STOR 664 Homework 2, Fall 2022

Richard L. Smith

August 29, 2022

Due: Thursday September 8

1. In section 2.3, we showed that for the simple linear regression, the least squares estimator $\hat{\beta}_1$ has the property that among all linear unbiased estimators of β_1, it uniquely has minimum variance (the BLUE property). Prove the corresponding statement for $\theta = A\beta_0 + B\beta_1$ with A and B arbitrary constants. In other words, if we define $\hat{\theta} = A\hat{\beta}_0 + B\hat{\beta}_1$ with $\hat{\beta}_0, \hat{\beta}_1$ the least squares estimators, and if $\tilde{\theta} = \sum c_i y_i$ is some other linear estimator which is also unbiased, then necessarily we have $\text{Var}(\tilde{\theta}) \geq \text{Var}(\hat{\theta})$.

2. For the simple linear regression model $y_i = \alpha + \beta x_i + \epsilon_i$, $i = 1, 2, \ldots n$, where the ϵ_i are independent, normally distributed with mean zero and variance σ^2, find the mean, variance and covariance of the least squares estimators $\hat{\alpha}$ and $\hat{\beta}$, where $\hat{\beta} = \sum(x_i - \bar{x})(y_i - \bar{y}) / \sum(x_i - \bar{x})^2$ and $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$. Show that the residual $e_i = y_i - \hat{\alpha} - \hat{\beta}x_i$ has mean zero and variance $\sigma^2 \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right]$. Hence show that the expected value of $\sum e_i^2$ is $(n-2)\sigma^2$.

3. For the regression model $y_i = \beta x_i + \epsilon_i$, i.e. a straight line through the origin with the usual error assumptions, calculate the least squares estimate $\hat{\beta}$ of β. Show that $\hat{\beta}$ is unbiased, find its variance and hence that it has the property of being BLUE. Define the residual $e_i = y_i - \hat{\beta}x_i$, show that $\text{E}\{e_i\} = 0$ and $\text{Var}\{e_i\} = \sigma^2 \left(1 - x_i^2 / \sum_j x_j^2\right)$ and find the covariance of e_i, e_j for $i \neq j$.

4. Suppose in the situation of the previous question, an experimenter informs you that exactly one of the residuals e_i is non-zero, the rest all being equal to 0. Do you believe her? Explain why or why not. How does the situation differ from the corresponding one in the usual model when an intercept is included in the regression?

5. Suppose we are trying to test the hypothesis $\rho = 0$ in the bivariate normal distribution. The method described in section 2.9.1 is based on the fact that when the null hypothesis is true, the standardized test statistic $\frac{\sqrt{n-3}}{2} \log\left(\frac{1+r}{1-r}\right)$ has an approximately standard normal distribution. Here r is the sample correlation coefficient. However, we also argued that the problem of testing $\rho = 0$ is equivalent to testing that the slope of the regression line is 0, when either variable is regressed upon the other. Show that the standardized test statistic obtained by testing the significance of the regression parameter is $\sqrt{n-2}(r^{-2} - 1)^{-1/2}$, irrespective of which variable is considered the x variable and which the y variable. Show that the two methods are asymptotically equivalent as $n \to \infty$ and $r \to 0$ (the case when r is small is most important for testing).