University of Ljubljana

Doctoral Programme in Statistics

Methodology of Statistical Research

Written examination

February 14st, 2025

Name and surname: ____________________ ID number: | | | | | | | | |

Instructions

Read carefully the wording of the problem before you start. There are four problems altogether. You may use a A4 sheet of paper and a mathematical handbook. Please write all the answers on the sheets provided. You have two hours.

Problem	a.	b.	c.	d.	Total
1.			•	•	
2.					
3.				•	
4.					
Total					

1. (25) A population of size N is divided into K groups of equal size $M = N/K$. A sample is selected in such a way that k groups are selected by simple random sampling, and then all the units in the selected groups are selected.

a. (10) Show that the sample average $\bar{Y}$ is an unbiased estimate of the population mean.

   Solution: let μ_i be the population mean in the i-th group. In the sampling procedure described we are choosing a simple random sample of groups and we observe μ_i for this group. The estimator $\bar{Y}$ is just a sample average of the μ_i selected. The expectation is therefore the average of all μ_is which is μ.

b. (15) Let μ_i be the population mean in group i for $i = 1, 2, \ldots, K$ and let μ be the population mean. Define

$$\sigma_b^2 = \frac{1}{K} \sum_{i=1}^{K} (\mu_i - \mu)^2 \,.$$

   Show that

$$\operatorname{se}\left(\bar{Y}\right) = \frac{\sigma_b}{\sqrt{k}} \cdot \sqrt{\frac{K-k}{K-1}} \,.$$

   Solution: think of groups as units selected and to each group assign the value μ_i. The formula is then the formula for the standard error of such a sample average. But $\bar{Y}$ is equal to this sample average.

2. (20) The Birnbaum-Saunders distribution has the density

$$f(x)=\frac{1}{2\gamma}\left(\frac{1}{x^{1/2}}+\frac{1}{x^{3/2}}\right)\exp\left(-\frac{1}{2\gamma^2}\left(\sqrt{x}-\frac{1}{\sqrt{x}}\right)^2\right)$$

for $x>0$ and $\gamma>0$. Assume that the observed values $x_1,\ldots,x_n$ are an i.i.d. sample from the density $f(x)$.

a. (5) Find the MLE estimate for the parameter γ.

Solution: the log-likelihood function is

$$\ell(\gamma,\mathbf{x})=-n\log 2-n\log\gamma+\sum_{k=1}^{n}\left(\frac{1}{x_k^{1/2}}+\frac{1}{x_k^{3/2}}\right)-\frac{1}{2\gamma^2}\sum_{k=1}^{n}\left(x_k^{1/2}-x_k^{-1/2}\right)^2 .$$

Take the derivative to get

$$\frac{\partial\ell}{\partial\gamma}=-\frac{n}{\gamma}+\frac{1}{\gamma^3}\sum_{k=1}^{n}\left(x_k^{1/2}-x_k^{-1/2}\right)^2 .$$

Set the derivative to zero and solve for γ to get

$$\hat{\gamma}=\sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(x_k^{1/2}-x_k^{-1/2}\right)^2}.$$

b. (5) Assume as known that

$$P(X\leq x)=\Phi\left(\frac{1}{\gamma}\left(\sqrt{x}-\frac{1}{\sqrt{x}}\right)\right),$$

where $\Phi(x)$ is the distribution function of the standard normal distribution. Show that the variable Y defined as

$$Y=\sqrt{X}-\frac{1}{\sqrt{X}}$$

has the $N(0,\gamma^2)$ distribution.

Solution: denote $f(x)=\sqrt{x}-1/\sqrt{x}$. The function $f(x)$ is increasing and

$$P(Y\leq y)=P(f(X)\leq y)=\Phi\left(\frac{1}{\gamma}f(f^{-1}(y))\right)=\Phi\left(\frac{y}{\gamma}\right).$$

c. (5) Is

$$\hat{\gamma}^2 = \frac{1}{n}\sum_{k=1}^{n}\left(\sqrt{X_k} - \frac{1}{\sqrt{X_k}}\right)^2$$

an unbiased estimator of γ^2?

Solution: compute

$$E\left[\left(\sqrt{X_k} - \frac{1}{\sqrt{X_k}}\right)^2\right] = \gamma^2\,.$$

It follows that $\hat{\gamma}^2$ is an unbiased estimate of γ^2.

d. (10) Compute the standard error for $\hat{\gamma}$.

Solution: compute the second derivative of the log-likelihood function for $n = 1$.

$$\frac{\partial^2 \ell}{\partial \gamma^2} = -\frac{1}{\gamma^2} + \frac{3}{\gamma^4}\left(\sqrt{x} - \frac{1}{\sqrt{x}}\right)\,.$$

It follows

$$-E\left(\frac{\partial^2 \ell}{\partial \gamma^2}\right) = \frac{2}{\gamma^2}\,.$$

hence

$$\mathrm{se}(\hat{\gamma}) = \frac{\gamma}{\sqrt{2n}}\,.$$

3. (25) Assume the observations $x_1, \ldots, x_n$ are an i.i.d.sample from the $\Gamma(2, \theta)$ distribution with density

$$f(x) = \theta^2 x e^{-\theta x}$$

for $x > 0$ and $\theta > 0$.

a. (5) Find the maximum likelihood estimator for the parameter θ.

Solution: the log-likelihood function is

$$\ell\left(\theta|\mathbf{x}\right) = 2n \log\theta + \sum_{k=1}^{n} \log x_k - \theta \sum_{k=1}^{n} x_k \,.$$

Equating the derivative to 0 *we get*

$$\hat{\theta} = \frac{2n}{\sum_{k=1}^{n} x_k} \,.$$

b. (10) For the testing problem $H_0\colon \theta = 1$ versus $H_1\colon \theta \neq 1$ find the Wilks's test statistic λ. Describe when you would reject H_0 given that the size of the test is $1 - \alpha$ with $\alpha \in (0, 1)$.

Solution: by definition

$$\lambda = 2\ell(\hat{\theta}) - 2\ell(1) \,.$$

Using the maximum likelihood estimator $\hat{\theta}$ *we get*

$$\lambda = -4n \log\left(\frac{\bar{x}}{2}\right) + 2n\left(\bar{x} - 2\right) \,.$$

By Wilks's theorem under H_0 *the distribution of the test statistic* λ *is approximately* $\chi^2(1)$. *The null-hypothesis is rejected when* $\lambda > c_\alpha$ *where* c_α *is such that* $P(\chi^2(1) \geq c_\alpha) = \alpha$.

c. (10) The function

$$f(y) = -4n \log\left(\frac{y}{2}\right) + 2n(y - 2)$$

is strictly decreasing on $(0, 2)$ and strictly increasing on $(2, \infty)$. Assume for all $c > \min_{y>0} f(y)$ you can find the two solutions of the equation $f(y) = c$. Can you use this information to give an exact test given $\alpha \in (0, 1)$? Describe the procedure. No calculations are required.

Hint: by properties of the gamma distribution $\bar{X} \sim \Gamma(2n, \theta/n)$.

Solution: given the assumptions we can find such a c_α *that under* H_0 *we have*

$$P_{H_0}\left(f(\bar{X}) \geq c_\alpha\right) = \alpha \,.$$

Let $x_1 < x_2$ *be the solutions of the equation* $f(x) = c_\alpha$. *The test that rejects* H_0 *when either* $\bar{X} < x_1$ *or* $\bar{X} > x_2$ *is exact.*

4. (25) Assume the regression equations are

$$Y_k = \alpha + \beta x_k + \epsilon_k$$

for $k = 1, 2, \ldots, n$. The error terms satisfy the assumptions that

$$E(\epsilon_k) = 0 \quad \text{and} \quad \text{var}(\epsilon_k) = \sigma^2(1+\tau^2)$$

for $k = 1, 2, \ldots, n$, and

$$\text{cov}(\epsilon_k, \epsilon_l) = \sigma^2\tau^2$$

for $k \neq l$, where τ^2 is assumed to be a known constant. Assume that $\sum_{k=1}^n x_k = 0$.

a. (5) Let

$$\hat{\alpha} = \frac{1}{n}\sum_{k=1}^n Y_k \qquad \text{and} \qquad \hat{\beta} = \frac{\sum_{k=1}^n x_k Y_k}{\sum_{k=1}^n x_k^2}$$

be the ordinary least squares estimators of the two regression parameters. Show that the estimators are unbiased.

Solution: from the assumptions we have

$$E\left(\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix}\right) = \begin{pmatrix}\frac{1}{n}\sum_{k=1}^n(\alpha+\beta+E(\epsilon_k))\\ \frac{\sum_{k=1}^n x_k(\alpha+\beta x_k+E(\epsilon_k))}{\sum_{k=1}^n x_k^2}\end{pmatrix} = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}.$$

b. (5) Let

$$\begin{pmatrix}\tilde{\alpha}\\ \tilde{\beta}\end{pmatrix} = \begin{pmatrix}\sum_{k=1}^n a_k Y_k\\ \sum_{k=1}^n b_k Y_k\end{pmatrix}$$

be a linear unbiased estimator of the regression parameters. Compute

$$\text{cov}(\tilde{\alpha} - \hat{\alpha}, \hat{\alpha}).$$

Solution: by assumption

$$E\left(\sum_{k=1}^n a_k E(Y_k)\right) = \sum_{k=1}^n a_k(\alpha + \beta x_k) = \alpha$$

which means $\sum_{k=1}^n a_k = 1$. *Compute*

$$\begin{aligned}
&\text{cov}(\tilde{\alpha} - \hat{\alpha}, \hat{\alpha})\\
&= \frac{1}{n}\text{cov}\left(\sum_{k=1}^n (a_k - \frac{1}{n})Y_k, \sum_{l=1}^n Y_l\right)\\
&= \frac{1}{n}\sum_{k=1}^n\left(a_k - \frac{1}{n}\right)\sigma^2(1+\tau^2)\\
&\qquad + \frac{1}{n}\sum_{k\neq l}\frac{1}{n}\left(a_k - \frac{1}{n}\right)\sigma^2\tau^2\\
&= 0\,.
\end{aligned}$$

c. (5) Find the best unbiased linear estimator of α.

Solution: compute

$$\mathrm{var}(\tilde{\alpha}) = \mathrm{var}(\tilde{\alpha} - \hat{\alpha} + \hat{\alpha}) + \mathrm{var}(\hat{\alpha}) + \mathrm{var}(\tilde{\alpha} - \hat{\alpha}) \geq \mathrm{var}(\hat{\alpha})\,.$$

The assertion follows.

d. (10) Find the best unbiased linear estimator of β.

Solution: by assumption

$$E\left(\tilde{\beta}\right) = \sum_{k=1}^{n} b_k(\alpha + \beta x_k)$$

which implies $\sum_{k=1}^{n} b_k = 0$ *and* $\sum_{k=1}^{n} b_k x_k = 1$. *Denote* $\sum_{k=1}^{n} x_k^2 = s$ *and compute*

$$\begin{aligned}
&\mathrm{cov}(\tilde{\beta} - \hat{\beta}, \hat{\beta}) \\
&\quad= \mathrm{cov}\left(\sum_{k=1}^{n}\left(b_k - \frac{x_k}{s}\right)Y_k, \sum_{l=1}^{n}\frac{x_l}{s}Y_l\right) \\
&\quad= \sum_{k=1}^{n}\left(b_k - \frac{x_k}{s}\right)\left(\frac{x_k}{s}\right)\sigma^2(1+\tau^2) \\
&\qquad\qquad \sum_{k\neq l}\left(b_k - \frac{x_k}{s}\right)\frac{x_l}{s}\sigma^2\tau^2 \\
&\quad= 0\,.
\end{aligned}$$

The assertion about the best unbiased linear estimator follows the same way as in c.