TAMS24: Statistisk teori — solutions to 14 frågor

$\boxed{1\sim}$

Solution. (a) First $E(X_i) = \int_0^\infty x f(x)dx = 5 - 4a$. From $E(X_i) = \bar{x}$, it follows that $5 - 4a = \bar{x}$. Thus $\hat{a} = (5 - \bar{x})/4$.

(b) Let Y be the number of elements which "har gått sönder under detta halvår", then from independence $Y \sim Bin(n_2, p)$ where

$$\begin{aligned} p &= P(\text{en element "har gått sönder under detta halvår"}) \\ &= P(X_i \leq 1/2) = \int_0^{1/2} f(x)dx = (1-a)(1-e^{-1/10}) + a(1-e^{-1/2}). \end{aligned}$$

It is reasonable to believe that $n_2 \cdot p = y$ (for instance momentmetoden), which gives

$$\hat{a} = (\frac{y}{n_2} - 0.0952)/0.2983.$$

(c) In practice, Metod 2 is much more useful because: in Metod 1 in order to get $\bar{x}$, one needs to collect the lifetime of each element, which means that one has to wait until all elements die (this may take a long time); in Metod 2, in order to get y one just needs to wait for a half year. □

$\boxed{2\sim}$

Solution. Let X be the number of "transistorer" which still work after en tidsenhet, then $X \sim Bin(400, p)$ where

$$\begin{aligned} p &= P(\text{a "transistorer" still works after en tidsenhet}) \\ &= P(Exp(\frac{1}{\mu}) > 1) = \int_1^\infty \frac{1}{\mu} e^{-x/\mu} dx = e^{-1/\mu}. \end{aligned}$$

Thus for mean, we use $400 \cdot p = 109$ to get $\hat{p} = 109/400$ and $\hat{\mu} = \frac{1}{\ln(400/109)} = 0.77$.

For the median, from $P(Exp(\frac{1}{\mu}) > \text{median}) = 1/2$ we have median$=\mu \ln 2$, that is

$$\widehat{\text{median}} = \hat{\mu} \cdot \ln 2 = 0.77 \cdot \ln 2 = 0.53.$$

□

$\boxed{3\sim}$

Solution. First

$$E(X_i) = \int_\theta^1 x \frac{1}{1-\theta} dx = \frac{1}{2}(1+\theta).$$

It follows from $E(X_i) = \bar{x}$ that $\frac{1}{2}(1+\theta) = \bar{x}$, thus $\hat{\theta} = 2\bar{x} - 1$. Den är väntevärdesriktig eftersom

$$E(\hat{\Theta}) = E(2\bar{X} - 1) = 2E(X_i) - 1 = 2 \cdot \frac{1}{2}(1+\theta) - 1 = \theta.$$

□

$\boxed{4\sim}$

Solution. Because of the normal assumption with the same variance, a point estimate of $\mu_i - \mu_j$ would be $\hat{\mu}_i - \hat{\mu}_j = \bar{x}_i - \bar{x}_j$ which is an observation of $\bar{X}_i - \bar{X}_j \sim N(\mu_i - \mu_j, \sigma\sqrt{\frac{1}{n_i} + \frac{1}{n_j}})$. Then we have the hjälpvariabeln

$$\frac{(\bar{X}_i - \bar{X}_j) - (\mu_i - \mu_j)}{S\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \sim t(df)$$

where S is an estimator of σ defined as

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2 + (n_4-1)s_4^2}{n_1+n_2+n_3+n_4-4},$$

and the degrees of freedom $df = n_1+n_2+n_3+n_4-4$. Based on this 99% confidence interval of $\mu_i - \mu_j$ is

$$I_{\mu_i-\mu_j} = (\bar{x}_i - \bar{x}_j) \mp t_{0.005}(df) \cdot s \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}.$$

Therefore we have (notice that $s = 0.1270$ and $df = 11$)

$$I_{\mu_1-\mu_2} = (-0.03, 0.48), \quad I_{\mu_1-\mu_3} = (-0.01, 0.59), \quad I_{\mu_1-\mu_4} = (0.02, 0.70).$$

What we can conclude: $\mu_1 > \mu_4$ (this means that Material 1 is worse than Material 4). Other differences are not significant. □

$\boxed{5\sim}$

Solution. The hjälpvariabeln is

$$\frac{(n_1+n_2+n_3-3)S^2}{\sigma^2} \sim \chi^2(n_1+n_2+n_3-3).$$

(a) A 95% confidence interval I_{σ^2} of type $(0, a^2)$ of σ^2 would be

$$I_{\sigma^2} = (0, a^2) = (0, \; \frac{(n_1+n_2+n_3-3)s^2}{\chi^2_{1-\alpha}(n_1+n_2+n_3-3)}) = (0, \; \frac{27 \cdot 39.45887}{\chi^2_{0.95}(27)}) = (0, 65.76)$$

where $\chi^2_{0.95}(27) = 16.2$ from the table, and

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2}{n_1+n_2+n_3-3} = 39.45887.$$

Thus a 95% confidence interval I_σ of type $(0, a)$ of σ is $I_\sigma = (\sqrt{0}, \sqrt{65.76}) = (0, 8.11)$.

(b) First we know that

$$c_i\bar{X}_i + c_j\bar{X}_j \sim N(c_i\mu_i + c_j\mu_j, \sigma\sqrt{\frac{c_i^2}{n_i} + \frac{c_j^2}{n_j}}),$$

then the hjälpvariabeln is

$$\frac{(c_i\bar{X}_i + c_j\bar{X}_j) - (c_i\mu_i + c_j\mu_j)}{S\sqrt{\frac{c_i^2}{n_i} + \frac{c_j^2}{n_j}}} \sim t(n_1+n_2+n_3-3).$$

(NOTE that the degrees of freedom involves all samples sizes n_1, n_2 and n_3, NOT just n_i and n_j.) In order to investigate whether $\mu_2 > 1.4\mu_3$, we construct a 95% confidence interval of $\mu_2 - 1.4\mu_3$ of the form $(a, +\infty)$. To this end,

$$(a, +\infty) = ((\bar{x}_2 - 1.4\bar{x}_3) - t_\alpha(n_1+n_2+n_3-3)\cdot s\cdot\sqrt{\frac{1^2}{n_2} + \frac{1.4^2}{n_3}},\ +\infty) = (0.17, +\infty).$$

Since $0.17 > 0$, we conclude that $\mu_2 > 1.4\mu_3$. □

$\boxed{6\sim}$

Solution. The idea of the solution (including the hjälpvariabeln) is exactly the same as $\boxed{5\sim}$, so we omit the details.

(a) $I_\sigma = (0, 5.86)$.

(b) $I_{\mu_1-\mu_2} = (2.89, 16.35), I_{\mu_1-\mu_3} = (-12.71, 1.95)$ and $I_{\mu_2-\mu_3} = (-22.67, -7.33)$. Thus we conclude that $\mu_1 > \mu_2$ and $\mu_3 > \mu_2$. Difference between μ_1 and μ_3 is not significant. □

$\boxed{7\sim}$

Solution. Let X be the antal felaktiga bland 500, then $X \sim Bin(500, p)$. From normal approximation, the hjälpvariabeln is

$$\frac{\hat{P} - p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}} \approx N(0,1),$$

where $\hat{p} = x/n = 87/500$. Thus a 95% confidence interval I_p of p is

$$I_p = \hat{p} \mp \lambda_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = (0.141, 0.207).$$

□

$\boxed{8\sim}$

Solution. From normal approximation, the hjälpvariabeln is

$$\frac{\bar{X}-\mu}{\mu/\sqrt{n}} \approx N(0,1).$$

(a) By solving μ, a 95% confidence interval I_μ of μ is

$$I_\mu = \left(\frac{\bar{x}}{1+\frac{\lambda_{\alpha/2}}{\sqrt{n}}},\ \frac{\bar{x}}{1-\frac{\lambda_{\alpha/2}}{\sqrt{n}}}\right) = \left(\frac{4.5}{1+\frac{1.96}{\sqrt{80}}},\ \frac{4.5}{1-\frac{1.96}{\sqrt{80}}}\right) = (3.69,\ 5.76).$$

(b) Be definition $p = P(Exp(\frac{1}{\mu}) > 10) = e^{-10/\mu}$. A 95% confidence interval should be $I_p = (?_1, ?_2)$ where

$$95\% = P(?_1 < p < ?_2) = P(?_1 < e^{-10/\mu} < ?_2) = P(-10/\ln(?_1) < \mu < -10/\ln(?_2)).$$

Thus $-10/\ln(?_1) = 3.69$ and $?_1 = e^{-10/3.69} = 0.067$. Similarly $?_2 = e^{-10/5.76} = 0.176$. $I_p = (0.067, 0.176)$. □

$\boxed{9\sim}$

Solution. From normal approximation, the hjälpvariabeln is

$$\frac{\bar{X}-\mu}{\sqrt{\bar{X}/n}} \approx N(0,1).$$

Thus a 95% confidence interval I_μ of μ is

$$I_\mu = \bar{x} \mp \lambda_{\alpha/2}\sqrt{\bar{x}/n} = 2.02 \mp 1.96\sqrt{2.02/500} = (1.90,\ 2.14).$$

□

$\boxed{10\sim \text{ och } 11\sim}$

Solution. From the definition of Type I error, it follows, $X \sim Po(\lambda t)$,

$$\begin{aligned} 0.01 = \alpha &= P(\text{reject } H_0 \text{ if } H_0 \text{ is true}) = P(X > ? \text{ if } \lambda = 5) \\ &= P\left(\frac{X-\lambda t}{\sqrt{\lambda t}} > \frac{?-\lambda t}{\sqrt{\lambda t}} \text{ if } \lambda = 5\right) = P\left(\frac{X-5t}{\sqrt{5t}} > \frac{?-5t}{\sqrt{5t}}\right) \\ &\approx P\left(N(0,1) > \frac{?-5t}{\sqrt{5t}}\right), \quad \text{thus } ? = 5t + \lambda_{0.01}\sqrt{5t}. \end{aligned}$$

From the definition of power, it follows

$$
\begin{aligned}
0.99 = \alpha = P(\text{reject } H_0 \text{ if } H_0 \text{ is false and } \lambda = 7.5) &= P(X >? \text{ if } \lambda = 7.5) \\
= P(\frac{X - \lambda t}{\sqrt{\lambda t}} > \frac{? - \lambda t}{\sqrt{\lambda t}} \text{ if } \lambda = 7.5) &= P(\frac{X - 7.5t}{\sqrt{7.5t}} > \frac{? - 7.5t}{\sqrt{7.5t}}) \\
\approx P(N(0,1) > \frac{? - 7.5t}{\sqrt{7.5t}}), \quad \text{thus } ? = 7.5t + \lambda_{0.99}\sqrt{7.5t}.
\end{aligned}
$$

Therefore we get

$$5t + \lambda_{0.01}\sqrt{5t} = 7.5t + \lambda_{0.99}\sqrt{7.5t}, \qquad \text{so } t = 21.428.$$

□

$\boxed{12\sim}$

Solution. (a) We do a testing $H_0 : \sigma_1^2 = \sigma_2^2$ mot $H_0 : \sigma_1^2 \neq \sigma_2^2$. We know the test statistic is $s_1^2/s_2^2 = 0.0478$, and the critical region is

$$(0, F_{1-\frac{\alpha}{2}}(7,7)) \cup (F_{\frac{\alpha}{2}}(7,7), +\infty) = (0, 0.1125) \cup (8.89, +\infty).$$

Since the test statistic is in the critical region, we reject H_0. Therefore variances are NOT the same!

(b) Try to get 95% confidence intervals $I_{\mu_2-\mu_1}, I_{\mu_2-\mu_3}$ and $I_{\mu_3-\mu_1}$. At the end, from these intervals one can see that μ_2 is the biggest. □

$\boxed{13\sim}$

Solution. (a) The test statistic is $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{11.2-10}{2.1/\sqrt{25}} = 2.857$, and the critical region is

$$(-\infty, -t_{\alpha/2}(n-1)) \cup (t_{\alpha/2}(n-1), +\infty) = (-\infty, -2.80) \cup (2.80, +\infty).$$

Since the test statistic is in the critical region, we reject H_0.

(b) A 99% confidence interval of μ is

$$I_\mu = \bar{x} \mp t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} = 11.2 \mp 2.80 \cdot \frac{2.1}{\sqrt{25}} = (10.024, \quad 12.376).$$

This is another way to reject H_0 since 10 is not in the 99% confidence interval. □

14~ a) Livslängderna för 50 elektronrör bestämdes och man fick $\bar{x}=38.5$. Man antar att livslängderna är oberoende och exponentialfördelade med ett okänt väntevärde μ. Konstruera ett tvåsidigt konfidensinterval för μ med konfidensgraden 95%

b) Då man såg hur observationera fördelat sig på tallinjen, blev man tveksam över exponentialfördelningsantagandet:

Intervall	Absolut frekvens
$0 \le x < 20$	14
$20 \le x < 40$	18
$40 \le x < 60$	7
$60 \le x < 80$	6
$80 \le x$	5

Undersök med ett χ^2-test på nivån 0.10 om antagandet om exponentialfördelning är rimligt

Solution a) 95% confidence interval of μ (large sample) is

$$\bar{x} \mp \lambda_{\frac{\alpha}{2}} \frac{\bar{x}}{\sqrt{n}} = 38.5 \mp 1.96 \frac{38.5}{\sqrt{50}} = (27.83, 49.17)$$

b) $\hat{\mu}=\bar{x}=38.5$ $\quad X \sim Exp\left(\frac{1}{\hat{\mu}}\right)$

$$\begin{cases} P_1 = p(0 \le X < 20) = \int_0^{20} \frac{1}{\hat{\mu}} e^{-\frac{1}{\hat{\mu}}x} dx = \cdots = 0.4052 \\ P_2 = p(20 \le X < 40) = \cdots = \cdots = 0.241 \\ P_3 = \cdots = \cdots = 0.1433 \\ P_4 = \cdots = 0.0853, \quad nP_4 = 4.265 \\ P_5 = \cdots = 0.1252, \quad nP_5 = 6.26 \end{cases}$$

Combine

After combine,

$$\begin{cases} P_1 = 0.4052, & nP_1 = 20.26 \\ P_2 = 0.241, & nP_2 = 12.05 \\ P_3 = 0.1433, & nP_3 = 7.165 \\ P_4 = 0.2105, & nP_4 = 10.525 \end{cases}$$

$$TS = \sum_{i=1}^{4} \frac{(N_i - nP_i)^2}{nP_i} = 4.897$$

$$C = (\chi^2_\alpha(k-1-\#\text{of unknowns}), +\infty) = (\chi^2_{0.1}(4-1-1), +\infty) = (4.6, \infty)$$