STAT 350

Assignment 1

1. Suppose $Z_1, \ldots, Z_{10}$ are independent random variables each having a $N(0,6)$ distribution. Let $\bar{Z} = \sum Z_i/10$, $U = \sum_{i=1}^{4} Z_i^2/6$, $V = \sum_{i=5}^{10} Z_i^2/6$, $X = Z_1/\sqrt{V}$ and $Y = (3U)/(2V)$. Give the names for the distributions of each of $\bar{Z}$, U, V, X and Y and use tables to find $P(|\bar{Z}| > 1)$, $P(U \le 9.49)$, $P(-1.2 \le X \le 1.2)$, $P(Y > 6.23)$, $P(U \le V)$, $P(Z_1 - 2Z_2 \ge 10)$.

2. A new process for measuring the concentration of a chemical in water is being investigated. A total of n samples are prepared in which the concentrations are the known numbers x_i for $i = 1, \ldots, n$; the new process is used to measure the concentrations for these samples. It is thought likely that the concentrations measured by the new process, which we denote Y_i, will be related to the true concentrations via

$$Y_i = \beta x_i + \epsilon_i$$

where the ϵ_i are independent, have mean 0 and all have the same variance σ^2 which is unknown.

(a) If this model is fitted by least squares, (that is by minimizing $\sum_i (Y_i - \beta x_i)^2$) show that the least squares estimate of β is

$$\hat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2}.$$

(b) Show that the estimator in part (a) is unbiased.

(c) Compute (give a formula for) the standard error of $\hat{\beta}$.

(d) The error sum of squares for this model is $\sum (Y_i - \hat{\beta} x_i)^2$ which may be shown to have $n-1$ degrees of freedom. If the x_i are the numbers 1, 2, 3 and 4, $\hat{\beta} = 1$ and the error sum of squares is 0.12 find a 95% confidence interval for β and explain what further assumptions you must make to do so.

(e) Show that the estimator

$$\tilde{\beta} = \frac{\sum Y_i}{\sum x_i}$$

is also unbiased.

(f) Compute (give a formula for) the standard error of $\tilde{\beta}$. Which is bigger, the standard error of $\hat{\beta}$ or that of $\tilde{\beta}$?

(g) Show that the mle of β in this model is $\hat{\beta}$, the least squares estimate, if the ϵ_i have normal distributions.

3. Consider the two-way layout without replicates. We have data Y_{ij} for $i = 1, \ldots I$ and $j = 1, \ldots, J$. We generally fit a so-called additive model

$$Y_{ij} = \mu + \rho_i + \gamma_j + \epsilon_{ij}$$

In the following questions consider the case $I = 2$ and $J = 3$.

(a) If we treat μ, ρ_1, ρ_2, γ_1, γ_2 and γ_3 as the entries in the parameter vector β what is the design matrix $X = X_a$ and what is the rank of X_a?

(b) What is the determinant of the matrix $X_a^T X_a$? Is this matrix invertible? How many solutions do the normal equations have?

(c) Usually we impose the restrictions $\rho_1 + \rho_2 = 0$ and $\gamma_1 + \gamma_2 + \gamma_3 = 0$. Use these restrictions to eliminate ρ_2 and γ_3 from the model equation and, for the parameter vector $\beta^T = (\mu, \rho_1, \gamma_1, \gamma_2)$ find the design matrix X_b.

(d) An alternate set of restrictions is called corner point coding where we assume $\rho_1 = \gamma_1 = 0$. With this restriction and the parameter vector $\beta^T = (\mu, \rho_2, \gamma_2, \gamma_3)$ what is the design matrix X_c?

(e) Show that the three design matrices have the same column space by finding a matrix A such that $X_a = X_b A$ and similarly for X_b and X_c and for X_a and X_c.

(f) Use the previous part to show that the vectors of fitted values $\hat{Y}$ will be the same for any solution of the normal equations for any of the three design matrices.

4. From the text question 1.19, 1.23, 2.13 a and b and 2.23 a, b and c. In 2.23 c give a P-value and interpret this P-value.

DUE: Friday, 15 January.