Week 8 Wednesday Worksheet

Valentino Aceves

May 22, 2024

Joint Hypothesis Testing

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

An example of a Joint Null Hypothesis could be

$$H_0 : \beta_1 = \beta_2 = 0 \quad , \quad H_1 : \text{at least one of the estimates} \neq 0$$

We can now create two different models:

Unrestricted Regression which is our original:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

Restricted Regression where we assume H_0 is true:

$$Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4$$

Let SSR_u, SSR_r be the SSR's of the unrestricted and restricted models respectively. Let $p = 4$ be the number of parameters in the unrestricted model. Let $q = 2$ be the number of parameters in the restricted model. We calculate our F statistic with

$$\hat{F} = \frac{(SSR_r - SSR_u)/(p-q)}{SSR_u/(n-p-1)} \sim F_{df_1, df_2}$$

where $df_1 = p - q$ and $df_2 = n - p - 1$. We reject H_0 when $\hat{F} > F^c_{df_1, df_2}$ or $p_{\hat{F}} < \alpha$.

Example 1

```r
library(POE5Rdata)
data("cocaine")
```

```r
names(cocaine)
```

```
## [1] "price" "quant" "qual"  "trend"
```

```r
unrestricted_model <- lm(price ~ quant + qual + trend, cocaine)
restricted_model <- lm(price ~ quant, cocaine)
```

Perform an F test with the restricted and unrestricted models from above. State H_0 and H_1.

```r
df1 <- 3 - 1 # p-q
df2 <- unrestricted_model$df.residual # n-p-1
c(df1, df2)
```

```
## [1]  2 52
```

```r
SSR_u <- sum(unrestricted_model$residuals^2)
SSR_r <- sum(restricted_model$residuals^2)
c(SSR_u, SSR_r)
```

```
## [1] 20920.35 22126.29
```

```r
Fc <- qf(0.95, df1, df2)
F_stat <- ((SSR_r - SSR_u)/df1) / (SSR_u/df2)
c(Fc, F_stat)
```

```
## [1] 3.175141 1.498765
```

```r
1 - pf(F_stat, df1, df2)
```

```
## [1] 0.2328972
```

F-tests on Single Variables

Consider these two model for an F-test:

Unrestricted Regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

Restricted Regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

```
unrestricted_model <- lm(price ~ quant + qual + trend, cocaine)
restricted_model <- lm(price ~ quant + qual, cocaine)
```

Perform an F test with the restricted and unrestricted models from above. State H_0 and H_1.

```
df1 <- 3 - 2 # p-q
df2 <- unrestricted_model$df.residual # n-p-1
SSR_u <- sum(unrestricted_model$residuals^2)
SSR_r <- sum(restricted_model$residuals^2)
```

```
Fc <- qf(0.95, df1, df2)
F_stat <- ((SSR_r - SSR_u)/df1) / (SSR_u/df2)
c(Fc, F_stat)
```

```
## [1] 4.026631 2.885524
```

Couldn't we also do a t-test on the single variable? Sure.

```
summary(unrestricted_model)
```

```
##
## Call:
## lm(formula = price ~ quant + qual + trend, data = cocaine)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -43.479 -12.014  -3.743  13.969  43.753
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 90.84669    8.58025  10.588 1.39e-14 ***
## quant       -0.05997    0.01018  -5.892 2.85e-07 ***
## qual         0.11621    0.20326   0.572   0.5700
## trend       -2.35458    1.38612  -1.699   0.0954 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.06 on 52 degrees of freedom
## Multiple R-squared:  0.5097, Adjusted R-squared:  0.4814
## F-statistic: 18.02 on 3 and 52 DF,  p-value: 3.806e-08
```

We can see the t-statistic for $H_0 : \beta_3 = 0, H_1 : \beta_3 \neq 0$ is $\hat{t} = -1.699$.

```
summary(unrestricted_model)$coefficients[4, 3] ^ 2 # t_hat squared
```

```
## [1] 2.885524
```

In a single variable hypothesis test, when we square the t-statistic, we get exactly the F-statistic.

F-Tests for Overall Significance

When we run a summary table, it gives us an F-stat along with its p-value. What is the F-test that the regression is running for us?

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \qquad \text{vs} \qquad H_1 : \text{at least one } \beta_i \neq 0$$

Which leaves us with a restricted model of

$$Y = \beta_0 \qquad \Longrightarrow \qquad \beta_0 = \bar{Y}$$

$$\Longrightarrow SSR_r = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 = TSS$$

Prove that, in this model, F statistic becomes

$$\hat{F} = \frac{R^2/p}{(1 - R^2)/(n - p - 1)}$$