

STA 711 Homework 7

Due: Friday, March 24, 12:00pm (noon) on Canvas.

Instructions: Submit your work as a single PDF. For this assignment, you may include written work by scanning it and incorporating it into the PDF. Include all R code needed to reproduce your results in your submission.

1. Let X_1, \dots, X_n be an iid sample from a population with mean μ and variance σ^2 , and suppose that σ^2 is known. We wish to test the hypotheses $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$.
 - (a) Write an expression for the (approximate) power function for the Wald test of these hypotheses.
 - (b) Plot power as a function of μ , using $\alpha = 0.05$, $\mu = \mu_0$, $n = 100$, and $\sigma^2 = 1$.
 - (c) Let $\beta(\mu)$ be the (approximate) power function for the Wald test. Show mathematically that for each $\mu \neq \mu_0$, $\beta(\mu) \rightarrow 1$ as $n \rightarrow \infty$.
 - (d) Suppose that $\alpha = 0.05$, $\mu_0 = 0$ and $\sigma^2 = 1$. What is the minimum sample size n needed such that $\beta(0.5) > 0.7$?
2. Suppose that $X_1, \dots, X_n \stackrel{iid}{\sim} N(0, \sigma^2)$. We wish to test the hypotheses $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_A : \sigma^2 = \sigma_1^2$, where $\sigma_0^2 < \sigma_1^2$.
 - (a) Show that the most powerful test of these hypotheses rejects when $\sum_{i=1}^n X_i^2 > c$, for some value c .
 - (b) Find c such that the test in part (a) has size α .
3. Suppose that $X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$, with both μ and σ^2 unknown. Our hypotheses are $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_A : \sigma^2 \neq \sigma_0^2$. Propose a test statistic and rejection region for testing these hypotheses, such that the resulting test is size α .
4. Suppose that $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Pareto}(\theta, \nu)$, with pdf

$$f(x|\theta, \nu) = \frac{\theta \nu^\theta}{x^{\theta+1}} \mathbb{1}\{x \geq \nu\},$$

where $\theta, \nu > 0$.

- (a) Find the maximum likelihood estimators of θ and ν .
- (b) We wish to test $H_0 : \theta = 1$ vs. $H_A : \theta \neq 1$, and ν is unknown. The likelihood ratio test rejects when

$$\frac{\sup_{\theta > 0, \nu > 0} L(\nu, \theta | \mathbf{X})}{\sup_{\theta = 1, \nu > 0} L(\nu, \theta | \mathbf{X})} > k.$$

Show that the likelihood ratio test is equivalent to rejecting when $T \leq c_1$ or $T \geq c_2$, where $0 < c_1 < c_2$ and

$$T = \log \left(\frac{\prod_{i=1}^n X_i}{X_{(1)}^n} \right).$$

5. Suppose we have two independent samples $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Exponential}(\theta)$ and $Y_1, \dots, Y_m \stackrel{iid}{\sim} \text{Exponential}(\mu)$.

(a) The likelihood ratio test rejects when

$$\frac{\sup_{\theta > 0, \mu > 0} L(\theta, \mu | \mathbf{X})}{\sup_{\theta = \mu} L(\theta, \mu | \mathbf{X})} > k.$$

Show that the LRT can be based on the statistic

$$T = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j}.$$

6. (Global F -test for linear regression) Suppose that $V_1 \sim \chi_{d_1}^2$ and $V_2 \sim \chi_{d_2}^2$ are independent χ^2 random variables. Then $F = \frac{V_1/d_1}{V_2/d_2} \sim F_{d_1, d_2}$, where F_{d_1, d_2} denotes the F -distribution with numerator degrees of freedom d_1 and denominator degrees of freedom d_2 .

The F -distribution is important for hypothesis testing in linear regression models. Suppose we observe independent data $(X_1, Y_1), \dots, (X_n, Y_n)$, where $Y_i = \beta^T X_i + \varepsilon_i$, with $\beta = (\beta_0, \dots, \beta_k)^T$ and $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$. We wish to test the hypotheses

$$H_0 : \beta_1 = \dots = \beta_k = 0 \quad H_A : \text{at least one of } \beta_1, \dots, \beta_k \neq 0.$$

The F -test for these hypotheses is based on the F -statistic

$$F = \frac{(SSTO - SSE)/k}{SSE/(n - k - 1)},$$

where $F \sim F_{k, n-k-1}$ under H_0 , and

$$SSTO = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad SSE = \sum_{i=1}^n (Y_i - \hat{\beta}^T X_i)^2$$

The goal of this problem is to demonstrate that, indeed, $F \sim F_{k, n-k-1}$ under H_0 .

- (a) Show that under H_0 , $\frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \beta_0)^2 \sim \chi_n^2$.

- (b) Find symmetric matrices A_1, A_2, A_3 such that under H_0 ,

$$\frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \beta_0)^2 = Z^T A_1 Z + Z^T A_2 Z + Z^T A_3 Z$$

where $Z \sim N(0, I)$, $\frac{1}{\sigma^2} SSE = Z^T A_1 Z$, and $\frac{1}{\sigma^2} (SSTO - SSE) = Z^T A_2 Z$.

- (c) Using the matrices A_1, A_2, A_3 from part (b), show that $\text{rank}(A_1) = n - k - 1$, $\text{rank}(A_2) = k$, and $\text{rank}(A_3) = 1$.

- (d) By applying Cochran's theorem, show that $F = \frac{(SSTO - SSE)/k}{SSE/(n - k - 1)} \sim F_{k, n-k-1}$ under H_0 .

7. When does the central limit theorem kick in?