

# Multiple Linear Regression: Global tests and Multiple Testing

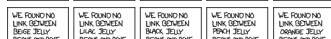
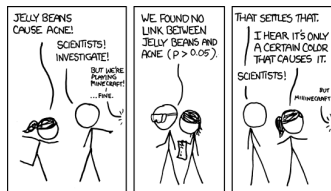
Author: Nicholas G Reich, Jeff Goldsmith

*This material is part of the **statsTeachR** project*

*Made available under the Creative Commons Attribution-ShareAlike 3.0 Unported License: [http://creativecommons.org/licenses/by-sa/3.0/deed.en\\_US](http://creativecommons.org/licenses/by-sa/3.0/deed.en_US)*

# Today's Lecture

- Multiple testing - preserving your Type I error rate.



## Inference about multiple coefficients

Our model contains multiple parameters; often we want ask a question about multiple coefficients simultaneously. I.e. “are any of these  $k$  coefficients significantly different from 0?” This is equivalent to performing multiple tests (or looking at confidence ellipses):

$$H_{01} : \beta_1 = 0$$

$$H_{02} : \beta_2 = 0$$

$$\vdots = \vdots$$

$$H_{0k} : \beta_k = 0$$

where each test has a size of  $\alpha$  *+ type I error*

- For any individual test,  $P(\text{reject } H_{0i} | H_{0i}) = \alpha$

## Inference about multiple coefficients

For any individual test,  $P(\text{reject } H_{0i} | H_{0i}) = \alpha$ .

But it DOES NOT FOLLOW that

$$P(\text{reject at least one } H_{0i} | \text{all } H_{0i} \text{ are true}) \neq \alpha.$$

This is called the Family-wise error rate (FWER). Ignore it at your own peril!

## Family-wise error rate

To calculate the FWER

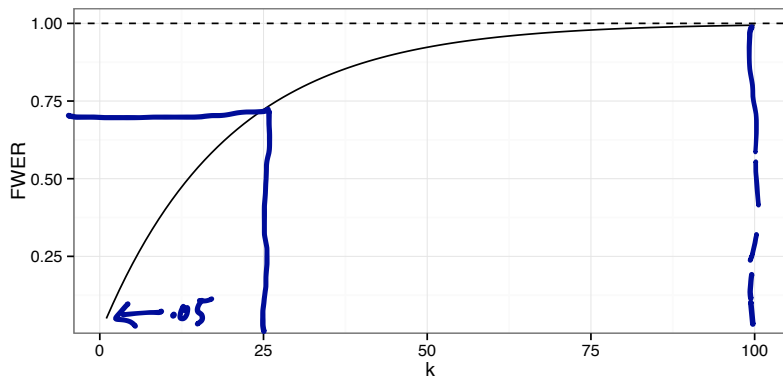
- First note  $P(\text{no rejections} | \text{all } H_{0i} \text{ are true}) = (1 - \alpha)^k$
- It follows that

$$\begin{aligned}\text{FWER} &= P(\text{at least one rejection} | \text{all } H_{0i} \text{ are true}) \\ &= 1 - (1 - \alpha)^k\end{aligned}$$

# Family-wise error rate


$$\text{FWER} = 1 - (1 - \alpha)^k$$

```
alpha <- .05  
k <- 1:100  
FWER <- 1-(1-alpha)^k  
qplot(k, FWER, geom="line") + geom_hline(yintercept = 1, lty=2)
```



# Addressing multiple comparisons

## Three general approaches

- Do nothing in a reasonable way
    - ▶ Don't trust scientifically implausible results
    - ▶ Don't over-emphasize isolated findings
  - Correct for multiple comparisons
    - ▶ Often, use the Bonferroni correction and use  $\alpha_i = \alpha/k$  for each test
    - ▶ Thanks to the Bonferroni inequality, this gives an overall  $FWER \leq \alpha$
  - Use a global test
- 

## Global tests

Compare a smaller “null” model to a larger “alternative” model

- Smaller model must be nested in the larger model
- That is, the smaller model must be a special case of the larger model
- For both models, the  $RSS$  gives a general idea about how well the model is fitting
- In particular, something like

$$\frac{RSS_S - RSS_L}{RSS_L}$$

Handwritten annotations: "Small" with an arrow pointing to  $RSS_S$  and "Large" with an arrow pointing to  $RSS_L$ .

compares the relative  $RSS$  of the models



## Nested models

- These models are nested:

Smaller = Regression of  $Y$  on  $X_1$

Larger = Regression of  $Y$  on  $X_1, X_2, X_3, X_4$

- These models are not:

Smaller = Regression of  $Y$  on  $X_2$

Larger = Regression of  $Y$  on  $X_1, X_3$

## Global $F$ tests

- Compute the test statistic

$$F_{obs} = \frac{(RSS_S - RSS_L) / (df_S - df_L)}{RSS_L / df_L}$$

- If  $H_0$  (the null model) is true, then  $F_{obs} \sim F_{df_S - df_L, df_L}$
- Note  $df_S = n - p_S - 1$  and  $df_L = n - p_L - 1$
- We reject the null hypothesis if the p-value is above  $\alpha$ , where

$$\text{p-value} = P(F_{df_S - df_L, df_L} > F_{obs})$$

especially valuable for testing significance of categorical/dummy variables.

# Global $F$ tests

There are a couple of important special cases for the  $F$  test

- The null model contains the intercept only
  - ▶ When people say ANOVA, this is often what they mean (although all  $F$  tests are based on an analysis of variance)
- The null model and the alternative model differ only by one term
  - ▶ Gives a way of testing for a single coefficient
  - ▶ Turns out to be equivalent to a two-sided  $t$ -test:  $t_{df_L}^2 \sim F_{1,df_L}$

# Lung data: multiple coefficients simultaneously

You can test multiple coefficients simultaneously using the  $F$  test

```
mlr_null <- lm(disease ~ nutrition, data=dat)
mlr1 <- lm(disease ~ nutrition+ airqual + crowding + smoking, data=dat)
anova(mlr_null, mlr1)
```

## Analysis of Variance Table

##

## Model 1: disease ~ nutrition

## Model 2: disease ~ nutrition + airqual + crowding + smoking

##	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
## 1	97	9192.7				
## 2	94	1248.0	3	7944.7	199.47	< 2.2e-16 ***

## ---

## Signif. codes:

## 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

This test shows that airqual, crowding, and smoking together significantly improve the fit of our model (assuming model diagnostics look good). Further analyses may be warranted to determine which, if any, coefficients are not different from 0.

## Lung data: single coefficient test

The  $F$  test is equivalent to the  $t$  test when there's only one parameter of interest

```
mlr_null <- lm(disease ~ nutrition, data=dat)
mlr1 <- lm(disease ~ nutrition + airqual, data=dat)
anova(mlr_null, mlr1)
```

```
## Analysis of Variance Table
##
## Model 1: disease ~ nutrition
## Model 2: disease ~ nutrition + airqual
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      97 9192.7
## 2      96 5969.5  1   3223.1 51.833 1.347e-10 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(mlr1)$coef
```

```
##           Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 37.62538251  2.43946243  15.423637 9.946294e-28
## nutrition   -0.03469855  0.01692446  -2.050202 4.307101e-02
## airqual      0.36114435  0.05016218   7.199535 1.346935e-10
```

# Today's Big Ideas

$F$  tests can control for multiple comparisons!

- hands-on example