Friday, January 24

Confidence Intervals and Significance Tests

A significance test can be used to derive a confidence interval, and a confidence interval can be used to conduct a significance test. If we have hypotheses for a two-sided test like

$$H_0\colon \beta_j = c \ \text{ and } \ H_a\colon \beta_j \neq c,$$

then we reject H_0 if and only if the confidence interval for β_j does not contain c, with a couple of caveats.

1. The confidence level must be $(1-\alpha)100\%$ (α is the *significance level*).
2. The test is two-sided (but one-sided tests match one-sided confidence intervals).

A confidence interval with confidence level $(1-\alpha)100\%$ effectively defines all values of the parameter that *would not be rejected* in a two-sided test with significance level α.

Note that this also applies to a linear function of model parameters (ℓ). So if we have the hypotheses

$$H_0\colon \ell = c \ \text{ and } \ H_a\colon \ell \neq c,$$

then we reject H_0 if and only if the confidence interval for ℓ does not contain c.

Example: Consider again the model for the `anorexia` data, but parameterized to compare the two treatment conditions against the control so that the model is

$$E(Y_i) = \begin{cases} \beta_0, & \text{if the } i\text{-th observation is under the control condition,} \\ \beta_0 + \beta_1, & \text{if } i\text{-th observation under cognitive behavioral therapy,} \\ \beta_0 + \beta_2, & \text{if the } i\text{-th observations is under family therapy.} \end{cases}$$

```r
library(MASS) # for anorexia data
anorexia$change <- anorexia$Postwt - anorexia$Prewt
anorexia$Treat <- relevel(anorexia$Treat, ref = "Cont")
m <- lm(change ~ Treat, data = anorexia)
cbind(summary(m)$coefficients, confint(m))
```

```
            Estimate Std. Error t value Pr(>|t|)   2.5 % 97.5 %
(Intercept)   -0.450      1.476 -0.3048 0.761447 -3.3954  2.495
TreatCBT       3.457      2.033  1.7001 0.093608 -0.5994  7.513
TreatFT        7.715      2.348  3.2854 0.001602  3.0302 12.399
```

We can produce the same inferences using `contrast`.

```r
library(trtools)
contrast(m,
  a = list(Treat = c("CBT","FT")),
  b = list(Treat = "Cont"),
  cnames = c("Cognitive vs Control", "Family vs Control"))
```

```
                     estimate    se   lower  upper tvalue df   pvalue
Cognitive vs Control    3.457 2.033 -0.5994  7.513  1.700 69 0.093608
Family vs Control       7.715 2.348  3.0302 12.399  3.285 69 0.001602
```

Joint Hypotheses

Example: Consider the following model and hypotheses for the `anorexia` data.

```
library(MASS) # for anorexia data
anorexia$change <- anorexia$Postwt - anorexia$Prewt
m.anorexia <- lm(change ~ Treat, data = anorexia)
summary(m.anorexia)$coefficients
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.007      1.398   2.151  0.03499
TreatCont     -3.457      2.033  -1.700  0.09361
TreatFT        4.258      2.300   1.852  0.06838
```

The model is therefore

$$E(Y_i) = \begin{cases} \beta_0, & \text{if the } i\text{-th observation is under cognitive behavioral therapy}, \\ \beta_0 + \beta_1, & \text{if } i\text{-th observation is under the control condition}, \\ \beta_0 + \beta_2, & \text{if the } i\text{-th observations is under family therapy}. \end{cases}$$

In some cases we might be testing hypothesis like $H_0 : \beta_2 = 0$ or $H_0 : \beta_1 - \beta_2 = 0$. But in other cases we might be testing what is sometimes called a *joint* hypothesis such as

$$H_0 \colon \beta_1 = 0 \text{ and } \beta_2 = 0 \quad \text{versus} \quad H_a \colon \text{not both } \beta_1 = 0 \text{ and } \beta_2 = 0.$$

What does it imply if both $\beta_1 = 0$ and $\beta_2 = 0$?

Example: Consider the following model for the `whiteside` data.

```
m.insulation <- lm(Gas ~ Insul + Temp + Insul:Temp, data = whiteside)
summary(m.insulation)$coefficients
```

```
                Estimate Std. Error t value  Pr(>|t|)
(Intercept)       6.8538    0.13596  50.409 7.997e-46
InsulAfter       -2.1300    0.18009 -11.827 2.316e-16
Temp             -0.3932    0.02249 -17.487 1.976e-23
InsulAfter:Temp   0.1153    0.03211   3.591 7.307e-04
```

The model is therefore

$$E(Y_i) = \begin{cases} \beta_0 + \beta_2 t_i, & \text{if } i\text{-th observation is before insulation}, \\ \beta_0 + \beta_1 + (\beta_2 + \beta_3) t_i, & \text{if } i\text{-th observation is after insulation}. \end{cases}$$

We might test a single null hypothesis that the rate of change in expected gas consumption with respect to temperature is the same before and after insulation — i.e., $H_0 : \beta_3 = 0$. But consider the joint hypothesis

$$H_0 \colon \beta_1 = 0 \text{ and } \beta_3 = 0 \quad \text{versus} \quad H_a \colon \text{not both } \beta_1 = 0 \text{ and } \beta_3 = 0.$$

What does it imply if both $\beta_1 = 0$ and $\beta_3 = 0$?

The "Analysis of Variance" Calculations

Calculations for inference for linear models is often based on the sums of squares decomposition

$$\underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{\text{total}} = \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{\text{model/regression}} + \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{\text{error/residual}},$$

where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$, and the degrees of freedom decomposition

$$\underbrace{n-1}_{\text{total}} = \underbrace{p-1}_{\text{model/regression}} + \underbrace{n-p}_{\text{error/residual}},$$

where p is the number of β_j parameters, and $p = k+1$ if the model includes a β_0. (Note: If the β_0 parameter is omitted from the model, the total degrees of freedom becomes n and the model/regression degrees of freedom becomes p.)

A *mean square* is a variance-like quantity that is a sum of squares divided by its corresponding degrees of freedom.

Tests can be conducted using the F test statistic which can be written as

$$F = \frac{(\text{RSS}_{\text{null}} - \text{RSS}_{\text{full}})/(\text{RDF}_{\text{null}} - \text{RDF}_{\text{full}})}{\text{RSS}_{\text{full}}/\text{RDF}_{\text{full}}},$$

where RSS and RDF refer to the *residual* sum of squares and degrees of freedom, respectively. The degrees of freedom for the F distribution are $\text{RDF}_{\text{null}} - \text{RDF}_{\text{full}}$ (numerator) and RSS_{full} (denominator). The *full* model is the model we are using, and the *null* (aka "reduced") model is what the full model reduces to *if the null hypothesis is true*. The F test statistic can be used for tests of individual and joint hypotheses in linear models.

Using the `anova` Function

The `anova` function is particularly useful for testing joint hypothesis, although it can also be used to test a hypothesis about a single parameter.

Applying `anova` to a single model will produce the RSS and RDF in the `Residuals` row.

```r
anova(m.anorexia)
```

```
Analysis of Variance Table

Response: change
          Df Sum Sq Mean Sq F value Pr(>F)
Treat      2    615   307.3    5.42 0.0065 **
Residuals 69   3911    56.7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To conduct a test, the recommended approach is to apply `anova` to a null model and the full model.

```r
m.full <- lm(change ~ Treat, data = anorexia)
m.null <- lm(change ~ 1, data = anorexia) # use ~ 1 if no explanatory variables
anova(m.null, m.full)
```

```
Analysis of Variance Table

Model 1: change ~ 1
Model 2: change ~ Treat
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1     71 4525
2     69 3911  2       615 5.42 0.0065 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
m.full <- lm(Gas ~ Insul + Temp + Insul:Temp, data = whiteside)
m.null <- lm(Gas ~ Temp, data = whiteside)
anova(m.null, m.full)
```

```
Analysis of Variance Table

Model 1: Gas ~ Temp
Model 2: Gas ~ Insul + Temp + Insul:Temp
  Res.Df  RSS Df Sum of Sq   F Pr(>F)
1     54 40.0
2     52  5.4  2      34.6 166 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The `anova` function can also do a test concerning a single parameter. Here are two approaches to testing the null hypothesis that $\beta_3 = 0$ in the model

$$E(Y_i) = \begin{cases} \beta_0 + \beta_2 t_i, & \text{if } i\text{-th observation is before insulation,} \\ \beta_0 + \beta_1 + (\beta_2 + \beta_3) t_i, & \text{if } i\text{-th observation is after insulation.} \end{cases}$$

```r
m.full <- lm(Gas ~ Insul + Temp + Insul:Temp, data = whiteside)
m.null <- lm(Gas ~ Insul + Temp, data = whiteside)
anova(m.null, m.full)
```

```
Analysis of Variance Table

Model 1: Gas ~ Insul + Temp
Model 2: Gas ~ Insul + Temp + Insul:Temp
  Res.Df  RSS Df Sum of Sq    F  Pr(>F)
1     53 6.77
2     52 5.43  1      1.34 12.9 0.00073 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(m.full)$coefficients
```

```
                Estimate Std. Error t value  Pr(>|t|)
(Intercept)       6.8538    0.13596  50.409 7.997e-46
InsulAfter       -2.1300    0.18009 -11.827 2.316e-16
Temp             -0.3932    0.02249 -17.487 1.976e-23
InsulAfter:Temp   0.1153    0.03211   3.591 7.307e-04
```

Comment: When conducting a test concerning one parameter (or a single linear function of the model parameters), the F and t test statistics have the relationship $t^2 = F$ and produce the same p-values.

Example: Three Approaches to One Test

Consider again the model for the `anorexia` data, but suppose we parameterized the model differently.

```r
anorexia$Treat <- relevel(anorexia$Treat, ref = "Cont")
m.anorexia <- lm(change ~ Treat, data = anorexia)
summary(m.anorexia)$coefficients
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.450      1.476 -0.3048 0.761447
TreatCBT       3.457      2.033  1.7001 0.093608
TreatFT        7.715      2.348  3.2854 0.001602
```

The model is therefore

$$E(Y_i) = \begin{cases} \beta_0, & \text{if the } i\text{-th observation is from the control group,} \\ \beta_0 + \beta_1, & \text{if the } i\text{-th observation is from the cognitive-behavioral therapy group,} \\ \beta_0 + \beta_2, & \text{if the } i\text{-th observations is from the family therapy group.} \end{cases}$$

Now consider a test of the null hypothesis that the expected weight change is the same regardless of which of the two therapies (i.e., cognitive-behavioral or family) is used. This is the null hypothesis that $\beta_1 = \beta_2$ or, equivalently, $\beta_1 - \beta_2 = 0$.

1. Using `lincon` we can test this null hypothesis as follows.

```
m <- lm(change ~ Treat, data = anorexia)
trtools::lincon(m, a = c(0, 1, -1))
```

```
             estimate  se  lower  upper tvalue df  pvalue
(0,1,-1),0   -4.258 2.3 -8.845 0.3299 -1.852 69 0.06838
```

   This is because the null hypothesis can be written as

$$\ell = 0 \times \beta_0 + 1 \times \beta_1 + (-1) \times \beta_2 = \beta_1 - \beta_2.$$

2. Using `contrast` we can test this null hypothesis as follows.

```
m <- lm(change ~ Treat, data = anorexia)
trtools::contrast(m, a = list(Treat = "CBT"), b = list(Treat = "FT"))
```

```
 estimate  se  lower  upper tvalue df  pvalue
   -4.258 2.3 -8.845 0.3299 -1.852 69 0.06838
```

3. Using `anova` we can test this null hypothesis as follows.

```
anorexia$therapy <- ifelse(anorexia$Treat == "Cont", "control", "therapy")
head(anorexia)
```

```
  Treat Prewt Postwt change therapy
1  Cont  80.7   80.2   -0.5 control
2  Cont  89.4   80.1   -9.3 control
3  Cont  91.8   86.4   -5.4 control
4  Cont  74.0   86.3   12.3 control
5  Cont  78.1   76.1   -2.0 control
6  Cont  88.3   78.1  -10.2 control
```

```
tail(anorexia)
```

```
   Treat Prewt Postwt change therapy
67    FT  82.1   95.5   13.4 therapy
68    FT  77.6   90.7   13.1 therapy
69    FT  83.5   92.5    9.0 therapy
70    FT  89.9   93.8    3.9 therapy
71    FT  86.0   91.7    5.7 therapy
72    FT  87.3   98.0   10.7 therapy
```

```
m.full <- lm(change ~ Treat, data = anorexia)
m.null <- lm(change ~ therapy, data = anorexia)
anova(m.null, m.full)
```

```
Analysis of Variance Table

Model 1: change ~ therapy
```

```
Model 2: change ~ Treat
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1     70 4105
2     69 3911  1       194 3.43  0.068 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note that the null model can be written as

$$E(Y_i) = \begin{cases} \beta_0, & \text{if the } i\text{-th observation is from the control group,} \\ \beta_0 + \beta_1, & \text{if the } i\text{-th observation is from the therapy group,} \end{cases}$$

or

$$E(Y_i) = \begin{cases} \beta_0, & \text{if the } i\text{-th observation is from the control group,} \\ \beta_0 + \beta_1, & \text{if the } i\text{-th observation is from the cognitive-behavioral therapy group,} \\ \beta_0 + \beta_1, & \text{if the } i\text{-th observations is from the family therapy group.} \end{cases}$$

So this model is effectively equivalent to the full model with $\beta_1 = \beta_2$.

The Trouble with ANOVA Tables

I **do not** recommended trying to produce tests by applying `anova` to a *single* model object. While it can produce desired tests *in some cases* and *if used correctly*, it often produces confusing results. For example, the following produces a test of the null hypothesis $H_0 : \beta_1 = 0, \beta_2 = 0$ for the `anorexia` model.

```
m <- lm(change ~ Treat, data = anorexia)
anova(m)
```

```
Analysis of Variance Table

Response: change
          Df Sum Sq Mean Sq F value Pr(>F)
Treat      2    615   307.3    5.42 0.0065 **
Residuals 69   3911    56.7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

But the tests shown here are maybe not what you think they are.

```
m <- lm(Gas ~ Insul + Temp + Insul:Temp, data = whiteside)
anova(m)
```

```
Analysis of Variance Table

Response: Gas
           Df Sum Sq Mean Sq F value  Pr(>F)
Insul       1   22.3    22.3   214.2 < 2e-16 ***
Temp        1   45.9    45.9   439.9 < 2e-16 ***
Insul:Temp  1    1.3     1.3    12.9 0.00073 ***
Residuals  52    5.4     0.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If you know what you are doing, there are alternatives to `anova` that are perhaps better (e.g., the `Anova` function from the **car** package), but there is often a more clear approach using two models in `anova`, using `contrast` or `lincon`, or using the **emmeans** package (which we will discuss later).

Note: Another potentially confusing test is one that appears at the bottom of `summary`. It tests the null hypothesis that all β_j (except β_0) equal zero. For the model for the `anorexia` data it is the same as the test conducted earlier.

```
m <- lm(change ~ Treat, data = anorexia)
summary(m)
```

```

Call:
lm(formula = change ~ Treat, data = anorexia)

Residuals:
   Min     1Q Median     3Q    Max 
-12.56  -4.54  -1.01   3.85  17.89 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)    -0.45       1.48   -0.30   0.7614   
TreatCBT        3.46       2.03    1.70   0.0936 . 
TreatFT         7.71       2.35    3.29   0.0016 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.53 on 69 degrees of freedom
Multiple R-squared:  0.136, Adjusted R-squared:  0.111 
F-statistic: 5.42 on 2 and 69 DF,  p-value: 0.0065
```

But for the model for the `whiteside` data the utility of this test is questionable.

```
m <- lm(Gas ~ Insul + Temp + Insul:Temp, data = whiteside)
summary(m)
```

```

Call:
lm(formula = Gas ~ Insul + Temp + Insul:Temp, data = whiteside)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.9780 -0.1801  0.0376  0.2093  0.6380 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)       6.8538     0.1360   50.41  < 2e-16 ***
InsulAfter       -2.1300     0.1801  -11.83  2.3e-16 ***
Temp             -0.3932     0.0225  -17.49  < 2e-16 ***
InsulAfter:Temp   0.1153     0.0321    3.59  0.00073 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.323 on 52 degrees of freedom
Multiple R-squared:  0.928, Adjusted R-squared:  0.924 
F-statistic:  222 on 3 and 52 DF,  p-value: <2e-16
```

Just because R gives you output does not mean it is useful!

Note: The `Residual standard error` shown by `summary` is the square root of the residual/error mean square (i.e., the square root of the ratio of the residual sum of squares to the residual degrees of freedom).

The degrees of freedom shown after `Residual standard error` is the residual degrees of freedom.