Tentamentsskrivning i **Mathematisk statistik TMA321, 3p.**

Tid: Onsdagen den 17 januari, 2007 kl 14.00-18.00
Examinator och jour: Serik Sagitov, tel. 772-5351, mob. 0736 907 613, MV-huset rum H3026.
Hjälpmedel: valfri räknare, egen formelsamling (4 sidor på 2 blad A4) samt utdelade tabeller.

There are five questions with the total number of marks 30. Attempt as many questions, or parts of the questions, as you can. Preliminary grading system (no bonus points this time):
grade "3" for 12 to 17 marks,
grade "4" for 18 to 23 marks,
grade "5" for 24 and more marks.

---

1. (6 marks) The Myers-Briggs personality system distinguishes among four types of temperament: traditionalist, action-oriented, knowledge-searcher, identity-searcher. A (fictitious) survey based on a Myers-Briggs personality test was aimed to compare homosexual men to women. It gave the following proportions for the four temperaments 40-20-20-20 (homosexual men) and 40-15-15-30 (women).

a. Are the observed differences between homosexual men and women statistically significant? Why?

b. Imagine an experiment where we choose at random a person from a population composed of two unequal strata: homosexual men (40%) and women (60%). For this person to be an identity-searcher is a random event. Being a homosexual man or a woman is another random event. Are these two events independent? Explain.

c. Suppose that a randomly chosen person in an experiment described above happens to be a knowledge-searcher. What is the probability that this person is a woman?

2. (6 marks) An investigation is conducted to study gasoline mileage in automobiles when used exclusively for urban driving. Ten properly tuned and serviced automobiles manufactured during the same year are used in the study. Each automobile is driven for 1000 miles, and the average number of miles per gallon obtained (Y) and the weight of the car in tons (X) are recorded.

Figure 1 is produced by the Matlab command *polytool*. It gives the data scatterplot together with the fitted regression line $y = 23.75 - 4.03x$. The dashed lines give a 95% confidence band for the regression line. According to this figure, for $X = 1.675$ the confidence interval for the mean value of Y is 17 ± 0.258.

a. For $X = 1.8$ the confidence interval for the mean value of Y has the half-width

$$2.306 \cdot 0.355 \cdot \sqrt{\frac{1}{10} + \frac{(1.8 - 1.675)^2}{0.581}} = 0.292.$$

Here 0.355 is the estimated standard deviation in Y due the external factors and 0.581 describes the observed variation in X values. Where the number 2.306 comes from? Write down the confidence interval in question.

b. Four out of ten observations fall off the 95% confidence band. Does it surprise you? Explain.

c. The standard error of the slope value (-4.03) is estimated as $\frac{0.355}{\sqrt{0.581}} = 0.466$. Find a 99% confidence interval for the slope parameter. What crucial assumption stands behind the formula you apply?

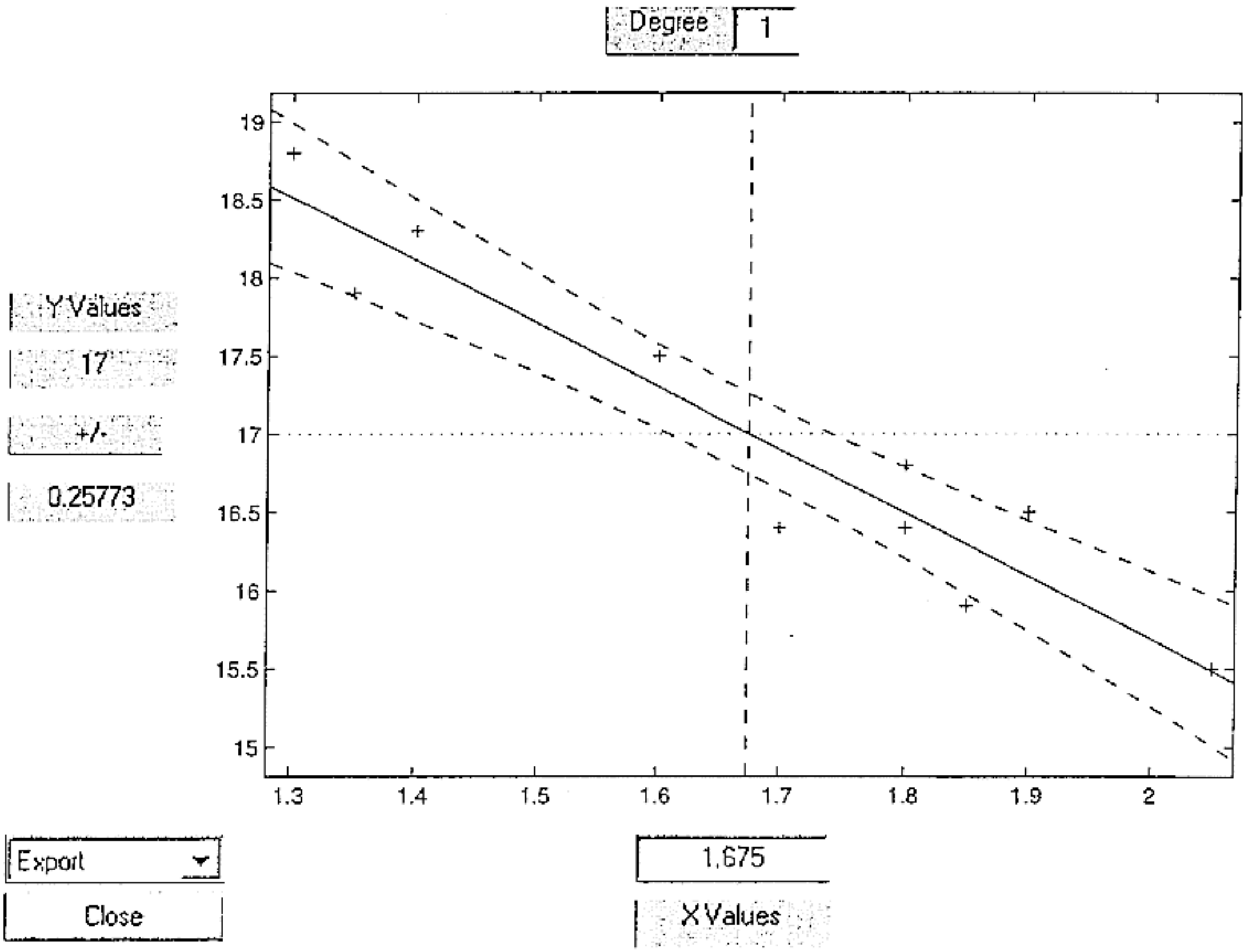

Figure 1: *The average number of miles per gallon Y against the weight of the car in tons X.*

3. (6 marks) Opponents of the construction of a dam on the New River claim

that less than half of the residents living along the river are in favor of its construction. A survey is conducted to gain support for this point of view.

a. Set up the appropriate null and alternative hypotheses.

b. Find the rejection region for the null hypothesis on 10% significance level.

c. Of 500 people surveyed, 230 favor the construction. Is this sufficient evidence to justify the claim of the opponents of the dam?

d. To what type of error are you now subject? Discuss the practical consequences of making such an error.

4. (6 marks) Computer terminals have a battery pack that maintains the configuration of the terminal. These packs must be replaced occasionally. Let X denote the life span in years of a such a battery. Assume that X has probability density function $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. The following data on the X values has been collected

1.7	4.0	1.9	2.0	1.7
2.1	2.7	4.2	1.8	2.2
3.1	1.5	2.4	6.2	7.0
3.6	1.4	5.0	3.8	1.6

a. Plot the data to see if it really fits the exponential model $f(x) = \lambda e^{-\lambda x}$.

b. Find a method of moment estimate for the parameter λ.

c. Compute an approximate 95% CI for λ. Hint: first find a CI for the mean $\mu = E(X)$.

5. (6 marks) Suppose you can repeatedly generate a random number X uniformly distributed between 0 and 1. It is easy to use X for simulating a fair coin tossing: if $X \leq 0.5$, we decide it is Heads; and if $X > 0.5$, then it is Tails.

a. How would you simulate the number of Heads Y in ten coin tossings with a single random number X?

b. Verify that $P(\ln X^{-1} \leq x) = 1 - e^{-x}$ and explain how this can be used to simulate an exponentially distributed random variable with mean 2.

c. Sketch the density curve of the random variable X^2.

Statistical tables supplied:
1. Normal distribution table
2. t-distribution table