Name: ____________________________

Directions: **Work only on this sheet** (on both sides, if needed); do not turn in any supplementary sheets of paper. There is actually plenty of room for your answers, as long as you organize yourself BEFORE starting writing.

SHOW YOUR WORK!

0. Advice: The first three problems have half-line answers and should be very quick. Most of the other problems also have very short answers, but may or may not be quick.

1. (15) Fill in the blank with a term from our course: In comparing two estimators of some population quantity, we might consider the better one to be the one with smaller ___________________________.

2. (15) Consider the R code on p.262, which consists of assignments to **md** and **lmout**. Suppose we wish to fit a model with no first-degree term, i.e. (9.12) would change to

$$m_{A;b}(b) = \beta_0 + \beta_1 b^2 \tag{1}$$

How should we change the code in the line on p.262 that assigns to **lmout**? Assume that the line assigning to **md** is retained.

3. (10) Consider the R code at the top of p.200. Give an approximate 95% confidence interval for the population value λ. Your answer will be in the form $(c \pm d$, where c and d are numerical expressions, e.g. $2.37^8/5$.

4. Suppose the random variable X is equal to 1, 2 and 3, with probabilities c, c and 1-2c. The value c is thus a population parameter. We have a random sample $X_1, ..., X_n$ from this population.

(a) (15) Show that the Method of Moments Estimator of c, which we will denote by $\hat{c}$, is $(3 - \overline{X})/3$.

(b) (15) Find the bias of $\hat{c}$. Cite mailing tubes and other reasons!

5. In the notation of Chapter 9, give matrix and/or vector expressions for each of the following in the linear regression model:

(a) (10) s^2, our estimator of σ^2

(b) (10) the standard error of the estimated value of the regression function $m_{Y;X}(t)$ at $t = c$, where $c = (c_0, c_1, ..., c_r)$.

6. (10) Suppose Jack and Jill each collect random samples of size n from a population having unknown mean μ but KNOWN variance σ^2. They each form an approximate 95% confidence interval for μ, using (6.21) but with s replaced by σ. Find the approximate probability that their intervals do not overlap. Express your answer in terms of Φ, the cdf of the N(0,1) distribution.

Solutions:

1. Mean squared error. Partial credit was given for some other answers, but it was emphasized that MSE is the major criterion, as it balances variance and bias.

2.

```
lm(md[,2] ~ md[,3])
```

3. $1.027 \pm 1.96 \cdot 0.117$

4a. $EX = c \cdot 1 + c \cdot 2 + (1 - 2c) \cdot 3 = 3 - 3c)$, so c = (3-EX)/3. So, set $\hat{c} = (3 - \overline{X})/3$.

4b.

$$
\begin{aligned}
E\hat{c} &= E[(3 - \overline{X})/3] \quad \text{(from (a))} && (2)\\
&= \frac{1}{3} \cdot (3 - E\overline{X}) \quad (3.13) && (3)\\
&= \frac{1}{3}[3 - EX] \quad (6.8) && (4)\\
&= \frac{1}{3}[3 - (3 - 3c)] && (5)\\
&= c && (6)
\end{aligned}
$$

So, the bias is 0.

5a.

$$\frac{1}{n}(V - Q\hat{\beta})'(V - Q\hat{\beta}) \tag{7}$$

Note that hats!

5b.

$$m_{Y;X}(c) = c'\beta \tag{8}$$

so

$$\widehat{m}_{Y;X}(c) = c'\hat{\beta} \tag{9}$$

Then (5.88) yields

$$Var[\widehat{m}_{Y;X}(c)] = c'Cov(\hat{\beta})c = \sigma^2 c'(Q'Q)^{-1}c \tag{10}$$

Thus the sample estimated variance is

$$s^2 c'(Q'Q)^{-1}c \tag{11}$$

so that the standard error is

$$s\sqrt{c'(Q'Q)^{-1}c} \tag{12}$$

6. The probability of nonoverlap is double the probability that Jack's interval is entirely to the left of Jill's. Since each interval has radius $1.96\sigma/\sqrt{n}$, Jack's interval will be entirely to the left of Jill's if

$$\overline{X} + 1.96\sigma/\sqrt{n} < \overline{Y} - 1.96\sigma/\sqrt{n} \tag{13}$$

we have

$$P(\text{nonoverlap}) = 2P(\overline{X} - \overline{Y} < -1.96\sigma/\sqrt{n}) \tag{14}$$

So, we need the distribution of $W = \overline{X} - \overline{Y}$. But $EW = \mu - \mu = 0$ and by independence

$$Var(W) = Var(\overline{X} - \overline{Y}) = 2\sigma^2/n \tag{15}$$

so W has standard deviation $\sigma\sqrt{2/n}$. Thus

$$P(\text{nonoverlap}) = 2\Phi\left(\frac{-1.96\sigma/\sqrt{n}}{\sigma\sqrt{2/n}}\right) \approx 2\Phi(-1.39) \tag{16}$$

You might be surprised to see that the answer is independent of n. The actual value is about 0.16. So, Jack and Jill have about a 16% chance of having nonoverlapping intervals.