Understanding the F-Test

January 16, 2021

1 F-Distribution

Definition 1. Suppose $X_1, \ldots, X_n \sim \mathcal{N}(\mu_X, \sigma_Y^2)$ and $Y_1, \ldots, Y_m \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Then the random variable,

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$$

where $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ and $S_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(Y_i - \bar{Y}\right)^2$. This random variable then has F distribution with $n-1$ and $m-1$ degrees of freedom.

Remark 2. Equivalently we can write,

$$F = \frac{X/d_1}{Y/d_2}$$

where $X \sim \chi^2_{d_1}$and $Y \sim \chi^2_{d_2}$.

2 F-Test under Homoscedasticity

Suppose now that we consider the regression,

$$Y = X'\beta_0 + \epsilon$$

where $X \in \mathbb{R}^{k \times n}$ and $Y \in \mathbb{R}^n$. Under homoscedasticity we are well aware of the fact that,

$$Var(\hat{\beta}_{OLS}) = \sigma^2 \left(X'X\right)^{-1}$$

Furthermore, we know,

$$\sqrt{n}\left(\hat{\beta}_{OLS} - \beta_0\right) \sim N\left(0, \sigma^2 \left(XX\right)^{-1}\right)$$

Define the linear transformation,

$$r(\beta) = R'\beta$$

where $R \in \mathbb{R}^{k \times q}$ which represents the k covariates in the above linear regression and q linear restrictions.

Consider the hypothesis test,

$$H_0 : R'\beta = \theta_0 \longleftrightarrow H_1 : R'\beta \neq \theta_0$$

By the delta method we have that,

$$R'\left(\hat{\beta}_{OLS} - \beta_0\right) \sim N\left(0, R'\sigma^2 \left(XX\right)^{-1} R\right)$$

So we have,

$$A = \frac{\left[R'\left(X'X\right)^{-1} R\right]^{-1/2}}{\sigma^2} R'\left(\hat{\beta}_{OLS} - \beta_0\right) \sim N\left(0, I_{q \times q}\right)$$

Claim 3. In the model defined above,

$$B = (n-k)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k}$$

where $\hat{\sigma}^2 = yM_X y/(n-k)$ is the unbiased error variance estimator.

Proof. Define the residuals as,

$$\hat{\epsilon} = y - X\beta'$$

Note we can define these with the residual maker matrix,

$$\hat{\epsilon} = M_X y = M_x\left(X\beta + \epsilon\right) = M_x\epsilon \equiv \left(I - X(X'X)^{-1}\right)\epsilon$$

Notice that,

$$tr\left(M_x\right) = n-k$$

This is obvious because of the diagonalization,

$$D = Q^T M_x Q = diag\left(1, \ldots, 1, 0, \ldots, 0\right)$$

Using the fact that,

$$\hat{\epsilon} \sim N(0, \sigma^2 M_x)$$

we have that by delta method,

$$Q\hat{\epsilon} \sim N(0, \sigma^2 D)$$

So clearly,

$$\frac{||Q\hat{\epsilon}||^2}{\sigma^2} \sim \chi^2_{n-k}$$

And to complete the proof notice that

$$\hat{\sigma}^2 = \frac{1}{n-k}\sum_i\left(y_i - x_i\hat{\beta}\right)^2 = \hat{\epsilon}^T M_X\hat{\epsilon} = [Q\hat{\epsilon}]^2$$

So we have,

$$\frac{||Q\hat{\epsilon}||^2}{\sigma^2} = \frac{\hat{\sigma}^2}{\sigma^2}$$

where we know that χ^2 has $n-k$ degree because $tr(M_x) = n-k$ as the trace is equal to the rank of the matrix. □

Now we can put this all together to get the **F-Statistic**,

$$F = \frac{A'A/q}{B/\left(n-k\right)} = \frac{R'\left(\hat{\beta}_{OLS} - \beta_0\right)\left[R'\left(X'X\right)^{-1}R\right]^{-1}R'\left(\hat{\beta}_{OLS} - \beta_0\right)/q}{\hat{\sigma}^2} \sim F_{q,n-k}$$

By definition of the F-distribution.
The F-statistic is often written as,

$$F = \frac{\left(TSS - RSS\right)/q}{RSS/\left(n-k\right)}$$

It is easy to see the equivalence in the denominator as,

$$RSS = yM_x y$$

Also define,

$$TSS = y^T\left(I - \frac{1}{n}\right)y$$

$$ESS = y^T\left(X'\left(X'X\right)^{-1}X^T - \frac{1}{n}\right)y$$

We know that if $\theta_0 = 0$ we have,

$$\frac{ESS}{\sigma^2} = \frac{y^T \left(X' \left(X'X \right)^{-1} X^T - \frac{1}{n} \right) y}{\sigma^2} \sim \chi^2_{q-1}$$

and,

$$\frac{ESS}{\sigma^2} = TSS - RSS$$

so it is easy to see that the usual formulation satisfies the F-distribution.

2.1 Asymptotic Power