Stats 100C: HW1 due Friday, April 3, 2009 (in class)

Exercise 1

For the regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, the residuals are defined as $e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the least squares estimates.

a. show that the least squares estimate of the slope $\hat{\beta}_1$ is also equal to

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

b. show that the least squares estimate of the slope $\hat{\beta}_1$ is also equal to

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

c. show that the sum of the residuals is always equal to zero, i.e $\sum_{i=1}^{n} e_i = 0$.

Exercise 2

Suppose that $y_i = \mu + \epsilon_i$, where $i = 1, 2, \ldots, n$, μ is an unknown parameter and the ϵ_i are independent errors with mean 0 and variance σ^2. Show that $\bar{y}$ is the least squares estimate of μ.

Exercise 3

The *coefficient of variation* for a sample of values $x_1, x_2, \cdots, x_n$ is defined by $C.V. = \frac{s}{\bar{x}}$, where s is the sample standard deviation and $\bar{x}$ is the sample mean. This term gives the standard deviation as a proportion of the mean, and it is sometimes an informative quantity. For example a value of $s = 10$ has little meaning unless we can compare it with something else. If $s = 10$ and $\bar{x} = 1000$ the amount of variation is small relative to the mean. However, if $s = 10$ and $\bar{x} = 5$ then the variation is quite large relative to the mean. If we were studying the precision (variation in repeated measurements) of a measuring instrument, the first case $C.V. = \frac{10}{1000}$ might give quite acceptable precision but the second case $C.V. = \frac{10}{5}$ would be quite unacceptable. Let $x_1, x_2, \cdots, x_{10}$ denote a random sample of size 10 from a normal distribution with mean 0 and variance σ^2.

a. Find the distribution of $\frac{10\bar{x}^2}{s^2}$.

b. Find the distribution of $\frac{s^2}{10\bar{x}^2}$.

c. Find the number c such that $P(-c < \frac{s}{\bar{x}} < c) = 0.95$.

Exercise 4

Suppose $y_i = \beta x_i + \epsilon_i$. In this equation x_i is non-random, β is an unknown parameter, ϵ_i are independent errors, and $\epsilon_i \sim N(0, \sigma^2)$.

a. Find the mean and variance of y_i.

b. What distribution does y_i follow? Write down the probability density function.

c. Write down the likelihood function based on n observations of y_i and x_i.

d. Show that the maximum likelihood estimate of β is

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

e. Show that the estimate of part (d) is an unbiased estimator of β, i.e., $E(\hat{\beta}) = \beta$.

f. Find the variance of this estimate.