



REGRESSION METHODS

MULTIPLE COMPARISONS



ANOVA: One-Way Model

- What are the groups with differences in means?

MULTIPLE COMPARISONS:

$$\left. \begin{array}{l} \mu_0 = \mu_1? \\ \mu_0 = \mu_2? \\ \mu_1 = \mu_2? \end{array} \right\} \text{Pairwise comparisons}$$

$(\mu_1 + \mu_2)/2 = \mu_0?$ ———> Non-pairwise comparison



Multiple Comparisons: Family-wise error rates

- Illustrating the multiple comparison problem
 - Truth: null hypotheses
 - Tests: pairwise comparisons - each at the 5% level.

What is the probability of rejecting at least one?

#groups = K	2	3	4	5	6	7	8	9	10
#pairwise comparisons $C = K(K-1)/2$	1	3	6	10	15	21	28	36	45
P(at least one sig) $= 1 - (1 - 0.05)^C$	0.05	0.143	0.265	0.401	0.537	0.659	0.762	0.842	0.901

That is, if you have three groups and make pairwise comparisons, each at the 5% level, your family-wise error rate (probability of making at least one false rejection) is over 14%!

Need to address this issue!

Several methods!!!



Multiple Comparisons

- Several methods:
 - None (no adjustment)
 - Bonferroni
 - Holm
 - Hochberg
 - Hommel
 - BH
 - BY
 - FDR
 - ...
- Available in R



Multiple Comparisons

- **Bonferroni** adjustment: for C tests performed, use level α/C (or multiply p-values by C).
 - Simple
 - Conservative
 - Must decide on number of tests beforehand
 - Widely applicable
 - Can be done without software!



Multiple Comparisons

- FDR (False Discovery Rate)
 - Less conservative procedure for multiple comparisons
 - Among rejected hypotheses, FDR controls the expected proportion of incorrectly rejected null hypotheses (that is, type I errors).

Multiple Comparisons

This option considers all pairwise comparisons

```
> ## call library for multiple comparisons
> library(multcomp)
>
> ## fit model
> fit1 = lm(chol ~ -1 + factor(rs174548))
>
> ## all pairwise comparisons
> ## -- first, define matrix of contrasts
> M = contrMat(table(rs174548), type="Tukey")
> M
```

Multiple Comparisons of Means: Tukey Contrasts

	0	1	2
1	-0	-1	1
2	-0	-1	0
2	-1	0	-1

```
>
> ## -- second, obtain estimates for multiple comparisons
> mc = glht(fit1, linfct = M)
```

Stands for general linear hypothesis testing



Multiple Comparisons

```
> ## -- third, adjust the p-values (or not) for multiple comparisons
> summary(mc, test=adjusted("none"))
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: `lm(formula = chol ~ -1 + factor(rs174548))`

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)	
1 - 0 == 0	6.802	2.321	2.930	0.00358	**
2 - 0 == 0	5.438	4.540	1.198	0.23167	
2 - 1 == 0	-1.364	4.665	-0.292	0.77015	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- none method)



Multiple Comparisons

```
> summary(mc, test=adjusted("bonferroni"))
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: `lm(formula = chol ~ -1 + factor(rs174548))`

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)	
1 - 0 == 0	6.802	2.321	2.930	0.0107	*
2 - 0 == 0	5.438	4.540	1.198	0.6950	
2 - 1 == 0	-1.364	4.665	-0.292	1.0000	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- bonferroni method)



Multiple Comparisons

```
> summary(mc, test=adjusted("fdr"))
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: `lm(formula = chol ~ -1 + factor(rs174548))`

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)	
1 - 0 == 0	6.802	2.321	2.930	0.0107	*
2 - 0 == 0	5.438	4.540	1.198	0.3475	
2 - 1 == 0	-1.364	4.665	-0.292	0.7702	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- fdr method)



Multiple Comparisons

- What about using other adjustment methods?
 - For example, we used:
 - > `summary(mc, test=adjusted("bonferroni"))`
(all pairwise comparisons, with Bonferroni adjustment)
 - > `summary(mc, test=adjusted("fdr"))`
(all pairwise comparisons, with FDR adjustment)
 - Other options are:
 - `summary(mc, test=adjusted("holm"))`
 - `summary(mc, test=adjusted("hochberg"))`
 - `summary(mc, test=adjusted("hommel"))`
 - `summary(mc, test=adjusted("BH"))`
 - `summary(mc, test=adjusted("BY"))`

Results, in this particular example, are basically the same, but they don't need to be! Different criteria could lead to different results!

Summary:

GOAL: Comparison of means across K groups

Relationships:

$$\mu_0 = \beta_0$$

$$\mu_1 = \beta_0 + \beta_1$$

$$\mu_2 = \beta_0 + \beta_2$$

...

$$\mu_{K-1} = \beta_0 + \beta_{K-1}$$

One-way ANOVA:

$$H_0: \mu_0 = \mu_1 = \dots = \mu_{K-1}$$

H_1 : not all means are equal

Multiple Regression:

Model: $E[Y|\text{groups}] = \beta_0 + \beta_1 \text{group}_2 + \dots + \beta_{K-1} \text{group}_K$
where group_1 is the reference group

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{K-1} = 0$$

H_1 : not all β_i are equal to zero

Rejected H_0 ?

YES

Multiple Comparisons
(control α overall)

e.g. Bonferroni: $\alpha/\#\text{comparisons}$



REGRESSION METHODS

Two-way ANOVA models



ANOVA: Two-Way Model

Motivation:

- Scientific question:
 - Assess the effect of rs174548 and sex on cholesterol levels.



ANOVA: Two-Way Model

- Factors: A and B
- Goals:
 - Test for main effect of A
 - Test for main effect of B
 - Test for interaction effect of A and B

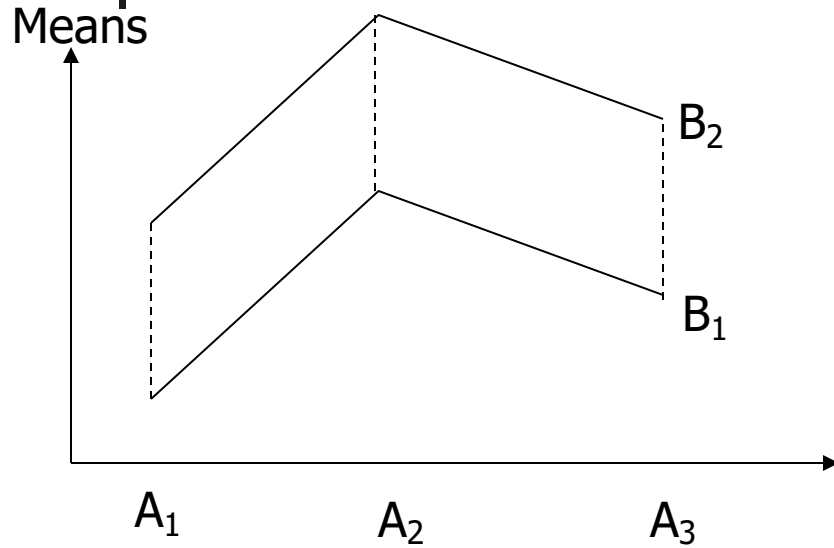


ANOVA: Two-Way Model

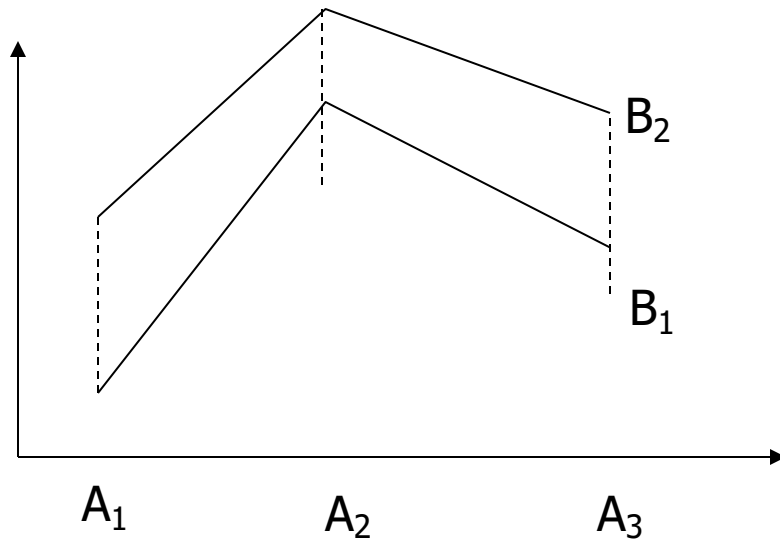
- To simplify discussion, assume that factor A has three levels, while factor B has two levels

		Factor A		
		A_1	A_2	A_3
Factor B	B_1	μ_{11}	μ_{21}	μ_{31}
	B_2	μ_{12}	μ_{22}	μ_{32}

ANOVA: Two-Way Model



Parallel lines = No interaction



Lines are not parallel = Interaction



ANOVA: Two-Way Model

- Recall:

- Categorical variables can be represented with “dummy” variables
- Interactions are represented with “cross-products”



ANOVA: Two-Way Model

- Model 1:

$$E[Y|A_2, A_3, B_2] = \beta_0 + \beta_1 A_2 + \beta_2 A_3 + \beta_3 B_2.$$

- What are the means in each combination-group?

	A_1	A_2	A_3
B_1	$\mu_{11} = \beta_0$	$\mu_{21} = \beta_0 + \beta_1$	$\mu_{31} = \beta_0 + \beta_2$
B_2	$\mu_{12} = \beta_0 + \beta_3$	$\mu_{22} = \beta_0 + \beta_1 + \beta_3$	$\mu_{32} = \beta_0 + \beta_2 + \beta_3$



ANOVA: Two-Way Model

- Model 1:

$$E[Y|A_2, A_3, B_2] = \beta_0 + \beta_1 A_2 + \beta_2 A_3 + \beta_3 B_2.$$

	A_1	A_2	A_3
B_1	$\mu_{11} = \beta_0$	$\mu_{21} = \beta_0 + \beta_1$	$\mu_{31} = \beta_0 + \beta_2$
B_2	$\mu_{12} = \beta_0 + \beta_3$	$\mu_{22} = \beta_0 + \beta_1 + \beta_3$	$\mu_{32} = \beta_0 + \beta_2 + \beta_3$

Model with no interaction:

- Difference in means between groups defined by factor B does not depend on the level of factor A.
- Difference in means between groups defined by factor A does not depend on the level of factor B.



ANOVA: Two-Way Model

- Model 2:

$$E[Y|A_2, A_3, B_2] = \beta_0 + \beta_1 A_2 + \beta_2 A_3 + \beta_3 B_2 + \beta_4 A_2 B_2 + \beta_5 A_3 B_2$$

- What are the means in each combination-group?

	A ₁	A ₂	A ₃
B ₁	$\mu_{11} = \beta_0$	$\mu_{21} = \beta_0 + \beta_1$	$\mu_{31} = \beta_0 + \beta_2$
B ₂	$\mu_{12} = \beta_0 + \beta_3$	$\mu_{22} = \beta_0 + \beta_1 + \beta_3 + \beta_4$	$\mu_{32} = \beta_0 + \beta_2 + \beta_3 + \beta_5$



ANOVA: Two-Way Model

- Three (possible) tests
 - Interaction of A and B (may want to start here)
 - Rejection would imply that differences between means of A depends on the level of B (and vice-versa) so stop
 - Main effect of A
 - Test only if no interaction
 - Main effect of B
 - Test only if no interaction

[Note: If you have one observation per cell, you cannot test interaction!]



ANOVA: Two-Way Model

- Model without interaction

$$E[Y|A_2, A_3, B_2] = \beta_0 + \beta_1 A_2 + \beta_2 A_3 + \beta_3 B_2.$$

How do we test for main effect of factor A?

$$H_0: \beta_1 = \beta_2 = 0 \quad \text{vs.} \quad H_1: \beta_1 \text{ or } \beta_2 \text{ not zero}$$

How do we test for main effect of factor B?

$$H_0: \beta_3 = 0 \quad \text{vs.} \quad H_1: \beta_3 \text{ not zero}$$



ANOVA: Two-Way Model

- Model with interaction:

$$E[Y|A_2, A_3, B_2] = \beta_0 + \beta_1 A_2 + \beta_2 A_3 + \beta_3 B_2 + \beta_4 A_2 B_2 + \beta_5 A_3 B_2$$

How do we test for interactions?

$$\begin{cases} H_0: \beta_4 = \beta_5 = 0 & \text{vs.} \\ H_1: \beta_4 \text{ or } \beta_5 \text{ not zero} \end{cases}$$

IMPORTANT:

If you reject the null, do not test main effects!!!

ANOVA: Two-Way Model (without interaction)

```
> fit1 = lm(chol ~ factor(sex) + factor(rs174548))
> summary(fit1)
Call:
lm(formula = chol ~ factor(sex) + factor(rs174548))

Residuals:
    Min       1Q   Median       3Q      Max
-66.6534 -14.4633  -0.6008  15.4450  57.6350

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      175.365      1.786   98.208 < 2e-16 ***
factor(sex)1       11.053       2.126    5.199 3.22e-07 ***
factor(rs174548)1    7.236       2.250    3.215 0.00141 **
factor(rs174548)2    5.184       4.398    1.179 0.23928
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.24 on 396 degrees of freedom
Multiple R-squared:  0.08458,    Adjusted R-squared:  0.07764
F-statistic: 12.2 on 3 and 396 DF,  p-value: 1.196e-07

> fit0 = lm(chol ~ factor(sex))
> anova(fit0, fit1)
Analysis of Variance Table

Model 1: chol ~ factor(sex)
Model 2: chol ~ factor(sex) + factor(rs174548)
  Res.Df  RSS Df Sum of Sq  F    Pr(>F)
1     398 183480
2     396 178681   2    4799.1 5.318 0.005259 **
```

ANOVA: Two-Way Model (without interaction)

```
> fit1 = lm(chol ~ factor(sex) + factor(rs174548))
> summary(fit1)
Call:
lm(formula = chol ~ factor(sex) + factor(rs174548))

Residuals:
    Min       1Q   Median       3Q      Max
-66.6534 -14.4633  -0.6008  15.4450  57.6350

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    175.365      1.786   98.208 < 2e-16 ***
factor(sex)1     11.053      2.126    5.199 3.22e-07 ***
factor(rs174548)1  7.236      2.250    3.215 0.00141 **
factor(rs174548)2  5.184      4.398    1.179 0.23928

Residual standard error: 21.24 on 396 degrees of freedom
Multiple R-squared:  0.08458,    Adjusted R-squared:  0.07764
F-statistic: 12.2 on 3 and 396 DF,  p-value: 1.196e-07

> anova(fit0, fit1)
Analysis of Variance Table

Model 1: chol ~ factor(sex)
Model 2: chol ~ factor(sex) + factor(rs174548)
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1     398 183480
2     396 178681  2    4799.1 5.318 0.005259 **
```

■ Interpretation of results:

- Estimated mean cholesterol for male C/C group: 175.37 mg/dl
- Estimated difference in mean cholesterol levels between females and males adjusted by genotype: 11.053 mg/dl
- Estimated difference in mean cholesterol levels between C/G and C/C groups adjusted by sex: 7.236 mg/dl
- Estimated difference in mean cholesterol levels between G/G and C/C groups adjusted by sex: 5.184 mg/dl
- There is evidence that cholesterol is associated with sex ($p < 0.001$).
- There is evidence that cholesterol is associated with genotype ($p = 0.005$)



ANOVA: Two-Way Model (without interaction)

- In words:
 - Adjusting for sex, the difference in mean cholesterol comparing C/G to C/C is 7.236 and comparing G/G to C/C is 5.184.
 - This difference does not depend on sex
 - (this is because the model does not have an interaction between sex and genotype!)

ANOVA: Two-Way Model (with interaction)

```
> fit2 = lm(chol ~ factor(sex) * factor(rs174548))
> summary(fit2)
```

Call:

```
lm(formula = chol ~ factor(sex) * factor(rs174548))
```

Residuals:

Min	1Q	Median	3Q	Max
-70.5286	-13.6037	-0.9736	14.1709	54.8818

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	178.1182	2.0089	88.666	< 2e-16 ***
factor(sex) 1	5.7109	2.7982	2.041	0.04192 *
factor(rs174548) 1	0.9597	3.1306	0.307	0.75933
factor(rs174548) 2	-0.2015	6.4053	-0.031	0.97492
factor(sex) 1:factor(rs174548) 1	12.7398	4.4650	2.853	0.00456 **
factor(sex) 1:factor(rs174548) 2	10.2296	8.7482	1.169	0.24297

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.07 on 394 degrees of freedom

Multiple R-squared: 0.1039, Adjusted R-squared: 0.09257

F-statistic: 9.14 on 5 and 394 DF, p-value: 3.062e-08



ANOVA: Model comparison

```
> anova(fit1,fit2)
```

```
Analysis of Variance Table
```

```
Model 1: chol ~ factor(sex) + factor(rs174548)
```

```
Model 2: chol ~ factor(sex) * factor(rs174548)
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	396	178681				
2	394	174902	2	3779	4.2564	0.01483 *

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA: Two-Way Model (with interaction)

```
> fit2 = lm(chol ~ factor(sex) * factor(rs174548))
> summary(fit2)
```

Call:

```
lm(formula = chol ~ factor(sex) * factor(rs174548))
```

Residuals:

Min	1Q	Median	3Q	Max
-70.5286	-13.6037	-0.9736	14.1709	54.8818

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	178.1182	2.0089	88.666	< 2e-16 ***
factor(sex)1	5.7109	2.7982	2.041	0.04192 *
factor(rs174548)1	0.9597	3.1306	0.307	0.75933
factor(rs174548)2	-0.2015	6.4053	-0.031	0.97492
factor(sex)1:factor(rs174548)1	12.7398	4.4650	2.853	0.00456 **
factor(sex)1:factor(rs174548)2	10.2296	8.7482	1.169	0.24297

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.07 on 394 degrees of freedom
Multiple R-squared: 0.1039, Adjusted R-squared: 0.09257
F-statistic: 9.14 on 5 and 394 DF, p-value: 3.062e-08

```
> anova(fit1, fit2)
```

Analysis of Variance Table

Model 1: chol ~ factor(sex) + factor(rs174548)						
Model 2: chol ~ factor(sex) * factor(rs174548)						
	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	396	178681				
2	394	174902	2	3779	4.2564	0.01483 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

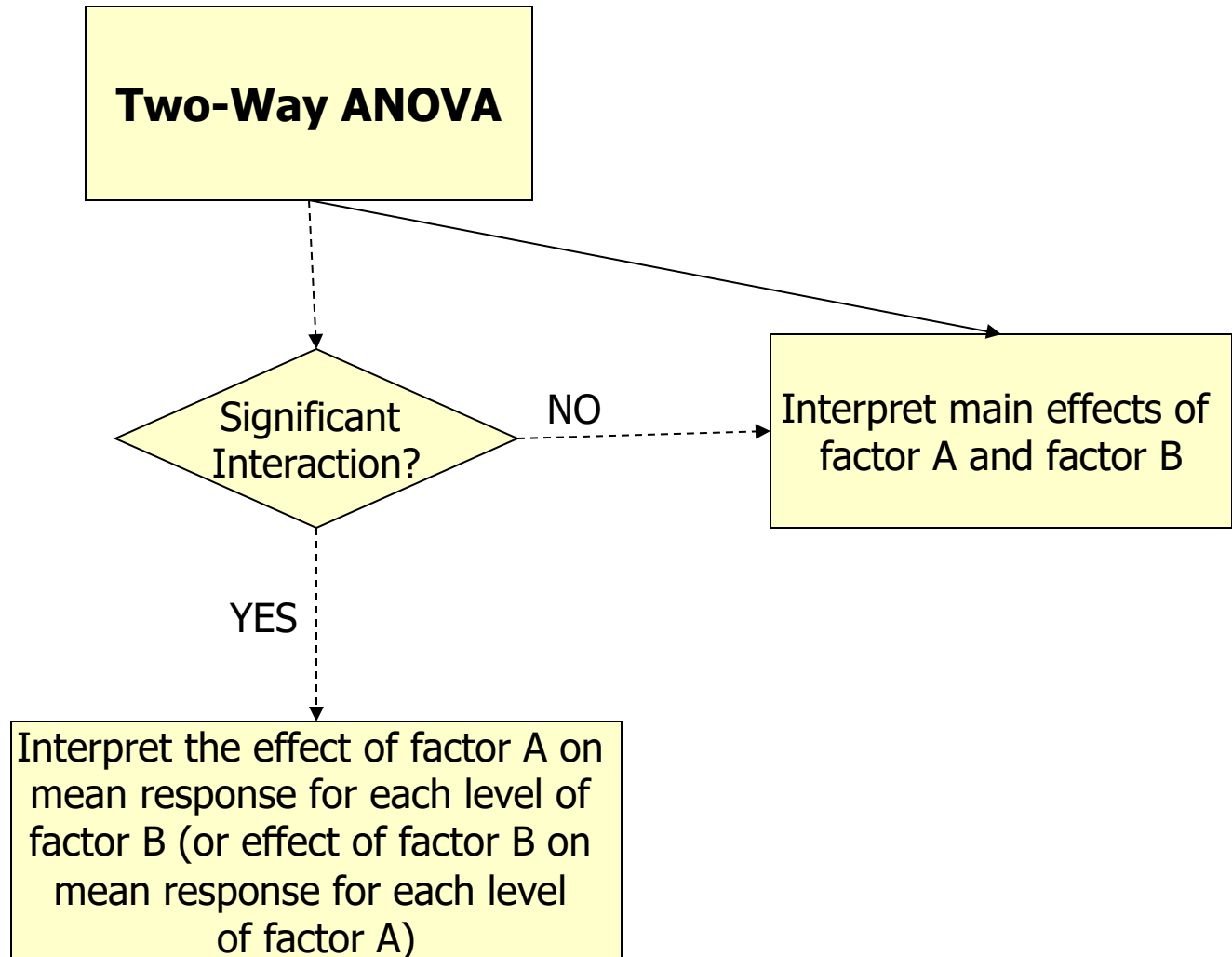
■ Interpretation of results:

- Estimated mean cholesterol for male C/C group: 178.12 mg/dl
- Estimated mean cholesterol for female C/C group? (178.12 + 5.7109) mg/dl
- Estimated mean cholesterol for male C/G group: (178.12 + 0.9597) mg/dl
- Estimated mean cholesterol for female C/G group: (178.12 + 5.7109 + 0.9597 + 12.7398) mg/dl
- ...

- There is evidence for an interaction between sex and genotype (p= 0.015)



SUMMARY:





ANalysis of COVAriance Models (ANCOVA)

Motivation:

- Scientific question:
 - Assess the effect of rs174548 on cholesterol levels adjusting for age



ANalysis of COVAriance Models (ANCOVA)

- ANOVA with one or more continuous variables
 - Equivalent to regression with “dummy” variables and continuous variables
 - Primary comparison of interest is across k groups defined by a categorical variable, but the k groups may differ on some other potential predictor or confounder variables [also called covariates].



ANalysis of COVAriance Models (ANCOVA)

- To facilitate discussion assume
 - Y: continuous response (e.g. cholesterol)
 - X: continuous variable (e.g. age)
 - Z: dummy variable (e.g. indicator of C/G or G/G versus C/C)
- Model: $Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon$

Interaction term

Note that:

$$Z = 0 \Rightarrow E[Y | X, Z = 0] = \beta_0 + \beta_1 X$$

$$Z = 1 \Rightarrow E[Y | X, Z = 1] = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X$$

This model allows for different intercepts/slopes for each group.



ANCOVA

- Testing coincident lines: $H_0 : \beta_2 = 0, \beta_3 = 0$

- Compares overall model with reduced model

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Testing parallelism: $H_0 : \beta_3 = 0$

- Compares overall model with reduced model

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon$$

ANCOVA

```
> fit0 = lm(chol ~ factor(rs174548))
> summary(fit0)
Call:
lm(formula = chol ~ factor(rs174548))

Residuals:
      Min       1Q   Median       3Q      Max
-64.06167 -15.91338  -0.06167  14.93833  59.13605

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      181.062      1.455 124.411  < 2e-16 ***
factor(rs174548)1       6.802      2.321   2.930  0.00358 **
factor(rs174548)2       5.438      4.540   1.198  0.23167
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.93 on 397 degrees of freedom
Multiple R-squared:  0.0221,    Adjusted R-squared:  0.01718
F-statistic: 4.487 on 2 and 397 DF,  p-value: 0.01184

> anova(fit0)
Analysis of Variance Table

Response: chol
              Df Sum Sq Mean Sq F value    Pr(>F)
factor(rs174548)    2   4314    2157   4.4865 0.01184 *
Residuals          397 190875     481
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANCOVA

```
> fit1 = lm(chol ~ factor(rs174548) + age)
> summary(fit1)
Call:
lm(formula = chol ~ factor(rs174548) + age)

Residuals:
    Min       1Q   Median       3Q      Max
-57.2089 -14.4293  0.4443  14.2652  55.8985

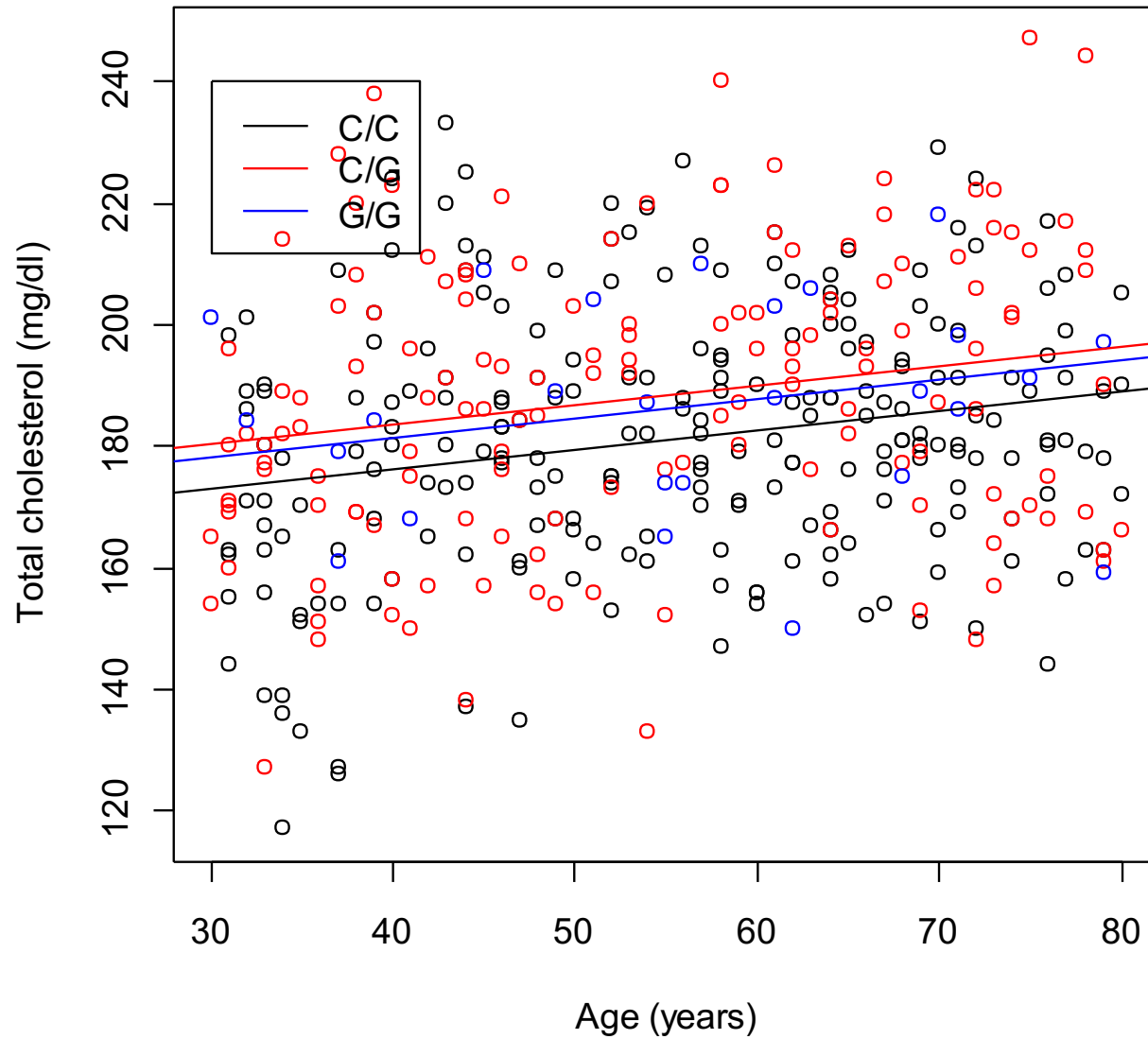
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      163.28125    4.36422   37.414 < 2e-16 ***
factor(rs174548)1    7.30137    2.27457    3.210 0.00144 **
factor(rs174548)2    5.08431    4.44331    1.144 0.25321
age                0.32140    0.07457    4.310 2.06e-05 ***

Residual standard error: 21.46 on 396 degrees of freedom
Multiple R-squared: 0.06592,    Adjusted R-squared: 0.05884
F-statistic: 9.316 on 3 and 396 DF,  p-value: 5.778e-06

> anova(fit0, fit1)
Analysis of Variance Table

Model 1: chol ~ factor(rs174548)
Model 2: chol ~ factor(rs174548) + age
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1     397 190875
2     396 182322  1    8552.9 18.577 2.062e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANCOVA



ANCOVA

```
> fit2 = lm(chol ~ factor(rs174548) * age)
> summary(fit2)
Call:
lm(formula = chol ~ factor(rs174548) * age)

Residuals:
    Min       1Q   Median       3Q      Max
-57.5425 -14.3002  0.7131  14.2138  55.7089

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    164.14677     5.79545   28.323  < 2e-16 ***
factor(rs174548)1     3.42799     8.79946    0.390  0.69707
factor(rs174548)2    16.53004    18.28067    0.904  0.36642
age              0.30576     0.10154    3.011  0.00277 **
factor(rs174548)1:age  0.07159     0.15617    0.458  0.64692
factor(rs174548)2:age -0.20255     0.31488   -0.643  0.52043

Residual standard error: 21.49 on 394 degrees of freedom
Multiple R-squared:  0.06777,    Adjusted R-squared:  0.05594
F-statistic: 5.729 on 5 and 394 DF,  p-value: 4.065e-05
```

ANCOVA

```
> fit0 = lm(chol ~ age)
> summary(fit0)
```

Call:

```
lm(formula = chol ~ age)
```

Residuals:

Min	1Q	Median	3Q	Max
-60.453	-14.643	-0.022	14.659	58.995

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	166.90168	4.26488	39.134	< 2e-16 ***
age	0.31033	0.07524	4.125	4.52e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.69 on 398 degrees of freedom

Multiple R-squared: 0.04099, Adjusted R-squared: 0.03858

F-statistic: 17.01 on 1 and 398 DF, p-value: 4.522e-05

```
> anova(fit0, fit2)
```

Analysis of Variance Table

Model 1: chol ~ age

Model 2: chol ~ factor(rs174548) * age

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	398	187187				
2	394	181961	4	5226.6	2.8293	0.02455 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Test of
coincident
lines

ANCOVA

```
> anova(fit1,fit2)
```

Analysis of Variance Table

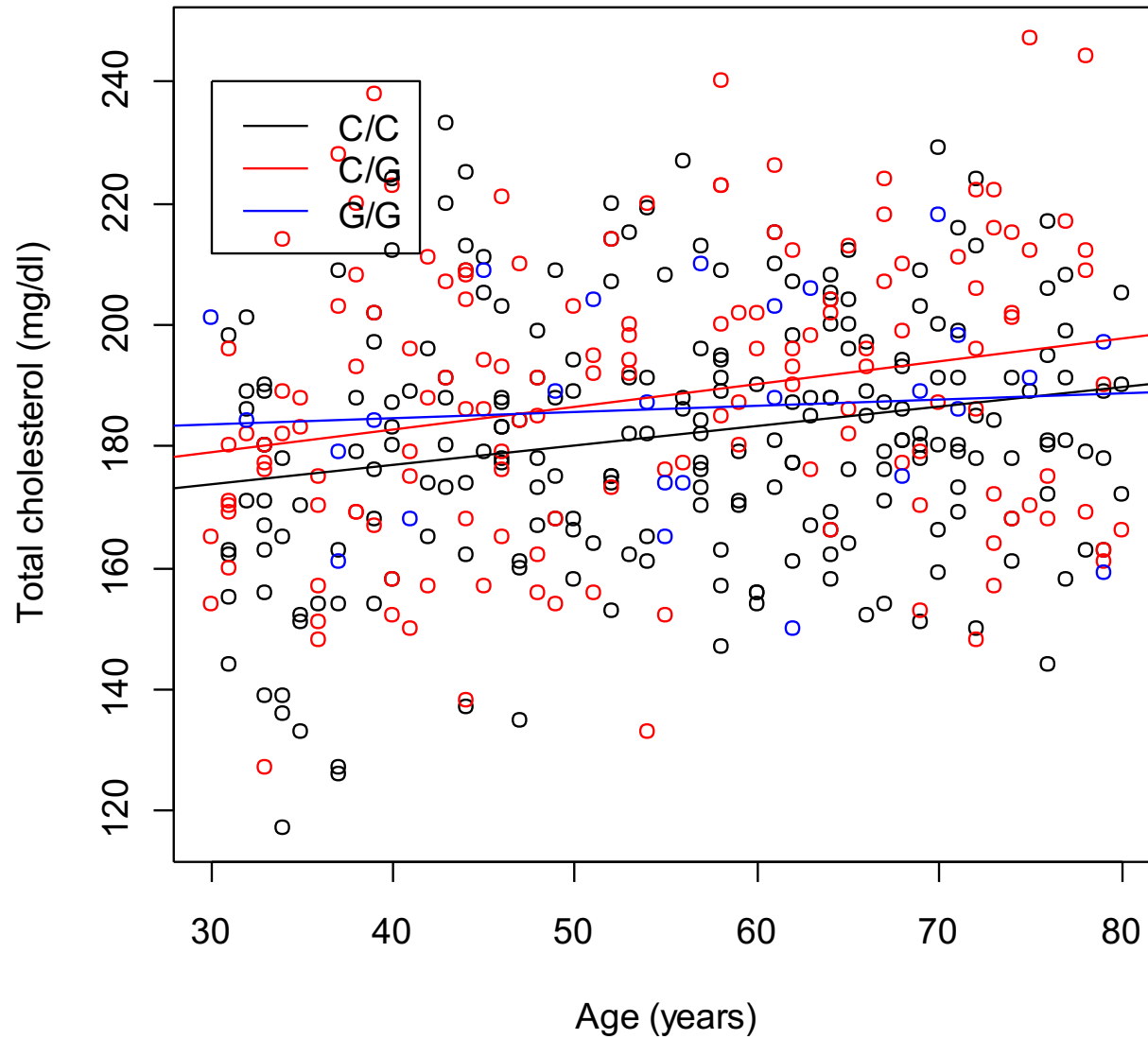
Model 1: chol ~ factor(rs174548) + age

Model 2: chol ~ factor(rs174548) * age

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	396	182322				
2	394	181961	2	361.11	0.391	0.6767

Test of
parallel lines

ANCOVA





ANCOVA

- In summary:

- If the slopes are not equal, then age is an effect modifier

$$E[Y | x, z] = \beta_0 + \beta_1 x + \beta_2 (CG) + \beta_3 (GG) + \beta_4 (x * CG) + \beta_5 (x * GG)$$

- If the slopes are the same,

$$E[Y | x, z] = \beta_0 + \beta_1 x + \beta_2 (CG) + \beta_3 (GG)$$



ANCOVA

- If the slopes are the same,

$$E[Y | x, z] = \beta_0 + \beta_1 x + \beta_2 (CG) + \beta_3 (GG)$$

- then one can obtain adjusted means for the three genotypes using the mean age over all groups
 - For example, the adjusted means for the three groups would be

$$\bar{Y}_1(\text{adj}) = \hat{\beta}_0 + \bar{x} \hat{\beta}_1$$

$$\bar{Y}_2(\text{adj}) = (\hat{\beta}_0 + \hat{\beta}_2) + \bar{x} \hat{\beta}_1$$

$$\bar{Y}_3(\text{adj}) = (\hat{\beta}_0 + \hat{\beta}_3) + \bar{x} \hat{\beta}_1$$



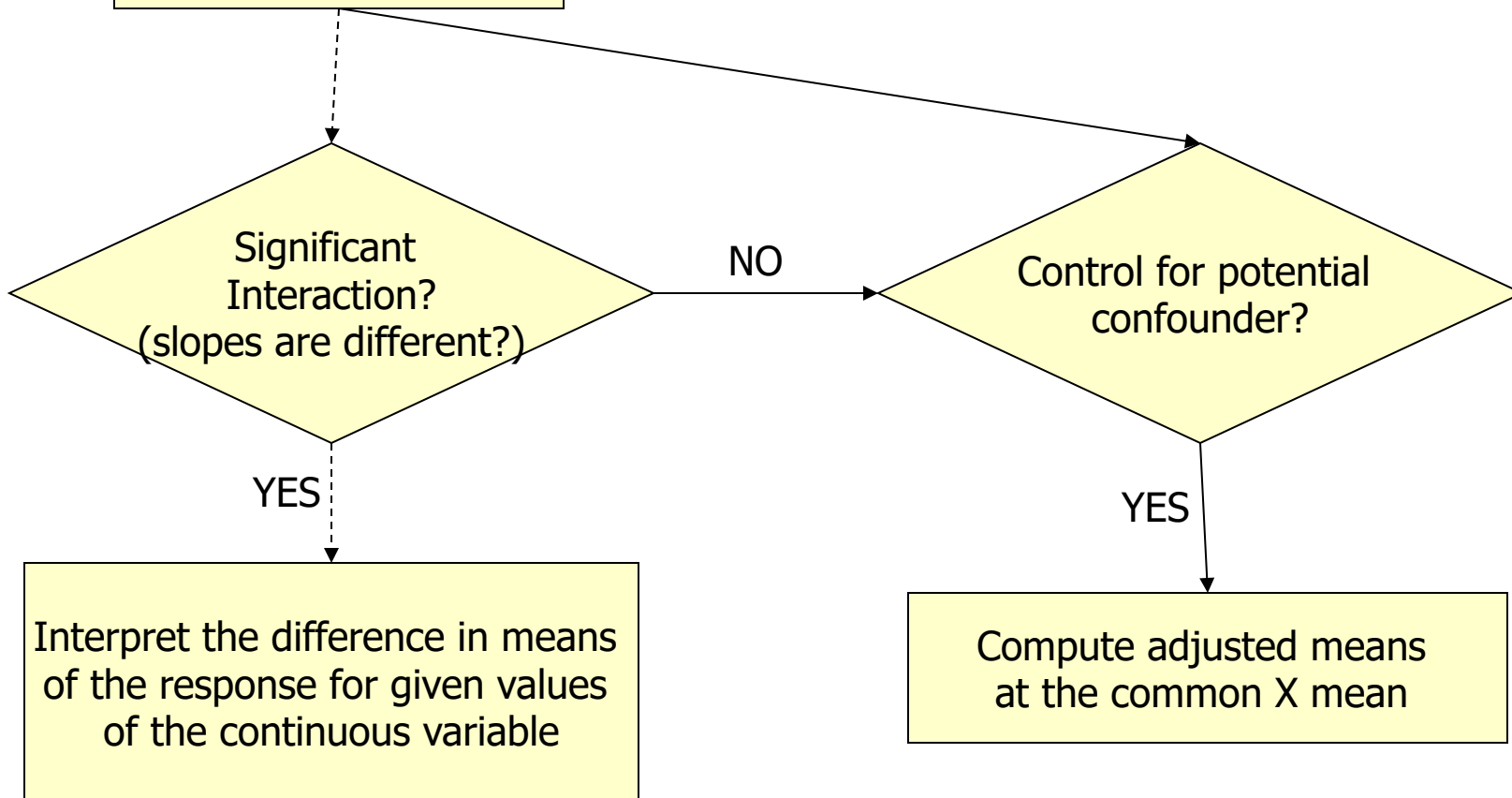
ANCOVA

```
> ## mean cholesterol for different genotypes adjusted by age
> predict(fit1, new=data.frame(age=mean(age),rs174548=0))
      1
180.9013
> predict(fit1, new=data.frame(age=mean(age),rs174548=1))
      1
188.2026
> predict(fit1, new=data.frame(age=mean(age),rs174548=2))
      1
185.9856
```

```
> ## mean cholesterol for different genotypes adjusted by age
> mean(predict(fit1, new=data.frame(age=age,rs174548=0)))
180.9013
> mean(predict(fit1, new=data.frame(age=age,rs174548=1)))
188.2026
> mean(predict(fit1, new=data.frame(age=age,rs174548=2)))
185.9856
```

SUMMARY:

ANCOVA





Summary

We have considered:

- ANOVA and ANCOVA
 - Interpretation
 - Estimation
 - Interaction

- Multiple comparisons



Exercise

- Work on **Exercise 9-12**
 - Try each exercise on your own
 - Make note of any questions or difficulties you have
 - At **3:30ET??** we will meet as a group to go over the solutions and discuss your questions