvärdena (dvs. grisarnas viktökningar) kallar vi $x_1, x_2, \cdots, x_{15}$ resp. $y_1, y_2, \cdots, y_{15}$ och för att underlätta de numeriska beräkningarna har följande mer eller mindre relevanta summor beräknats:

$$\sum_{j=1}^{15} x_j = 94.5, \quad \sum_{j=1}^{15} y_j = 91.5, \quad \sum_{j=1}^{15} x_j^2 = 665.4, \quad \sum_{j=1}^{15} y_j^2 = 626.7, \quad \sum_{j=1}^{15} x_j y_j = 287.75,$$

Som statistisk modell ansattes följande: x_j:na är utfall av $N(m_A, \sigma^2)$-fördelade stokastiska variabler och y_j:na är utfall av $N(m_B, \sigma^2)$-fördelade stokastiska variabler. Alla variabler antas vara statistiskt oberoende.

a) Beräkna ett (exakt) 95%-igt konfidensintervall för skillnaden i genomsnittlig tillväxt mellan de två fodermedlen (dvs för $m_A - m_B$). (5 p)

b) Beräkna ett 95%-igt konfidensintervall för σ. (5 p)

Uppgift 4

Inom många verksamheter är det viktigt att jämföra provningsresultat med varandra. Man talar om olika precisionsmått, repeterbarhet som anger överensstämmelse mellan två provningsresultat som erhålls från samma plats t.ex laboratorium och reproducerbarhet som anger överensstämmelse mellan två provningsresultat som erhålls från olika platser t.ex olika laboratorier. För att beräkna repeterbarhet och reproducerbarhet för en viss metod skickade man runt ett referensprov till fem slumpmässigt utvalda laboratorier bland en stor mängd tänkbara laboratorier och bad dem att göra 4 analyser var på referensprovet med metoden ifråga. Data:

Lab 1	9.7	9.6	9.8	9.3
Lab 2	11.0	11.5	11.7	12.0
Lab 3	9.7	10.0	9.7	10.3
Lab 4	12.4	11.1	11.4	12.2
Lab 5	9.9	9.6	10.7	10.5

Räknehjälp: Totalkvadratsumma $\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = 18.5495$.
Lab-medelvärden $\bar{y}_{1.} = 9.6$ $\bar{y}_{2.} = 11.55$ $\bar{y}_{3.} = 9.925$ $\bar{y}_{4.} = 11.775$ $\bar{y}_{5.} = 10.175$

a) Ställ upp en lämplig statistisk modell och testa på lämpligt sätt hypotesen att variansen mellan laboratorier är noll. (4 p)

b) <u>Repeterbarhets</u>standardavvikelsen är standardavvikelsen för skillnaden mellan två oberoende observationer gjorda inom ett och samma laboratorium. Skatta denna repeterbarhetsstandardavvikelse. (3 p)

c) Reproducerbarhetsstandardavvikelsen är standardavvikelsen för skillnaden mellan två oberoende observationer erhållna från två slumpmässigt utvalda laboratorier. Skatta denna reproducerbarhetsstandardavvikelse. (3 p)

Uppgift 5

I ett försök för att utröna tillförlitligheten hos kullager bedömdes tre faktorer vara väsentliga; värmebehandling av innerringen (A), ytterringens s.k. oskulation (B) samt hållarkonstruktionen (C). De tre faktorerna undersöktes vid standardutseende (–) och vid modifierad konstruktion (+). Resultaten från livslängdstestet (två observationer (y) för varje konstruktion):

Försök nr (i)	A	B	C	y_{i1}	y_{i2}	Medelvärde $\bar{y}_{i.}$	Varians s_i^2
1	−	−	−	16	18	17	2
2	+	−	−	21	31	26	50
3	−	+	−	20	30	25	50
4	+	+	−	96	88	92	32
5	−	−	+	19	19	19	0
6	+	−	+	13	19	16	18
7	−	+	+	18	24	21	18
8	+	+	+	99	93	96	18

$\sum_i \sum_j (y_{ij} - \bar{y}..)^2 = 16\ 508$ Som räknehjälp ges att med

$$\mathbf{M} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{pmatrix} \text{ så erhålles } \mathbf{M}\bar{\mathbf{y}} = \begin{pmatrix} 312 \\ 148 \\ 156 \\ -8 \end{pmatrix}$$

där $\bar{\mathbf{y}}$ är medelvärdeskolumnen ovan.

a) Ställ upp en lämplig modell samt testa vilka effekter som är signifikanta (nivå 5%). (6 p)

b) Reducera modellen för de icke-signifikanta effekterna i a). Beräkna i denna reducerade modell en ny σ-skattning samt ange hur många frihetsgrader den har. (4 p)

Kungl
Tekniska
Högskolan

Matematiska institutionen
avd matematisk statistik

LÖSNING FÖR TENTAMEN I 5B1503 STATISTIK MED FÖRSÖKSPLANERING FÖR B OCH K FREDAGEN DEN 11 JANUARI 2002 KL 14.00–19.00.

Uppgift 1

a) Enligt Gauss' approximationsformler får vi

$$E(g(X)) \approx g(E(X)) \text{ och } V(g(X)) \approx \left(g'(E(X))\right)^2 V(X)$$

och vi har här $g(x) = xe^x$ och erhåller alltså $g'(x) = (1+x)e^x$ som ger $E(Xe^X) \approx E(X)e^{E(X)} = 3e^3 \approx 60.26$ och $V(Xe^X) \approx ((1+3)e^3)^2 \cdot 2$ som ger $D(Xe^X) \approx \sqrt{16e^6} \cdot \sqrt{2} \approx \underline{113.62}$.

b)

$$P(X > 2 \text{ och } Y > 3) = (\text{p g a oberoendet}) = P(X > 2)P(Y > 3) =$$

$$P\left(\frac{X-3}{2} > \frac{2-3}{2}\right) P\left(\frac{Y-2.5}{1.5} > \frac{3-2.5}{1.5}\right) =$$

$$(1 - \Phi(-0.5))\,(1 - \Phi(1/3)) \approx \Phi(0.5)\,(1 - \Phi(0.33)) \approx 0.6915 \cdot (1 - 0.6293) \approx \underline{0.256}.$$

Vidare är $X+Y$ normalfördelad eftersom X och Y är det och de är oberoende. $E(X+Y) = E(X) + E(Y) = 3 + 2.5 = 5.5$ och $V(X+Y) = V(X) + V(Y) = 2^2 + 1.5^2 = 6.25$ dvs $D(X+Y) = \sqrt{6.25} = 2.5$ och vi ser att $X+Y$ är $N(5.5, 2.5)$ vilket ger

$$P(X+Y > 5) = P\left(\frac{X+Y-5.5}{2.5} > \frac{5-5.5}{2.5}\right) = 1 - \Phi(-0.2) = \Phi(0.2) \approx 0.5793 \approx \underline{0.580}.$$

Uppgift 2

Man har att

$$\overline{x} = \frac{1}{13}\sum_{i=1}^{13} x_i = 119, \quad \overline{y} = \frac{1}{13}\sum_{i=1}^{13} y_i = 8.61538.$$

Enligt formelsamlingen fås

$$\beta^* = \frac{\sum_{i=1}^{13} x_i y_i - 13 \cdot \overline{x} \cdot \overline{y}}{\sum_{i=1}^{13} x_i^2 - 13 \cdot \overline{x}^2} = \frac{17978 - 13 \cdot 119 \cdot 8.61538}{251139 - 13 \cdot 119^2} = 0.069364.$$

Minstakvadratskattningarna är:

$$\alpha^* = \overline{y} - \beta^* \overline{x} = 8.61538 - 0.069364 \cdot 119 = 0.361064, \text{ och } \beta^* = 0.0694.$$

Den skattade regressionslinjen är med andra ord

$$\widehat{y}(x) = \alpha^* + \beta^* x = 0.361064 + 0.0694 \cdot x$$

Med $x_0 = 100$ ser vi att vi skattar förväntad papperskvalité med $\alpha^* + \beta^* \cdot 100 \approx 7.297 \approx \underline{7.3}$.

För att skatta s beräknas residualkvadratsumman Q_0. Enligt formelsamlingen gäller att

$$Q_0 = \sum_{i=1}^{13} y_i^2 - 13\overline{y}^2 - (\beta^*)^2 (\sum_{i=1}^{13} x_i^2 - 13\overline{x}^2) =$$

$$= 1368 - 13 \cdot 8.61538^2 - 0.0694^2 \left(251139 - 13 \cdot 119^2\right) = 80.1603.$$

Därmed skattas σ^2 av

$$s^2 = \frac{Q_0}{13-2} = \frac{80.1603}{11} = 7.2873.$$

$\alpha^* + \beta^* \cdot 100 = \overline{Y} + \beta^*(100 - \overline{x})$ ser vi har variansen ($\overline{Y}$ och β^* är ju oberoende)

$$V(\overline{Y}) + (100 - \overline{x})^2 V(\beta^*) = \frac{\sigma^2}{13} + (100-119)^2 \frac{\sigma^2}{\sum_{i=1}^{13}(x_i - \overline{x})^2} = \sigma^2 \cdot \left(\frac{1}{13} + \frac{(100-119)^2}{\sum_{i=1}^{13}(x_i - \overline{x})^2} \right) =$$

$$\sigma^2 \cdot \left(\frac{1}{13} + \frac{19^2}{251139 - 13 \cdot 119^2} \right) \approx \sigma^2 \cdot 0.0823$$

och vi får alltså att medelfelet är $s\sqrt{0.0823} = \sqrt{7.2873 \cdot 0.0823} \approx \underline{0.774}$.

Svar: Skattningen är 7.3 med medelfelet 0.78.

Uppgift 3

Vi har fallet med två oberoende stickprov. Av de beräknade summorna följer att

$$\bar{x} = \frac{94.5}{15} = 6.3, \ \bar{y} = \frac{91.5}{15} = 6.1,$$

$$s_A^2 = \frac{1}{14}(665.4 - 15(6.3)^2) = 5.004, \ s_B^2 = \frac{1}{14}(626.7 - 15(6.1)^2) = 4.896,$$

som ger

$$s^2 = \frac{14 s_A^2 + 14 s_B^2}{28} = 4.95.$$

a) Vi får ett 95%-igt konfidensintervall för $m_A - m_B$ till

$$\bar{x} - \bar{y} \pm t_{0.025}(28) s \sqrt{\frac{1}{15} + \frac{1}{15}} = \underline{0.2 \pm 1.66}.$$

b) Eftersom $\dfrac{28s^2}{\sigma^2}$ är $\chi^2(28)$-fördelad så får vi att

$$\left(\frac{28s^2}{\chi^2_{0.025}}, \frac{28s^2}{\chi^2_{0.9755}} \right) = (3.115, \ 9.059)$$

är ett 95%-igt konfidensintervall för σ^2. Alltså blir $(\sqrt{3.115}, \sqrt{9.059}) = \underline{(1.76,\ 3.01)}$ ett 95%-igt konfidensintervall för σ.

Uppgift 4

a) Ensidig variansanalys, slumpmässig faktor. σ^2 varians mellan analayser inom lab, σ_l^2 varians mellan lab;
$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ där $\alpha \in N(0, \sigma_l^2)$ och $\varepsilon \in N(0, \sigma^2)$. Variansanalystabell:

Källa	Fg	Kvs	Mkvs	E(Mkvs)
Mellan lab	4	15.6770	3.9192	$\sigma^2 + 4\sigma_l^2$
Inom lab	15	2.8725	0.1915	σ^2
Totalt	19	18.5495		

Testkvot är $F_l = \dfrac{3.9192}{0.1915} = 20.5 > 3.06$.
Variansen är signifikant skild från 0.

b) Låt Y_1 och Y_2 vara två observationer från ett visst lab. Variansen $V(Y_i) = V(\varepsilon) = \sigma^2$, $i = 1, 2$ varför $V(Y_1 - Y_2) = 2\sigma^2$ och standardavvikelsen $D(Y_1 - Y_2) = \sigma\sqrt{2}$. σ skattas med $\sqrt{0.1915}$ varför repeterbarhetsstandardavvikelsen skattas med $\sqrt{2 \cdot 0.1915} = 0.619$.

c) Låt nu Y_{11} och Y_{21} vara två observationer från olika lab. Då är variansen

$$V(Y_{11} - Y_{21}) = V\big(\mu + \alpha_1 + \varepsilon_{11} - (\mu + \alpha_2 + \varepsilon_{21}\big) = V\big(\alpha_1 - \alpha_2 + \varepsilon_{11} - \varepsilon_{21}\big) =$$
$$V(\alpha_1) + V(\alpha_2) + V(\varepsilon_{11}) + V(\varepsilon_{21}) = 2\sigma_l^2 + 2\sigma^2$$

σ_l^2 skattas med $\dfrac{1}{4}(3.9192 - 0.1915)$=0.9319.
Reproducerbarhetsstandardavvikelsen skattas med $\sqrt{2(0.1915 + 0.9319)} = 1.499$

Uppgift 5

Vi har här en 2^3-modell med två observationer per försökspunkt. Vi får med hjälp av ledningen att $\widehat{\mu} = \dfrac{312}{8} = 39$, $\widehat{A} = \dfrac{148}{8} = 18.5$, $\widehat{B} = \dfrac{156}{8} = 19.5$, $\widehat{C} = \dfrac{-8}{8} = -1$. Vi får vidare med hjälp av teckenschemat från formelsamlingen avsnitt 15.2 att

$$\widehat{AB} = \frac{1}{8}(17 - 26 - 25 + 92 + 19 - 16 - 21 + 96) = 17$$

$$\widehat{AC} = \frac{1}{8}(17 - 26 + 25 - 92 - 19 + 16 - 21 + 96) = -0.5$$

$$\widehat{BC} = \frac{1}{8}(17 + 26 - 25 - 92 - 19 - 16 + 21 + 96) = 1$$

$$\widehat{ABC} = \frac{1}{8}(-17 + 26 + 25 - 92 + 19 - 16 - 21 + 96) = 2.5$$

Vi får vidare

$$\widehat{\sigma}^2 = \frac{1}{8}(s_1^2 + s_2^2 + s_3^2 + s_4^2 + s_5^2 + s_6^2 + s_7^2 + s_8^2) = \frac{188}{8} = 23.50$$

som är baserad på 8 frihetsgrader ($\widehat{\sigma} = \sqrt{23.50} \approx 4.85$). De enskilda effektskattningarna har alla variansen

$$\frac{\sigma^2}{2^3 \cdot 2} = \frac{\sigma^2}{16}$$

vilket gör att vi får konfidensintervallen effektskattning $\pm t_{0.025}(8)\widehat{\sigma}\frac{1}{\sqrt{16}} = \text{skattn.} \pm 2.80$. Alltså är de effektskattningar som är till beloppet större än 2.80 signifikanta, vilket innebär A, B, AB förutom μ.

b) Vi bildar en ny kvadratsumma genom att till residualkvadratsumman (Inom celler) lägga

$$2^3 \cdot 2(\widehat{C}^2 + \widehat{AC}^2 + \widehat{BC}^2 + \widehat{ABC}^2)$$

Alltså får vi den nya σ-skattningen till

$$\frac{1}{12}\left(8 \cdot \widehat{\sigma}^2 + 16\left((-1)^2 + (-0.5)^2 + 1^2 + (2.5)^2\right)\right) = \frac{1}{12}(188 + 16 \cdot 8.5) = 27$$

dvs den nya σ-skattningen blir $\sqrt{27} \approx \underline{5.20}$ och denna skattning är baserad på $8 + 4 = \underline{12}$ frihetsgrader.

För övrigt kan man inse denna nya reducerade modell faktiskt innebär ett fullständigt 2^2-försök med 4 observationer per försökspunkt, där man alltså bortser från faktorn C.