STAT 135, Spring 2008

Final

There are 6 problems in the exam. The order in which they are given **does not** necessarily indicate their level of difficulty. It might be a good idea to look at all the problems before starting to work on any one of them so you can gauge which one you prefer. Within a problem, even if you get stuck on one question, you should be able to answer some of the questions which follow. Unless otherwise noted, each question is worth 3 points.
Partial credit will be given. Short answers are ok; they just need to be to the point. Show your work and justify your answers. Good luck!

Name:

Student ID:

Score:

Problem 1

1. What is a confidence interval? How can you use the bootstrap (what is it and what does it mean?) to create them? What are the advantages and disadvantages (if any) of the bootstrap in this context?

2. When is the Neyman-Pearson lemma used and what does it say? How would you use it to design a statistical procedure? What good properties would your procedure have if you used the Neyman-Pearson lemma to design it?

3. Describe in as much detail as you can the Mann-Whitney test. When would you use it? What is its advantage over the corresponding parametric method? What is its disadvantage?

4. What is a qq-plot? Describe two types of problems/settings where you would use them.

5. What is the regression effect?

6. True or false and explain: $\mathbf{E}(X) = \bar{X}$

Problem 2

Suppose we observe a random sample $X_1, \ldots, X_n$ from the distribution with density

$$f(x;\theta) = \exp(-(x-\theta)) \, , \theta \le x < \infty \, .$$

1. Show that $f(x;\theta)$ is a probability density

2. (4 points) What is the maximum likelihood estimator of θ?

3. (4 points) Call $Y_1 = \min(X_1, \ldots, X_n)$. Show that the density of Y_1 is $g(y) = n\exp(-n(y-\theta))$, for $\theta \le y < \infty$.

4. Find a $100(1-\alpha)\%$ confidence interval for θ based on Y_1

5. Find a method of moment estimator $\widehat{\theta}_{MM}$ for θ.

6. What is the asymptotic distribution of $\widehat{\theta}_{MM}$ as $n \to \infty$?

7. What is the efficiency of $\widehat{\theta}_{MM}$ relative to Y_1?

8. Suppose you are given $n = 100$ observations drawn for $f(x;\theta)$. Which estimator of θ would you choose? Y_1 or $\widehat{\theta}_{MM}$?

Problem 3

Suppose you observe a sequence of coin flips and count the number of times it takes before you see a head. Suppose the probability of heads is θ, with $0 < \theta < 1$. Then it is known that X, the number of tails you'll see before you see a head has a geometric distribution:

$$P(X = k) = (1-\theta)^k \theta \, , k = 0, 1, \ldots \, .$$

1. Suppose θ has a prior distribution that is uniform on the three values $\{1/4, 1/2, 3/4\}$. What is the posterior distribution of θ given $X = 2$?

2. (5 points) Suppose now that you observe $X = k$. What would be your estimate of θ if you used the prior given above? (Partial credit will be given if you answer just for $X = 2$.)

3. (4 points) Find the posterior distribution of θ given that $X = k$ when the prior is a beta distribution $\beta(r, s)$

Problem 4

A driver decides to test the quality of three types of fuel sold in her area based on mpg (i.e miles per gallon). The data she collects is the following:

Type 1 :	38.7	39.2	40.1	38.9
Type 2 :	41.9	42.3	41.3	41.5
Type 3 :	40.8	41.2	39.5	38.9

1. (12 points) Analyze the data and explain in detail your strategy, reasoning, and assumptions you make.
2. (3 points) What conclusions can you draw about the quality of the different types of fuel?

Note: the means of each type are $(39.225, 41.75, 40.1)$. The standard estimates of variance for each types are $(0.38, 0.20, 1.17)$.

Problem 5

Suppose we observe data $X_1, \ldots, X_n$, generated through

$$X_i = \theta X_{i-1} + \epsilon_i, \ \ i = 1, \ldots, n \ .$$

We assume that $X_0 = 0$, $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and ϵ_i's are independent.

1. What is the distribution of X_1?
2. Assuming that $X_{i-1} \sim \mathcal{N}(\mu_{i-1}, \sigma_{i-1}^2)$, what is the distribution of X_i? *(You can use μ_{i-1} and σ_{i-1} in your answer and not simplify any further.)*
3. What is the joint density of the ϵ_i's?
4. Suppose for this question that $\theta = 0$. Compute $\text{cov}(X_i, X_j)$ in that case for $j \neq i$ and $j = i$. Are X_i and X_j independent if $i \neq j$?
   In what follows you can assume that the joint density of the X_i's is given by
   $$p(x_1, \ldots, x_n, \theta) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta x_{i-1})^2)$$
5. (5 points)What is the maximum likelihood estimate of θ?
6. (5 points)We want to test $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$. Show that the generalized likelihood ratio test rejects for large values of
   $$T = \frac{(\sum_{i=1}^{n} X_i X_{i-1})^2}{\sum_{i=1}^{n} X_{i-1}^2} \ .$$
   (*If you have not found the MLE in the previous question and you think you need it to answer this one, it's ok to derive your answer to this question with a parameter $\widehat{\theta}_{MLE}$ and give your answer as a function of this parameter*)
7. (4 points)The statistic T found in the previous question does not have a known distribution. Explain how you could nevertheless find a cut-off point to design a level α test by using questions 1) to 6) and a computer. Be as precise as you can.
   (Explain in words what you would do; no R code is necessary)

Problem 6(12 points)

The last page of the exam contains four scatter plots of data to which a linear regression line was fitted via least squares. For each of these plots, explain if you think it is appropriate or not to fit a (least-squares) linear regression to the corresponding data. Explain briefly why each time. (Each explanation is worth 3 points)

(Of course, you can tear the last page apart if that makes things easier for you. Please refer to the plots by the same numbering scheme I used)

Reminders

The density of the $\mathcal{N}(\mu, \sigma^2)$ is

$$\frac{1}{\sqrt{2\pi}\sigma} \exp(-(x-\mu)^2/(2\sigma^2)) \; .$$

If Z has the corresponding distribution, $\mathrm{E}(Z) = \mu$ and $\mathrm{var}(Z) = \sigma^2$, $\mathrm{E}((Z-\mu)^3) = 0$, and $\mathrm{E}((Z-\mu)^4) = 3\sigma^4$. In the case where $\mu = 0$ and $\sigma = 1$, $P(Z \leq 2) \simeq 2.5\%$ and $P(Z \leq 1.68) \simeq 5\%$.

The beta distribution with parameters a and b has density

$$h(\theta) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} & \text{for } 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Its mean and variance are

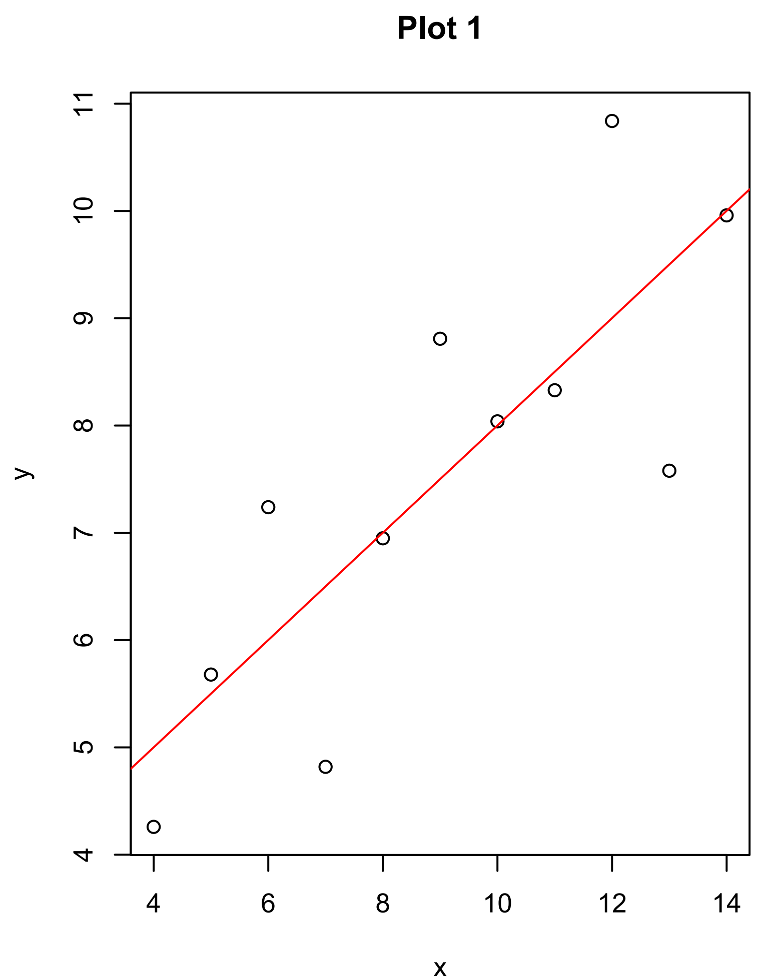
$$\mu = \frac{a}{a+b} \; , \qquad \sigma^2 = \frac{ab}{(a+b)^2(a+b+1)} \; .$$

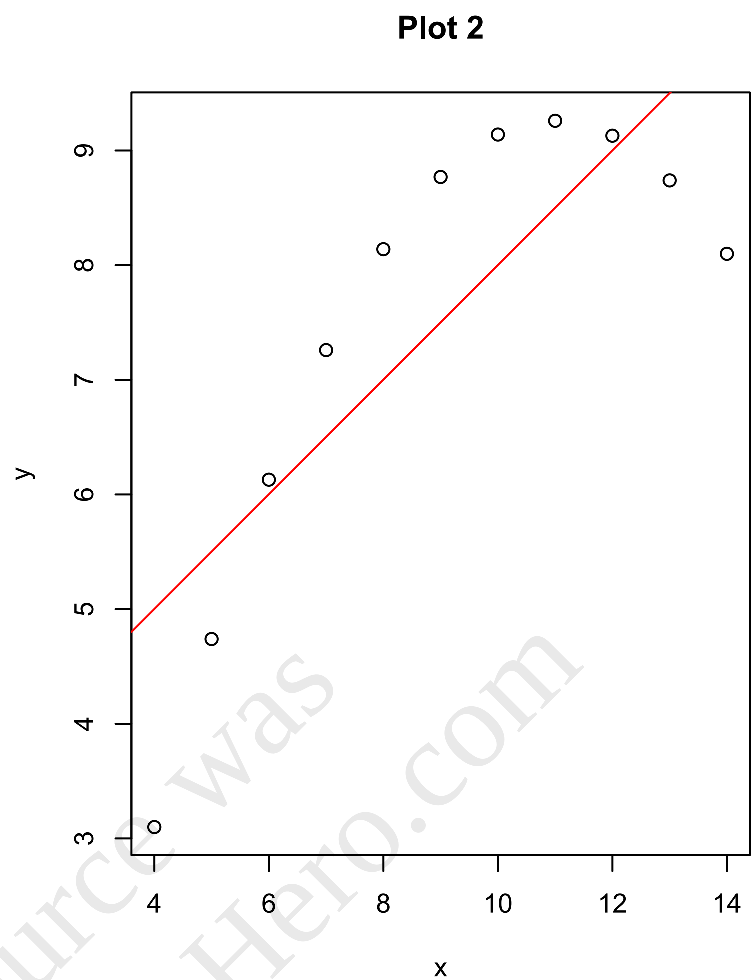
The Gamma function, Γ is defined, for $t \in \mathbb{R}$ as

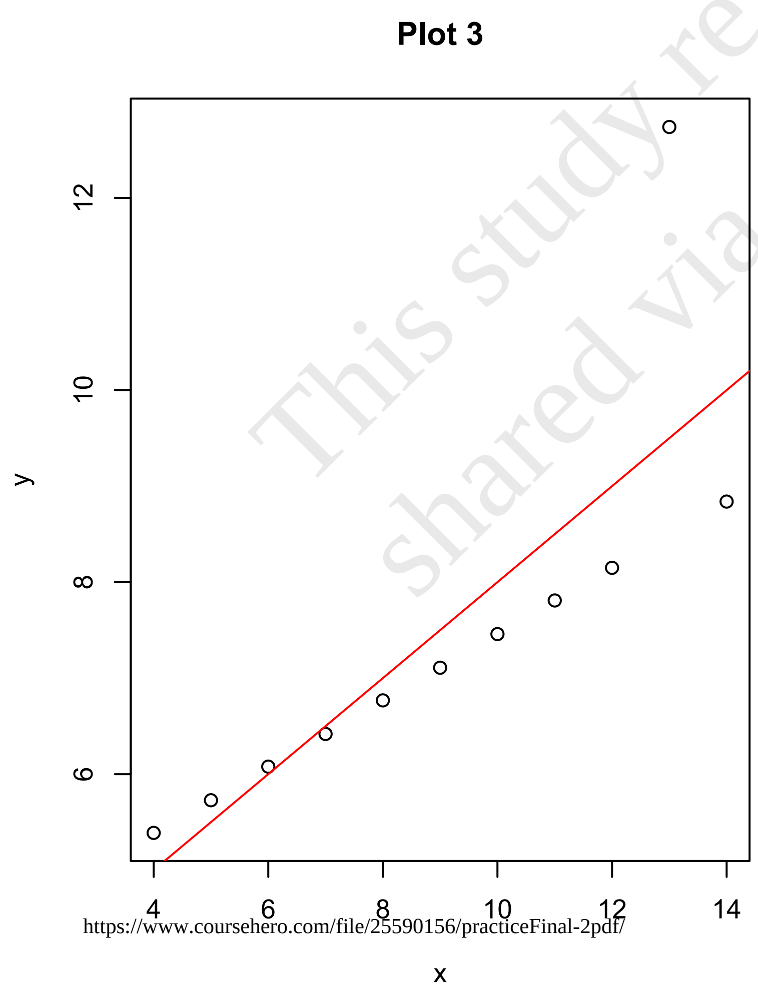
$$\Gamma(t) = \int_0^\infty y^{t-1} \exp(-y) dy \; .$$

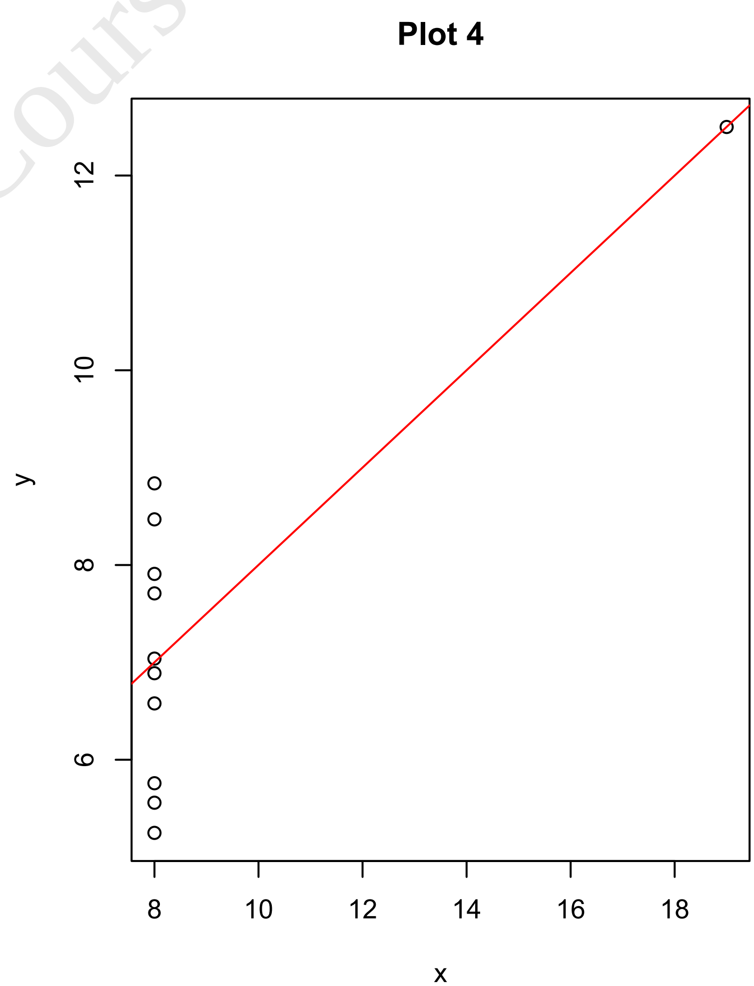
If t is integer,

$$\Gamma(t) = (t-1)!$$


Plot 1
y
x
4
5
6
7
8
9
10
11
4
6
8
10
12
14


Plot 2
y
x
3
4
5
6
7
8
9
4
6
8
10
12
14


Plot 3
y
x
6
8
10
12
4
6
8
10
12
14


Plot 4
y
x
6
8
10
12
8
10
12
14
16
18