**Kurskod: TAMS65**
**Provkod: TEN1**

# MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

## Tentamen måndagen den 4 juni 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. Ett företag som både har butik i city och i ett köpcentrum strax utanför staden vill undersöka om åldersfördelningen för kunderna är densamma i de båda butikerna. De undersöker därför åldern vid 100 oberoende besök vid respektive butik

| | Ålder | | |
|---|---|---|---|
| | Yngre än 30 | 30-55 | Äldre än 55 |
| City | 20 | 40 | 40 |
| Köpcentrum | 30 | 50 | 20 |

Avgör med lämpligt test om det är samma åldersfördelning vid butikerna. (2p)

2. Vid tillverkning av takstolar prövade man fyra olika skarvningsmetoder. Vid provbelastning erhölls följande knäckhållfasthet (enhet: $10^4$ N).

| | | | | | Medelvärde | Standardavvikelse |
|---|---|---|---|---|---|---|
| Metod 1: | 1.3800 | 1.4977 | 1.4234 | 1.3811 | 1.4206 | 0.0553 |
| Metod 2: | 2.0950 | 1.9907 | 2.0162 | 2.0477 | 2.0374 | 0.0449 |
| Metod 3: | 1.7326 | 1.8165 | 1.8201 | 1.8033 | 1.7931 | 0.0410 |
| Metod 4: | 1.4036 | 1.3881 | 1.4163 | 1.3809 | 1.3972 | 0.0158 |

Modell: Vi har fyra oberoende stickprov från $N(\mu_i, \sigma_i), i = 1, 2, 3, 4$. Vi betraktar också observationerna inom varje stickprov som oberoende.

(a) Har de fyra olika stickproven samma standardavvikelse? Genomför lämpliga test på nivån 5%. Det räcker att du gör ett test men motivera hur du drar dina slutsatser. (1p)

(b) Gör tvåsidiga konfidensintervall för de olika differenserna $\mu_i - \mu_j$. Den simultana konfidensgraden ska vara minst 94%. Finns det skillnader mellan skarvningsmetoderna? (2p)

3. Ett tillverkningsföretag har samlat data över kostnader för leveranser till olika kunder dels då leveranserna har skett i egen regi (= alternativ I) och dels via ombud

(= alternativ II). I utskriften nedan finns resultatet av en analys enligt regressionsmodellen

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon,$$

där $\varepsilon \sim N(0, \sigma)$, $Y$ = kostnaden, $x_1$ = transportsträckan, och $x_2 = 0$ för alternativ I samt $x_2 = 1$ för alternativ II.

Resultat: $y = -2.34 + 0.595 x_1 + 5.01 x_2 + 0.258 x_1 x_2$,

| $i$ | $\widehat{\beta}_i$ | $d(\widehat{\beta}_i)$ |
|---|---|---|
| 0 | -2.343 | 2.047 |
| 1 | 0.59497 | 0.04341 |
| 2 | 5.011 | 2.719 |
| 3 | 0.25807 | 0.07996 |

VARIANSANALYS

| | Frihetsgrader | Kvadratsumma |
|---|---|---|
| REGR | 3 | 2414.39 |
| RES | 16 | 110.56 |
| TOT | 19 | 2524.95 |

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.60659 & -0.01175 & -0.60659 & 0.01175 \\ -0.01175 & 0.00027 & 0.01175 & -0.00027 \\ -0.60659 & 0.01175 & 1.07004 & -0.02715 \\ 0.01175 & -0.00027 & -0.02715 & 0.00093 \end{pmatrix}$$

(a) Skriv ut väntevärdet $E(Y)$ för de båda alternativen I och II. Förklara kort hur de två linjerna skiljer sig åt. (1p)

(b) Beräkna förklaringsgraden $R^2$ för modellen. Verkar modellen passa till observationerna? (1p)

(c) Har transportsträckan betydelse för kostnaden? Genomför lämpligt test på nivån 5% (1p)

(d) Beräkna ett 95% konfidensintervall för skillnaden i $E(Y)$ mellan alternativ II och I då transportsträckan är 50. (2p)

4. Betrakta en stokastisk variabel med följande täthetsfunktion

$$f_X(x) = \theta(x-1) + \frac{1}{2}, \quad 0 \leq x \leq 2.$$

Skatta $\theta$ med mommentmetoden baserat på ett stickprov om $n$ oberoende observationer av $X$. (1p)

5. Vid en hastighetskontroll antar vi att antalet bilar som kör för fort följer en poissonprocess med den okända intensiteten $\lambda$ fortkörare/timme. Under 3 timmar har man noterar 57 fortkörare. Beräkna ett tvåsidigt konfidensintervall för $\lambda$ med approximativt 95% konfidensgrad. Approximationer ska motiveras. (3p)

6. Låt $(X_1, ..., X_n)$ är ett stickprov på den stokastiska variabeln $X$ som är $N(\mu, 0.5)$, där standardavvikelsen $\sigma = 0.5$ är känd. Givet $n = 9$ observationer med $\bar{x} = 1.3$ önskar man testa hypotesen

$$H_0 : \mu = 1 \quad \text{mot} \quad H_1 : \mu > 1,$$

på signifikansnivån 5%.

(a) Genomför ett lämpligt test för hypotesen ovan. (1p)

(b) Beräkna styrkan för testet i (a) då $\mu = 1.5$. (1.5p)

(c) Beräkna hur många observationer $n$ man behöver för att få styrkan minst 99% då $\mu = 1.5$. (1.5p)

**Kurskod: TAMS65**
**Provkod: TEN2**

# MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

## Lösningsförslag till tentamen måndagen den 4 juni 2018 kl 8–12

1. Testar hypotesen $H_0$ : 'Samma åldersfördelning för City och Köpcentrum' mot $H_1$ : inte samma fördelning.

   Tabell med förväntade värden givet $H_0$

| | Ålder | | | |
|---|---|---|---|---|
| | Yngre än 30 | 30-55 | Äldre än 55 | |
| City | 20 | 40 | 40 | 100 |
| | 25 | 45 | 30 | |
| Köpcentrum | 30 | 50 | 20 | 100 |
| | 25 | 45 | 30 | |
| | 50 | 90 | 60 | 200 |
| $\hat{p}_i$ | 0.5 | 0.45 | 0.3 | |

   Teststorhet $T = \dfrac{(20-25)^2}{25} + \cdots + \dfrac{(20-30)^2}{30} = 9.778$

   Förkasta $H_0$ om $T > c = \chi^2_{0.95}((3-1)(2-1)) = \chi^2_{0.95}(2) = 9.22$. Alltså, förkasta $H_0$, det verkar vara olika åldersfördelning.

2. (a) Testa hypotesen $H_0 : \sigma_1 = \sigma_4$ mot $H_1 : \sigma_1 \neq \sigma_4$ på nivån $\alpha = 5\%$.

   Teststorhet $V = \dfrac{S_1^2}{S_4^2} \sim F(3,3)$ under $H_0$. Förkasta $H_0$ om $v < a < 1$ eller $1 < b < v$. Vi har att $v = \dfrac{s_1^2}{s_4^2} = 12.25$ och $b = F_{0.975}(3,3) = 15.44$ samt $a = \dfrac{1}{15.44} = 0.065$. Alltså, förkasta inte $H_0$. Eftersom vi inte kan förkasta $H_0$ när vi jämför den största och minsta stickprovsstandardavvikelsen (lika många observationer) så kan vi heller inte förkasta för någon annan jämförelse. Vi kan alltså inte förkasta hypotesen om lika standardavvikelser för de fyra stickproven.

   (b) Antag att $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma$. Bilda $I_{\mu_i - \mu_j}$ med konfidensgraden 99%.

   $\hat{\mu}_i - \hat{\mu}_j = \bar{x}_i - \bar{x}_j$ som är en observation från $\bar{X}_i - \bar{X}_j \sim N\left(\mu_i - \mu_j, \sigma\sqrt{\dfrac{1}{4} + \dfrac{1}{4}}\right)$.

   Hjälpvariabel $\dfrac{\bar{X}_i - \bar{X}_j - (\mu_i - \mu_j)}{S\sqrt{\dfrac{1}{4} + \dfrac{1}{4}}} \sim t(12)$,

   där $S^2 = \dfrac{(4-1)S_1^2 + (4-1)S_2^2 + (4-1)S_3^2 + (4-1)S_4^2}{12}$ med $s^2 = 0.001751$.

Intervallen blir nu

$$I_{\mu_i-\mu_j} = \left(\bar{x}_i - \bar{x}_j \mp t\,\frac{s}{\sqrt{2}}\right) = \left(\bar{x}_i - \bar{x}_j \mp 0.090\right),$$

där $t = t_{0.995}(12) = 3.05$. Vidare är

$$\bar{x}_i - \bar{x}_j = \begin{cases} -0.6168 & i=1,\ j=2, \\ -0.3725 & i=1,\ j=3, \\ 0.0234 & i=1,\ j=4, \\ 0.2443 & i=2,\ j=3, \\ 0.6402 & i=2,\ j=4, \\ 0.3959 & i=3,\ j=4 \end{cases}$$

vilket ger att $\underline{\underline{\mu_1,\ \mu_4 < \mu_3 < \mu_2}}$, dvs. $\mu_2$ är störst och vi kan inte säga något mellan $\mu_1$ och $\mu_4$.

3. (a) Momentmetoden ger

$$E(X) = \int_0^2 xf(x)dx = ... = \frac{2}{3}\theta + 1 = \mu(\theta),$$

$$\mu(\hat{\theta}) = \bar{x} \quad \Rightarrow \quad \underline{\underline{\hat{\theta} = \frac{3(\bar{x}-1)}{2}}}.$$

(b) $E(\hat{\Theta}) = \dfrac{3}{2}(E(\bar{X}) - 1) = \dfrac{3}{2}(E(X) - 1) = \dfrac{3}{2}\left(\left[\dfrac{2}{3}\theta + 1\right] - 1\right) = \theta$, <u>dvs. vvr.</u>

Vidare gäller

$$E(X^2) = \int_0^2 x^2 f(x)dx = ... = \frac{4(\theta+1)}{3},$$

$$var(X) = E(X^2) - (E(X))^2 = \frac{1}{3} - \frac{4\theta^2}{9}$$

och

$$var(\hat{\Theta}) = \frac{9}{4}\,\mathrm{var}(\bar{X}) = \frac{9}{4}\frac{\mathrm{var}(X)}{n} = \underline{\underline{\frac{1}{n}\left(\frac{3}{4} - \theta^2\right)}}.$$

4. $x = 57$ är en observation från $X \sim Po(3\lambda) \approx N(3\lambda, \sqrt{3\lambda})$, där approximationen är ok eftersom $\hat{\lambda} = \dfrac{x}{3} = 19$. Vidare har vi att

$\hat{\lambda} = \dfrac{X}{3} \approx N\left(\lambda, \sqrt{\dfrac{\lambda}{3}}\right)$ vilket ger hjälpvariabeln

$$\frac{\hat{\lambda} - \lambda}{\sqrt{\frac{\hat{\lambda}}{3}}} \approx N(0,1) \quad \text{och intervallet} \quad I_\lambda = \left(\hat{\lambda} \mp z_{0.975}\sqrt{\frac{\hat{\lambda}}{3}}\right) = \underline{\underline{(14.07\ ;\ 23.93)}}.$$

Vi kan alltså förvänta oss mellan 14 och 24 fortkörare under 3 timmar.

5. (a) Teststorhet $Z = \dfrac{\bar{X}-1}{0.5/\sqrt{9}} \sim N(0,1)$ under $H_0$ med $z = 1.8$ från data. Förkasta $H_0$ om $z > c = z_{0.95} = 1.645 \quad \Rightarrow$ Förkasta $H_0$.

(b) Styrkan ges av

$$\begin{aligned} h(1.5) &= P\left(\frac{\bar{X}-1}{0.5/\sqrt{9}} > 1.645 | \mu = 1.5\right) \\ &= P\left(\frac{\bar{X}-1.5}{0.5/\sqrt{9}} > \underbrace{1.645 - \frac{0.5}{0.5/\sqrt{9}}}_{=-1.346} | \bar{X} \sim N(1.5, 0.5/\sqrt{9})\right) \\ &= 1 - \Phi(-1.346) = \Phi(1.346) = 91\%. \end{aligned}$$

(c) Vi har att

$$5\% = P\left(\frac{\bar{X}-1}{0.5/\sqrt{n}} > c | \mu = 1\right) \quad \Rightarrow c = 1.645$$

och

$$\begin{aligned} 99\% &\leq P\left(\frac{\bar{X}-1}{0.5/\sqrt{n}} > c | \mu = 1.5\right) \\ &= P\left(\frac{\bar{X}-1.5}{0.5/\sqrt{n}} > c - \sqrt{n} | \bar{X} \sim N(1.5, 0.5/\sqrt{9})\right) = \Phi(\sqrt{n} - c) \end{aligned}$$

vilket ger att $2.33 \leq \sqrt{n} - c$ och $n \geq 16$.