Exam Probability and Stochastics, August 25, 2009 (12.00-14.45)

It is allowed to use a *basic* calculator, *not* a graphical one. All exercises have equal weight.

1. (a) Give the Five-Number Summary of the following dataset:

4.2 5.1 5.8 6.0 6.6 7.1 7.8 7.8 8.3 9.9 47.4

(b) Draw the corresponding boxplot.
(c) Compute the mean and the median.
(d) Compute the mean and the median discarding the highest value of 47.4.
(e) Can you draw conclusions about the sensitivity of the mean and median?

2. (a) Explain in your own words the so called "68-95-99.7 rule" for a normal distribution.
(b) The heights of American men aged 18-24 are normally distributed with mean 75 inches, and standard deviation 3.1 inch. Apply the "68-95-99.7 rule" to this situation.
(c) What percentage of men have height between 70 and 75 inch?

3. (a) Formulate the Central Limit Theorem.
(b) The mean salary of 16,000 employees is 26.400 euro with a standard deviation of 2.420. Someone takes a random sample of 1000 employees, and finds that the sample mean is 26.500 euro. Does the discrepancy between the sample mean and the population mean suggest that the results are suspect?
(c) The same question as in (b) but now assume that the sample has size 10 (and the same sample mean).

4. Suppose we have an ordinary deck of 52 cards. We shuffle and take the cards one by one from the deck.
(a) What is the probability that we see a spade before a 6
(b) What is the probability Queen hearts is the first hearts to be seen?
(c) What is the probability that we start by seeing a 4, a 5, a Queen and a 6 (in that order)?
(d) What is the probability that the first four cards are a 4, a 5, a 6 and a Queen (the order is not important now)?
(e) If the first card is a 3, what is the conditional probability that the one but last card is a 5?

5. Comment on the following statement: It is found that there is a correlation between the value of a car and its weight. Using the equation of the best-fit line, it is concluded that a car with weight at least 10,000 kilo is worth a fortune.

6. Consider a sample with 2,012 respondents, 49% of which answered the question asked with"yes".
(a) Construct a 95% confidence interval for the population proportion of people that would answer the question by "yes".
(b) How large should the sample be in order to obtain (roughly) a confidence interval of length 2 percentage points?

7. Suppose the null hypothesis has the form $H_0 : \mu = \mu_0$, where μ is some population mean. Comment on the following statements:
(a) If we reject H_0 with a right-tailed test, then we also reject H_0 with a two-sided test with twice the significance level.
(b) If we reject H_0 with a right-tailed test with significance level 10%, then we also reject H_0 with a two-sided test with significance level 5%.
(c) If we reject H_0 with significance level 5% with a right-tailed test, then we also reject H_0 with a right-tailed test and significance level 10%.
(d) Not rejecting H_0 is the same as accepting H_0.

8. The mean household income in a poll of 10,267 households is 41,045 euro, with a standard deviation of 1,605 euro. Test the claim that the mean household income in the population differs from 40,100. Take significance level 5%.

9. 15.1% out of a random sample of 4,342 people live in poverty. Test the claim that the poverty in the full population is higher than 13.3%.

Table 5.1 Standard Scores and Percentiles for a Normal Distribution (cumulative values from the *left*)

Standard score	%	Standard score	%	Standard score	%	Standard score	%
−3.5	0.02	−1.0	15.87	0.0	50.00	1.1	86.43
−3.0	0.13	−0.95	17.11	0.05	51.99	1.2	88.49
−2.9	0.19	−0.90	18.41	0.10	53.98	1.3	90.32
−2.8	0.26	−0.85	19.77	0.15	55.96	1.4	91.92
−2.7	0.35	−0.80	21.19	0.20	57.93	1.5	93.32
−2.6	0.47	−0.75	22.66	0.25	59.87	1.6	94.52
−2.5	0.62	−0.70	24.20	0.30	61.79	1.7	95.54
−2.4	0.82	−0.65	25.78	0.35	63.68	1.8	96.41
−2.3	1.07	−0.60	27.43	0.40	65.54	1.9	97.13
−2.2	1.39	−0.55	29.12	0.45	67.36	2.0	97.72
−2.1	1.79	−0.50	30.85	0.50	69.15	2.1	98.21
−2.0	2.28	−0.45	32.64	0.55	70.88	2.2	98.61
−1.9	2.87	−0.40	34.46	0.60	72.57	2.3	98.93
−1.8	3.59	−0.35	36.32	0.65	74.22	2.4	99.18
−1.7	4.46	−0.30	38.21	0.70	75.80	2.5	99.38
−1.6	5.48	−0.25	40.13	0.75	77.34	2.6	99.53
−1.5	6.68	−0.20	42.07	0.80	78.81	2.7	99.65
−1.4	8.08	−0.15	44.04	0.85	80.23	2.8	99.74
−1.3	9.68	−0.10	46.02	0.90	81.59	2.9	99.81
−1.2	11.51	−0.05	48.01	0.95	82.89	3.0	99.87
−1.1	13.57	0.0	50.00	1.0	84.13	3.5	99.98

NOTE: The % column gives the percentage of values in the distribution less than the corresponding standard score.