STAT 8311/PUBH 8401 HW 8

Due Nov. 16, 2022

Exercise 1. Suppose that $y_i \sim \text{Ber}(\pi)$ for $i = 1, \ldots, n$ and the y_i are independent. Derive an expression for the joint distribution of the set of y_i and show that $\sum_i y_i$ is sufficient for π. Show that y_1 is an unbiased estimate of π. Derive the conditional expectation of y_1 given the sufficient statistic and argue that it is an unbiased minimum variance estimator.

Exercise 2. For the ANOVA table for a 1-way layout with p groups (e.g., Table 6.5 in the course notes), show how to derive the expressions for the expected mean square for each of the sources.

Exercise 3. For a 1-way ANOVA model

$$y_{ij} = \beta_i + \epsilon_{ij}$$

for $i = 1, \ldots, p$ and $j = 1, \ldots, n_i$ find an expression for the power of the test of the null hypothesis $\beta_p = \frac{1}{(p-1)} \sum_{i=1}^{p-1} \beta_i$. If $n_i = 10$ for all i, $p = 3$ and you have an estimate of σ of 5, draw a power curve over a range of values for the difference $\beta_p - \sum_{i=1}^{p-1} \beta_i/(p-1)$ using a 5% significance level. Values for the x-axis should be selected so that your curve crosses or intersects the values of 0.05 and 0.95.

Exercise 4. Consider a 1-way ANOVA with m treatment groups and a single control group. Suppose there are t observations in each of the treatment groups and c observations in the control group. If the goal is to minimize the variance of comparisons of each treatment to the control group for a fixed number of observations show that $c/t = \sqrt{m}$.

Exercise 5. A quantity called "R squared" is frequently used to assess the fit of a linear model. Suppose we are fitting a multiple regression model with p variables and a sample size of n (with $p < n$), and we are interested in testing the null hypothesis that all regression coefficients are zero other than the intercept. Define

$$SSTO = ||y||^2 - ||P_{\mathcal{E}_0} y||^2.$$

First show that

$$SSTO = ||(I - \frac{1}{n} J_n J_n^T) y||^2.$$

Next show that

$$SSTO = ||P_{\mathcal{E}-\mathcal{E}_0}y||^2 + ||Q_{\mathcal{E}}y||^2.$$

If we use the notation

$$SSR = ||P_{\mathcal{E}-\mathcal{E}_0}y||^2$$

then we can define R^2 by $\frac{SSR}{SSTO}$. Next show how to express R^2 in terms of the F-statistic.

Finally show that the distribution of R^2 under the null hypothesis follows a certain beta distribution (find the parameters of that distribution), and use this to show how the mean of R^2 depends on p for fixed n. What happens as p approaches n?