Nested model comparison

Peter Hoff

STAT 423

Applied Regression and Analysis of Variance

University of Washington
Main effects model
Interaction model
fit_int<-lm( salary ~ year + rank + year:rank, data=salary)

summary(fit_int)$coef

# (Intercept) 16416.5723 816.0186 20.1178895 1.967510e-24
# year 324.5027 141.9312 2.2863379 2.688729e-02
# rankAssoc 5354.2430 1492.5574 3.5872945 8.063338e-04
# rankProf 8176.4105 1418.1287 5.7656336 6.493300e-07
# year:rankAssoc -129.7345 205.7747 -0.6304686 5.315079e-01
# year:rankProf 151.1750 171.7437 0.8802364 3.833070e-01

Q: How can we test for interactions?
Multiparameter hypotheses

\[ E[\text{salary}|x_y, x_a, x_p] = \beta_0 + \beta_y x_y + \beta_a x_a + \beta_p x_p + \beta_{a:y} x_y x_a + \beta_{p:y} x_y x_p \]

**Test of interaction:**

\( H_0: (\beta_{a:y}, \beta_{p:y}) = (0, 0) \)

\( H_1: (\beta_{a:y}, \beta_{p:y}) \neq (0, 0) \)
Multiparameter hypotheses

\[ E[\text{salary}|x_y, x_a, x_p] = \beta_0 + \beta_y x_y + \beta_a x_a + \beta_p x_p + \beta_{a:y} x_y x_a + \beta_{p:y} x_y x_p \]

**Test of interaction:**

\[ H_0: (\beta_{a:y}, \beta_{p:y}) = (0, 0) \]
\[ H_1: (\beta_{a:y}, \beta_{p:y}) \neq (0, 0) \]

**Q:** How can we test two parameters simultaneously?
Old Faithful eruption data

fit<-lm(waiting~eruptions,data=faithful)
Old Faithful eruption data

fit<-lm(waiting~eruptions, data=faithful)
fit<-lm(waiting~eruptions,data=faithful)
plot(fit$res ~ faithful$eruptions) ; abline(h=0)
Polynomial regression

\[ y = \text{waiting} \]
Polynomial regression

\[ y = \text{waiting} \]
\[ x = \text{eruptions} \]
Polynomial regression

\[ y = \text{waiting} \]
\[ x = \text{eruptions} \]

Consider the following model:

\[ E[y|x] = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \]

- This model is \textit{nonlinear} in \( x \):
  - it is a polynomial.
Polynomial regression

\[ y = \text{waiting} \]
\[ x = \text{eruptions} \]

Consider the following model:

\[ E[y|x] = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \]

- This model is \textit{nonlinear} in \( x \):
  it is a polynomial.
- This model is \textit{linear} in \( \beta \):
  the mean is a linear combination of \( \beta \)-coefficients.
Polynomial regression

\( y = \text{waiting} \)
\( x = \text{eruptions} \)

Consider the following model:

\[
E[y|x] = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3
\]

- This model is \textit{nonlinear} in \( x \):
  - it is a polynomial.

- This model is \textit{linear} in \( \beta \):
  - the mean is a linear combination of \( \beta \)-coefficients.
Polynomial regression

\[ y = \text{waiting} \]
\[ x = \text{eruptions} \]

Consider the following model:

\[ E[y|x] = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \]

- This model is \textit{nonlinear} in \( x \):
  - it is a polynomial.
- This model is \textit{linear} in \( \beta \):
  - the mean is a linear combination of \( \beta \)-coefficients.

We can define \( x = (x_0, x_1, x_2, x_3) = (1, x, x^2, x^3) \).
Polynomial regression

\[ y = \text{waiting} \]
\[ x = \text{eruptions} \]

Consider the following model:

\[ E[y|x] = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \]

- This model is \textit{nonlinear} in \( x \):
  - it is a polynomial.
- This model is \textit{linear} in \( \beta \):
  - the mean is a linear combination of \( \beta \)-coefficients.

We can define \( x = (x_0, x_1, x_2, x_3) = (1, x, x^2, x^3) \).

Then the mean is in linear-model form: \( E[y|x] = \beta^T x \).
Polynomial regression

```r
fit3 <- lm(waiting ~ eruptions + I(eruptions^2) + I(eruptions^3), data=faithful)
fit3b <- lm(waiting ~ poly(eruptions, 3, raw=TRUE), data=faithful)
fit3c <- lm(waiting ~ poly(eruptions, 3), data=faithful)
```

```r
code
```
Polynomial regression

```r
fit3<-lm( waiting ~ eruptions + I(eruptions^2) + I(eruptions^3) ,data=faithful)
fit3b<-lm( waiting ~ poly(eruptions,3,raw=TRUE),data=faithful)
fit3c<-lm( waiting ~ poly(eruptions,3),data=faithful)

sum( fit3$res^2)
## [1] 8656.627

sum( fit3b$res^2)
## [1] 8656.627

sum( fit3c$res^2)
## [1] 8656.627
```
summary(fit3)$coef

|            | Estimate | Std. Error | t value | Pr(>|t|)  |
|------------|----------|------------|---------|----------|
| (Intercept)| 71.822814| 17.9066644 | 4.010954| 7.848652e-05 |
| eruptions  | -32.640220| 17.6875966 | -1.845373| 6.608630e-02 |
| I(eruptions^2)| 15.212251| 5.4134533 | 2.810083| 5.318008e-03 |
| I(eruptions^3)| -1.658674| 0.5269041 | -3.147962| 1.829923e-03 |

summary(fit3b)$coef

|            | Estimate | Std. Error | t value | Pr(>|t|)  |
|------------|----------|------------|---------|----------|
| (Intercept)| 71.822814| 17.9066644 | 4.010954| 7.848652e-05 |
| poly(eruptions, 3, raw = TRUE)1| -32.640220| 17.6875966 | -1.845373| 6.608630e-02 |
| poly(eruptions, 3, raw = TRUE)2| 15.212251| 5.4134533 | 2.810083| 5.318008e-03 |
| poly(eruptions, 3, raw = TRUE)3| -1.658674| 0.5269041 | -3.147962| 1.829923e-03 |

summary(fit3c)$coef

|            | Estimate | Std. Error | t value | Pr(>|t|)  |
|------------|----------|------------|---------|----------|
| (Intercept)| 70.89706 | 0.3446057  | 205.733831| 5.316699e-297 |
| poly(eruptions, 3)1| 201.60290| 5.6833834 | 35.472339| 3.103641e-103 |
| poly(eruptions, 3)2| -21.60253| 5.6833834 | -3.800998| 1.784403e-04 |
| poly(eruptions, 3)3| -17.89108| 5.6833834 | -3.147962| 1.829923e-03 |
Discuss: Model fit, prediction and extrapolation.
Discuss: Model fit, prediction and extrapolation.
Testing linearity in $x$

$E[y|x] = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

**Test of linearity in $x$:**

- $H_0$: $(\beta_2, \beta_3) = (0, 0)$
- $H_1$: $(\beta_2, \beta_3) \neq (0, 0)$
Testing linearity in $x$

\[ E[y|x] = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \]

**Test of linearity in $x$:**

- $H_0$: $(\beta_2, \beta_3) = (0, 0)$
- $H_1$: $(\beta_2, \beta_3) \neq (0, 0)$

**Q:** How can we test two parameters simultaneously?
ANNOVA for faithful data

```r
fit1 <- lm(waiting ~ eruptions, data = faithful)
fit3 <- lm(waiting ~ poly(eruptions, 3, raw = TRUE), data = faithful)

anova(fit1, fit3)
```

```r
## Analysis of Variance Table
## Model 1: waiting ~ eruptions
## Model 2: waiting ~ poly(eruptions, 3, raw = TRUE)
## Res.Df   RSS  Df  Sum of Sq     F  Pr(>F)
## 1   270 9443.4
## 2   268 8656.6  2  786.76 12.179 8.662e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The hypothesis \( H_0 \) of linearity in eruptions is strongly rejected.
ANOVA for faithful data

\[ \text{fit1} \leftarrow \text{lm}(\text{waiting} \sim \text{eruptions}, \text{data} = \text{faithful}) \]
\[ \text{fit3} \leftarrow \text{lm}(\text{waiting} \sim \text{poly}(\text{eruptions}, 3, \text{raw} = \text{TRUE}), \text{data} = \text{faithful}) \]

\[ \text{anova(fit, fit3)} \]

```
## Analysis of Variance Table
## Model 1: waiting ~ eruptions
## Model 2: waiting ~ poly(eruptions, 3, raw = TRUE)
## Res.Df RSS Df Sum of Sq F Pr(>F)
## 1 270 9443.4
## 2 268 8656.6 2 786.76 12.179 8.662e-06 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The hypothesis \( H_0 \) of linearity in eruptions is strongly rejected.
Main effects model for salary data
Interaction model for salary data

![Interaction model for salary data](image-url)
summary(lm(salary~year+rank+year:rank,data=salary))

## Call:
## lm(formula = salary ~ year + rank + year:rank, data = salary)
##
## Residuals:
## Min 1Q Median 3Q Max
## -3687.8 -1123.6 -392.1 720.9 9646.6
##
## Coefficients:
##                  Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 16416.6       816.0   20.118 < 2e-16 ***
## year         324.5        141.9    2.286  0.026887 *
## rankAssoc    5354.2       1492.6   3.587  0.000806 ***
## rankProf     8176.4       1418.1   5.766  6.49e-07 ***
## year:rankAssoc -129.7      205.8   -0.630  0.531508
## year:rankProf  151.2       171.7    0.880  0.383307
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2386 on 46 degrees of freedom
## Multiple R-squared:  0.8534, Adjusted R-squared:  0.8375
## F-statistic: 53.56 on 5 and 46 DF,  p-value: < 2.2e-16
ANOVA for salary data

```r
fit <- lm(salary ~ year + rank, data = salary)
fitint <- lm(salary ~ year + rank + year:rank, data = salary)

anova(fit, fitint)
```

```
## Analysis of Variance Table

## Model 1: salary ~ year + rank
## Model 2: salary ~ year + rank + year:rank

## Res.Df RSS Df Sum of Sq F Pr(>F)
## 1 48 276992734
## 2 46 261777280 2 15215454 1.3368 0.2727
```

The hypothesis $H_0$ of a constant year effect is not rejected.
F-statistic

The F-statistic calculated by the `anova` command is defined as follows:

- \( RSS_f, \nu_f \) = residual sum of squares and dof for “full” (larger) model;
- \( RSS_r, \nu_r \) = residual sum of squares and dof for “reduced” submodel.

\[
F = \frac{(RSS_r - RSS_f) / (\nu_r - \nu_f)}{RSS_f / \nu_f} = \frac{\Delta RSS / \Delta \nu}{\hat{\sigma}^2_f}
\]

**Interpretation**

- \( \Delta RSS \) = improvement in model fit going from reduced to full
- \( \Delta \nu \) = change in number of parameters going from reduced to full
The $F$-statistic calculated by the `anova` command is defined as follows:

- $RSS_f, \nu_f =$ residual sum of squares and dof for “full” (larger) model;
- $RSS_r, \nu_r =$ residual sum of squares and dof for “reduced” submodel.

$$F = \frac{(RSS_r - RSS_f)/(\nu_r - \nu_f)}{RSS_f/\nu_f} = \frac{\Delta RSS}{\Delta \nu} \frac{\hat{\sigma}^2_f}{\hat{\sigma}^2_f}$$

**Interpretation**

- $\Delta RSS =$ improvement in model fit going from reduced to full
- $\Delta \nu =$ change in number of parameters going from reduced to full

The $F$-statistic is large, and the reduced model is rejected, if

*the improvement in fit per parameter added is large compared to the estimated error variance.*
**F-statistic**

The F-statistic calculated by the `anova` command is defined as follows:
- $RSS_f$, $\nu_f$ = residual sum of squares and dof for “full” (larger) model;
- $RSS_r$, $\nu_r$ = residual sum of squares and dof for “reduced” submodel.

$$F = \frac{(RSS_r - RSS_f)/(\nu_r - \nu_f)}{RSS_f/\nu_f} = \frac{\Delta RSS}{\Delta \nu} \hat{\sigma}_f^2$$

**Interpretation**
- $\Delta RSS$ = improvement in model fit going from reduced to full
- $\Delta \nu$ = change in number of parameters going from reduced to full

The F-statistic is large, and the reduced model is rejected, if

the improvement in fit per parameter added is large compared to the estimated error variance.

We will derive the null distribution for $F$ on the board.
Simulation study

```r
n<-50
x1<-rnorm(n) ; x2<-cbind( rbinom(n,1,.5) , rbinom(n,1,.5) )
b0<-1 ; b1<-.4 ; b2<-0 ; b3<-0
y<-b0 + b1*x1 + b2*x2[,1] + b3*x2[,2] + rnorm(n)
fit<-lm( y~x1+x2)
anova(fit)

## Analysis of Variance Table
## Response: y
## Df Sum Sq Mean Sq F value Pr(>F)
## x1  1  8.534  8.5343  7.2259 0.009972 **
## x2  2  0.109  0.0545  0.0461 0.954965
## Residuals 46 54.329  1.1811
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(fit)[2,4]
## [1] 0.04612703
```
F.sim<-NULL
for(i in 1:2000)
{
  y<-b0 + b1*x1 + b2*x2[,1] + b3*x2[,2] + rnorm(n)
  fit<-lm( y~x1+x2)
  F.sim<-c(F.sim, anova(fit)[2,4])
}

Density
0 2 4 6 8 10
0.0 0.2 0.4 0.6

F.sim

Density
F.sim
0 2 4 6 8 10
0.0 0.2 0.4 0.6
Importance of DOF

Under $H_0$,

- $E[RSS_0 - RSS_1/\Delta p] = \sigma^2$
- $E[RSS_1/(n - p)] = \sigma^2$

so it seems that under $H_0$, $F \approx 1$. 

Why would we only reject in this case if $F_{\text{obs}}$ is much larger than 1?
Importance of DOF

Under $H_0$,

- $\mathbb{E}[RSS_0 - RSS_1 / \Delta p] = \sigma^2$
- $\mathbb{E}[RSS_1 / (n - p)] = \sigma^2$

so it seems that under $H_0$, $F \approx 1$.

Why would we only reject in this case if $F_{\text{obs}}$ is much larger than 1?
F.sim<-NULL
for(i in 1:2000)
{
y<-b0 + b1*x1 + b2*x2[,1] + b3*x2[,2] + rnorm(n)
afit<-anova(lm( y~x1+x2))
F.sim<-rbind(F.sim, c( afit[2,3],afit[3,3],afit[2,4] ) )
}
Importance of DOF

If $\Delta p$ is small, then $RSS_0 - RSS_1 / \Delta p$ is highly variable around $\sigma^2$. 

Critical value can be quite large.

If $\Delta p$ is large, then $RSS_0 - RSS_1 / \Delta p$ is less variable around $\sigma^2$.

Critical value is closer to 1.

Remember: For $n \gg \Delta p$, the critical value of the $F$-test is highly dependent on the numerator dof.
Importance of DOF

If $\Delta p$ is small, then $RSS_0 - RSS_1/\Delta p$ is highly variable around $\sigma^2$. Critical value can be quite large.
Importance of DOF

If $\Delta p$ is small, then $RSS_0 - RSS_1 / \Delta p$ is highly variable around $\sigma^2$. Critical value can be quite large.

If $\Delta p$ is large, then $RSS_0 - RSS_1 / \Delta p$ is less variable around $\sigma^2$. 
Importance of DOF

If $\Delta p$ is small, then $RSS_0 - RSS_1/\Delta p$ is highly variable around $\sigma^2$. Critical value can be quite large.

If $\Delta p$ is large, then $RSS_0 - RSS_1/\Delta p$ is less variable around $\sigma^2$. Critical value is closer to 1.
Importance of DOF

If $\Delta p$ is small, then $RSS_0 - RSS_1 / \Delta p$ is highly variable around $\sigma^2$. Critical value can be quite large.

If $\Delta p$ is large, then $RSS_0 - RSS_1 / \Delta p$ is less variable around $\sigma^2$. Critical value is closer to 1.

**Remember:** For $n \gg \Delta p$, the critical value of the $F$-test is highly dependent on the numerator dof.
p2<-8
x2<-matrix(rbinom(n*(p2-1),1,.5),n,p2-1)

F.sim<-NULL
for(i in 1:2000)
{
y<-b0 + b1*x1 + rnorm(n)
afit<-anova(lm( y~x1+x2))
F.sim<-rbind(F.sim, c( afit[2,3],afit[3,3],afit[2,4] ) )
}

```r
# F.sim[, 1]
Density
0 1 2 3 4
0.0 0.2 0.4 0.6 0.8

# F.sim[, 2]
Density
0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8
0.0 0.5 1.0 1.5

# F.sim
Density
0 1 2 3 4 5
0.0 0.2 0.4 0.6 0.8
```
Reconsider a model for salary: \( \text{salary} \sim \text{year} + \text{sex} \)
Reconsider a model for salary: \( \text{salary} \sim \text{year} + \text{sex} \)

Q: How can we evaluate the effect of sex?
Reconsider a model for salary: salary $\sim$ year + sex

Q: How can we evaluate the effect of sex?

- fit salary $\sim$ year
- fit salary $\sim$ year + sex
- compare models with $F$-test
**Reconsider a model for salary:** \( \text{salary} \sim \text{year} + \text{sex} \)

**Q:** How can we evaluate the effect of sex?

- fit \( \text{salary} \sim \text{year} \)
- fit \( \text{salary} \sim \text{year} + \text{sex} \)
- compare models with \( F \)-test

- fit \( \text{salary} \sim \text{year} + \text{sex} \)
- do a \( t \)-test on the coefficient for sex
Reconsider a model for salary: \( \text{salary} \sim \text{year} + \text{sex} \)

Q: How can we evaluate the effect of sex?

- fit \( \text{salary} \sim \text{year} \)
- fit \( \text{salary} \sim \text{year} + \text{sex} \)
- compare models with \( F \)-test

- fit \( \text{salary} \sim \text{year} + \text{sex} \)
- do a \( t \)-test on the coefficient for sex

Could we get two different results?
summary(lm(salary ~ year + sex, data=salary))$coef

# Analysis of Variance Table

anova(fit0, fit1)

.1384514^2

## [1] 0.01916879