Exam Probability and Stochastics, August 20, 2008 (12.00-14.45)

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

1. Suppose that after seven exams, the median of your scores is 8 (an exam score can be any number between 0 and 10 inclusive).
(a) What is the maximal mean score you can have?
(b) What is the minimal mean score you can have?

Suppose now that I tell you that the median of my exam score so far is 9, but I do not tell you how many exams I have had.
(c) If I score a 10 on my next exam, can you compute my new median? Explain your answer.

2. Consider the following set of values: 3, 6, 7, 7, 9, 9, 10, 100.
(a) Compute the standard deviation of the set.
(b) Do the same, but now without the last value (100).
(c) Is the standard deviation affected by outliers? Motivate your answer.
(d) Give the five-number summary of the original set and draw a boxplot.

3. Certain scores are normally distributed with a mean of 120 and a standard deviation of 20. Suppose that many samples of size n are taken from a large population of people, and the mean score is computed for each sample. Describe the mean and standard deviation of the resulting distribution of sample means for $n = 10$ and $n = 500$. Comment on the difference.

4. Blood pressures (lower value) of men 18 to 24 years of age are normally distributed with a mean of 80 and a standard deviation of 10.
(a) What is the standard score of a 20-year old man with a blood pressure of 85?
(b) What percentage of men 18 to 24 years of age have a higher blood pressure?

5. An MP3 player is loaded with 50 musical selctions: 10 rock selections, 15 jazz selections and 25 classical music selections. The player is set on 'random play', so selections are played randomly and can be repeated. What is the probability of each of the following events:
(a) The second to fifth selections are all jazz.
(b) Among the first three selections, exactly one is rock.
(c) The fourth selection is a different genre as the second selection.
(d) There is no classical music among the first ten selections.

6. We want to estimate a population mean. The sample size is $n = 200$,

the sample mean is 80, and the sample standard deviation is 5. Give a 95% confidence interval for the population mean. Explain your method.

7. Give an estimate of the sample size needed to obtain a margin of error of 1 in the 95% confidence interval, if the standard deviation is equal to 15 and we are estimating a population mean.

8. The Bureau of Labour Statistics estimates unemployment rates in the US monthly by surveying 10,000 individuals.
(a) In one month, 2% of the 10,000 individuals are found to be unemployed. Find the margin of error for the 95% confidence interval.
(b) If the number of individuals surveyed would be increased to 40,000, by how much would the margin of error change? Give the answer without doing any computation.

9. Consider a test for a population mean μ, where the null-hypothesis is $H_0 : \mu = 30$, and the alternative hypothesis is $H_a : \mu \neq 30$. The sample gives $\bar{x} = 26$ and the sample standard deviation s is equal to $s = 9$. The sample size is $n = 100$.
(a) Is H_0 rejected when we take significance level 0.05?
(b) Does this change if instead we consider $H_a : \mu < 30$? Why?

10. Determine about each statement below whether it is sensible or not. Motivate your answers.
(a) If the mean of a dataset is 200, and the median is 100, then there must be a small outlier.
(b) In a right-tail hypothesis test, the null-hypothesis $H_0 : \mu = \mu_0$ is never rejected at the 0.05 level when the sample mean is above μ_0.
(c) When you reject a hypothesis at the 0.05 level, you also reject it at the 0.01 level.
(d) There is a positive correlation between exercise and health. Hence, when you do exercises, you can be sure that your health will improve.