Examination: **Statistics II** July 20, 2004

Examiner: Priv.-Doz. Dr. T. Müller-Gronbach

Please note the following:

- The exam consists of 5 problems for solution; the points you can get for each problem are given in brackets next to the number of the problem. You do not have to solve the individual problems completely, partial solutions are also possible. It is not enough, however, to state the result only, but you should clearly display your approach and way to solution.
- You can achieve 50 points. For passing the exam you have to achieve a total of (at least) **22 points** from all problems.
- You are allowed to use: Pocket calculators, text books, mathematical and/or statistical tables, manuscripts and notes from lectures and/or exercises.

Problem 1. *(5 points)*
The following statements are TRUE or FALSE. So just answer TRUE or FALSE without giving any explanation.

(a) If $\widehat{\theta}$ is an unbiased estimator of the one-dimensional paramter θ then $(\widehat{\theta})^2$ is an unbiased estimator of θ^2.
(b) Let φ_1 and φ_2 be tests for H_0 versus H_1 with respective significance levels $\alpha_1 < \alpha_2$. If H_0 is rejected by φ_1 then H_0 is also rejected by φ_2
(c) If the p-value is greater than the significance level α then the null hypothesis has to be rejected.
(d) If the coefficient of determination is 0 then the estimated slope of the fitted linear regression line is also zero.
(e) A fitted linear regression line always contains the point $(2\overline{x}, 2\overline{y})$, where $\overline{x}$ and $\overline{y}$ are the means of the influence factor values and the observed data, respectively.

Problem 2. *(9 points)*
A certain waiting time is modelled by a random variable X with density function

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} \cdot e^{-\frac{(x-1)}{\theta}} & \text{if } x \geq 1 \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is unknown. Note that $E(X) = \theta + 1$.

(a) Determine the maximum-likelihood estimator of θ based on a random sample $X_1, \ldots, X_n$ i.i.d. as X.

Assume $n \geq 2$ and consider the two estimators $\widehat{\theta}_0 = \overline{X} - 1$ and $\widehat{\theta}_1 = 2X_2 - X_1 - 1$ of the parameter θ.

(b) Show that both estimators are unbiased.
(c) Which of the two estimators is more efficient?

Problem 3. *(10 points)*
In a hospital the weights (in grams) of 526 newborns were recorded. The 270 male newborns showed the mean weight $\overline{x} = 3350$ and the standard deviation $s_x = 480$. For the 256 female newborns the mean weight was $\overline{y} = 3100$ and the standard deviation $s_y = 470$.
State a suitable model and solve the following problems:

(a) Construct a 95%-confidence interval for the difference in the mean weight of newborn males and females.
(b) Do the data provide significant evidence, at level 0.05, that the mean weight of a newborn depends on its sex? (Hint: Use the result from part (a)).
(c) Can one confirm the common believe that on the average the weight of a male newborn is greater than the weight of a female newborn? Use the level 1%.

Problem 4. *(10 points)*
A random number generator produced the five numbers

$$0.08,\ -0.91,\ 1.75,\ -0.41,\ -0.01,$$

which are supposed to be the outcome of a normal random sample $X_1, \ldots, X_5$ with mean $\mu = 0$ and variance $\sigma^2 = 1.2$
The arithmetic mean and the standard deviation of the five numbers are given by $\overline{x} = 0.1$ and $s_x \approx 1$, respectively.

(a) Check, under the assumption of normality, whether the data is consistent with the hypotheses of $\sigma^2 = 1.2$ at level 0.1.
(b) Check whether the data is consistent with the hypothesis of a normal distribution at level 10%.

Problem 5. *(16 points)*
The following table shows the rounded salaries (in 1000 €) of 5 graduates at the start of their professional career and ten years later.

graduate	1	2	3	4	5
salary at start	34	38	24	38	51
salary 10 years later	56	63	60	41	70

(a) Do the data provide sufficient evidence at level 0.1, that on the average the "ten years later" salary is at least 20000 € higher than the salary at the start of the career? State the model you are using in answering this question.

Use linear regression to model the dependence of the salary after 10 years on the starting salary.

(b) Compute the least squares estimates of the intercept a and the slope b of the regression line. Interpret the estimate of the slope.
(c) Assume normal errors. Do the data provide significant evidence, at level 0.01, that on the average a higher starting salary leads to a higher salary 10 years later?