Kurskod: TAMS65
Provkod: TEN1

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Tentamen onsdagen den 22 augusti 2018 kl 8–12

Hjälpmedel: Formelsamling i matematisk statistik utgiven av matematiska institutionen och/eller formelsamling ”Formel- och tabellsamling i matematisk statistik TAMS65 (Martin Singull)”. Inga anteckningar i formelsamlingarna är tillåtet. Miniräknare med tömda minnen.
Betygsgränser: 8-11 poäng ger betyg 3, 11.5-14.5 ger betyg 4 och 15-18 poäng ger betyg 5.
Examinator: Martin Singull, Matematisk statistik, MAI
Resultatet meddelas *normalt* via LADOK inom 12 arbetsdagar.
Tydliga svar och motiveringar krävs till varje uppgift.

1. Antag att

   -0.36 1.46 0.15 0.67 1.55 2.04 -0.12 2.26

   är ett observerat stickprov från den stokastiska variabeln $X \sim N(\mu, \sigma)$.

   (a) Skatta väntevärdet μ och standardavvikelsen σ baserat på de åtta observationerna ovan. (1p)

   (b) Beräkna ett nedåt begränsat 95% konfidensintervall för väntevärdet μ. (2p)

2. Man vill undersöka hur trivseln förändras hos personalen vid ett företag när ledningen ändrar till en ny personalpolitik. Man genomför därför en medarbetarundersökning innan man inför den nya politiken och sen samma undersökning igen efter att man har infört den nya politiken. Resultat för tio slumpmässiga individer i personalen där högre värde är bättre trivsel

Individ	1	2	3	4	5	6	7	8	9	10
Före	27.6	44.3	37.7	62.8	32.4	76.1	23.5	14.1	31.2	71.2
Efter	33.2	51.7	38.2	76.2	35.1	75.1	32.4	29.3	25.3	75.5

   Vilken personalpolitik verkar bäst? Motivera ditt svar med hjälp av ett lämpligt 95% konfidensintervall. Normalfördelning kan förutsättas. (3p)

3. En stor konsultfirma vill undersöka lönsamheten (y) hos olika kunder och studerar olika variabler som skulle kunna påverka. Man har undersökt följande

   $x_1 =$ antal egna konsulter hos kunden,
   $x_2 =$ antal år i branchen för kunden,
   $x_3 =$ medelåldern för konsulterna uthyrda till kunden,
   $x_4 =$ totala antalet konsulter hos kunden (även konkurrerande),
   $x_5 =$ 1 om det finns 2 eller flera konkurrerande konsulter hos kunden, 0 annars.

   Man undersökte 30 kunder och fick följande observationer

y	20	9	...	142	194
x_1	5	10	...	19	15
x_2	1	1	...	9	10
x_3	28.6	38.1	...	54.0	61.9
x_4	20	21	...	44	40
x_5	1	1	...	1	1

Data har analyserats enligt två olika modeller

Modell 1: $\ln Y = \beta_0 + \beta_1 x_1 + \beta_2 \ln x_2 + \varepsilon,$

Modell 2: $\ln Y = \beta_0 + \beta_1 x_1 + \beta_2 \ln x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \tilde{\varepsilon},$

där ε och $\tilde{\varepsilon}$ antas vara normalfördelade.

Modell 1: Skattad regressionslinje:

$$\ln y = 2.77 - 0.03 x_1 + 1.22 \ln x_2$$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	2.7669	0.1908
1	-0.0298	0.0210
2	1.2212	0.1269

Variansanalys:

	Frihetsgrader	Kvadratsumma
REGR	2	11.5879
RES	27	2.0676
TOT	29	13.6555

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.4753 & -0.0297 & -0.0460 \\ -0.0297 & 0.0057 & -0.0246 \\ -0.0460 & -0.0246 & 0.2103 \end{pmatrix}$$

Modell 2: Skattad regressionslinje:

$$\ln y = 2.49 - 0.08 x_1 + 1.18 \ln x_2 + 0.01 x_3 + 0.018 x_4 - 0.064 x_5$$

i	$\widehat{\beta}_i$	$d(\widehat{\beta}_i)$
0	2.4886	0.3248
1	-0.0758	0.0454
2	1.1819	0.1552
3	0.0099	0.0096
4	0.0175	0.0208
5	-0.0643	0.1343

Variansanalys:

	Frihetsgrader	Kvadratsumma
REGR	5	11.7141
RES	24	1.9414
TOT	29	13.6555

(a) Undersök med hjälp av lämpligt test eller konfidensintervall om båda förklaringsvariablerna i **Modell 1** gör nytta. Använd nivån $\alpha = 0.05$ i varje test/konfidensintervall. (1p)

(b) Man planerar att hyra ut nio konsulter till ett nytt företag som har varit etablerade i ett år. Vilken lönsamhet kan man räkna med? Konstruera ett lämpligt 95% intervall. Använd **Modell 1**. (2p)

(c) Beskriver **Modell 2** datamaterialet bättre än **Modell 1**? Genomför ett lämpligt test på nivån 0.05. (1p)

4. Ett mindre finansbolag har utvecklat ett program som känner av onaturlig handel på börsen. Om handeln har varit onaturlig en dag så skickas ett meddelande till mäklarna på kvällen. Antag att sannolikheten att ett sådant meddelande skickas är p och att dessa händelser är oberoende av varandra. Antag att $x_1, ..., x_n$ är n oberoende observationer på antalet dagar innan meddelande skickas, vilka är observationer från en *'för första gången-fördelning'* med

$$p_X(x) = (1-p)^{x-1}p, \quad x = 1, 2, 3, \ldots .$$

Skatta p med maximum-likelihood-metoden baserat på stickprovet om n oberoende observationer av X. (2p)

5. Låt x vara en observation från $X \sim Bin(n, p)$ med en okänd parameter p. Vi vill testa följande hypotes

$$H_0 : p = 0.4 \quad \text{mot} \quad H_1 : p > 0.4,$$

på nivån högst 5%.

(a) Låt $n = 10$ och använd X som teststorhet. Ange det kritiska området. (1p)

(b) Beräkna styrkan för testet i (a) då $p = 0.8$. (1p)

(c) Låt $n = 100$ och använd X som teststorhet. Ange det kritiska området genom att utnyttja lämplig normalapproximation. (1p)

(d) Beräkna styrkan för testet i (c) då $p = 0.8$. (1p)

6. För att bestämma en fysikalisk konstant μ görs ett försök som ger två mätvärden x och y. Dessa mätvärden kan antas vara observationer från de stokastiska variablerna X och Y med $\mathrm{E}(X) = \mathrm{E}(Y) = \mu$, $\mathrm{var}(X) = 1, \mathrm{var}(Y) = 4$ och känd korrelationskoefficient ρ.

Som skattning av μ används en linjär skattningsfunktion $\hat{\mu} = aX + bY$.

(a) Bestäm a och b så att skattningen $\hat{\mu}$ blir väntevärdesriktig och så effektiv som möjligt. (1.5p)

(b) Vad blir skattningen av μ då X och Y är oberoende. (0.5p)

Kurskod: TAMS65
Provkod: TEN2

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Lösningsförslag till tentamen onsdagen den 22 augusti 2018 kl 8–12

1. (a) $\hat{\mu} = \bar{x} = \underline{\underline{0.96}}$ och $s = \sqrt{\frac{1}{8-1}\sum_{i=1}^{8}(x_i - \bar{x})^2} = \underline{\underline{1.01}}$

   (b) $I_\mu = \left(\bar{x} - \underbrace{t_{0.95}(7)}_{=1.89}\frac{s}{\sqrt{8}}\ ;\ \infty\right) = \underline{\underline{(0.28\ ;\ \infty)}}$

2. Före: x_i är en observation från $X_i \sim N(\mu_i, \sigma_1)$, $i = 1, ..., 10$, samt
   Efter: y_i är en observation från $Y_i \sim N(\mu_i + \Delta, \sigma_2)$, $i = 1, ..., 10$.
   Bilda differenserna $Z_i = Y_i - X_i \sim N(\Delta, \sigma)$, $i = 1, ..., 10$. Vi vill testa hypotesen

$$H_0 : \Delta = 0 \quad \text{mot} \quad H_1 : \Delta \neq 0, \quad \text{på nivån } 5\%,$$

   genom att bilda ett 95% konfidensintervall för Δ.
   $\widehat{\Delta} = \bar{z} = 5.11$ som är en observation från $\bar{Z} \sim N\left(\Delta, \frac{\sigma}{\sqrt{10}}\right)$, där vi skattar σ^2 med
   $s^2 = \frac{1}{10-1}\sum_{i=1}^{1} 0(z_i - \bar{z})^2 = 41.85.$
   Stäng in hjälpvariabeln $\frac{\bar{Z} - \Delta}{s/\sqrt{10}} \sim t(7)$ enligt $95\% = P\left(-t_{0.975}(9) \leq \frac{\bar{Z} - \Delta}{S/\sqrt{10}} \leq t_{0.975}(9)\right)$
   vilket ger intervallet

$$I_\Delta = \left(\bar{z} - t_{0.975}(9)\frac{s}{\sqrt{10}}\ ;\ \bar{z} + t_{0.975}(9)\frac{s}{\sqrt{10}}\right) = \underline{\underline{(0.49\ ;\ 9.73)}} > 0.$$

   Alltså, förändringen i personalpolitiken ökar trivseln med stor sannolikhet.

3. (a)

$$L(p) = \prod_{i=1}^{n}(1-p)^{x_i - 1}p = p^n(1-p)^{\sum_{i=1}^{n} x_i - n}$$
$$l(p) = \ln L(P) = n \ln p + n(\bar{x} - 1)\ln(1-p)$$
$$l'(p) = \frac{n}{p} - \frac{n(\bar{x}-1)}{1-p} = 0$$
$$\Rightarrow \quad \hat{p} = \frac{1}{\bar{x}}$$

   Max eftersom $l''(p) = -\left(\frac{n}{p^2} + \frac{n(\bar{x}-1)}{(1-p)^2}\right) < 0$ för alla p.

(b) $\mathrm{E}\left(\frac{1}{\hat{p}}\right) = \mathrm{E}(\bar{x}) = \mathrm{E}(x) = \frac{1}{p}$, eftersom ffg-fördelning. Ja, skattningen $\frac{1}{\hat{p}}$ är vvr.

4. (a) Förkasta H_0 om $x \geq c$.

$$5\% \geq P(X \geq c \mid X \sim Bin(10, 0.4))$$

Tabell ger $c = 8$, dvs. kritiskt område ges av $\{x \geq 8\}$ eller $\{x > 7\}$.

(b) Styrkan för $p = 0.8$ ges av

$$h(0.8) = P(X \geq 8 \mid X \sim Bin(10, 0.8)) = 67.8\% \quad \text{(använd tabell)}$$

(c)

$$5\% \geq P(X > c \mid X \sim Bin(100, 0.4)) = 1 - P(X \leq c \mid X \sim Bin(100, 0.4))$$
$$\Rightarrow \quad 95\% = P(X \leq c \mid X \approx N(40, \sqrt{24})) = P\left(\frac{X-40}{\sqrt{24}} \leq \frac{c-40}{\sqrt{24}}\right)$$

vilket ger att $\frac{c-40}{\sqrt{24}} = 1.645 \quad \Rightarrow \quad c = 40 + 1.645\sqrt{24} = 48$ och kritiskt område ges av $\{x > 48\}$ eller $\{x \geq 49\}$.

(d) Styrkan för $p = 0.8$ ges nu av

$$\begin{aligned} h(0.8) &= P(X > 48 \mid X \sim Bin(100, 0.8)) \\ &= 1 - P(X \leq 48 \mid X \approx N(80, 4)) \\ &= 1 - \Phi\left(\frac{48-80}{4}\right) = \Phi\left(\frac{80-48}{4}\right) = \Phi(8) = 100\%. \end{aligned}$$

Väldigt hög styrka.

5. (a) vvr ges av ekvationen $\mathrm{E}(\hat{\mu}) = (a+b)\mu$, dvs., $a + b = 1$. Den effektivaste skattningen har minst varians. Variansen ges av

$$\begin{aligned} \mathrm{var}(\hat{\mu}) &= a^2\,\mathrm{var}(X) + b^2\,\mathrm{var}(y) + 2ab\,\mathrm{kov}(X,Y) = a^2 + 4b^2 + 4ab\rho \\ &= \Big/ b = 1 - a \Big/ = \ldots = (5-4\rho)a^2 + 4(\rho-2)a + 4 \end{aligned}$$

som minimeras för $a = 2\frac{2-\rho}{5-4\rho}$ och $b = 1 - a = \frac{1-2\rho}{5-4\rho}$.

(b) Om X och Y är oberoende så är $\rho = 0$. Skattningen ges då av

$$\hat{\mu} = \frac{4}{5}X + \frac{1}{5}Y.$$