Examinator/Examiner: Xiangfeng Yang (Tel: 070 0896661). 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. (3 points) English

A baker bakes three large chocolate cakes every working day. If the cakes are not sold during the day, then they will be given to a so-called 'food Bank'. The following results are obtained for one year:

Number of sold cakes	Number of days
0	1
1	16
2	55
3	228

Test the hypothesis that the number of sold cakes during one working day is a Binomial distribution $Bin(3,p)$. Use significance level 1%.

Solution. There is one unknown parameter p in the hypothesis $Bin(3,p)$, so first we must estimate this unknown parameter using the given sample:

$$\hat{p} = \frac{\text{cakes to the 'food Bank'}}{\text{total baked cakes}} = \frac{(0\times 1 + 1\times 16 + 2\times 55 + 3\times 228)}{(3\times 1 + 3\times 16 + 3\times 55 + 3\times 228)} = \frac{810}{900} = 0.9.$$

Therefore, we can compute the theoretical probabilities as

$$p_0 = P(Bin(3,\hat{p}) = 0), \quad \ldots, \quad p_3 = P(Bin(3,\hat{p}) = 3).$$

But it turned out that $np_0 < 5$, thus we need to combine the two groups 0 and 1 to get only three groups:

$$\begin{aligned} 0 \text{ and } 1: &\quad p_0 + p_1 = 0.028; \\ 2: &\quad p_2 = 0.243; \\ 3: &\quad p_3 = 0.729. \end{aligned}$$

Thus the test statistics is (keep in mind $n = 300, N_1 = 1 + 16 = 17, N_2 = 55, N_3 = 228$)

$$TS = \sum_{i=1}^{3} \frac{(N_i - np_i)^2}{np_i} = 13.6.$$

On the other hand, the rejection region is

$$C = (\chi^2_\alpha(3-1-1), \quad +\infty) = (6.64, \quad +\infty).$$

Since $TS \in C$, reject H_0, namely, we don't think that it is a Binomial distribution $Bin(3,p)$. □

2. (3 points) English

The following data are 40 observations from a Poisson distribution with a parameter μ (which is the mean):

Value	0	1	2	3	4
Frequency	21	0	11	6	2

Find the Maximum-Likelihood estimate for μ given this information.

Solution. The likelihood function is

$$L(\mu) = (e^{-\mu})^{21} \cdot (\mu e^{-\mu})^{0} \cdot (\frac{\mu^2}{2} e^{-\mu})^{11} \cdot (\frac{\mu^3}{6} e^{-\mu})^{6} \cdot (\frac{\mu^4}{24} e^{-\mu})^{2} = e^{-40\mu} \cdot u^{48} \cdot constant.$$

The logarithmic function is

$$\ln L(\mu) = -40\mu + 48 \ln \mu + constant.$$

By setting $0 = \ln' L(\mu) = -40 + 48/\mu$, we have $\mu = 1.2$. The Maximum-Likelihood estimate for μ is

$$\hat{\mu}_{ML} = 1.2.$$

□

3. (3 points) English

Let X_1, X_2, X_3 be random variables such that

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 4 \\ 4 \\ 4 \end{pmatrix}, \quad \begin{pmatrix} 3 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 3 \end{pmatrix} \right).$$

(3.1). (1p) Find $P(X_2 > X_1 + 1)$.
(3.2). (2p) One wants to do a linear combination

$$Y = a_1 X_1 + a_2 X_2 + a_3 X_3$$

such that $E(Y) = 4$ and $V(Y)$ is minimized. Find a_1, a_2 and a_3.

Solution. (3.1). Using the 'An important theorem' on the formula paper, one can easily see that

$$(X_2 - X_1 - 1) \sim N(-1, \sqrt{7}).$$

So,

$$P(X_2 > X_1 + 1) = P(X_2 - X_1 - 1 > 0) = P(N(-1, \sqrt{7}) > 0) = P(N(0,1) > 1/\sqrt{7}) = 0.352.$$

(3.2). The first condition $E(Y) = 4$ gives:

$$4a_1 + 4a_2 + 4a_3 = 4. \tag{1}$$

To use the second condition, we first compute the variance:

$$V(Y) = \text{...use covariance matrix...} = 3a_1^2 - 2a_1a_2 - 2a_1a_3 + 2a_2^2 - 2a_2a_3 + 3a_3^2.$$

Now we replace every a_3 in $V(Y)$ by (1): $a_3 = 1 - a_1 - a_2$, and regard $V(Y) =: f(a_1, a_2)$ as a function of a_1 and a_2. To get a minimal value, we must take the partial derivatives: as follows:

$$0 = \frac{\partial f(a_1, a_2)}{\partial a_1} = 16a_1 + 8a_2 - 8; \tag{2}$$

$$0 = \frac{\partial f(a_1, a_2)}{\partial a_2} = 8a_1 + 14a_2 - 8. \tag{3}$$

Solving the above two equations gives $a_1 = 0.3$ and $a_2 = 0.4$, which in turn yields $a_3 = 0.3$.

□

4. (3 points) English

The following two questions (4.1) and (4.2) are independent of each other.

(4.1). (1.5p) For a number of coal plants with similar purification techniques one has measured the sulfur dioxide emissions y (ppm) and the power x (gigawatt). An analysis according to the model

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon, \qquad \varepsilon \sim N(0, \sigma),$$

gave results in the following output. Find a 95% confidence interval for $E(Y)$ when the power is 0.5 gigawatt.

i	$\hat{\beta}_i$	$d(\hat{\beta}_i)$
0	204.46	82.95
1	-638.0	298.5
2	959.2	263.6

	Degrees of freedom	Sum of squares
REGR	2	15049.3
RES	6	254.3
TOT	8	15303.6

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 162.38 & -581.90 & 508.70 \\ -581.90 & 2102.17 & -1850.58 \\ 508.70 & -1850.58 & 1639.75 \end{pmatrix}.$$

(4.2). (1.5p) Consider the simple linear regression model

$$Y_j = \beta_0 + \beta_1 x_j + \varepsilon_j, \qquad j = 1, \ldots, n$$

where $\varepsilon_j, j = 1, \ldots, n$, are independent $N(0, \sigma)$. The least-square-method's estimator $\begin{pmatrix} \hat{B}_0 \\ \hat{B}_1 \end{pmatrix}$ of the parameter $\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ is a normal vector $N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \sigma^2 (X'X)^{-1}\right)$. Prove that $\hat{B}_0$ and $\hat{B}_1$ are independent if and only if $\sum_{j=1}^n x_j = 0$.

Solution. (4.1). The confidence interval is

$$I_\mu = \hat{\mu} \mp t_{\alpha/2}(n-k-1) \cdot s \cdot \sqrt{\mathbf{x}' \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}} = (118.02, 132.52).$$

The values are as follows:
$\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 \cdot 0.5 + \hat{\beta}_2 \cdot 0.5^2$;
$t_{\alpha/2}(n-k-1) = t_{0.05/2}(9-2-1)$;
$s^2 = SS_E/(n-k-1) = 15049.3/(9-2-1)$;
$\mathbf{x}' = (1, 0.5, 0.5^2)$.

(4.2). It is clear that $X'X = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \cdot \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum_{j=1}^n x_j \\ \sum_{j=1}^n x_j & \sum_{j=1}^n x_j^2 \end{pmatrix}$. Thus,
$(X'X)^{-1} = \frac{1}{n\sum_{j=1}^n x_j^2 - (\sum_{j=1}^n x_j)^2} \begin{pmatrix} \sum_{j=1}^n x_j^2 & -\sum_{j=1}^n x_j \\ -\sum_{j=1}^n x_j & n \end{pmatrix}$. The covariance of $\hat{B}_0$ and $\hat{B}_1$ is

$$cov(\hat{B}_0, \hat{B}_1) = \frac{-\sigma^2}{n\sum_{j=1}^n x_j^2 - (\sum_{j=1}^n x_j)^2} \sum_{j=1}^n x_j,$$

which means that $\hat{B}_0$ and $\hat{B}_1$ are independent if and only if $\sum_{j=1}^n x_j = 0$. □

5. (3 points) English

In order to compare the efficacy of three different antihypertensive medications, one treated three groups (each group has ten patients) with the different medicines. After a month, the blood pressure reductions are measured. Results are:

	Sample mean $\bar{x}_i$	Sample standard deviation s_i
Medicine 1:	17.3	6.19
Medicine 2:	21.1	7.26
Medicine 3:	10.8	5.23

Model: We take these three samples from $N(\mu_i, \sigma), i = 1, 2, 3$. Does it seem like $\mu_2 > 1.4\mu_3$? Answer the question by constructing an appropriate confidence interval with confidence level 0.95.

Solution. The unknown parameter σ^2 can be estimated as

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

Since σ is unknown, the help variable 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 $\mu_2 > 1.4\mu_3$.
This problem can be also solved by constructing a two-sided confidence interval. □

6. (3 points) English

For a sample with ten observations $\{x_1, \ldots, x_{10}\}$ from $\{X_1, \ldots, X_{10}\}$ where $X_j \sim N(\mu, \sigma), j = 1, \ldots, 10$, one has estimated the sample standard deviation and got $s = 3.21$.
(6.1). (1p) Test the hypothesis $H_0 : \sigma = 2.5$ against $H_1 : \sigma > 2.5$ with a level $\alpha = 0.05$.
(6.2). (2p) Find the σ-value for which the test in (6.1) has a power 0.9.

Solution. (6.1). It is easy to get

$$TS = \frac{(n-1)s^2}{\sigma_0^2} = \frac{9\cdot 3.21^2}{2.5^2} = 14.84, \quad C = (\chi^2_{0.05}(9), \quad +\infty) = (16.919, \quad +\infty).$$

Since $TS \notin C$, don't reject H_0.
(6.2).

$$\begin{aligned}
0.9 &= P(\text{reject } H_0,\ \text{given the real value } \sigma) \\
&= P(\frac{9\cdot S^2}{2.5^2} > 16.919,\ \text{given the real value } \sigma) \\
&= P(\frac{9\cdot S^2}{\sigma^2} > 16.919\cdot 2.5^2/\sigma^2,\ \text{given the real value } \sigma) \\
&= P(\chi^2(9) > 16.919\cdot 2.5^2/\sigma^2).
\end{aligned}$$

It is from the χ^2 table that

$$16.919\cdot 2.5^2/\sigma^2 = 4.168.$$

thus

$$\sigma = 5.04.$$

□

Svensk Version

1. (3 poäng) Svenska

En bagare bakar varje arbetsdag tre stora chokladtårtor. De tårtor som inte säljs under dagen ges bort till en sk 'food bank'. Följande resultat erhölls efter ett arbetsår:

Antal sålda tårtor	Antal dagar
0	1
1	16
2	55
3	228

Pröva hypotesen att antalet sålda tårtor under en dag är Binomialfördelat $Bin(3, p)$. Använd signifikansnivå 1%.

2. (3 poäng) Svenska

Följande datamaterial består av 40 observationer från en Poissonfördelning med parameter μ (som är väntevärdet):

Värde	0	1	2	3	4
Frekvens	21	0	11	6	2

Beräkna Maximum-Likelihood skattningen för μ givet denna information.

3. (3 poäng) Svenska

Låt X_1, X_2, X_3 vara stokastiska variabler sådana att

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 4 \\ 4 \\ 4 \end{pmatrix}, \quad \begin{pmatrix} 3 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 3 \end{pmatrix}\right).$$

(3.1). (1p) Beräkna $P(X_2 > X_1 + 1)$.
(3.2). (2p) Man vill göra en linjärkombination

$$Y = a_1X_1 + a_2X_2 + a_3X_3$$

sådan att $E(Y) = 4$ och $V(Y)$ minimiseras. Bestäm a_1, a_2 och a_3.

4. (3 poäng) Svenska

Deluppgifterna (4.1) och (4.2) är fristående.

(4.1). (1.5p) För ett antal kolkraftverk med likartad reningsteknik har man mätt svaveldioxidutsläpp y (ppm) samt effekten x (gigawatt). En analys enligt modellen

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon, \qquad \varepsilon \sim N(0, \sigma),$$

gav ett resultat som finns i utskriften nedåt. Beräkna ett 95% konfidensintervall för $E(Y)$ då kraftverkets effekt är 0.5 gigawatt.

i	$\hat{\beta}_i$	$d(\hat{\beta}_i)$
0	204.46	82.95
1	-638.0	298.5
2	959.2	263.6

	Frihetsgrader	Kvadratsumma
REGR	2	15049.3
RES	6	254.3
TOT	8	15303.6

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 162.38 & -581.90 & 508.70 \\ -581.90 & 2102.17 & -1850.58 \\ 508.70 & -1850.58 & 1639.75 \end{pmatrix}.$$

(4.2). (1.5p) Betrakta den enkla linjära regressionsmodellen

$$Y_j = \beta_0 + \beta_1 x_j + \varepsilon_j, \qquad j = 1, \ldots, n$$

där $\varepsilon_j, j = 1, \ldots, n$, är oberonde $N(0, \sigma)$. Minst-kvadrat-metodens skattningar $\begin{pmatrix} \hat{B}_0 \\ \hat{B}_1 \end{pmatrix}$ av parametrarna $\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ är normalfördelad $N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \sigma^2 (X'X)^{-1}\right)$. Visa att $\hat{B}_0$ och $\hat{B}_1$ är oberoende om och endast om $\sum_{j=1}^{n} x_j = 0$.

5. (3 poäng) Svenska

För att jämföra effekten hos tre olika blodtryckssänkande mediciner behandlas tre grupper om vardera tio patienter med de olika medicinerna. Efter en månad mättes sänkningen av blodtrycket. Resultat:

	Medelvärde $\bar{x}_i$	Stickprovsstand.avvik s_i
Medicin 1:	17.3	6.19
Medicin 2:	21.1	7.26
Medicin 3:	10.8	5.23

Modell: Vi har tre stickprov från $N(\mu_i, \sigma), i = 1, 2, 3$. Förefaller det troligt att $\mu_2 > 1.4\mu_3$? Besvara frågan genom att konstruera ett lämpligt konfidensintervall med konfidensgraden 0.95.

6. (3 poäng) Svenska

För ett slumpmässigt stickprov med tio observationer $\{x_1, \ldots, x_{10}\}$ från $\{X_1, \ldots, X_{10}\}$ där $X_j \sim N(\mu, \sigma), j = 1, \ldots, 10$, har man beräknat stickprovsstandardavvikelsen och fått $s = 3.21$.

(6.1). (1p) Pröva $H_0 : \sigma = 2.5$ mot $H_1 : \sigma > 2.5$ på nivån $\alpha = 0.05$.

(6.2). (2p) Bestäm det σ-värde för vilket testet i (6.1) har styrkan 0.9.