**STATISTICS 608**
**Homework 03**

**Question 1** [**4**] A regression analysis involving an intercept term and *two* covariate measurements on each of $n = 22$ subjects produced an estimate $\hat{\beta}_1 = 1.0$ and a 90% confidence interval $(-2.36, 4.36)$ for $\beta_1$. Find the standard error of $\hat{\beta}_1$.

**Question 2** [**4+2+3+2+4=15**] (Chapter 5, Section 5.2) The output below was obtained from fitting a three - covariate linear model,

$$Y_i = \beta_0 + \sum_{j=1}^{3} \beta_j x_{ij} + e_i,$$

to observed responses:

| **covariate** | **est.coef.** | **st.err.** | **sig.** |
|---|---|---|---|
| $x_1$ | $\hat{\beta}_1$ | 0.5 | 0.02 |
| $x_2$ | 1.0 | 0.25 | 0.00 |
| $x_3$ | 0.4 | 0.4 | 0.33 |

$$RSS = 100;\ n = 30;\ R^2 = 0.9$$

where *df* is short for "degrees of freedom", **est.coef.** denotes the least squares estimates of $\beta_j,\ j = 1, 2, 3$, and **sig.** denotes the $p-$value (two-sided) of the $t-$statistic for testing the hypothesis that the coefficient of the respective covariate is zero.

**2.1** Find an unbiased estimate of the error variance, $\sigma^2$.

**2.2** Find the two possible numerical values of $\hat{\beta}_1$.

**2.3** Find a 95% confidence interval for $\beta_2$, the coefficient of $X_2$ in the model.

**2.4** Find the regression sum of squares, $SS_{reg}$, for these data.

**2.5** Test at the 5% level the hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

**NOTE:** Here you could use either the Analysis of Variance method (page 135) or the Partial F-test method (page 137)

**Question 3 [2+1+1+2+2+1+2+1+1=13]** (Chapter 5, Section 5.2) The *Question 3 data* (attached) show the numerical values of a predictor $y$ and three covariates $x_1$, $x_2$ and $x_3$. Assume that a valid linear regression

$$E[y|x_1, x_2, x_3] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

is in force.

**3.1** Write down the least squares estimates of the four regression coefficients.

$\hat{\beta}_0 =$ ; $\hat{\beta}_1 =$ ; $\hat{\beta}_2 =$ ; $\hat{\beta}_3 =$ .

**3.2** Test the null hypothesis $H_0 : \beta_1 = \beta_2, \ \beta_3 = 0$ at the 10% level of significance by writing your answers to (a) through (f) below. Follow the same METHOD as the Partial F-test, but make allowance for the fact that the null hypothesis is compound, not simple. This is equivalent to, but simpler to implement, than Model Reduction Method 2 on slide 45 of the Chapter 5 slides and 46:54 min. into Lecture 14.

**(a)** $RSS_{full} =$

**(b)** $df_{full} =$

**(c)** $RSS_{null} =$

**(d)** $df_{null} =$

**(e)** Write an expression for your test statistic in terms of the symbols in parts (a) through (d).

**(f)** The numerical value of the test statistic is

**(g)** State the critical value for the hypothesis test OR state the $p$-value associated with (f).

**(h)** State your decision regarding the validity of $H_0$ and justify it with reference to part (g).

**Question 4** [4]

A botanist is interested in the efficacy on predator bugs in reducing pests on garden plants. In particular, two species ($A$ and $B$) of praying mantis are to be compared to see which devours potato beetles at a higher rate. One hundred grams of potato beetles are released into a cage containing potato foliage, and one praying mantis of each species is introduced into the cage for one week. At the end of the week, the reduction (in grams) of potato beetles is measured. Another identical cage is prepared, but this time, there are two praying mantis of Species $A$ and one of Species $B$. For the third cage, there are two of Species $B$ and one of Species $A$, and for the fourth cage of the study, there are two of each species introduced. Let $\beta_A$ be the average grams of potato beetles eaten per week per praying mantis for Species $A$, and let $\beta_B$ be the average for Species $B$. Write down a linear model in matrix notation that can be used to estimate $\beta_A$ and $\beta_B$, giving your design matrix $X$. Assume that the consumption of each praying mantis is independent of others in the cage.

**Question 5**$^*$ **[6+6=12]** We wish to *predict* the response $Y^*$ at a *covariate vector* $\mathbf{u}$ in a general linear model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$. The predicted value at $\mathbf{u}$ is $\hat{\mathbf{Y}}^* = \mathbf{u}'\hat{\beta}$ where $\hat{\beta}$ denotes the least squares estimate of $\beta$ based upon the data $(\mathbf{Y}, \mathbf{X})$. The prediction error is defined to be $\epsilon = \mathbf{Y}^* - \hat{\mathbf{Y}}^*$.

**5.1** Show, taking care to *justify all the steps in your derivation, that*

$$var(\epsilon) = \{1 + \mathbf{u}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{u}\}\sigma^2 \tag{1}$$

where $\sigma^2$ is the common variance of the components of the error vector $\mathbf{e}$.

**5.2** Show that formula (2.16) on page 26 is a special case of (1). In other words, show that if you substitute

$$\mathbf{u} = [1\ x^*]' \text{ and } \mathbf{X} = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix}'$$

into (1), then you get

$$var(\epsilon) = \left[1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{SXX}\right]\sigma^2_.$$