Otto–von–Guericke–Universität Magdeburg **February 27, 2004**

Examination: **Statistics II (Bachelor in Economics & Management)**

Examiner: PD Dr. Th. 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 parentheses next to the number of the problem. You need not solve the individual problems completely, partial solutions are also accepted. For problems 2–5 it is not enough, however, to state the result only, but you should clearly display your approach and your way to the solution. Also, *conclusions* are to be drawn where applicable.
- For passing the exam you have to achieve a total of (at least) **20 points** out of all problems.
- You are allowed to use: Pocket calculators, text books, mathematical and/or statistical tables, manuscripts and notes from lectures and/or exercises.

Good luck!

Problem 1. (7 points)

In the following $X_1, \ldots, X_n \sim \mathrm{B}(1, p)$ is an i.i.d. sample of $n > 1$ Bernoulli variables with probability of success p, i.e. $P(X_i = 1) = p = 1 - P(X_i = 0)$. Are the following statements correct or wrong? (Just answer *Yes* or *No*.)

a) The larger the sample size n, the larger the sample total $\sum_{i=1}^{n} X_i$.

b) The larger the sample size n, the larger the variance of the sample mean, $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

c) For n large, $n \cdot \dfrac{(\overline{X} - p)^2}{p \cdot (1-p)}$ is approximately χ^2-distributed with 1 degree of freedom.

d) $1 - \overline{X}$ is an unbiased estimator of the probability of failure, $1 - p$.

e) The larger the parameter p, the smaller the probability $P\left(\sum_{i=1}^{n} X_i \geq 1\right)$.

f) If φ is a level-α-test (for a testing problem H_0 versus H_1), then the probability of a type-II-error is larger than α.

g) If a χ^2-test of independence does not reject the zero hypothesis $H_0 : X, Y$ *are independent*, then the random variables X and Y are indeed independent.

Problem 2. (9 points)

Let $X_1, \ldots, X_{n_x}$ and $Y_1, \ldots, Y_{n_y}$ be two independent i.i.d. samples of sizes n_x and n_y, respectively, from a normally distributed population with (unknown) mean μ and variance $\sigma^2 = 8$. Consider the two estimators of μ,

$$\widehat{\mu}_1 \ := \ \frac{1}{2}\overline{X} + \frac{1}{2}\overline{Y} \qquad \text{and} \qquad \widehat{\mu}_2 \ := \ \frac{1}{n_y}\left((n_x + n_y)\overline{X} - n_x\overline{Y}\right).$$

We are given the following relevant data from the samples:

	sample X	sample Y
sample size	$n_x = 10$	$n_y = 25$
sample mean	$\overline{x} = 1.1282$	$\overline{y} = -1.3440$

a) Show that $\widehat{\mu}_1$ and $\widehat{\mu}_2$ are unbiased for μ. Calculate the according estimates.

b) Calculate the variances of $\widehat{\mu}_1$ and $\widehat{\mu}_2$. Which of the estimators is more efficient?

Problem 3. (15 points)

A company producing washing machines is interested in comparing the water consumption of their own product with that of a similar washing machine (having the same capacity) produced by a competitor. The water consumptions of the own and the competitor's washing machine are assumed to be independent and normally distributed with unknown means μ_x, μ_y and variances σ_x^2, σ_y^2, respectively.

The following data have been observed from two samples of sizes 10, each:

i	x_i	x_i^2	y_i	y_i^2
1	33.6	1128.96	41.3	1705.69
2	40.9	1672.81	43.0	1849.00
3	37.8	1428.84	46.6	2171.56
4	37.7	1421.29	44.1	1944.81
5	40.5	1640.25	39.6	1568.16
6	39.0	1521.00	34.1	1162.81
7	42.9	1840.41	37.6	1413.76
8	39.7	1576.09	42.7	1823.29
9	43.0	1849.00	36.0	1296.00
10	38.6	1489.96	42.5	1806.25
$\sum$	393.7	15568.60	407.5	16741.30

a) Test, at a level of 10%, the hypothesis that the two population variances are equal.

In the following, assume that the population variances are equal, $\sigma_x^2 = \sigma_y^2$.

b) Calculate a 95%-confidence interval estimate of the difference of the two population means, $\mu_y - \mu_x$. Interpret this confidence interval.

c) At level $\alpha = 0.05$, is there enough evidence that the water consumption of the competitor's washing machine is larger than that of the company's own washing machine?

Problem 4. (12 points)

Does reading books on Statistics have a significant influence on the performance in the Statistics examination? A sample of $n = 75$ students yielded the following data:

	frequency	# of read books 0	1	2
grade	A or B	6	7	10
in the	C or D	8	9	8
exam	F	13	11	3

Test whether

a) *'grade'* is independent of *'number of read books'*, at a level of 10%;

b) the probability of *'passing the exam'*, i.e. grades A to D, is 0.6, at a level of 5%.

Problem 5. (7 points)

A liquor wholesaler is interested in assessing the effect of the price of a premium scotch whiskey on the quantity sold. The table below shows a sample of 8 weeks of sales records with the corresponding prices.

sales (in cases)	19.2	20.5	19.7	21.3	20.8	19.9	17.8	17.2
price (in \$)	25.4	14.7	18.6	12.4	11.1	15.7	29.2	35.2

The sample standard deviations are 1.43 and 8.70 for the sales and the prices, respectively. The empirical correlation coefficient is -0.97.

a) Determine the least squares estimates of the parameters for linear regression of sales on price.

b) Calculate a 95%-confidence interval for the slope of the regression line under suitable assumptions. State these assumptions.