Indian Statistical Institute, Bangalore
B. Math (III)
First Semester 2013-2014
Mid-Semester Examination : Statistics (III)

Date: 10-09-2013 | Maximum Score ~~8~~70 | Duration: 3 Hours

1. If $\mathbf{Y} \sim \mathbf{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$, $\sigma^2 > 0$ and $\rho^2 < 1$, then show that the random variable $W = \frac{(Y_1 - Y_2)^2}{(Y_1 + Y_2)^2} \times \frac{1+\rho}{1-\rho}$ has F distribution. Do $\frac{1}{W}$ and W have the same distribution? Justify.

[$9 + 3 = 12$]

2. Let $\boldsymbol{A}$ be an $n \times n$ real symmetric matrix. Let $(n - r)$ be the number of its zero eigenvalues. Then show that all its eigenvalues belong to the set $\{0, 1\}$ if and only if $\boldsymbol{A}$ is idempotent, nonnegative definite and $rank(\boldsymbol{A}) = r$.

[8]

3. Write down simple linear regression model with *uncorrelated* errors, explaining clearly the set-up and terms that you use. Prove that $c_0\widehat{\beta}_0 + c_1\widehat{\beta}_1$ is *best linear unbiased estimator (BLUE)* for $c_0\beta_0 + c_1\beta_1$, in the sense of minimum variance.

[8]

4. Under the simple linear regression model, if in addition we have the assumption that for given x the errors are *iid* normal with mean 0 and variance $\sigma^2 > 0$, then show that

   (a) $\begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix}$ is distributed as $N\left[\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \sigma^2 \begin{pmatrix} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_x^2}\right) & -\frac{\bar{x}}{S_x^2} \\ -\frac{\bar{x}}{S_x^2} & \frac{1}{S_x^2} \end{pmatrix}\right]$.

   (b) $\sum_{i=1}^n e_i^2$ is distributed independently of $\begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix}$ and has $\sigma^2\chi^2_{(n-2)}$ distribution.

   (c) Derive a test for testing the significance of the slope β_1 at $\alpha = 0.05$. Obtain $p - value$.

   (d) Derive 95% *confidence interval* for the intercept β_0.

[24]

5. Under the simple linear regression model with uncorrelated errors, if β_0 is known then obtain *least squares estimator* for β_1. Find *residual sum of squares.* Hence or otherwise find an unbiased estimator for σ^2, the error variance.

[$3 + 4 + 5 = 12$]

[*PTO*]

6. Suppose that a safety expert is interested in the relationship between the number of licensed vehicles in a community (X) and the number of accidents per year in that community (Y). In particular, the expert wishes to use the number of licensed vehicles to predict the number of accidents per year. A random sample of 10 communities yields the data shown in the following table.

Sr. No. i	1	2	3	4	5	6	7	8	9	10
Licensed vehicles (thousands) x_i	4	10	15	12	8	16	5	7	9	10
No. of accidents (hundreds) y_i	1	4	5	4	3	4	2	1	4	2

$\sum x_i = 96$, $\sum x_i^2 = 1060$, $\sum y_i = 30$, $\sum y_i^2 = 108$, $\sum x_i y_i = 328$.

(a) Write down a linear regression model stating clearly the assumptions you make.

(b) Suppose that another community chosen at random has 12,000 licensed vehicles. Estimate the expected number of accidents per year in that community and construct 99% confidence interval (CI).

(c) Construct 99% confidence interval for the future observation on number of accidents for the same community with $x = 12,000$.

(d) What do you notice about the midpoints of the above two intervals and their relative sizes? Comment.

(e) Use analysis of variance ($ANOVA$) to conduct an *F-test* for testing the significance of regression model. Obtain *p value*.

(f) Compute R^2, the coefficient of determination. Comment.

$[3 + 4 + 5 + 3 + 8 + 3 = 26]$