STA 242/ENV 255 Midterm **Name:**_____

February 22, 2001

1. This test is CLOSED BOOK; however, you may use one 8.5" × 11" sheet (front and back) of notes.
2. With this exam, we will provide tables of t, F.
3. You may use a scientific calculator for this exam.
4. **To receive partial credit, you must show all steps.**
5. You have 1 hour and 15 minutes to complete the exam; you must finish by the end of the period (1:55 pm).

Duke University

Nicholas School of the Environment

Honor Code

I HAVE NOT GIVEN OR RECEIVED, NOR WILL I GIVE, INFORMATION BEFORE, DURING OR AFTER THIS EXAM RELATING TO ITS CONTENTS OR COMPLETION THAT WOULD CONFER AN ADVANTAGE OVER STUDENTS NOT PRIVY TO THIS INFORMATION, UNLESS SPECIFICALLY TOLD BY THE INSTRUCTOR THAT SUCH CONDUCT IS PERMISSIBLE. I HAVE NOT USED DESIGNATED STUDY MATERIALS FOR THIS COURSE IN A MANNER THAT WOULD PREVENT THEM FROM BEING EQUALLY AVAILABLE TO OTHER STUDENTS. I WILL REPORT ANY VIOLATIONS OF THIS CODE TO THE INSTRUCTOR IN CHARGE OF THE COURSE AND TO THE DEAN.

Signature: _____ Date:_____

Pages	2	3	4	5	8	9	Total
Possible Points	25	30	30	5	30	20	140
Points Earned	____	____	____	____	____	____	____

Questions (1) through (11) apply to the gamma radiation data below.

- The article "Effects of Gamma Radiation on Juvenile and Mature Cuttings of Quaking Aspen" (*Forest Science* (1967): 240-245) reported data on the effect of exposure to radiation (kR per 16 hours) on the dry weight of roots (milligrams). (A kiloroentgen, abbreviated kR, is a unit of gamma radiation.)
- The dataset analyzed here has 20 observations; 4 measured dry weights at each of 5 levels of radiation exposure.
- Summary statistics for the variables *radiation* and *weight*:

Variable	25%	Median	Mean	75%	90%	s
radiation (kR per 16 hours)	2	4	4	6	8	2.9
weight (mg)	82.5	100.6	98.7	120.2	123.3	22.7

- Partial Splus output and a plot are given below for the simple linear regression model: *weight* $\sim$ *radiation*. In the Splus output, dashes indicate omitted information.

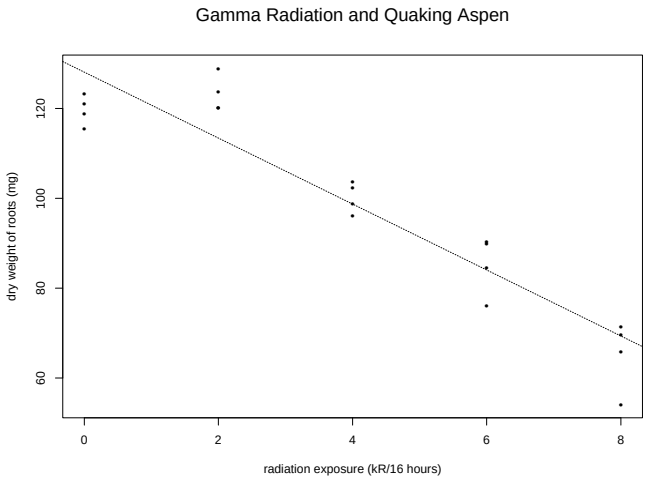
```
Call: lm(formula = weight ~ radiation)
Residuals:
    Min     1Q Median    3Q  Max
 -15.21 -5.437 0.3482 5.931 15.5

Coefficients:
                Value Std. Error   t value  Pr(>|t|)
(Intercept)       --     3.1466    40.7031     --
  radiation   -7.3387     0.6423      --        --

Residual standard error: 8.125 on -- degrees of freedom
Multiple R-Squared: 0.8788
F-statistic: 130.5 on -- and -- degrees of freedom, the p-value is 1.107e-09

Correlation of Coefficients:
          (Intercept)
radiation -0.8165

          Df Sum of Sq  Mean Sq  F Value  Pr(F)
radiation  1  8617.000  8617.00   --       --
Residuals 18  1188.138    66.01
```



1. Write out in precise statistical notation the fitted model for the simple linear regression. [5 points]

2. Explain in a sentence what the slope term means as applied to this problem. [5 points]

3. Give the value of the linear correlation coefficient between *radiation* and *height*. [5 points]

4. **True or False. Circle one.** A correlation coefficient near -1 always implies a large negative slope. [5 points]

5. What percent of variation remains unexplained by the regression of *weight* on *radiation*? [5 points]

6. At what value of *radiation* will we get our most precise prediction of *weight*? (*Hint: You do not need to do any calculations here.*)[5 points]

7. Determine whether there is evidence of a linear association between radiation exposure and dry weight. Give hypotheses, test statistic, a p-value for your test and a conclusion. [15 points]

8. Determine whether there is evidence of a negative association between radiation exposure and dry weight. Give hypotheses, test statistic, a p-value for your test and a conclusion. [10 points]

9. The partial Splus output for the model states:

```
F-statistic: 130.5 on -- and -- degrees of freedom, the p-value is 1.107e-09
```

   (a) [5 points] Give the null and alternative hypotheses associated with this F-statistic in formal statistical terms.

   (b) What is the rejection region for a test of the hypotheses in part (9a) at $\alpha = 0.05$? [5 points]

10. We wish to perform an F-test for lack of fit to the simple linear regression model.

   (a) Why can we perform a lack of fit test in this example? [5 points]

   (b) What are the hypotheses being tested? Give the hypotheses in formal statistical notation. [5 points]

   (c) Give the degrees of freedom associated with the Sum of Squares for Lack of Fit in this case. [5 points]

   (d) Imagine that the null hypothesis in (10b) is rejected; give an example of a model you might investigate next. [5 points]

11. Imagine that we need to rescale our radiation exposure measurements to units of *kR per hour* (rather than *kR per 16 hours*) and then refit the simple linear regression model. Which of the following would change? **Circle all that apply.** [5 points]

    (a) the estimated coefficient for radiation in the regression, $\hat{\beta}_{radiation}$

    (b) the overall conclusion of a hypothesis test of $H_o : \beta_{radiation} = 0$ vs. $H_a : \beta_{radiation} \neq 0$

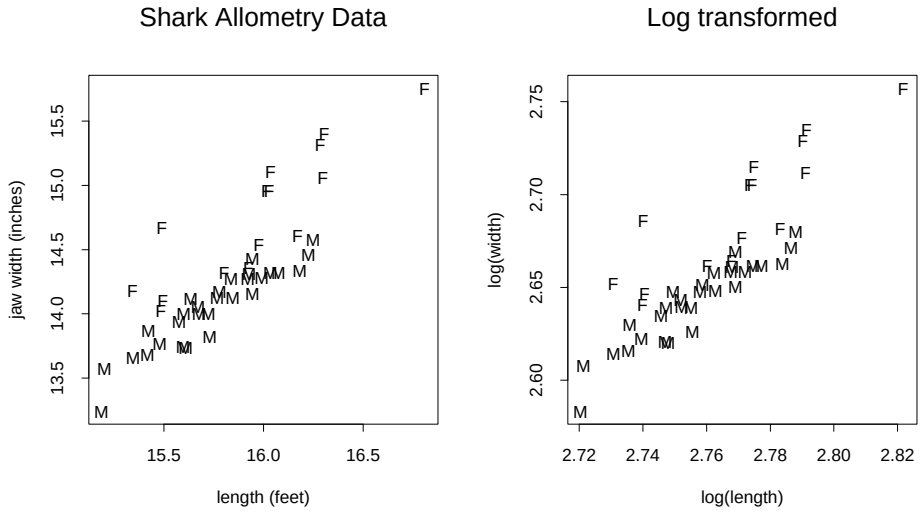
    (c) the residual sum of squares for the regression

    (d) the degrees of freedom for $\hat{\sigma}^2$

    (e) R^2

Questions (12) through (17) apply to the shark growth data described below.

- Allometry studies focus on how the growth of one part of an organism relates to the growth of the whole organism. The data below relate overall body length (in feet) and jaw width (in inches) for 44 male and female sharks. Of interest is whether it is possible to estimate jaw width from body length, and whether growth patterns vary by gender.
- A common allometric model in this context is a regression of log(width) on log(length). The slope parameter for such a regression is called a *growth coefficient*. Researchers are often interested in whether the growth coefficient is greater than 1 (positive allometry), equal to 1 (isometric), or less than 1 (negative allometry).
- Overall length (feet) and jaw width (inches) were measured for 15 female sharks and 29 male sharks of different ages.
- Variable names are *length*, *width*, *gender*. The variable *gender* takes on two values: 0=male and 1=female.
- Splus output for a particular model, "Model A", is provided on the following page. In terms of Splus language, the model is: *log(width) ~ log(length)+gender+log(length):gender*



- Partial Splus output for regression **Model A**, *log(width) ~ log(length) + gender + log(length):gender*.

```
Call: lm(formula = log.width ~ log.length + gender + log.length:gender)

Coefficients:
                    Value Std. Error t value Pr(>|t|)
      (Intercept) -0.4778  0.3520    -1.3575  0.1822
       log.length  1.1320  0.1277     8.8640  0.0000
           gender -0.2465  0.5074    -0.4858  0.6298
log.length:gender  0.1010  0.1836     0.5502  0.5852

Residual standard error: 0.0122 on -- degrees of freedom
Multiple R-Squared: 0.8906
F-statistic: 108.6 on -- and -- degrees of freedom, the p-value is 0

Correlation of Coefficients:
                  (Intercept) log.length  gender
       log.length -1.0000
           gender -0.6937      0.6937
log.length:gender  0.6955     -0.6955    -1.0000

Terms Added Sequentially
                  Df  Sum of Sq    Mean Sq  F Value     Pr(F)
log.length        -- 0.03885729 0.03885729 261.2426 0.0000000
gender            -- 0.00953727 0.00953727  64.1203 0.0000000
log.length:gender -- 0.00004503 0.00004503   0.3027 0.5852297
Residuals         -- 0.00594961 0.00014874
```

12. Write out the fitted model for Model A, using precise statistical notation. Then write out the separate simplified fitted model expressions for males and for females. [10 points]

13. In allometry studies, one area of interest is whether the linear equation relating width and length is the same for males and females. Use the output of Model A to answer this question using a formal hypothesis test with $\alpha = 0.05$. What conclusion would you make?[15 points]

14. **True or False. Circle one.** The p-values for the t-tests given in the Model A output indicate that we should drop both *gender* and *log.length:gender* terms. [5 points]

15. Four 95% prediction intervals are calculated independently for future observations of jaw width for male sharks with overall lengths of 11, 13, 15, and 17 feet. (You do not need to calculate the four 95% intervals.)

    (a) **True or False. Circle one.** The value of SE[Pred$\{Width|Length, Gender = 0\}$] should be the same regardless of whether the researcher believes the separate lines or parallel lines model is true. [5 points]

    (b) **True or False. Circle one.** The four intervals simultaneously cover their respective true population values with 95% probability. [5 points]

16. Assuming Model A is correct, what is the effect of a tripling of length for females? Give your answer on the original scale of measurement. [5 points]

17. Further examination of the data indicates that the sharks can be classified into four groups according to species. Let these species be denoted by "Species A", "Species B", "Species C" and "Species D".

    How might we include a species effect? Describe (1) how many variable(s) you would create to represent a species effect and (2) what values the variables would take. [5 points]