# Lecture 8: Statistical Inference

Ailin Zhang

2023-05-30

# Agenda

- t-Test of a Single $\beta_j$
- Confidence Intervals for Coefficients
- Confidence Intervals for Prediction
- Prediction Intervals for Prediction
- Sum of Squares
- Model Comparison (F-test, Today)

# Nested Models

- We say Model 2 is nested in Model 1 if Model 2 is a submodel of Model 1 (and Model 1 is an extension of Model 2)
    - Model A: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$
    - Model B: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$
    - Model C: $Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \epsilon$
    - Model D: $Y = \beta_0 + \beta_1 (X_1 + X_2) + \epsilon$
- B is nested in A since A reduces to B when $\beta_3 = 0$
- C is nested in A since A reduces to C when $\beta_2 = 0$
- D is nested in B since B reduces to D when $\beta_2 = \beta_1$
- B and C are NOT nested in either way
- D is not nested in C

# Motivating Question: Model Comparison

When two models are nested (Model 1 is nested in Model 2)

- The simpler model is called the reduced model (Model 1)
- The more general model is called the full model (Model 2)

Which model is better?

- SST? Full and ruduced model have equal SST.
- SSE?

  $\text{SSE}_{full} \leq \text{SSE}_{reduced}$
- SSR?

  $\text{SSR}_{full} \geq \text{SSR}_{reduced}$

# General Framework for Testing Nested Models

$H_0$: Reduced model is true v.s. $H_1$: Full model is true

- Simplicity or Accuracy?
    - The full model fits the data better (with a smaller SSE) but it is more complex.
    - The reduced model doesn't fit as well but it is simpler.
    - If $\text{SSE}_{reduced} \approx \text{SSE}_{full}$: one can sacrifice a bit of accuracy in exchange for simplicity.
    - If $\text{SSE}_{reduced} \gg \text{SSE}_{full}$:it would sacrifice too much in accuracy in exchange for simplicity. The full model is preferred.
- How to quantify?
    - F statistic!

# Construction of the F-statistic

- Consider the complexity (difference in the number of parameters)

  $df_{H_0} - df_{H_A}$

- and the fit

  $\text{SSE}_{H_0} - \text{SSE}_{H_A} = \text{RSS}_{H_O} - \text{RSS}_{H_A}$

> The F distribution with a,b degrees of freedom is defined to be the distribution of the ratio $\frac{\chi_a^2/a}{\chi_b^2/b}$ when $\chi_a^2$ and $\chi_b^2$ are independent.

- Hence a reasonable test statistic is $\dfrac{(\text{SSE}_{reduced} - \text{SSE}_{full})/(df_{reduced} - df_{full})}{\text{SSE}_{full}/df_{full}}$

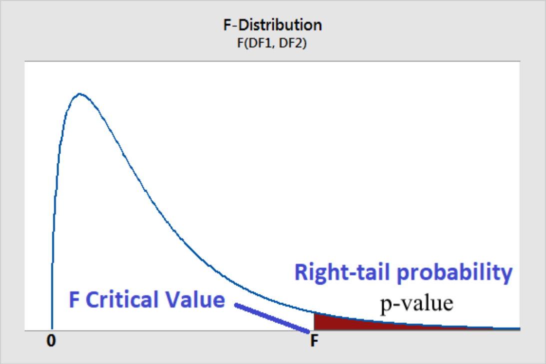
For denominator $\text{SSE}_{full}/df_{full} = \text{MSE} = \hat{\sigma}^2$ we use $\text{MSE}_{full}$ rather than $\text{MSE}_{reduced}$ since the full model is always true as the reduced model is a special case of the full model.

# The F-Statistic

$$F = \frac{(\text{SSE}_{reduced} - \text{SSE}_{full})/(df_{reduced} - df_{full})}{\text{MSE}_{full}}$$

- Under $H_0$, F statistic has an F-distribution with $(df_{reduced} - df_{full}, df_{full})$ degrees of freedom
- $F \geq 0$ since $\text{SSE}_{reduced} \geq \text{SSE}_{full}$
- The smaller the F-statistic, the more the reduced model is favored

# Decision Rule




- Reject $H_0$ if p-value $< \alpha$ (usually $\alpha = 5\%$)
- Accept $H_0$ otherwise

# Example 1: Testing All Coefficients Equal to Zero

$H_0 : \beta_1 = \beta_2 \cdots = \beta_p = 0$ v.s. $H_a$: not all $\beta_1, \dots, \beta_p = 0$

- The is a test to evaluate the **overall significance** of a model.
  - Full: $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i$
  - Reduced: $y_i = \beta_0 + \epsilon_i$ (All predictors are unnecessary)
- The OLS estimate for $\beta_0$ in the reduced model is $\hat{\beta}_0 = \bar{y}$, so $\text{SSE}_{reduced} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \text{SST}_{full}$
- $F = \frac{(\text{SSE}_{reduced} - \text{SSE}_{full})/(df_{reduced} - df_{full})}{\text{MSE}_{full}} = \frac{(\text{SST}_{full} - \text{SSE}_{full})/(n - 1 - (n - p - 1))}{\text{MSE}_{full}} = \frac{\text{SSR}_{full}/p}{\text{MSE}_{full}} = \frac{\text{MSR}_{full}}{\text{MSE}_{full}}$
- Moreover, $F \sim F_{p,n-p-1}$ under $H_0 : \beta_1 = \beta_2 \cdots = \beta_p = 0$.

# Example 1: Testing All Coefficients Equal to Zero

In R, the F statistic and p-value are displayed in the last line of the output of the `summary()` command.

```
data(trees)
trees$Diameter = trees$Girth
lmtrees = lm(log(Volume) ~ log(Diameter) + log(Height), data=trees)
summary(lmtrees)
```

```
##
## Call:
## lm(formula = log(Volume) ~ log(Diameter) + log(Height), data = trees)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.168561 -0.048488  0.002431  0.063637  0.129223
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.63162    0.79979  -8.292 5.06e-09 ***
## log(Diameter)  1.98265    0.07501  26.432  < 2e-16 ***
## log(Height)    1.11712    0.20444   5.464 7.81e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08139 on 28 degrees of freedom
## Multiple R-squared:  0.9777, Adjusted R-squared:  0.9761
## F-statistic: 613.2 on 2 and 28 DF,  p-value: < 2.2e-16
```

# ANOVA (Analysis of Variance) and F-Test

The test of all coefficients $= 0$ is summarized in an ANOVA table:

| Source | df | Sum of Squares | Mean Squares | F |
|---|---|---|---|---|
| Regression | p | SSR | MSR | $F = \frac{MSR}{MSE}$ |
| Error | n-p-1 | SSE | MSE | |
| Total | n-1 | SST | | |

- ANOVA is the shorthand for **an**alysis **o**f **va**riance.
- It decomposes the total variation in the response (SST) into separate pieces that correspond to different sources of variation,

  like SST $=$ SSR $+$ SSE in the regression setting.

# Example 2: General Test on Full and Reduced Models

anova(model1,model2)

```r
lmfull = lm(log(Volume) ~ log(Diameter) + log(Height),
            data=trees)
lmreduced = lm(log(Volume) ~ 1, data=trees)
anova(lmreduced, lmfull)
```

```
## Analysis of Variance Table
##
## Model 1: log(Volume) ~ 1
## Model 2: log(Volume) ~ log(Diameter) + log(Height)
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 8.3087
## 2     28 0.1855  2    8.1232 613.19 < 2.2e-16 ***
## ---
```

# Example 3: Testing Some Coefficients Equal to Non-Zero Values

In the model for the trees data,

$$\log(\text{Volume}) = \beta_0 + \beta_1 \log(\text{Diameter}) + \beta_2 \log(\text{Height}) + \epsilon$$

recall we think that $\beta_1 = 2$ and$\beta_2 = 1$.

We can test both coefficients in one test.

$H_0 : \beta_1 = 2, \beta_2 = 1$

- Note in the reduced model, the coefficients of log(Diameter) and log(Height) are both known
- Terms with known coefficients in an MLR model are called offsets. One can add an `offset` term in an `lm()` model

```
lmreduced = lm(log(Volume) ~ 1, offset=2*log(Diameter)+log(Hei
                 data=trees)
```

# Example 3: Testing Some Coefficients Equal to Non-Zero Values

```r
anova(lmreduced, lmfull)
```

```
## Analysis of Variance Table
##
## Model 1: log(Volume) ~ 1
## Model 2: log(Volume) ~ log(Diameter) + log(Height)
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     30 0.18769
## 2     28 0.18546  2 0.0022224 0.1678 0.8464
```

# Example 4: Testing the Equality of Coefficients

Testing $H0 : \beta_1 = \beta_2 = \beta_3$ under the model
$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$, the reduced model is
$Y = \beta_0 + \beta_1 (X_1 + X_2 + X_3) + \beta_4 X_4 + \epsilon$

- Make a new variable $W = X_1 + X_2 + X_3$
- Fit the reduced model by regressing Y on W and X4
- What is $df_{reduced} - df_{full}$? 2

```
lmfull = lm(Y ~ X1 + X2 + X3 + X4)
W = X1 + X2 + X3
lmreduced = lm(Y ~ W + X4)
# or simply
lmreduced = lm(Y ~ I(X1 + X2 + X3) + X4)
anova(lmreduced, lmfull)
```

# Other Scenarios

- Testing Equality of Coefficients

```
lmfull = lm(Y ~ X1 + X2 + X3 + X4)
lmreduced = lm(Y ~ X1 + I(X2 + X3) + I(X2 + X4))
anova(lmreduced, lmfull)
```

- Testing Coefficients under Constraints

```
lmfull = lm(Y ~ X1 + X2 + X3 + X4)
lmreduced = lm(Y ~ I(X1 + 2*X2) + X3 + X4)
anova(lmreduced, lmfull)
```

- Testing Coefficients under Constraints

```
lmfull = lm(Y ~ X1 + X2)
lmreduced = lm(Y ~ I(X1 -X2), offset = X2)
anova(lmreduced, lmfull)
```

# F-Test on a Single $\beta_j$ in equivalent to t-Test

If you want to test a single $\beta_3 = 0$ in the model

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

- You can accomplish the hypothesis test with a t-statistics for $X_3$ from the output for `summary(lm(Y~ X1 + X2 + X3))`
- Alternatively, you can conduct an F-test comparing the models

  Full model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$

  Reduced model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

```r
anova(lm(Y ~ X1 + X2 + X3), lm(Y ~ X1 + X2))
```

One can show that the F-statistic = (t-statistic)$^2$ and the P-values are the same, thus the two tests are equivalent.

The proof involves complicate matrix algebra and it hence omitted.

# Example

```r
lm1 = lm(log(Volume) ~ log(Diameter) + log(Height), data = trees)
lmreduced = lm(log(Volume) ~ log(Diameter), data = trees)
anova(lmreduced,lm1)
```

```
## Analysis of Variance Table
##
## Model 1: log(Volume) ~ log(Diameter)
## Model 2: log(Volume) ~ log(Diameter) + log(Height)
##   Res.Df     RSS Df Sum of Sq     F    Pr(>F)
## 1     29 0.38324
## 2     28 0.18546  1   0.19778 29.86 7.805e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(lm1)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)   -6.631617 0.79978973 -8.291701 5.057138e-09
## log(Diameter)  1.982650 0.07501061 26.431592 2.422550e-21
## log(Height)    1.117123 0.20443706  5.464388 7.805278e-06
```

- $(\text{t-statistics})^2 = 5.464^2 \approx 29.86 = \text{F-statistic}$
- The P-values are both $7.8e-6$ (The slight difference is due to rounding)