**Matematisk statistik**
**Matematikcentrum**
**Lunds universitet**

**Tentamen: 2024–08–30 kl $8^{00}$–$13^{00}$**
**FMSF30 & FMSF32**
**Matematisk statistik**

# Lösningsförslag

---

1. *Notering:* Denna uppgift liknar Vännman uppgift 2.29, där definitionen av oberoende skall användas.

   Vi betecknar utfallet med $a$ prickar på första tärningen och $b$ på den andra som $(a, b)$. Vi får då

$$\Omega = \left\{ \begin{array}{llllll} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\}$$

$$A = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$$

$$B = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

   och sannolikheterna blir $P(A) = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$ och $P(B) = \frac{|B|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$. (0.3)

   (b) Om $A$ och $B$ är oberoende gäller att $P(A \cap B) = P(A)\,P(B)$. Vi har vänsterled $P(A \cap \; B) = \frac{A \cap B}{|\Omega|} = \frac{|(3,4)|}{|\Omega|} = \frac{1}{36}$ och högerled $P(A)\,P(B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$. Händelserna A och B är således oberoende. (0.3)

   (c) Vi har händelsen $C = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$ med $P(C) = 5/36$. Vi får då $P(A \cap C) = \frac{1}{36} \neq \; P(A)P(C)$; A och C är <u>inte</u> oberoende. Utfallet av första tärningen påverkar sannolikheten att få totalt 6 prickar. (0.4)

2. *Notering*: Detta är samma uppgift (med andra värden) som tas upp i föreläsningsmaterialet.

   Med definitionen för väntevärde och varians får vi

$$E(\xi) = \sum_x x \; P(\xi = x) = 0 \cdot 0.35 + 1 \cdot 0.5 + 2 \cdot 0.1 + 3 \cdot 0.05 = 0.85$$

$$E(\xi^2) = \sum_x x^2 \; P(\xi = x) = 1^2 \cdot 0.5 + 2^2 \cdot 0.1 + 3^2 \cdot 0.05 = 1.35$$

$$D(\xi) = \sqrt{E(\xi^2) - E(\xi)^2} = \sqrt{1.35 - 0.85^2} = 0.7921$$

   Eftersom $\eta$ är summan av ett stort antal ($n = 1000$) likafördelade och oberoende slumpvariabler kan denna normalapproximeras med väntevärde $nE(\xi) = 850$ och standardavvikelse $\sqrt{n}D(\xi) = 25.05$, dvs $\eta \in N(850, 25.05)$. Vi beräknar $P(\eta < 1000) = \Phi(\frac{1000-850}{25.05} = \Phi(5.99) = 1.00$ ($= 0.999999998938049$). 1000 parkeringsplatser lär räcka mer än väl. (0.5)

   (b) Vi söker kvantilen $P(\eta > x_{0.01}) = 0.01$. Vi standardiserar olikheten och sätter lika med $\lambda_{0.01} = 2.3263$ från tabell;

$$\frac{x_{0.01} - nE(\xi)}{\sqrt{n}D(\xi)} = \lambda_{0.01} \quad \Rightarrow \quad x_{0.01} = nE(\xi) + \sqrt{n}D(\xi) \cdot \lambda_{0.01} = 908.2737.$$

909 parkeringsplatser vore tillräckligt för att täcka allas behov med sannolikheten (minst) 99 %. (0.5)

3. *Notering:* Denna uppgift liknar Vännman uppgifterna 9.5 och 9.6.

För medel gäller att $\bar{\xi} \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ där $\sigma = 15$ är känd. Vi beräknar kritisk nivå $k$ som 5 % (övre) kvantil under förutsättning att $\mu = \mu_0 = 100$ (ingen förbättring):

$$P(\bar{\xi} > k \mid \mu = \mu_0) = 0.05 \quad \Leftrightarrow \quad P\left(\overbrace{\frac{\bar{\xi} - \mu_0}{\sigma/\sqrt{n}}}^{N(0,1)} > \frac{k - \mu_0}{\sigma/\sqrt{n}}\right) = 0.05 \quad \Rightarrow$$

$$\frac{k - \mu_0}{\sigma/\sqrt{n}} = \lambda_{0.05} \quad \Leftrightarrow \quad k = \mu_0 + \lambda_{0.05} \cdot \sigma/\sqrt{n}.$$

För $\lambda_{0.05} = 1.6449$ blir således $k = 100 + 1.6449 \cdot 15/\sqrt{4} = 112.4$. Vi beräknar $\bar{x} = 109.68$ och vi har $\bar{x} < k$: Vi kan med felrisken 5 % <u>inte</u> påstå att livslängden har förbättrats. (0.5)

(b) Vi ansätter $k = \mu_0 + \lambda_{0.05} \cdot \sigma/\sqrt{n}$ och beräknar styrkan i $\mu = \mu_1 = 110$ för godtyckligt n:

$$P(\bar{\xi} > \mu_0 + \lambda_{0.05} \cdot \sigma/\sqrt{n} \mid \mu = 110) = 0.99 \quad \Leftrightarrow$$

$$P\left(\overbrace{\frac{\bar{\xi} - \mu_1}{\sigma/\sqrt{n}}}^{N(0,1)} > \frac{\mu_0 + \lambda_{0.05} \cdot \sigma/\sqrt{n} - \mu_1}{\sigma/\sqrt{n}}\right) = 0.99 \quad \stackrel{\text{sym.}}{\Leftrightarrow}$$

$$P\left(\overbrace{\frac{\bar{\xi} - \mu_1}{\sigma/\sqrt{n}}}^{N(0,1)} > -\left(\frac{\mu_0 + \lambda_{0.05} \cdot \sigma/\sqrt{n} - \mu_1}{\sigma/\sqrt{n}}\right)\right) = 0.01 \quad \Rightarrow$$

$$\frac{\mu_0 + \lambda_{0.05} \cdot \sigma/\sqrt{n} - \mu_1}{\sigma/\sqrt{n}} = -\lambda_{0.01} \quad \Rightarrow$$

$$\mu_1 - \mu_0 = \frac{\sigma}{\sqrt{n}}(\lambda_{0.01} + \lambda_{0.05}) \quad \Rightarrow$$

$$n = \sigma^2 \left(\frac{\lambda_{0.01} + \lambda_{0.05}}{\mu_1 - \mu_0}\right)^2$$

Med $\lambda_{0.01} = 2.3263$ behövs minst $n = 35.4835 = 36$ prototyper. (0.5)

*Var god vänd!*

4. *Notering*: Denna uppgift liknar uppgift 14.2 i regressionsmaterialet.
Som hjälp för skattningarna har följande kvantiteter beräknats:
$\bar{x} = 176.9$, $\bar{y} = 179.5$, $S_{xx} = 7124.6$, $S_{xy} = 3747.3$, samt $S_{yy} = 5541.7$.

   (a) Skattningarna blir $\beta^*_{obs} = \frac{S_{xy}}{S_{xx}} = 0.67$, $\alpha^*_{obs} = \bar{y} - \beta^*_{obs}\bar{x} = 59.82$ och
   $s = \sqrt{\frac{S_{yy} - S_{xy}^2/S_{xx}}{71-2}} = 7.19$. (0.3)

   (b) Förklaringsgraden är $R^2 = \frac{S_{xy}^2}{S_{xx}Syy} = 0.36$. Modellens residualer ser både normalfördelade och oberoende ut: helt enligt antagande. Förälderns längd förklarar dock enbart 36 % av barnens längd, vilket säger att det finns andra faktorer som också avgör barnets längd (exempelvis den andra föräldern, äldre ättlingar, kost m.m.) (0.1)

   (c) Galton ansåg (som andra) att sambandet är positivt, alltså $\beta > 0$. Men specifikt ansåg han att det sker en viss återgång till gruppens medelvärde, dvs. att en förändring med $x$ cm inte fullt ut motsvarar en förändring med $y$ cm, alltså $\beta < 1$. Vi formulerar hypoteserna (0.2)

$$
\begin{aligned}
&H_0 : \beta = 1 \\
&H_1 : 0 \leq \beta \leq 1 \quad \text{eller möjligen} \quad H_1 : \beta < 1
\end{aligned}
$$

   (d) Eftersom mothypotsen är formulerad som ett intervall gör vi ett tvåsidigt konfidensintervall för $\beta$, men även ett ensidigt övre begränsat intervall accepteras. Skattningen har fördelningen $\beta^* \in N\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right)$ där $\sigma$ skattas med $s$, vilket ger t-kvantilen $t_{0.005}(69) < t_{0.005}(60) = 2.660$. Vi bildar intervallet

$$
I_\beta = \beta^*_{obs} \pm t_{0.005}(60)\frac{s}{\sqrt{S_{xx}}} = 0.67 \pm 0.2367 = [\ 0.44, 0.90\ ]. \tag{1}
$$

   Eftersom konfidensintervallet inte innehåller $\beta = 1$ (eller $\beta = 0$) kan vi förkasta nollhypotesen på signifikansnivå 0.01. Det verkar finnas en återgång till medelvärdet (*regression to the mean*). (0.4)

5. *Notering:* Denna uppgift liknar Vännman uppgift 4.2, med annan frekvensfunktion.
För en frekvensfunktionen gäller att $f(x) \geq 0,\ \forall x$ och $\int_{-\infty}^{\infty} f(x)dx = 1$. Det första villkoret ger att $k > 0$. För det andra villkoret får vi att

$$
\begin{aligned}
1 &= \int_{-\infty}^{0} \frac{k}{(x-\lambda)^2}dx + \int_{0}^{\infty} \frac{k}{(x+\lambda)^2}dx = \left[\frac{(-1)\cdot k}{x-\lambda}\right]_{-\infty}^{0} + \left[\frac{(-1)\cdot k}{x+\lambda}\right]_{-\infty}^{0} \\
&= \left(\frac{-k}{-\lambda} + 0\right) + \left(0 + \frac{k}{\lambda}\right) = \frac{2k}{\lambda} \quad \Rightarrow \quad k = \frac{\lambda}{2}.
\end{aligned} \tag{2}
$$

Samt, för $\lambda = 1$:

$$
\begin{aligned}
P(\epsilon < 0.5) &= 1 - P(\epsilon > 1/2) = 1 - \int_{1/2}^{\infty} \frac{\lambda}{2(x+\lambda)^2}dx = 1 - \left[\frac{-\lambda}{2(x+\lambda)}\right]_{1/2}^{\infty} \\
&= 1 - \left(-0 + \frac{\lambda}{1+2\lambda}\right) = \frac{1+2\lambda-\lambda}{1+2\lambda} = \left.\frac{1+\lambda}{1+2\lambda}\right|_{\lambda=1} = \frac{2}{3}.
\end{aligned}
$$

*Alternativ lösning*: För att slippa integrera över hela intervallet, kan vi se att $f(x)$ är symmetrisk runt $x = 0$, dvs att $f(-x) = f(x),\ \forall x$. Det gör att $\int_{-\infty}^{0} f(x)dx = \int_{0}^{\infty} f(x)dx = 1/2$. Vi integrerar för $x < 0$ och får (1.0)

$$\frac{1}{2} = \int_{-\infty}^{0} \frac{k}{(x-\lambda)^2} dx = \left[\frac{(-1)\cdot k}{x-\lambda}\right]_{-\infty}^{0} = \left(\frac{-k}{-\lambda} - 0\right) = \frac{k}{\lambda} \quad \Rightarrow \quad k = \frac{\lambda}{2}.$$

6. *Notering:* Denna uppgift liknar tenta 230825 uppgift 5, eller tenta 230412 uppgift 6(a).

(a) Vi får $q = 1 - F(50) = 1 - \Phi\left(\frac{ln(50)-3.5}{0.25}\right) = 1 - \Phi(1.65) = 1 - 0.9505 = 0.0495.$ (0.2)

(b) Vi har att en förare kör alltför fort med sannolikhet $q$, eller gör det ej med sannolikhet $1-q$. Förarnas hastigheter är oberoende och $\eta$ avser antal fortkörare. Alltså är villkoren uppfyllda för att $\eta \in Bin(n, 0.0495)$. (0.2)

(c) För $n = 75$ kontrollerade fordon förväntar vi oss $E(\eta) = nq = 3.7$ fortkörare men vi observerade endast $x_0 = 1$. Vi beräknar p-värdet exakt som

$$\begin{aligned} p &= P(\eta \leq 1) = P(\eta = 0) + P(\eta = 1) = \binom{75}{0} q^0 (1-q)^{75} + \binom{75}{1} q^1 (1-q)^{74} \\ &= \frac{75!}{0! \cdot 75!}(1-q)^{75} + \frac{75!}{1! \cdot 74!} q(1-q)^{74} = (1-q)^{75} + 75 \cdot q \cdot (1-q)^{74} \\ &= 0.0222 + 75 \cdot 0.0012 = 0.1122. \end{aligned}$$

Eftersom sannolikheten att få vår observation eller en ännu mer extrem (alltså ännu färre) är $p = 0.11$ är vår observation inte särskilt extrem. Utan att acceptera stor felrisk har vi inte fog för fartkameran missar fortkörare.

. (0.6)

---

Slut!