Examination: **Statistical Analysis I** Wintersemester 2000/2001

Examiner: Professor Dr. B. Heiligers

Please note the following:

- The exam consists of 8 problems for solution; for each problem you can get at most 10 points. 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.
- For passing the exam you have to fulfill **all** of the following three requirements. You should achieve
  - a total of (at least) **25 points** from all problems, **among those**
  - (at least) **10 points from problems 1 – 4**, and
  - (at least) **10 points from problems 5 – 8.**
- 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 (10 pts)

The stem-leaf diagram gives the average daily costs (in USD) of hospital care last year in 50 hospitals.

23	05
24	22345778
25	159
26	57789
27	122456689
28	2567899
29	2358889
30	112567
31	
32	458

(stem units are 10 dollars, leaves units are dollars)

(a) Find the mode, the median, and the range of the data.
Up to what average daily costs operated (at least) 80 % of these hospitals?

(b) Draw a corresponding relative frequency histogram with four classes.

Problem 2 (10 pts)

An airline was concerned about the number of ticked passengers who fail to keep their reservations ("no–shows") on a particular flight. The number of no-shows from $n = 50$ flights were as follows:

number of no–shows	0	1	2	3	4	5	6
number of flights with that many no–shows	10	18	11	8	2	0	1

(a) Find the sample mean $\overline{x}$ and the sample standard deviation s of "no-shows" .

(b) The number of no-shows found in a second sample of 30 flights had mean 1.5 and variance 2.50 . Find the overall mean and overall variance of the combined sample .

Problem 3 (10 pts)

It was conjectured that income is one of the primary determinants of an individual's satisfaction with the job. To investigate this idea, 300 employees were asked to classify their job satisfaction, and their gross incomes.

		income (in 1000 USD)		
		< 20	$20 - 50$	> 50
job satisfaction	outstanding	24	39	21
	average	53	74	31
	poor	23	24	11

(a) Find all marginal and all conditional frequencies from the crosstab. Among the employees with an income less than 20000 USD, what is the percentage of those rating their job satisfaction as "poor" ?

(b) Calculate Cramer's V from the crosstab and interpret your result.

Problem 4 (10 pts)

A supermarket wants to study the effect on sales of the store brand grocery item by varying the price of the major competing brand. The table gives the sales y_i of the store brand item from a 7-week period, with varying prices x_i (in cent) of the competitor brand, (all other prices were held fixed); the other columns are given for convenience.

i	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
1	37	122	1369	14884	4514
2	32	107	1024	11449	3424
3	29	99	841	9801	2871
4	35	110	1225	12100	3850
5	33	113	1089	12769	3729
6	31	104	961	10816	3224
7	25	116	625	13456	2900
total	222	771	7134	85275	24512

(a) Find the correlation coefficient between "price of competitor" and "store brand sales" .

(b) Fit a regression line to the data, taking "price of competitor" as regressor, and "store brand sales" as response variable. Find the coefficient of determination, and comment upon the goodness of fit.

(c) Predict the store brand sales when the price of the competitor is 34 (cents) .

Problem 5 (10 pts)

A call center knows from history the following probabilities for the number X of incoming calls between 12 midnight and 6 a.m.

x	0	1	2	3	4	5	6
$f(x) = P(X = x)$	.10	.10	.15	.20	.25	.15	.05

(a) Find the cumulative probability function of X .

(b) Calculate the respective probabilities of the two events

A: "there will be 3 or more incoming calls",

B: "there will be not more than 5 incoming calls".

(c) Find the probabilities of the complement of A, and of intersection of A and B .

(d) Are A and B mutually exclusive? Are they exhaustive?

Problem 6 (10 pts)
A marketing firm found that 50 % of the customers who saw the advertisement for a product on TV will buy that product. 60 % of all customers saw the advertisement.

(a) What is the probability that a customer will buy the product?

(b) A customer did not buy the product. What is the probability that he did not see the advertisement?

(c) Are the two events "customer buys the product" and "customer saw the advertisement" independent?

Problem 7 (10 pts)
From experience, an investor knows the probabilities for the return X per invested USD:

x	0.8	0.9	1.0	1.1	1.2	1.3
$f(x) = P(X = x)$	0.05	0.15	0.25	0.30	0.20	0.05

(a) Find the expectation μ and the standard deviation σ of X.

(b) What is the probability that X falls into the interval $[\mu - 1.5\,\sigma\,,\ \mu + 1.5\,\sigma\,]$? What is the lower bound for this probability from Chebyshev's Theorem?

(c) 10000 USD are invested. What are the expected value and the standard deviation of the total return from this investment?

Problem 8 (10 pts)
For reducing drunken driving on New Year's Eve, police intended to stop 20 cars passing a certain control point, (the same car could be stopped more than once). If 10 % of all drivers have more than the legal limit to drink,

(a) what is the probability of catching at least 3 drunken drivers?

(b) what are expectation and variance of the number of drunken drivers stopped by the police?