Eindhoven University of Technology
Faculty of Mathematics and Computer Science

Examination “Mathematical Statistics” (2S990)
January 21, 2005, 14:00–17:00.

You are allowed to use the “Statistisch Compendium” as well as a pocket calculator, but not the textbook. All questions have equal weight. All answers must be fully motivated.

==

1. The length of Dutch university students is assumed to be $N(\mu, 100)$-distributed, where μ itself is assumed to be $N(175, 60)$-distributed. A sample $L_1, \ldots, L_{12}$ of student lengths is drawn, with sample mean $\bar{L} = 178$. Determine a 95% confidence interval for μ using:

(a) the distribution of μ only;

(b) the outcome of $\bar{L}$ only;

(c) both.

For (c), use that μ given $\bar{L} = 178$ is $N(a, b^2)$-distributed with

$$a = \frac{\frac{100}{12} \times 175 + 60 \times 175}{\frac{100}{12} + 60}, \qquad \frac{1}{b^2} = \frac{1}{\frac{100}{12}} + \frac{1}{60}.$$

Comment on the comparison of the answers to (a–c).

2. Let $X_1, X_2, \ldots$ be an i.i.d. sample drawn from the uniform distribution on $(0, \theta)$ with $\theta > 0$ unknown. Let $M_n = \max\{X_1, \ldots, X_n\}$.

(a) Show that

$$\lim_{n \to \infty} P\left(\frac{n(\theta - M_n)}{\theta} \leq x \right) = \begin{cases} 1 - e^{-x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

(b) A sample with $n = 100$ yields $M_{100} = 2.17$. Use (a) to determine an approximate 95% confidence interval for θ.

3. Draw a sample Y_1, Y_2, Y_3 from a population that is Poisson distributed with unknown parameter $\lambda > 0$. Test H_0: $\lambda = 2$ against H_a: $\lambda = 1$.

(a) Show that a critical region based on the Neyman-Pearson lemma has the form

$$G(c) = \{(n_1, n_2, n_3) \in (\mathbb{N} \cup \{0\})^3 \colon n_1 + n_2 + n_3 \leq c\}, \quad c > 0.$$

(b) Determine the size α of $G(2)$.

(c) Determine the power $1 - \beta$ of $G(2)$.

(d) Is there a critical region of the same size but with more power? If yes, then give an example. If no, then explain why.

4. A factory claims that its cassettes have a total playing time that is $N(92, 1)$-distributed. A sample of 100 cassettes is drawn and their playing times are partitioned into four catagories:

range	$(0, 91]$	$(91, 92]$	$(92, 93]$	$(93, \infty)$
number	20	30	26	24

(a) Determine the expected frequencies in the four categories under H_0.

(b) Carry out a goodness of fit test at significance level $\alpha = 0.05$.

(c) Is it possible that the test leads to a different decision when a different partition is chosen? If so, then give an example by clumping two successive categories.

5. $Z_1, \ldots, Z_n$ is an i.i.d. sample drawn from a Gamma-distribution with parameters α and β. Is the sample mean $\bar{Z}$ a sufficient statistic:

(a) for β when α is known?

(b) for α when β is known?