STA244 **1/08/2003**

Homework 1

Due 1/15/2003.

Please provide concise, neatly written or typed solutions. All work should be your own and not copied from other texts or sources. Do feel free to discuss questions with me, the TA, others in class, or post a question for clarification on the Course Info Discussion Board.

1. Assume that we have a sample of size n where
$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$
and the errors ϵ_i are iid $N(0, \sigma^2)$.
   (a) Find the maximum likelihood estimator of σ^2, $\hat{\sigma}^2$.
   (b) Under the assumption of normal errors as above, find the $E(\hat{\sigma}^2)$.
   (c) Is $\hat{\sigma}^2$ an unbiased estimate of σ^2? If not, find an unbiased estimate of σ^2.

2. **Invariance** Suppose that we recode the data above so that $Y_i^* = a + bY_i$ and $X_i^* = c + dX_i$ for known a, b, c, and d. Consider the model where Y_i^*s are independent normal random variables with mean $\gamma_0 + X_i^* \gamma_1$ and variance τ^2.
   (a) What is the relationship between the mles $\hat{\beta}_0$ and $\hat{\gamma}_0$?
   (b) What is the relationship between the mles $\hat{\beta}_1$ and $\hat{\gamma}_1$?
   (c) What is the relationships between $\hat{\sigma}^2$ and $\hat{\tau}^2$?
   (d) Are hypothesis tests 1) that the intercept equals zero versus non-zero and 2) that the slope is zero versus non-zero the same or different?

3. **Consistency** Recall that the variance of $\hat{\beta}_1$ is $\sigma^2 / \sum_{i=1}^n (X_i - \bar{X})^2$.
   (a) Under what conditions on the values of X will this variance approach zero as the sample size n goes to infinity?
   (b) Construct a sequence of values $(X_1, X_2, \ldots)$ such that as the sample size goes to infinity, the variance of $\hat{\beta}_1$ *does not* approach zero. What does this imply about convergence of $\hat{\beta}_1$?

4. In the simple linear regression model, as in Exercise (1),
   (a) show that the correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ is
$$\rho(\hat{\beta}_0, \hat{\beta}_1) = -\bar{X} \frac{\sqrt{\mathrm{Var}(\hat{\beta}_1)}}{\sqrt{\mathrm{Var}(\hat{\beta}_0)}}$$
   How can the values of X be chosen 1) so that the correlation is arbitrarily close to 0 and 2) be chosen so that the correlation is close to +1 or -1?

(b) *In your own words*, describe what $\rho(\hat{\beta}_0, \hat{\beta}_1)$ means.

(c) Let $Z_i = X_i - \bar{X}$ be the predictor variable centered to have average zero. An alternative parameterization of the regression model is to have $Y_i = \alpha + \beta_1 Z_i + \epsilon_i$, where $\alpha = \beta_0 + \beta_1 \bar{X}$. Write $\text{Var}(\hat{\alpha})$ as a function of n and σ^2, and find the value of $\rho(\hat{\alpha}, \hat{\beta}_1)$ for this model.

5. Consider the simple linear regression model in exercise 1) with residuals $\hat{e}_i = Y_i - \hat{Y}_i$ and fitted values $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$.

(a) Prove that $\sum \hat{e}_i = 0$ and $\sum \hat{Y}_i = \sum Y_i$.

(b) Show that the sample covariance between $\hat{e}$ and $\hat{Y}$ equals zero, and hence the residuals and fitted values are uncorrelated. *Hint: it is useful to write* $\hat{Y}_i = \bar{Y} + \hat{\beta}_1(X_i - \bar{X})$. *You will also need the result that* $\sum(a_i - \bar{a})(b_i - \bar{b}) = \sum(a_i - \bar{a})b_i$.

6. **Relationship between t and F** Let t be a random variable with a Student t distribution with ν degrees of freedom,

$$f(t|\nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \frac{1}{\sqrt{\nu\pi}} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{(\nu+1)}{2}} \qquad -\infty < t < \infty, \quad \nu \geq 1$$

(a) Prove that t^2 has a $F(\nu_n, \nu_d)$ distribution with $\nu_n = 1$ and $\nu_d = \nu$ degrees of freedom, where the density of a F is

$$f(f|\nu_n, \nu_d) = \frac{\Gamma((\nu_n + \nu_d)/2)}{\Gamma(\nu_n/2)\Gamma(\nu_d/2)} \left(\frac{\nu_n}{\nu_d}\right)^{\nu_n/2} \left(1 + \frac{\nu_n}{\nu_d} f\right)^{-\frac{(\nu_n+\nu_d)}{2}} f^{(\nu_n-2)/2} \qquad 0 < f < \infty$$

$\nu_n, \nu_d \geq 1$.

(b) Show that the square of the t-statistic for testing $H_o : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ is equal to the F statistic from the ANOVA table, F = MSReg/MSError.