Examinator/Examiner: Zhenxia Liu (Tel: 070 0895208). Please answer in ENGLISH if you can.

a. You are permitted to bring:
- a calculator;
- formel -och tabellsamling i matematisk statistik (from MAI);
- TAMS24: Notations and Formulas (by Xiangfeng Yang)

b. Scores rating: 8-11 points giving rate 3; 11.5-14.5 points giving rate 4; 15-18 points giving rate 5.

English Version

1 (4 points)

A population $X \sim N(\mu, \sigma)$ has unknown μ and σ. A sample $\{x_1, x_2, \ldots, x_n\}$ is from this population X.
(1.1). (1p) Find a point estimate $\hat{\mu}_{ML}$ of μ using Maximum-Likelihood method.
(1.2). (2p) Find a point estimate $\hat{\sigma}^2_{ML}$ of σ^2 using Maximum-Likelihood method.
(1.3). (1p) Is the point estimate $\hat{\sigma}^2_{ML}$ in (1.2) unbiased? Why?

Solution. (1.1) $\hat{\mu}_{MM} = \bar{x}$ (2.2) $\hat{\sigma^2}_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$. (2.3). No! Since $E(\hat{\Sigma^2}_{ML}) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$. □

2 (3 points)

The number of calls X to a telephone exchange during the fastest hour in the dag is a $Po(\mu)$. During 50 days, people have got the following independent observations on X :

Day i	1	2	...	50	$\bar{x}$
Number	90	86	...	120	100.8

Find an approximate 95% confidence interval for μ.

Solution. By CLT, we have $\frac{\bar{X}-\mu}{\sqrt{\mu}/\sqrt{n}} \approx N(0,1)$, so we get the confidence interval for μ is

$$I_\mu = \bar{x} \pm \lambda_{\alpha/2}\frac{\sqrt{\bar{x}}}{\sqrt{n}} \approx (98, \quad 103.6).$$

□

3 (3 points)

10 persons measure their own heights in morning and evening. Results are:

Person	1	2	3	4	5	6	7	8	9	10
Morning	172	168	180	181	160	163	165	177	169	175
Evening	172	167	177	179	159	161	166	175	168	176

The differences between morning and evening can be assumed to be a random sample from $N(\mu, \sigma)$.
(3.1). (2p) Test the hypothesis $H_0 : \mu = 0$ against $H_1 : \mu > 0$ with the significance level 0.05 .
(3.2). (1p) If $\sigma = 1.6$, calculate the power $h(0.5)$ when $\mu = 0.5$.

Solution. (3.1) Let z be the difference between morning and evening, then we get $\bar{z} = 1$ and the sample variance for the difference is $s^2 = \frac{16}{9}$. Since σ is unknown, we apply t distribution.

$$TS = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1 - 0}{\sqrt{\frac{16}{9}}\sqrt{1/10}} \approx 2.37.\text{and } C = (t_\alpha(n-1), \infty) = (t_{0.05}(9), \infty) = (1.83, \infty).$$

Because $TS \in C$, we reject H_0.
(3.2) According to the definition of power, we have

$$\begin{aligned} h(0.5) &= P(\text{reject } H_0 \text{ when } H_0 \text{ is false if } \mu = 0.5) \\ &= P(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \lambda_\alpha \text{ if } \mu = 0.5) \\ &(\text{need to change } \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \text{ to } \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ since } \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)) \\ &= P(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \lambda_\alpha - \frac{\mu}{\sigma/\sqrt{n}}) \\ &= P(N(0,1) > 0.66) = 0.2546. \end{aligned}$$

□

4 (3 points)

$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$ is a normal vector. Assume that $X_i \sim N(5, \sqrt{10})$ for $i = 1, 2, 3$, and $cov(X_1, X_2) = 8$, $cov(X_1, X_3) = 2$, $cov(X_2, X_3) = 1$.

(4.1). (2p) Determine the mean vector and the covariance matrix for $\begin{pmatrix} X_1 - X_2 \\ X_2 \\ X_2 + X_3 \end{pmatrix}$.

(4.2). (1p) Let $Y = X_1 - X_2 + X_3$, find $P(Y > 1)$.

Solution. (4.1). The mean vector and covariance matrix for $\mathbf{X}$ is

$$\mu_{\mathbf{X}} = \begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix} \text{ and } C_{\mathbf{X}} = \begin{pmatrix} 10 & 8 & 2 \\ 8 & 10 & 1 \\ 2 & 1 & 10 \end{pmatrix}.$$

We have

$$\begin{pmatrix} X_1 - X_2 \\ X_2 \\ X_2 + X_3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

Then the mean vector and covariance matrix for $\begin{pmatrix} X_1 - X_2 \\ X_2 \\ X_2 + X_3 \end{pmatrix}$ is

$$\mu = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \mu_{\mathbf{X}} = \begin{pmatrix} 0 \\ 5 \\ 10 \end{pmatrix} \text{ and } C = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} C_{\mathbf{X}} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}' = \begin{pmatrix} 4 & -2 & -1 \\ -2 & 10 & 11 \\ -1 & 11 & 22 \end{pmatrix}.$$

(4.2). Notice that

$$Y = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \mathbf{X} \sim N(5, 4).$$

So we get

$$P(Y > 1) = P(N(0,1) > \frac{1-5}{4}) = 0.8413.$$

□

5 (3 points)

The waiting times from day openings to their first entering customers are observed, and one got $\bar{x} = 3$. These observations are in the following four groups

Interval	Frequency
$0 \le x < 2$	15
$2 \le x < 3$	20
$3 \le x < 4$	8
$4 \le x$	6

Use a χ^2-test with a level 0.10 to determine if these observations are from an exponential distribution $Exp(\frac{1}{\mu})$.

Solution.

$$H_0 : \text{The population can be assumed as } Exp(\frac{1}{\mu})$$

$$H_1 : \text{The population can NOT be assumed as} Exp(\frac{1}{\mu})$$

$N_1 = 15, N_2 = 20, N_3 = 8, N_4 = 6,$ and $p_1 \approx 0.487, p_2 \approx 0.146, p_3 \approx 0.104, p_4 \approx 0.263$. Therefore,

$$TS = \sum_{i=1}^{4} \frac{(N_i - np_i)^2}{np_i} \approx 31.7, \text{where } n = 49$$

$$C = (\chi^2_\alpha((k-1-1)), \quad +\infty) = (\chi^2_{0.1}(2), \quad +\infty) = (4.6, \quad +\infty).$$

Since $TS \in C$, we reject H_0. Namely, this population can NOT be assumed as exponential distribution. □

6 (2 points)

One wants to utilize a regression model to estimate the prices of passenger aircrafts. The dependent variable is

$$Y = \text{aircraft price/number of seats (unit: 1000-tal dollars)},$$

and explanatory variables are

$$x_1 = \text{takeoff weight/number of seats},$$
$$x_2 = \ln \text{(speed)}.$$

<u>Model</u> is: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, where ε is assumed to be $N(0, \sigma)$. Observation values give:

Estimated regression line: $y = -283 + 0.688x_1 + 50.3x_2$

i	$\hat{\beta}_i$	$d(\hat{\beta}_i)$
0	-282.7	142.4
1	0.6881	0.2765
2	50.29	21.73

	Degrees of freedom	Sum of squares
REGR	2	5119.3
RES	14	6605.7
TOT	16	11725.0

(6.1). (1p) Find a 95% confidence interval for β_2.
(6.2). (1p) Can we remove x_1 with $\alpha = 5\%$? Why?

Solution. (6.1). The 95% confidence interval of β_2 which is

$$I_{\beta_2} = \hat{\beta}_2 \mp t_{0.025}(14) \cdot d(\hat{\beta}_2) = 50.29 \mp 2.14 \cdot (21.73) = (3.8, \quad 96.8).$$

(6.2).We can do the hypothesis testing

$$H_0 : \beta_1 = 0 \qquad mot \qquad H_1 : \beta_1 \neq 0$$

$TS = \frac{\hat{\beta}_1 - 0}{d(\hat{\beta}_1)} = 0.6881/0.2765 \approx 2.48$, and
$C = (-\infty, -t_{\alpha/2}(n-k-1)) \cup (t_{\alpha/2}(n-k-1), +\infty) = (-\infty, -2.14) \cup (2.14, +\infty)$. Since $TS \in C$, we reject H_0. So we believe that $\beta_1 \neq 0$, i.e. the variable x_1 is useful/important, so we can't remove x_1. □

Svensk Version

1 (4 poäng)

En population $X \sim N(\mu, \sigma)$ har okända μ och σ. Ett stickprov $\{x_1, x_2, \ldots, x_n\}$ är från denna population X.
(1.1). (1p) Hitta en punktskattning $\hat{\mu}_{ML}$ av μ genom att använda Maximum Likelihood metoden.
(1.2). (2p) Hitta en punktskattning $\hat{\sigma}^2_{ML}$ av σ^2 genom att använda Maximum Likelihood metoden.
(1.3). (1p) Är den punktskattningen $\hat{\sigma}^2_{ML}$ i (1.2) väntevärdesriktig? Varför?

2 (3 poäng)

Antalet anrop X till en telefonväxel under den brådaste timmen på dagen är $Po(\mu)$. Under 50 dagar har man fått följande oberoende observationer på X :

Dag i	1	2	...	50	$\bar{x}$
Antal	90	86	...	120	100.8

Beräkna ett approximativt 95% konfidensintervall för μ.

3 (3 poäng)

10 personer mäter sin egen längd (enhet: cm) morgon och kväll. Resultat:

Person	1	2	3	4	5	6	7	8	9	10
Morgon	172	168	180	181	160	163	165	177	169	175
Kväll	172	167	177	179	159	161	166	175	168	176

Skillnaderna mellan morgon och kvällsvärdena antas vara ett slumpmässigt stickprov från $N(\mu, \sigma)$.
(3.1). (2p) Pröva hypotesen $H_0 : \mu = 0$ mot $H_1 : \mu > 0$ på signifikansnivån 0.05 .
(3.2). (1p) Om $\sigma = 1.6$, beräkna styrkan $h(0.5)$ när $\mu = 0.5$.

4 (3 points)

$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$ är en normal vektor. Antag att $X_i \sim N(5, \sqrt{10})$ för $i = 1, 2, 3$, och $cov(X_1, X_2) = 8$, $cov(X_1, X_3) = 2$, $cov(X_2, X_3) = 1$.

(4.1). (2p) Bestäm väntevärdesmatris och kovariansmatris för $\begin{pmatrix} X_1 - X_2 \\ X_2 \\ X_2 + X_3 \end{pmatrix}$.

(4.2). (1p) Låt $Y = X_1 - X_2 + X_3$, beräkna $P(Y > 1)$.

5 (3 poäng)

Väntetiden X (enhet: minut) från öppningsdags till dess frösta kunden kommer in i en affär observeras, och man fick $\bar{x} = 3$. Dessa observationer är i följande fyra grupper

Intervall	Frekvens
$0 \leq x < 2$	15
$2 \leq x < 3$	20
$3 \leq x < 4$	8
$4 \leq x$	6

Undersök med ett χ^2-test på nivån 0.10 om dessa observationer är från exponentialfördelning $Exp(\frac{1}{\mu})$.

6 (2 poäng)

Man vill utnyttja en regressionsmodell för att beräkna priser på passagerarplan. Som beroende variabel har man

$$Y = \text{flygplanets pris/antal passagerarplatser (enhet: 1 000-tal dollar)},$$

och som förklaringsvariabler

$$x_1 = \text{startvikten/antalet passagerarplatser},$$
$$x_2 = \ln\text{ (hastigheten)}.$$

Modell är: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, där ε antas vara $N(0, \sigma)$. Observerade värden ger:

$$\text{Skattad regressionslinje: } y = -283 + 0.688x_1 + 50.3x_2$$

i	$\hat{\beta}_i$	$d(\hat{\beta}_i)$
0	-282.7	142.4
1	0.6881	0.2765
2	50.29	21.73

	Frihetsgrader	Kvadratsumma
REGR	2	5119.3
RES	14	6605.7
TOT	16	11725.0

(6.1). (1p) Beräkna ett 95% konfidensintervall för β_2.
(6.2). (1p) Kan vi ta bort x_1 med $\alpha = 5\%$? Varför?