KUNGL TEKNISKA HÖGSKOLAN

Matematiska institutionen
avd matematisk statistik

TENTAMEN I 5B1503 STATISTIK MED FÖRSÖKSPLANERING FÖR K OCH B
MÅNDAGEN DEN 19 AUGUSTI 2002 KL 14.00–19.00.

Examinatorer: Gunnar Englund, 790 7416.

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

Varje korrekt löst uppgift ger 10 poäng så maximalpoängen blir 50 poäng. För godkänt krävs 20 poäng. Resultatet anslås senast måndagen den 9 september 2002 på matematisk statistiks anslagstavla i entréplanet, Lindstedtsvägen 25.

Införda beteckningar skall förklaras och definieras. Resonemang och räkningar skall vara så utförliga att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet.

Uppgift 1

I en kurs för kemiteknologer ska var och en av teknologerna genomföra ett laborationsexperiment som man av erfarenhet vet kommer att misslyckas med sannolikhet 0.05 (oavsett teknologens skicklighet).

(a) Antag att 100 teknologer gör, oberoende av varandra, experimentet. Vad är sannolikheten att minst 95 teknologer lyckas? Välmotiverade approximationer får användas. (5 p)

(b) Misslyckas en teknolog med experimentet får hon/han göra om det ytterligare en gång. Skulle detta experiment också misslyckas finns ingen chans till ett tredje försök vid detta kurstillfälle. Låt X vara totala antalet experiment som utföres av 100 teknologer på kursen. Beräkna väntevärde och varians för X. (5 p)

Uppgift 2

I ett miljöövervakningssystem studeras övergödningen av våra vattendrag. I en viss å har man under en längre period gjort mätningar av bl a total fosforhalt. Under denna period införde man i avrinningsområdet en kemisk-biologisk rening av hushållens och industriernas avloppsvatten. För att undersöka vilken effekt dessa åtgärder haft på fosformängden i vattendraget beräknas årsmedelvärdena av total fosforhalt (mg/l) före och efter införandet av ny rening:

Fosforhalt mg/l) före införandet:	0.12	0.14	0.07	0.09	0.15	0.09	0.10	
Fosforhalt (mg/l) efter införandet:	0.09	0.03	0.07	0.09	0.07	0.08	0.11	0.07

a) Gör ett 95% konfidensintervall för den genomsnittliga effekten av den nya reningen. Redogör för dina modellantaganden. (8 p)

b) Gav åtgärderna upphov till en signifikant förändring av total fosforhalt i vattendraget? Motivera ditt svar! (2 p)

Uppgift 3

Vid ett laboratorieförsök tillverkades papper under två olika tryck. Sam tidigt undersöktes två olika pappersvarianter, P_1 och P_2, med varierande fiberlängd. Hos de tillverkade pappersarken registrerades i tre försök för vardera kombination den så kallade slitfaktorn. Resultat:

Tryck	Papper			
20	P_1	144	142	149
20	P_2	136	132	134
100	P_1	126	132	132
100	P_2	135	131	127

Antag att slitfaktorbestämningarna har oberoende normalfördelade fel med konstant varians. Gör lämpliga analyser för att undersöka om trycket och/eller fiberlängden har signifikant inverkan på slitfaktorn. (10 p)

Uppgift 4

På två olika fiskarter i Mississippifloden mättes mängden kvicksilver (ppm) hos 5 respektive 6 exemplar av arterna.

Fiskart 1: 2.35 2.44 2.70 2.48 2.44
Fiskart 2: 2.06 1.93 2.12 2.16 1.89 1.95

Eftersom de studerade fiskarna har ungefär samma vikt och eftersom samma mätinstrument används vid alla mätningar antas följande modell: De n_i mätningarna på fiskart $i, x_{i1}, \ldots x_{in}$, är observationer från $\mathrm{N}(\mu_i, \sigma^2)$.

a) Skatta σ^2. (3 p)

b) Gör ett 95% konfidensintervall för medelmängden kvicksilver i fiskart 1. (Ledning: Man vill även utnyttja mätningarna från fisk art 2.) (4 p)

c) På en tredje fiskart kunde man endast fånga ett exemplar så endast en kvicksilvermätning, 3.13 (ppm), kunde noteras. Gör ett 95% konfidensintervall för medelmängd kvicksilver hos denna fiskart. (3 p)

Uppgift 5

I miljöövervakningssystemet från uppgift 2 mättes samtidigt kvävehalten i vattendraget. Nedan ges medelvärdet av sommarmånadernas kvävehalt för ett antal år:

År	1965	1966	1969	1972	1975	1977	1980
Kvävehalt (mg/l)	0.37	0.47	0.64	0.52	0.99	0.97	1.08

Antag att kvävehalten ändras linjärt under den studerade perioden. Gör ett 95% konfidensintervall för den årliga förändringen av kvävehalt i vattendraget. (10 p)

KUNGL TEKNISKA HÖGSKOLAN

Matematiska institutionen
avd matematisk statistik

LÖSNING TENTAMEN I 5B1503 STATISTIK MED FÖRSÖKSPLANERING 02-08-19

Uppgift 1

a) Låt $Y=$ antalet som misslyckas. Y är $Bin(100, 0.05)$ ty vi har en typsituation för binomialfördelning. Antalet som lyckas är $100 - Y$ och vi söker $P(100 - Y \geq 95) = P(Y \leq 5)$. Eftersom $p = 0.05 \leq 0.1$ är Poisson-approximation tillåten och $Bin(100, 0.05) \approx Po(100 \cdot 0.05) = Po(5)$. Alltså får vi med hjälp av tabell 7 i formelsamlingen att $P(Y \leq 5) \approx \underline{0.61596}$.

b) Låt X_i =antalet försök teknolog nr i utför. Enligt texten är
$P(X_i = 1) = 0.95$ och $P(X_i = 2) = 0.05$. Vi får $X = \sum_1^{100} X_i$ och alltså att
$E(X) = E(\sum_1^{100} X_i) = \sum_1^{100} E(X_i) = 100E(X_1)$ och
$V(X) = V(\sum_1^{100} X_i) = \sum_1^{100} V(X_i) = 100V(X_1)$.
Vi har $E(X_1) = 1 \cdot P(X_1 = 1) + 2 \cdot P(X_1 = 2) = 0.95 + 2 \cdot 0.05 = 1.05$ och
$E(X_1^2) = 1^2 \cdot P(X_1 = 1) + 2^2 \cdot P(X_1 = 2) = 0.95 + 4 \cdot 0.05 = 1.15$.
Detta ger $V(X_1) = E(X_1^2) - (E(X_1))^2 = 1.15 - 1.05^2 = 0.0475$ vilket ger $E(X) = 100 \cdot 1.05 = \underline{105}$ och $V(X) = 100 \cdot V(X_1) = 100 \cdot 0.0475 = \underline{4.75}$.

Uppgift 2

a) Låt $x_i=$ värde före rening, $i = 1, 2, \cdots, 7$ och $y_i=$ värde efter rening, $i = 1, 2, \cdots, 8$.
x_i ses som utfall av X_i som är $N(m_f, \sigma)$ och y_i som utfall av Y_i som är $N(m_e, \sigma)$.
$X_1, X_2, \cdots, X_7, Y_1, Y_2, \cdots, Y_8$ antages vara oberoende stokastiska variabler.

Vi vill göra konfidensintervall för $m_e - m_f$ som skattas med $\bar{y} - \bar{x}$. Motsvarande stokastiska variabel $\bar{X} - \bar{Y}$ är alltså $N(m_e - m_f, \sigma\sqrt{1/7 + 1/8})$ och med t-metoden (Formelsamlingen 11.2) erhåller vi konfidensintervallet $\bar{y} - \bar{x} \pm t_{0.025}(7 + 8 - 2)s\sqrt{\dfrac{1}{7} + \dfrac{1}{8}}$ där $\bar{x} = \sum_1^7 x_i/7 = 0.1086$ och $\bar{y} = \sum_1^8 y_i/8 = 0.0763$ samt $s^2 = \dfrac{(7-1)s_x^2 + (8-1)s_y^2)}{7+8-2}$.
Vi får

$$s_x^2 = \frac{1}{7-1}\sum_{i=1}^{7}(x_i - \bar{x})^2 = \frac{1}{6}\left(\sum_1^7 x_i^2 - 7(\bar{x})^2\right) = \frac{1}{6}(0.0876 - 7 \cdot 0.1086^2) = \frac{0.0051}{6}$$

och

$$s_y^2 = \frac{1}{8-1}\sum_{i=1}^{8}(y_i - \bar{y})^2 = \frac{1}{7}\left(\sum_1^8 y_i^2 - 8(\bar{y})^2\right) = \frac{1}{7}(0.0503 - 8 \cdot 0.0763^2) = \frac{0.0038}{7}$$

som ger $s^2 = \dfrac{0.0051 + 0.0038}{13} = 6.8365 \cdot 10^{-4}$, dvs $s = 0.0261$. Detta ger intervallet

$$\bar{y} - \bar{x} \pm t_{0.025}(13)s\sqrt{\frac{1}{7} + \frac{1}{8}} = 0.0763 - 0.1086 \pm 2.16 \cdot 0.0261\sqrt{\frac{1}{7} + \frac{1}{8}} = \underline{-0.0323 \pm 0.292}.$$

b) $H_0 : m_e = m_f$ förkastas på nivån 5% eftersom ovanstående intervall ej innehåller 0.
<u>Slutsats:</u> Förändringen av fosforhalt är statistiskt säkerställd.

Uppgift 3

Vi har ett 2^2-försök med 3 observationer/cell. Låt y_{ijk} = mätvärde för upprepning nr k för kombinationen i (för tryck) och j (för papperstyp) där $k = 1, 2, 3$, $i = 20, 100$ och $j = P_1, P_2$. Vi ser y_{ijk} som utfall av Y_{ijk} som är $N(\mu_{ij}, \sigma)$ där $\mu_{ij} = I \pm B \pm A \pm AB$. Notera att ordningen av data indikerar att papperstyp är faktor A och tryck faktor B med beteckningar enligt formelsamlingens teckenschema (FS 15.2).
Vi har medelvärdena 145, 134, 130 resp 131 för de 4 faktorkombinationerna. Vi får skattningarna

$$\widehat{I} = \frac{1}{4}(145 + 134 + 130 + 131) = 135, \ \widehat{A} = \frac{1}{4}(-145 + 134 - 130 + 131) = -2.5,$$

$$\widehat{B} = \frac{1}{4}(-145 - 134 + 130 + 131) = -4.5, \ \widehat{AB} = \frac{1}{4}(145 - 134 - 130 + 131) = 3.$$

Spridningen σ^2 skattas med

$$\widehat{\sigma}^2 = \frac{1}{2^2(3-1)} \sum_{i,j,k} (y_{ijk} - \bar{y}_{ij\cdot})^2 = \frac{1}{8} \left(\sum_{i,j,k} y_{ijk}^2 - 3 \cdot \sum_{i,j} \bar{y}_{ij\cdot}^2 \right) = \frac{1}{8}(219216 - 3 \cdot 73042) = \frac{90}{8}.$$

Vi får

$$\pm t_{0.025}(8)\widehat{\sigma}\frac{1}{\sqrt{2^2 \cdot 3}} = \pm 2.31\sqrt{\frac{90}{8 \cdot 12}} = \pm 2.23$$

och alltså är samtliga effekter (inklusive samspelet) statistiskt säkerställda eftersom de till beloppet är större än 2.23.

Uppgift 4

Vi får $\bar{x}_{1\cdot} = \sum_j x_{1j}/5 = 2.482$, $\bar{x}_{2\cdot} = \sum_j x_{2j}/6 = 2.0182$ och

$$s_1^2 = \frac{1}{5-1} \sum_{j=1}^{5} (x_{1j} - \bar{x}_{1\cdot})^2 = \frac{1}{4} \left(\sum_{j=1}^{5} x_{1j}^2 - 5(\bar{x}_{1\cdot})^2 \right) = \frac{1}{4}(30.8701 - 5 \cdot 2.482^2) = \frac{0.0685}{4} = 0.0171$$

och

$$s_2^2 = \frac{1}{6-1} \sum_{j=1}^{6} (x_{2j} - \bar{x}_{2\cdot})^2 = \frac{1}{5} \left(\sum_{j=1}^{6} x_{2j}^2 - 6(\bar{x}_{2\cdot})^2 \right) = \frac{1}{5}(24.5031 - 6 \cdot 2.0183^2) = \frac{0.0611}{5} = 0.0122$$

som ger σ^2-skattningen

$$\widehat{\sigma}^2 = s^2 = \frac{(5-1)s_1^2 + (6-1)s_2^2}{5+6-2} = \frac{0.0685 + 0.0611}{9} = \underline{0.0144}.$$

b) Vi skattar μ_1 med $\widehat{\mu}_1 = \bar{x}_{1\cdot}$ som är ett utfall av $\bar{X}_{1\cdot}$ som är $N(\mu_1, \sigma/\sqrt{5})$ och vi får alltså med t-metoden enligt Formelsamlingen 11.2 (notera att vår σ-skattning är baserad på 9 frihetsgrader eftersom vi utnyttjat båda dataserierna) det 95%-iga konfidensintervallet till

$$\bar{x}_{1\cdot} \pm t_{0.025}(9)\frac{s}{\sqrt{5}} = 2.482 \pm 2.26\sqrt{\frac{0.0144}{5}} = \underline{2.482 \pm 0.121}.$$

c) Vi skattar μ_3 med $x_{31} = 3.13$ som är ett utfall av X_{31} som är $N(\mu_3, \sigma)$ och vi får alltså konfidensintervallet $x_{31} \pm t_{0.025}(9)s = \underline{3.13 \pm 0.27}$.

Uppgift 5

Enkel linjär regression (formelsamlingen 12). Vi låter y_i = kvävehalt år nr i, $i = 1, 2, \cdots, 7$ som vi ser som utfall av Y_i som är $N(\alpha + \beta x_i, \sigma)$. Vi skattar β med

$$\beta^* = \frac{\sum_{i=1}^{7} x_i y_i - 7 \cdot \bar{x}\bar{y}}{\sum_{i=1}^{7}(x_i - \bar{x})^2}$$

och α med $\alpha^* = \bar{y} - \beta^*\bar{x}$. Vi får

$$\bar{x} = 1972,\ \bar{y} = 0.72,\ \sum_i x_i y_i = 9948.01,\ \sum_i (x_i - \bar{x})^2 = 192 \text{ och } \sum_i y_i^2 = 4.1252$$

som ger $\beta^* = \dfrac{9948.01 - 7 \cdot 1972 \cdot 0.72}{192} = \dfrac{9.13}{192} = 0.0476$ och $\alpha^* = 0.72 - 0.0476 \cdot 1972 = -93.0527$. Den årliga förändringen av kvävehalten är β och vi får alltså konfidensintervallet (enligt formelsamlingen 12.2 andra intervallet) till $\beta^* \pm t_{0.025}(7-2)\dfrac{s}{\sqrt{192}}$. Vi får σ-skattningen

$$s^2 = \frac{Q_0}{7-2} = \frac{1}{5}\left(\sum_i (y_i - \bar{y})^2 - (\beta^*)^2 \sum_i (x_i - \bar{x})^2\right) = \frac{1}{5}\left(\sum_i y_i^2 - 7 \cdot (\bar{y})^2 - (\beta^*)^2 \cdot 192\right) =$$

$$\frac{1}{5}(4.1252 - 7 \cdot 0.72^2 - 0.0476^2 \cdot 192) = 0.0124$$

och konfidensintervallet för β blir

$$0.0476 \pm 2.57\sqrt{\frac{0.0124}{192}} = \underline{0.0476 \pm 0.0207}.$$