Tentamen TMSB18 Matematisk statistik IL 091015
Tid: 08.00-13.00
Telefon: 036-101620 (Abrahamsson),
Examinator: F Abrahamsson

TEKNISKA HÖGSKOLAN
HÖGSKOLAN I JÖNKÖPING

1. Livslängden för en viss tvättmaskin är exponentialfördelad med en genomsnittlig livslängd 10 år. Tillverkaren lovar att gratis reparera maskiner som går sönder innan garantitiden löpt ut.

   (a) Om tillverkaren endast vill reparera högst 8% av maskinerna som går sönder utan ersättning, hur lång garantitid skall han erbjuda köparen? (2p)

   (b) Om en hyresvärd köper 7 tvättmaskiner av denna typ, vad är då sannolikheten att minst 2 av dessa kommer att gå sönder inom en 3-årsperiod? (2p)

   Lösning: (a) Om ξ är maskinens livslängd och om T är garantitiden så vill vi att

$$P(\xi < T) = 0.08 \Leftrightarrow \int_0^T \frac{1}{10} e^{-\frac{1}{10}x}\,dx = 0.08 \Leftrightarrow 1 - e^{-\frac{1}{10}T} = 0.08$$

   vilket ger $T = -10 \ln 0.92 \approx 0.83$ år ≈ 10 månader.

   (b) Om η är antalet maskiner som går sönder inom en 3-årsperiod så är $\eta \in \mathrm{Bin}(7, p)$ med

$$p = P(\xi < 3) = 1 - e^{-\frac{1}{10}\cdot 3} \approx 0.259.$$

   Alltså är

$$P(\eta \geq 2) = 1 - P(\eta = 0) - P(\eta = 1) = 1 - (1-p)^7 - 7p(1-p)^6 \approx 0.58$$

2. En producent av frukostflingor påstår att förpackningen innehåller 240 g torkad frukt. En konsumentorganisation misstänker att den faktiska mängden torkad frukt är mindre än 240 g. För att undersöka saken undersöks 28 förpackningar och man finner att medelvärdet av fruktmängden är 227 g. Kan konsumentorganisationen påstå att fruktmängden är mindre än 240 g om man är villig att acceptera 1% felrisk?

   (a) Besvara frågan genom att göra att lämpligt hypotestest under antagande att fruktmängden är normalfördelad med standardavvikelse $\sigma = 27$ g. (1p)

   (b) Bestäm testets styrka om den sanna fruktmängden är 222 g. (2p)

   (c) Besvara frågan genom att konstruera ett lämpligt konfidensintervall. Samma normalfördelningsantagande som i a-uppgiften. (1p)

   Lösning: (a) Vi vill testa $H_0 : \mu = 240$ mot $H_1 : \mu < 240$ på signifikansnivån 1%. Testvariabeln är

$$Z_0 = \frac{\overline{\xi} - 240}{27/\sqrt{28}}$$

   och vi förkastar H_0 om $z_0 < -z_\alpha = -z_{0.01} \approx -2.32$ Vi får här

$$z_0 = \frac{227 - 240}{27/\sqrt{28}} \approx -2.54$$

   Alltså förkastar vi H_0 på signifikansnivån 1% dvs. konsumentorganisationen kan påstå att de har rätt.

   (b) Om $\mu = 222$ är

$$Z_0 = \frac{\overline{\xi} - 240}{27/\sqrt{28}} = \underbrace{\frac{\overline{\xi} - 222}{27/\sqrt{28}}}_{\in N(0,1)} - \underbrace{\frac{18}{27/\sqrt{28}}}_{\approx 3.53} \in N(-3.53, 1)$$

   så att testets styrka blir

$$1 - \beta = P(Z_0 < -2.32 | Z_0 \sim N(-3.53, 1)) = \Phi\left(\frac{-2.32 - (-3.53)}{1}\right) = \Phi(1.21) \approx 0.89$$

(c) Vi kan ej förkasta H_0 på signifikansnivån 1% om

$$Z_0 \geq -\lambda_{0.01} \Leftrightarrow \mu_0 \leq \overline{\xi} + \lambda_{0.01}\frac{27}{\sqrt{28}}$$

där $\mu_0 = 240$. Vi bestämmer alltså ett uppåt begränsat konfidensintervall för μ och förkastar H_0 om μ_0 ligger utanför detta intervall. Här fås konfidensintervallet

$$[-\infty, \overline{x} + \lambda_{0.01}\frac{27}{\sqrt{28}}] = [-\infty, 238.9]$$

och eftersom detta intervall inte täcker över $\mu_0 = 240$ så förkastar vi H_0 på signifikansnivån 1%.

3. Antalet timmar som ett hushåll dammsuger under en vecka är en kontinuerlig stokastisk variabel ξ med frekvensfunktion

$$f(x) = \begin{cases} \frac{3}{4}x(2-x)\,, & \text{om } 0 < x < 2\,, \\ 0\,, & \text{övriga } x \end{cases}$$

(a) Bestäm den kumulativa fördelningsfunktionen $F(x)$ för ξ. (2p)

(b) Vad är sannolikheten att ett hushåll dammsuger mer än 90 minuter under en vecka? (1p)

Lösning: (a)

$$F(x) = P(\xi \leq x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0\,, & x \leq 0 \\ \frac{3x^2 - x^3}{4}\,, & 0 < x \leq 2 \\ 1\,, & x \geq 2 \end{cases}$$

(b)

$$P(\xi > \frac{3}{2}) = 1 - P(\xi \leq \frac{3}{2}) = 1 - F(\frac{3}{2}) = 1 - \frac{27}{32} = \frac{5}{32}$$

4. En sträcka med längd L skall mätas som en summa av två delsträckor dvs. vi skriver $L = L_1 + L_2$. Mätningarna av delsträckorna kan ses som observationer av normalfördelade oberoende stokastiska variabler med väntevärden L_1 resp. L_2 och varians $\sigma = 0.1$ m. Vad är sanolikheten att det uppmätta värdet kommer att avvika högst 0.1 m från det korrekta värdet? (2p)

Lösning: Låt ξ_i för $i = 1, 2$ vara mätresultatet för delsträcka i. Då är $\xi_i \in N(L_i, 0.1)$ och den uppmätta längden är $\xi = \xi_1 + \xi_2 \in N(L, \sqrt{2} \cdot 0.1)$. Vi får

$$\begin{aligned} P(|\xi - L| \leq 0.1) &= P(-0.1 \leq \xi - L \leq 0.1) = P\left(-\frac{1}{\sqrt{2}} \leq \underbrace{\frac{\xi - L}{\sqrt{2} \cdot 0.1}}_{\in N(0,1)} \leq \frac{1}{\sqrt{2}}\right) \\ &= \Phi\left(\frac{1}{\sqrt{2}}\right) - \Phi\left(-\frac{1}{\sqrt{2}}\right) = 2\Phi\left(\frac{1}{\sqrt{2}}\right) - 1 \approx 0.52 \end{aligned}$$

5. En robot är programmerad att förflytta sig enligt följande: den skall gå 50 steg och inför varje nytt steg är det en slumpgenerator som bestämmer i vilken riktning steget sker; med sannolikheten 0.6 går roboten ett steg i nordlig riktning och med sannolikheten 0.4 går roboten i östlig riktning. Om vi lägger in robotens rörelser i ett koordinatsystem med start i origo, vad är då sannolikheten att roboten avslutar sin promenad i punkten med koordinater $(25, 25)$? (2p)

Lösning: Vi noterar att roboten kommer att stanna i punkten $(25, 25)$ om och endast om den har tagit 25 steg i nordlig riktning och resterande 25 steg i östlig riktning. Låt ξ vara antalet steg i nordlig riktning som roboten tar under de 50 stegen. Då är $\xi \in \text{Bin}(50, 0.6)$ och

$$P\,(\text{roboten hamnar i } (25, 25)) = P(\xi = 25) = \binom{50}{25} 0.6^{25} \cdot 0.4^{25} \approx 0.04$$

6. Ett handelsföretag i Jönköpingsregionen har behov av att rekrytera en logistikkunnig person som ska ansvara för kravställning när man inom en snar framtid ska byta ut sitt egenutvecklade IT-system mot ett standardsystem av ERP-typ. Ledningen på företaget har dock varit bekymrade över att den teoretiska logistikkompetensen hos de sökande varit relativt låg varför man utformat ett antal testfrågor som man ska använda i andra omgången av urvalsprocessen. Frågorna utgår från en av företagets volymprodukter vilken man prognostiserar med exponentiell utjämning och styr med beställningspunktssystem. Ledtiden för den här typen av produkter, som beställs från ett lågkostnadsland, är ca 2 månader. Prognos och parametrar i beställningspunktssystemet uppdateras varje månad och de senaste månadernas prognoser (F) och efterfrågehistorik (D) har tagits ur det nuvarande IT-systemet och framgår av tabell 1.

t	Dec	Jan	Feb	Mar	Apr
D_t	156	152	165	158	152
F_t	155.0	155.2	154.6	156.6	156.9

Tabell 1: Försäljningshistorik och prognos för de senaste månaderna.

I dagsläget använder man en säkerhetslagernivå, för alla artiklar av den aktuella typen, som är satt till 20 st men målet är att dimensionera säkerhetslagret baserat på SERV1 (dvs sannolikhet för att brist inte uppstår under en ordercykel). Den önskade servicenivån är satt till 95% och i prognostiseringen har man valt att sätta utjämningsfaktorn α till 0.2 som en lagom avvägning mellan följsamhet och stabilitet. För att underlätta beräkningen tillåter man att standardavvikelsen skattas med 1.25 multiplicerat med medelabsolutavvikelsen (MAD) i de senaste två perioderna.

(a) Beräkna prognosen för nästa period (Maj). (1p)

(b) Bestäm lämplig beställningspunkt. (1p)

(c) Förklara hur du i (b)-uppgiften använde antagandet om oberoende stokastiska variabler. (1p)

Lösning: (a) Prognostisering med exponentiell utjämning baseras på ekvationen:

$$F_t = \alpha D_{t-1} + (1-\alpha)F_{t-1}\,.$$

I det här fallet blir det: $F_{\text{Maj}} = 0.2 \cdot 152 + (1-0.2) \cdot 156.9 = 155.9$ st.

(b) Beställningspunkten beräknas som vanligt med $BP = D \cdot L + SS$ där D fås från prognosen i a-uppgiften och $L = 2$ månader är given i uppgiften. Säkerhetslagret SS beräknas enligt SERV1 som $SS = k \cdot \sigma_L$ där det är viktigt att komma ihåg att σ_L är standardavvikelsen över ledtid. Denna kan beräknas baserat på osäkerheten i prognos (vilken skattas enligt uppgiftstexten) som:

$$\sigma_L = \sqrt{L} \cdot \sigma = \sqrt{2} \cdot 1.25 \cdot \frac{|156.6 - 158| + |156.9 - 152|}{2} = 5.54\,.$$

Säkerhetsfaktorn k tas ur en $N(0,1)$-tabell:

$$k = \Phi^{-1}(SERV1) = \Phi^{-1}(0.95) = 1.64\,.$$

Beställningspunkten kan då beräknas som $BP = 155.9 \cdot 2 + 5.54 \cdot 1.64 = 320.9$ st.

(c) Vid dimensionering av beställningspunkten utgick man från behov under ledtid och för beräkning av säkerhetslagret från osäkerhet i behov under ledtid. Prognosperioden avvek dock ifrån ledtiden varför en linjärkombination av osäkerheten skapades vilket i det här fallet är summan av två prognosperioder som motsvarar en ledtid. I den beräkningen adderades varianser vilket endast kan göras om osäkerheten i olika prognosperioder kan antas vara oberoende av varandra.

7. Det hävdas att andra gången man gör högskoleprovet så förbättras resultatet. Man har följande resultat för 10 personer som har gjort provet två gånger.

Person	1	2	3	4	5	6	7	8	9	10
Första gången	1.2	1.5	0.9	1.3	1.4	1.8	1.7	0.6	1.0	0.9
Andra gången	1.3	1.6	1.3	1.6	1.7	1.7	1.8	0.8	1.1	1.1

(a) Genomför ett teckentest på signifikansnivån 1% för att se om det går att styrka påståendet ovan. (2p)

(b) Med antagandet att det individuella resultatet på högskoleprovet följer en normalfördelning, genomför ett relevant hypotestest på 1% signifikansnivå för att se om det går att styrka påståendet ovan. (2p)

Lösning: (a) Vi använder en modell där stickprovet är parvis ordnat $(\xi_1, \eta_1), \ldots, (\xi_{10}, \eta_{10})$ med individens resultat första, resp. andra gången. Differenserna $\eta_i - \xi_i$ har här resulterat i $r_+ = 9$ positiva resultat. Vi vill testa

$$\begin{array}{rl} H_0: & \text{testresultatet är inte bättre andra gången} \\ H_1: & \text{testresultatet är bättre andra gången} \end{array}$$

Vi beräknar därför

$$P - \text{värdet} = P\left(R_+ \geq 9 \mid R_+ \in \text{Bin}(10, \frac{1}{2})\right) = \binom{10}{9}\left(\frac{1}{2}\right)^9 \frac{1}{2} + \binom{10}{10}\left(\frac{1}{2}\right)^{10} \approx 0.011$$

På signifikansnivån 1% kan vi alltså inte förkasta H_0.

(b) Med antagandet om normalfördelning är $\xi_i \in N(\mu_i, \sigma_1)$ och $\eta_i \in N(\mu_i + \Delta, \sigma_2)$ så att $Z_i = \eta_i - \xi_i \in N(\Delta, \sigma)$ där $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. Vi vill nu testa

$$\begin{array}{rl} H_0: & \Delta = 0 \\ H_1: & \Delta > 0 \end{array}$$

på signifikansnivån 1%. Observerade värden $z_i = y_i - x_i$ beräknas;

Person	1	2	3	4	5	6	7	8	9	10
Första gången x_i	1.2	1.5	0.9	1.3	1.4	1.8	1.7	0.6	1.0	0.9
Andra gången y_i	1.3	1.6	1.3	1.6	1.7	1.7	1.8	0.8	1.1	1.1
Differens $z_i = y_i - x_i$	0.1	0.1	0.4	0.3	0.3	-0.1	0.1	0.2	0.1	0.2

Vi får $\overline{z} = 0.17$ och $s_z = 0.14184$. Testvariabelns värde blir

$$t_0 = \frac{\overline{z} - 0}{s_z/\sqrt{10}} \approx 3.79.$$

Med $\alpha = 0.01$, $n = 10$ är $t_\alpha(n-1) = t_{0.01}(9) = 2.82$. Eftersom $t_0 > 2.82$ så förkastas H_0 på signifikansnivån 1%.

8. Ett visst sjukhus får $2/5$ av sitt influensavaccin från företag A och resten från företag B. Från företag A är 3% av doserna verkningslösa, och från företag B är 2% av doserna verkningslösa. Sjukhuset testar 25 doser vaccin, valda på måfå från ett av företagen och finner att 2 av doserna var verkningslösa. Vad är då sannolikheten att de testade vaccinen kom från företag A? (3p)

Lösning: Låt X vara antalet verkningslösa bland 25 testade. Låt A och B vara händelserna att vi har valt de 25 testvaccinen från företag A resp. B. Vi söker då $P(A|X = 2)$. Enligt Bayes sats är

$$P(A|X = 2) = \frac{P(X = 2|A)P(A)}{P(X = 2|A)P(A) + P(X = 2|B)P(B)}.$$

Vidare så är

$$P(X = 2|A) = P(X = 2|X \in \text{Bin}(25, 0.03)) = \binom{25}{2} 0.03^2 0.97^{23} \approx 0.134$$

och

$$P(X = 2|B) = P(X = 2|X \in \text{Bin}(25, 0.02)) = \binom{25}{2} 0.02^2 0.98^{23} \approx 0.0.075\,.$$

Insättning ovan ger då att

$$P(A|X = 2) \approx \frac{0.134 \cdot \frac{2}{5}}{0.134 \cdot \frac{2}{5} + 0.075 \cdot \frac{3}{5}} \approx 0.542.$$