Interactions and Factorial ANOVA

STA312 Fall 2022

See last slide for copyright information

Interactions

- Interaction between explanatory variables means "It depends."
- Relationship between one explanatory variable and the response variable *depends* on the value of the other explanatory variable.
- Can have
  - Quantitative by quantitative
  - Quantitative by categorical
  - Categorical by categorical

Quantitative by Quantitative

$$\begin{aligned} Y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon \\ E(Y|\mathbf{x}) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \end{aligned}$$

For fixed x_2

$$E(Y|\mathbf{x}) = (\beta_0 + \beta_2 x_2) + (\beta_1 + \beta_3 x_2) x_1$$

Both slope and intercept depend on value of x_2

And for fixed x_1, slope and intercept relating x_2 to E(Y) depend on the value of x_1

Quantitative by Categorical

- One regression line for each category.
- Interaction means slopes are not equal
- Form a product of quantitative variable by each dummy variable for the categorical variable
- For example, three treatments and one covariate: x_1 is the covariate and x_2, x_3 are dummy variables

$$\begin{aligned} Y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &\quad + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \epsilon \end{aligned}$$

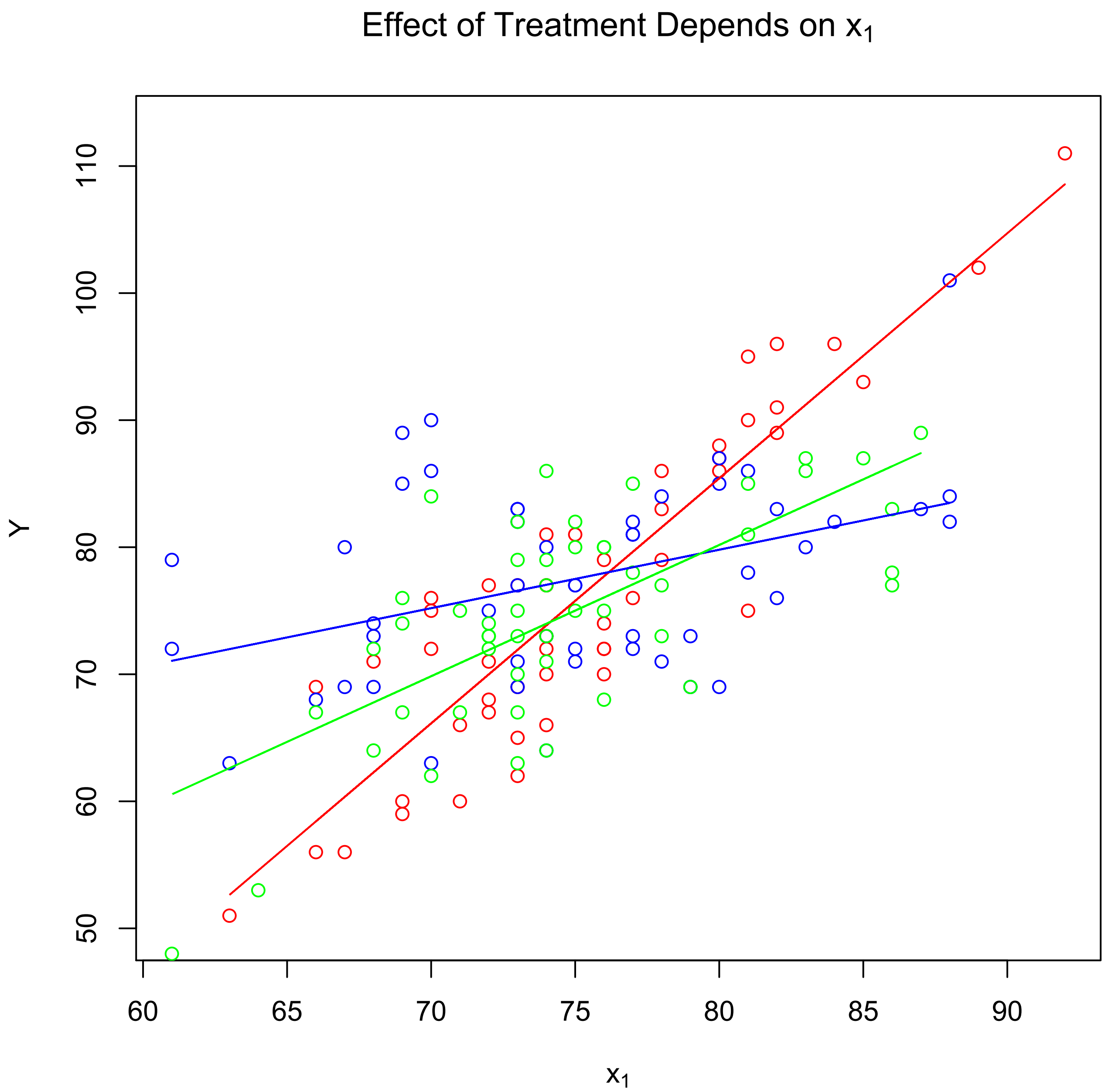
General principle

- Interaction between A and B means
  - Relationship of A to Y depends on value of B
  - Relationship of B to Y depends on value of A
- The two statements are formally equivalent

Make a table

$$E(Y|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3$$

Group	x_2	x_3	$E(Y\|\mathbf{x})$
1	1	0	$(\beta_0 + \beta_2) + (\beta_1 + \beta_4)x_1$
2	0	1	$(\beta_0 + \beta_3) + (\beta_1 + \beta_5)x_1$
3	0	0	$\beta_0 + \beta_1 x_1$



Effect of Treatment Depends on x_1
Y
110
100
90
80
70
60
50
60
65
70
75
80
85
90
x_1

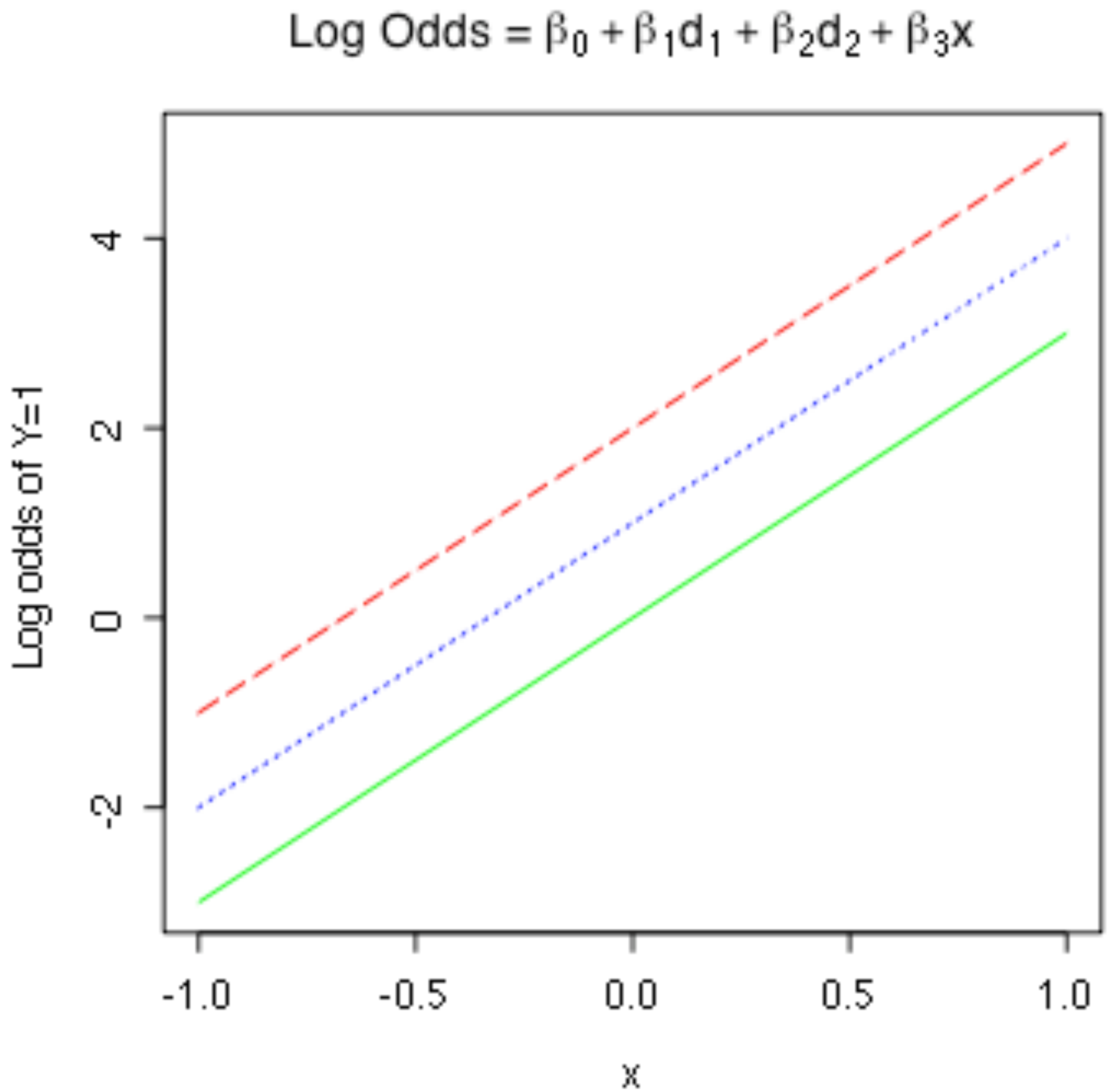
Group	x_2	x_3	$E(Y\|\mathbf{x})$
1	1	0	$(\beta_0 + \beta_2) + (\beta_1 + \beta_4)x_1$
2	0	1	$(\beta_0 + \beta_3) + (\beta_1 + \beta_5)x_1$
3	0	0	$\beta_0 + \beta_1 x_1$

What null hypothesis would you test for

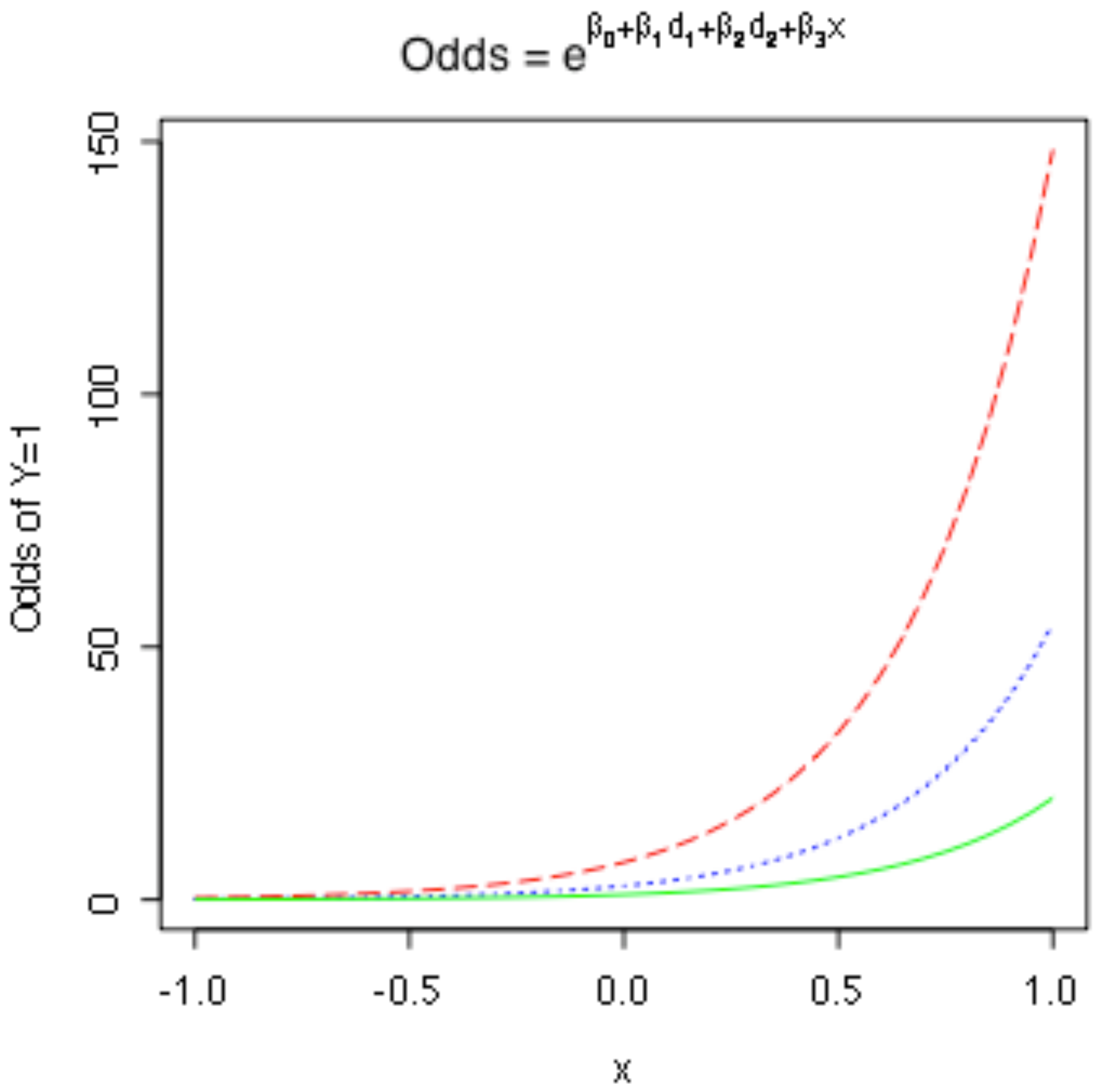
- Equal slopes
- Comparing slopes for group one vs three
- Comparing slopes for group one vs two
- Equal regressions
- Interaction between group and x_1

What about logistic regression?

Equal slopes in the log odds scale

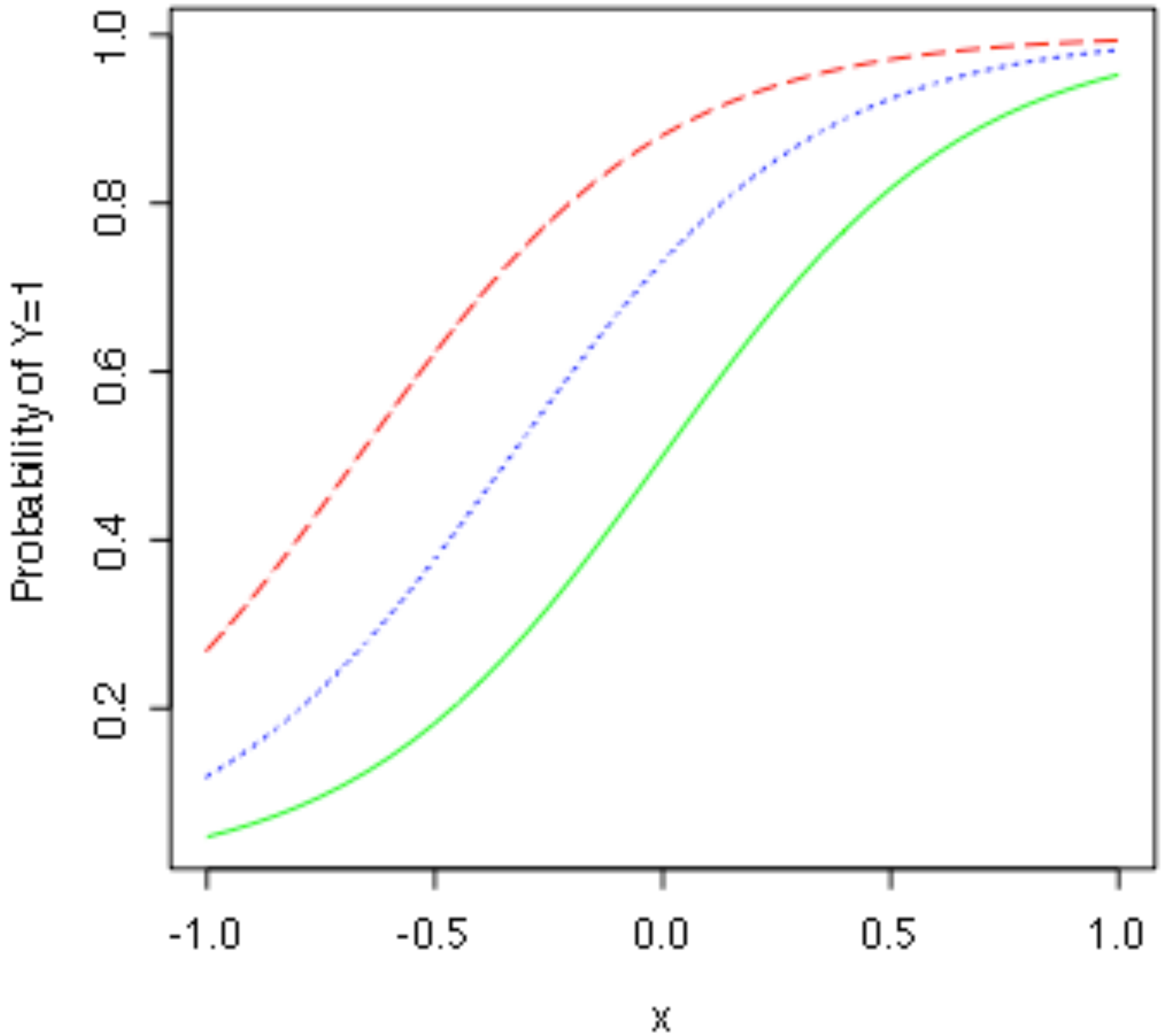


Equal slopes in the log odds scale means proportional odds



Proportional Odds in Terms of Probability

$$\text{Probability} = \frac{e^{\beta_0+\beta_1 d_1+\beta_2 d_2+\beta_3 x}}{1 + e^{\beta_0+\beta_1 d_1+\beta_2 d_2+\beta_3 x}}$$



Interactions

- With equal slopes in the log odds scale, *differences* in odds and *differences* in probabilities do depend on x.
- Regression coefficients for product terms still mean something.
- If zero, they mean that the *odds ratio* does not depend on the value(s) of the covariate(s).
- Odds ratio has odds of Y=1 for the reference category in the denominator.
- Categorical by categorical interactions are meaningful.

Categorical by Categorical

- Naturally part of factorial ANOVA in experimental studies
- Also applies to purely observational data

Factorial ANOVA

More than one categorical explanatory variable

Factorial ANOVA

- Categorical explanatory variables are called **factors**
- More than one at a time
- Primarily for true experiments, but also used with observational data

- If there are observations at all combinations of explanatory variable values, it's called a *complete* factorial design (as opposed to a fractional factorial).

The potato study

- Cases are potatoes
- Inoculate with bacteria, store for a fixed time period.
- Response variable is percent surface area with visible rot.
- Two explanatory variables, randomly assigned
  - Bacteria Type (1, 2, 3)
  - Temperature (1=Cool, 2=Warm)

Two-factor design

	Bacteria Type		
Temp	1	2	3
1=Cool			
2=Warm			

Six treatment conditions

Factorial experiments

- Allow more than one factor to be investigated in the same study: Efficiency!
- Allow the scientist to see whether the effect of an explanatory variable *depends* on the value of another explanatory variable: Interactions
- Thank you again, Mr. Fisher.

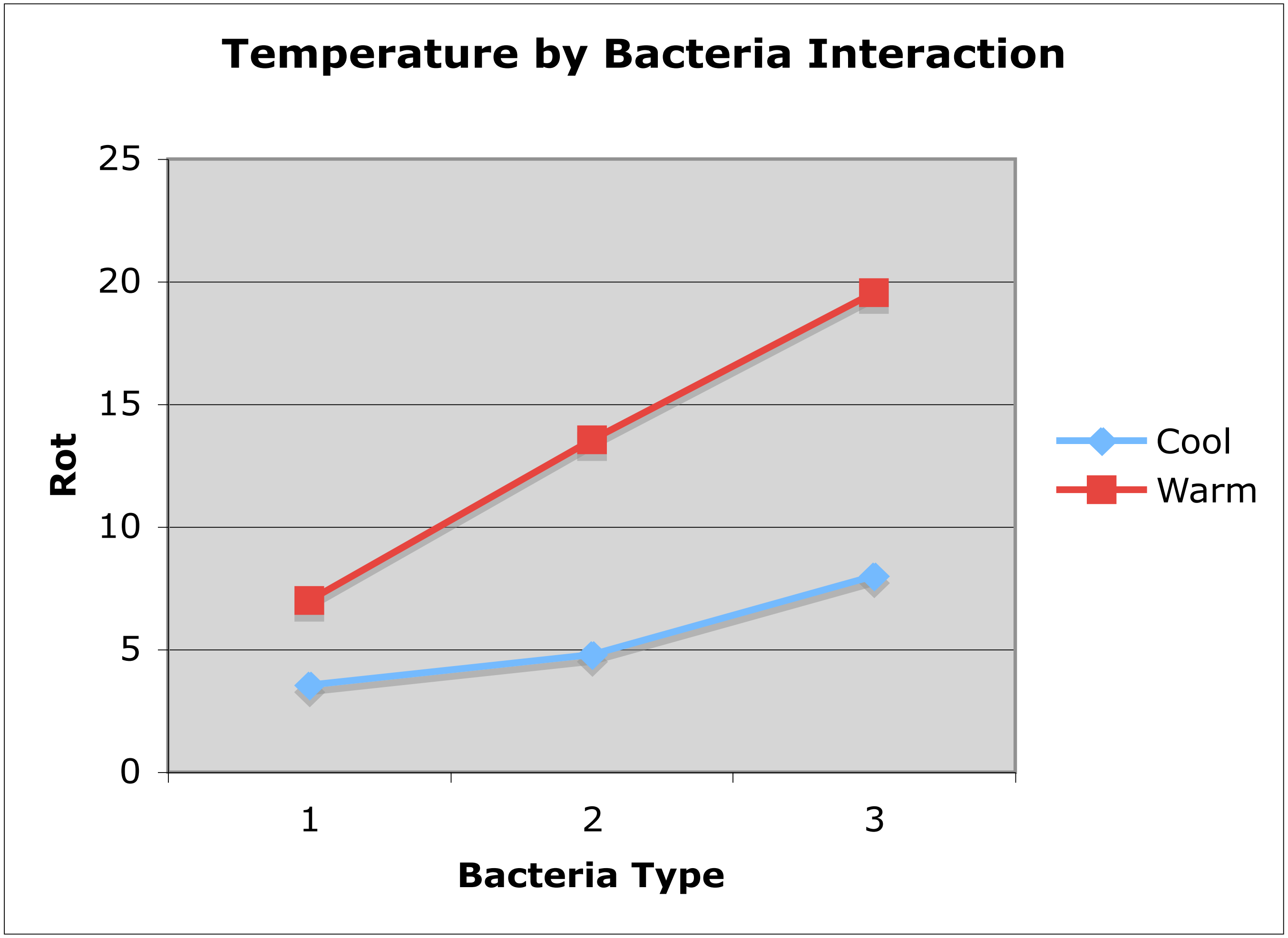
Normal with equal variance and conditional (cell) means $\mu_{i,j}$

	Bacteria Type			
Temp	1	2	3	
1=Cool	$\mu_{1,1}$	$\mu_{1,2}$	$\mu_{1,3}$	$\frac{\mu_{1,1}+\mu_{1,2}+\mu_{1,3}}{3}$
2=Warm	$\mu_{2,1}$	$\mu_{2,2}$	$\mu_{2,3}$	$\frac{\mu_{2,1}+\mu_{2,2}+\mu_{2,3}}{3}$
	$\frac{\mu_{1,1}+\mu_{2,1}}{2}$	$\frac{\mu_{1,2}+\mu_{2,2}}{2}$	$\frac{\mu_{1,3}+\mu_{2,3}}{2}$	μ

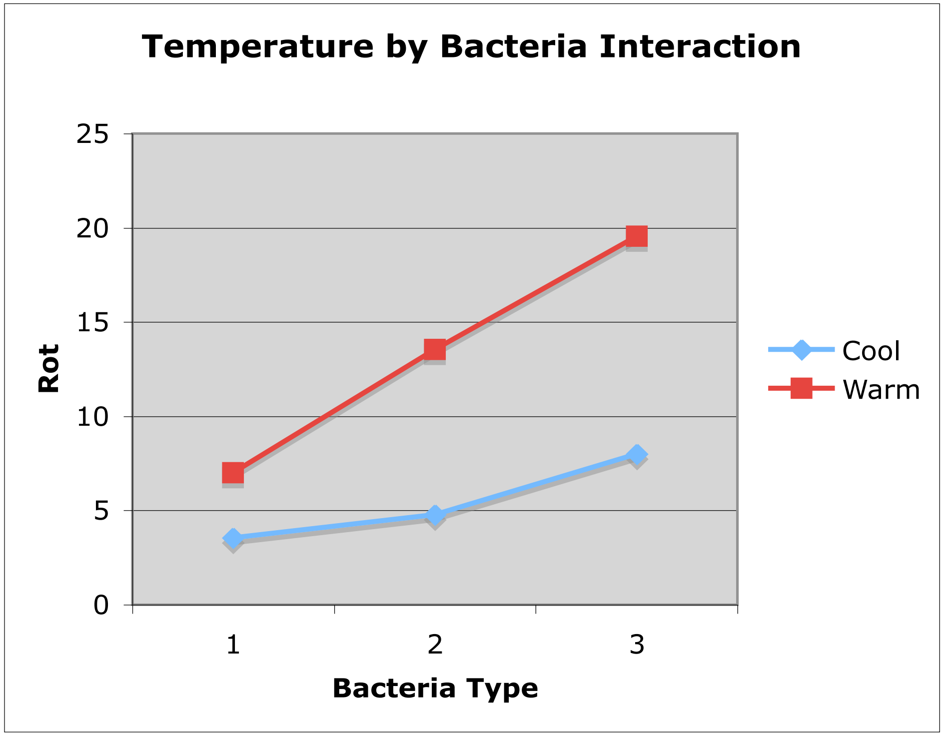
Tests

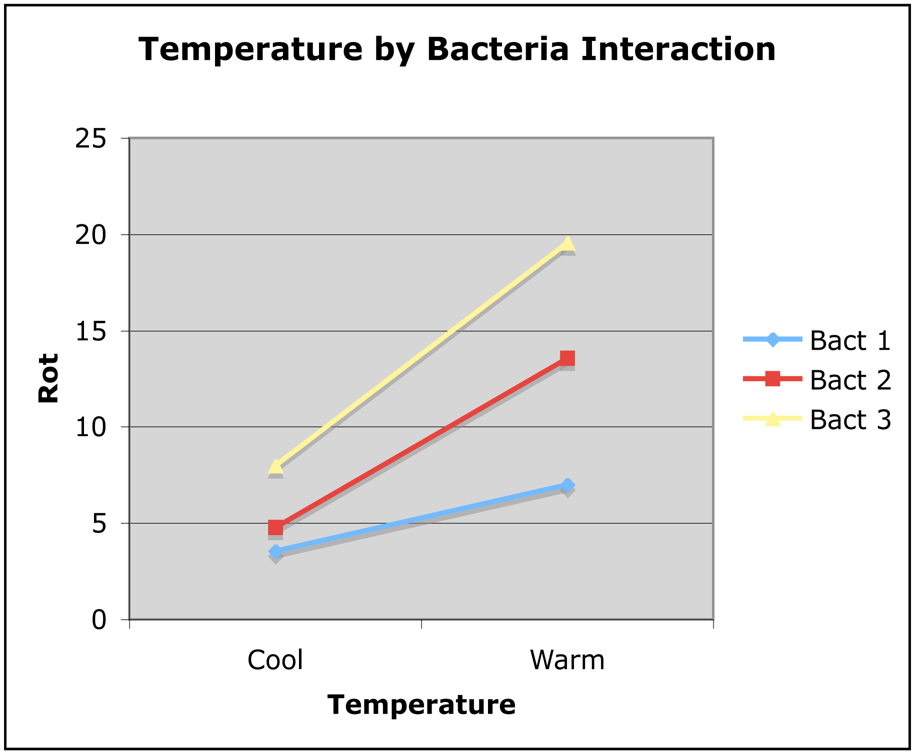
- Main effects: Differences among marginal means
- Interactions: Differences between differences (What is the effect of Factor A? **It depends** on the level of Factor B.)

To understand the interaction, plot the means

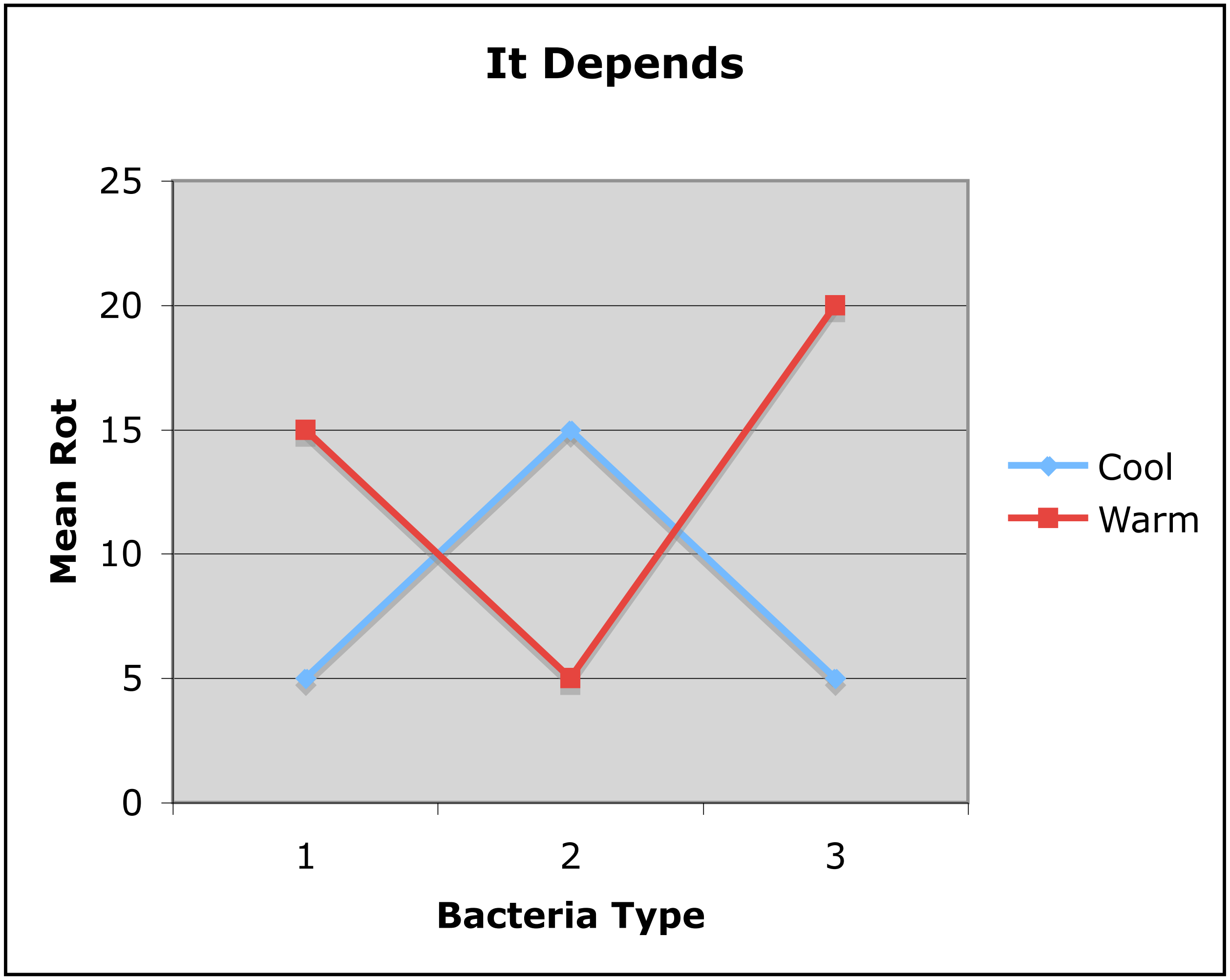


Either Way

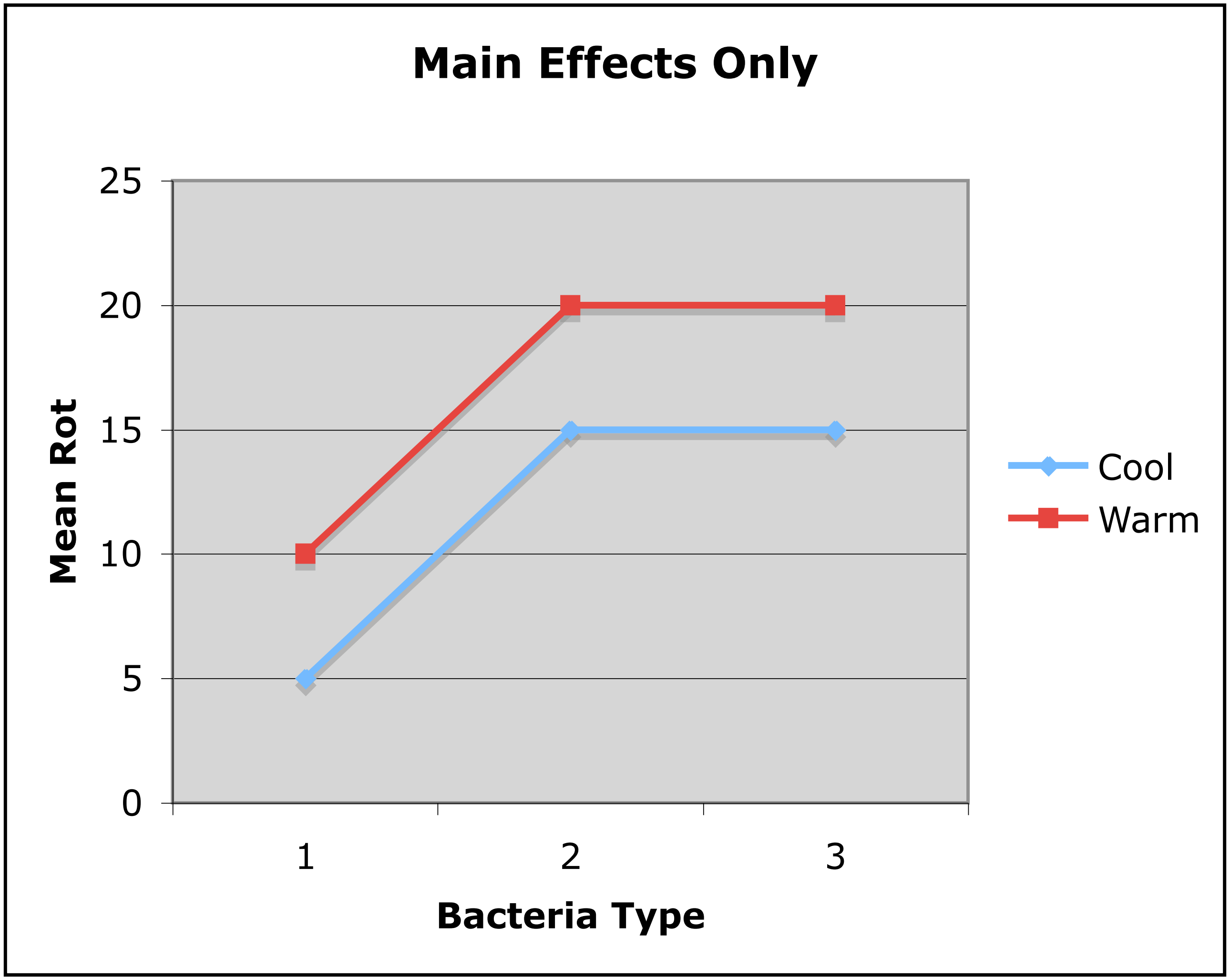

Temperature by Bacteria Interaction
25
20
15
10
5
0
Rot
1
2
3
Bacteria Type
Cool
Warm


Temperature by Bacteria Interaction
25
20
15
10
5
0
Rot
Cool
Warm
Temperature
Bact 1
Bact 2
Bact 3

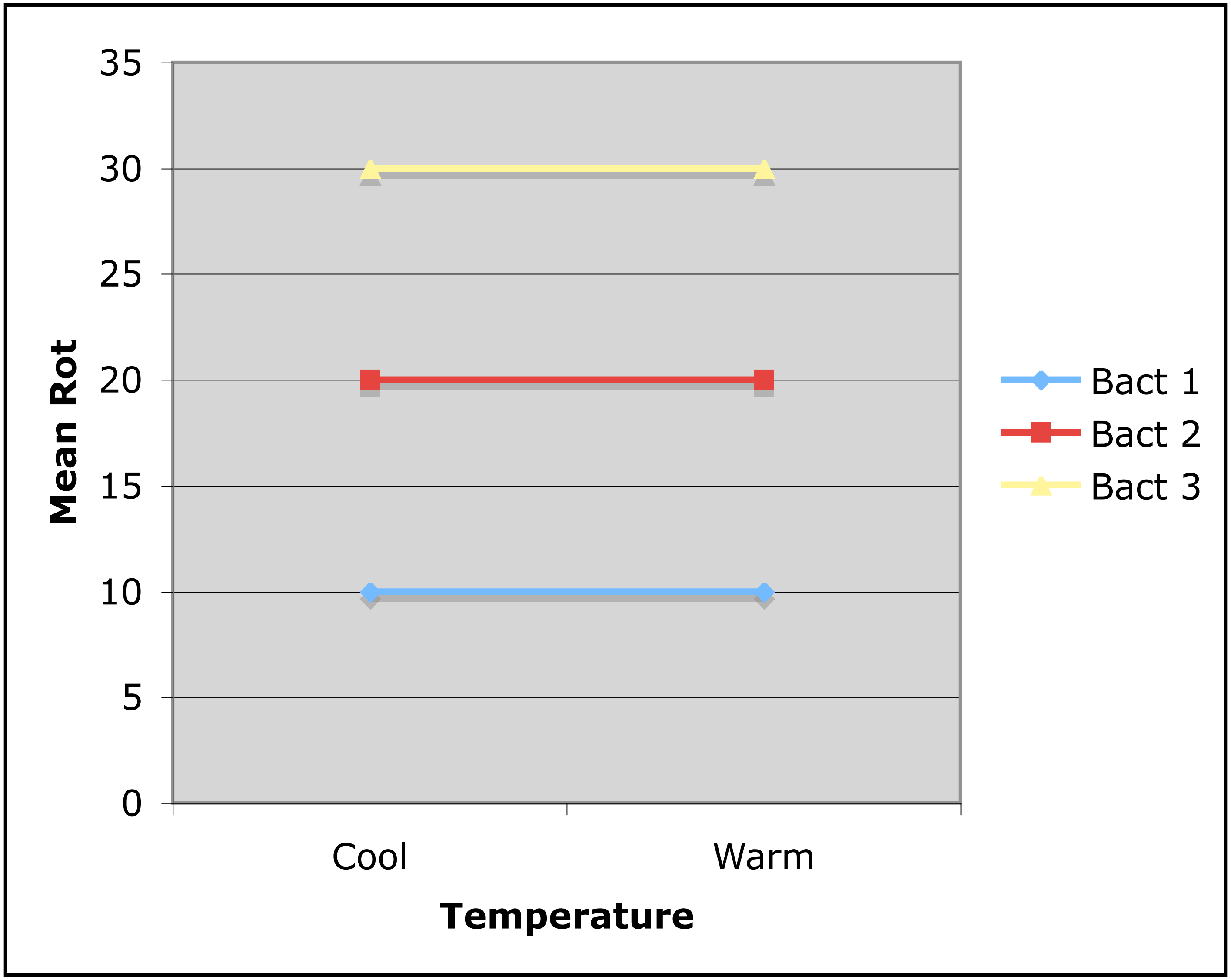
Non-parallel profiles = Interaction



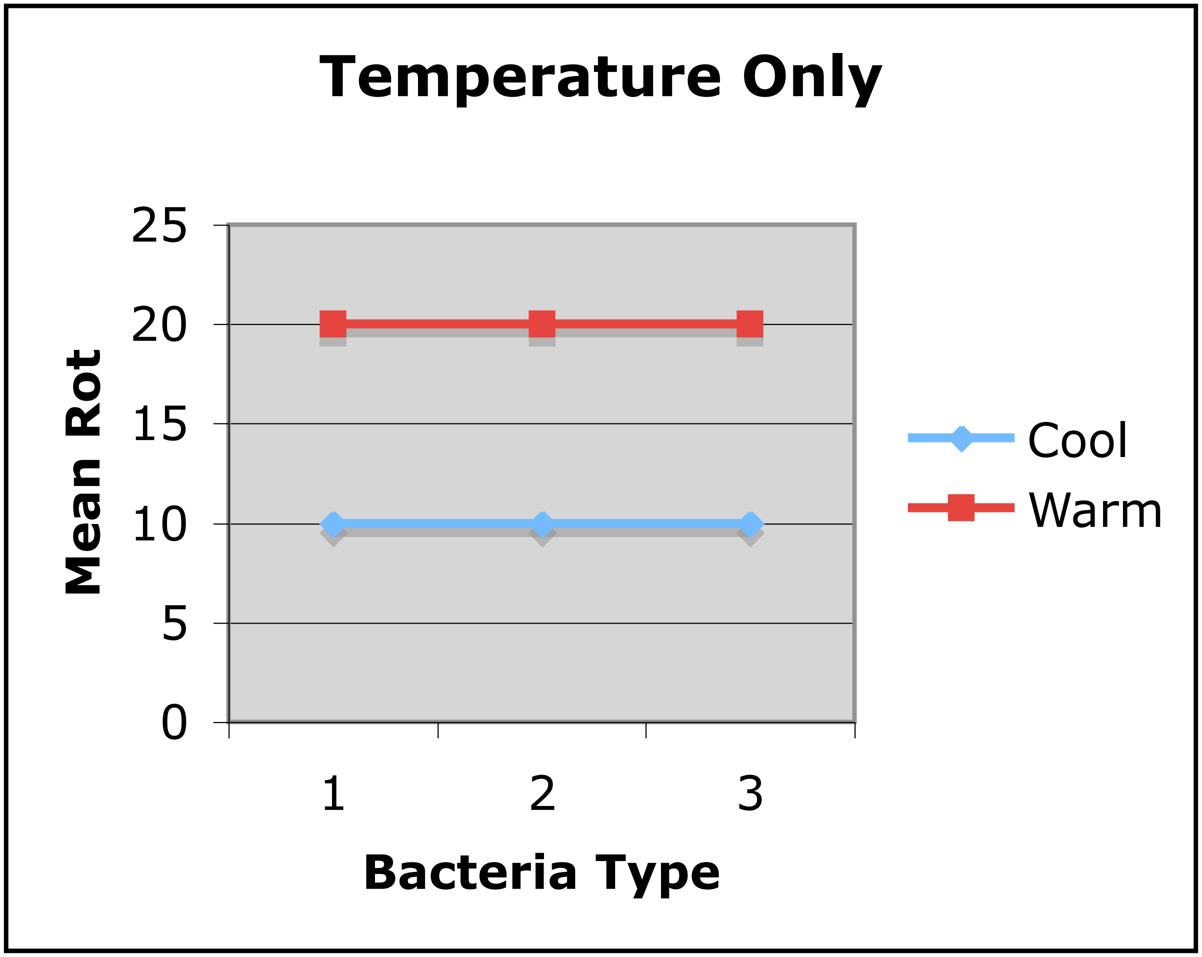
Main effects for both variables, no interaction



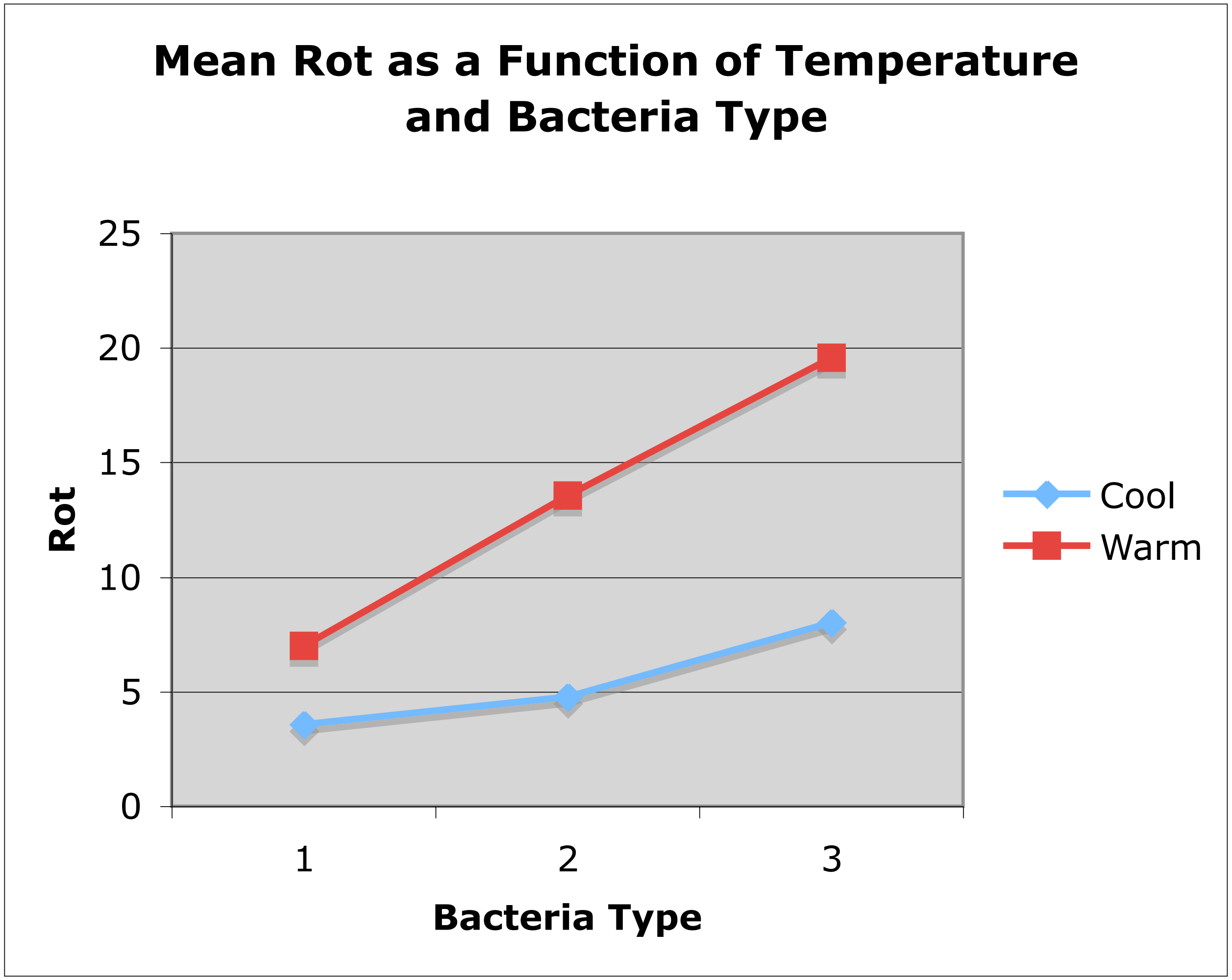
Main effect for Bacteria only



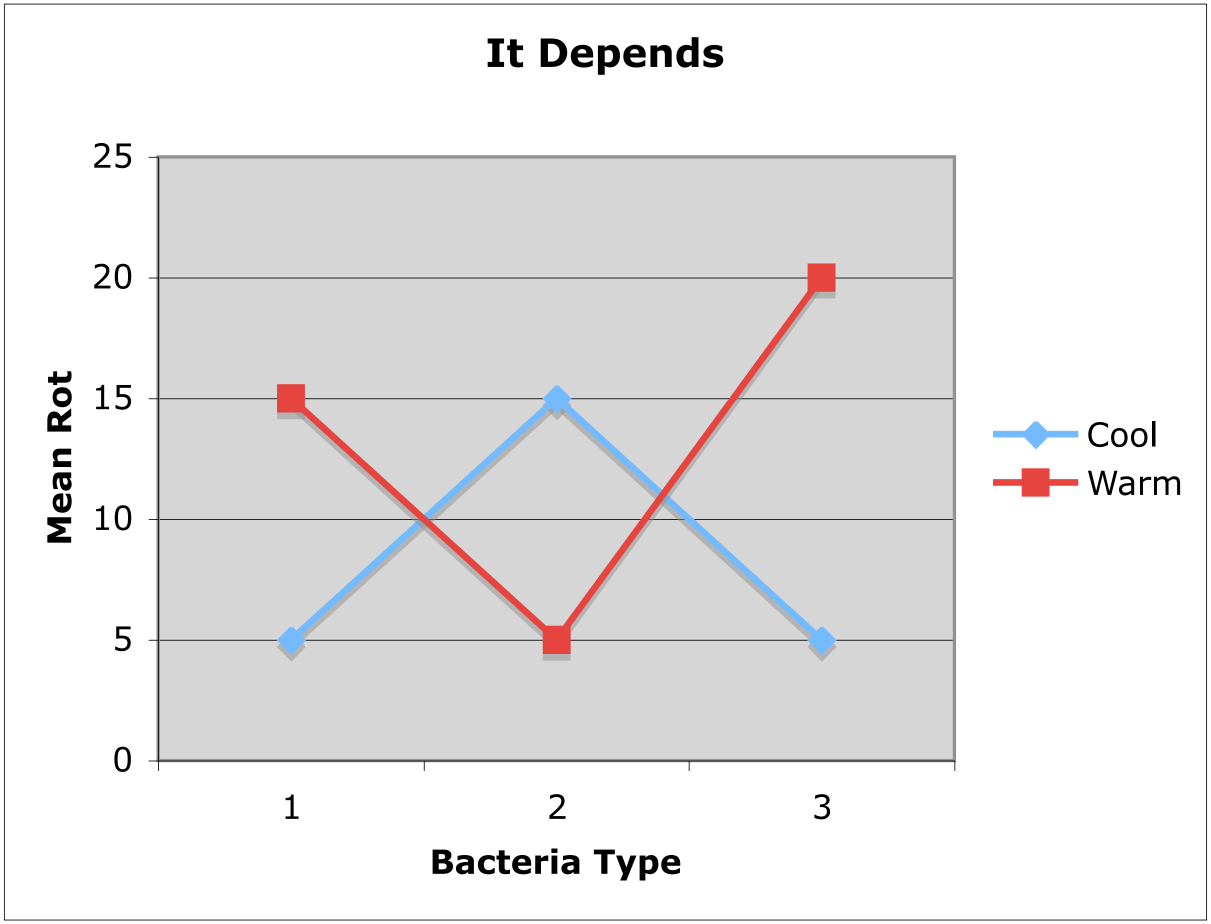
Main Effect for Temperature Only



Both Main Effects, and the Interaction



Should you interpret the main effects?



A common error

- Categorical explanatory variable with *k* categories
- *p* dummy variables (rather than *k-1*)
- And an intercept

- There are *k* population means represented by *k+1* regression coefficients - not unique

But suppose you leave off the intercept

- Now there are *k* regression coefficients and *k* population means
- The correspondence is unique, and the model can be handy -- less algebra
- Called **cell means coding**

Less algebra

In a model with an intercept, test whether the average response to Drug A and Drug B is different from response to the placebo, controlling for age. What is the null hypothesis?

Drug	x_2	x_3	$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
A	1	0	$(\beta_0 + \beta_2) + \beta_1 x_1$
B	0	1	$(\beta_0 + \beta_3) + \beta_1 x_1$
Placebo	0	0	$\beta_0 \quad + \beta_1 x_1$

$$H_0 : \beta_2 + \beta_3 = 0$$

Show your work

$$\tfrac{1}{2}\left[\,(\beta_0 + \beta_2 + \beta_1 x_1) + (\beta_0 + \beta_3 + \beta_1 x_1)\right] = \beta_0 + \beta_1 x_1$$

$$\iff \quad \beta_0 + \beta_2 + \beta_1 x_1 + \beta_0 + \beta_3 + \beta_1 x_1 = 2\beta_0 + 2\beta_1 x_1$$

$$\iff \quad 2\beta_0 + \beta_2 + \beta_3 + 2\beta_1 x_1 = 2\beta_0 + 2\beta_1 x_1$$

$$\iff \quad \beta_2 + \beta_3 = 0$$

We want to avoid this kind of thing

Cell means coding: *k* indicators and no intercept

$$E[Y|\boldsymbol{X} = \boldsymbol{x}] = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Drug	x_1	x_2	x_3	$\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
A	1	0	0	$\mu_1 = \beta_1$
B	0	1	0	$\mu_2 = \beta_2$
Placebo	0	0	1	$\mu_3 = \beta_3$

Easy to test H_0: $\beta_1 + \beta_2 = 2\beta_3$

Add a covariate: x_4

$$E[Y|\boldsymbol{X} = \boldsymbol{x}] = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

Drug	x_1	x_2	x_3	$\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$
A	1	0	0	$\beta_1 + \beta_4 x_4$
B	0	1	0	$\beta_2 + \beta_4 x_4$
Placebo	0	0	1	$\beta_3 + \beta_4 x_4$

Contrasts

$$c = a_1\mu_1 + a_2\mu_2 + \cdots + a_p\mu_p$$

$$\widehat{c} = a_1\overline{Y}_1 + a_2\overline{Y}_2 + \cdots + a_p\overline{Y}_p$$

where $a_1 + a_2 + \cdots + a_p = 0$

In a one-factor design

- Mostly, what you want are tests of contrasts,
- Or collections of contrasts.
- You could do it with any dummy variable coding scheme.
- Cell means coding is often most convenient.
- With $\boldsymbol{\beta}=\boldsymbol{\mu}$, test H_0: $\mathbf{L}\boldsymbol{\beta}=\mathbf{h}$

- Can get a confidence interval for any single contrast using the *t* distribution.

Testing Contrasts in Factorial Designs

	Bacteria Type			
Temp	1	2	3	
1=Cool	$\mu_{1,1}$	$\mu_{1,2}$	$\mu_{1,3}$	$\frac{\mu_{1,1}+\mu_{1,2}+\mu_{1,3}}{3}$
2=Warm	$\mu_{2,1}$	$\mu_{2,2}$	$\mu_{2,3}$	$\frac{\mu_{2,1}+\mu_{2,2}+\mu_{2,3}}{3}$
	$\frac{\mu_{1,1}+\mu_{2,1}}{2}$	$\frac{\mu_{1,2}+\mu_{2,2}}{2}$	$\frac{\mu_{1,3}+\mu_{2,3}}{2}$	μ

- Differences between marginal means are definitely contrasts
- Interactions are also sets of contrasts

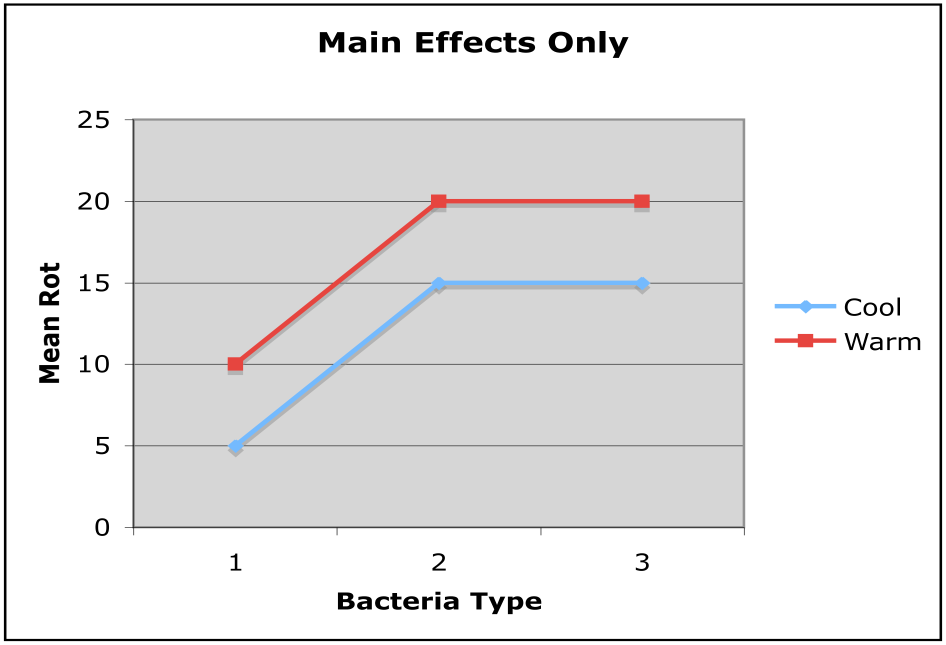
Interactions are sets of Contrasts

	Bacteria Type			
Temp	1	2	3	
1=Cool	$\mu_{1,1}$	$\mu_{1,2}$	$\mu_{1,3}$	$\frac{\mu_{1,1}+\mu_{1,2}+\mu_{1,3}}{3}$
2=Warm	$\mu_{2,1}$	$\mu_{2,2}$	$\mu_{2,3}$	$\frac{\mu_{2,1}+\mu_{2,2}+\mu_{2,3}}{3}$
	$\frac{\mu_{1,1}+\mu_{2,1}}{2}$	$\frac{\mu_{1,2}+\mu_{2,2}}{2}$	$\frac{\mu_{1,3}+\mu_{2,3}}{2}$	μ

- $H_0 : \mu_{1,1} - \mu_{2,1} = \mu_{1,2} - \mu_{2,2} = \mu_{1,3} - \mu_{2,3}$

- $H_0 : \mu_{1,2} - \mu_{1,1} = \mu_{2,2} - \mu_{2,1}$ and
  $\mu_{1,3} - \mu_{1,2} = \mu_{2,3} - \mu_{2,2}$

Interactions are sets of Contrasts



- $H_0 : \mu_{1,1} - \mu_{2,1} = \mu_{1,2} - \mu_{2,2} = \mu_{1,3} - \mu_{2,3}$

- $H_0 : \mu_{1,2} - \mu_{1,1} = \mu_{2,2} - \mu_{2,1}$ and

  $\mu_{1,3} - \mu_{1,2} = \mu_{2,3} - \mu_{2,2}$

Equivalent statements

- The effect of A depends upon B
- The effect of B depends on A

$$H_0 : \mu_{1,1} - \mu_{2,1} = \mu_{1,2} - \mu_{2,2} = \mu_{1,3} - \mu_{2,3}$$

$$H_0 : \mu_{1,2} - \mu_{1,1} = \mu_{2,2} - \mu_{2,1} \text{ and}$$
$$\mu_{1,3} - \mu_{1,2} = \mu_{2,3} - \mu_{2,2}$$

Three factors: A, B and C

- There are three (sets of) main effects: One each for A, B, C
- There are three two-factor interactions
  - A by B (Averaging over C)
  - A by C (Averaging over B)
  - B by C (Averaging over A)
- There is one three-factor interaction: AxBxC

Meaning of the 3-factor interaction

- The form of the A x B interaction depends on the value of C
- The form of the A x C interaction depends on the value of B
- The form of the B x C interaction depends on the value of A
- These statements are equivalent. Use the one that is easiest to understand.

To graph a three-factor interaction

- Make a separate mean plot (showing a 2-factor interaction) for each value of the third variable.

- In the potato study, a graph for each type of potato

Four-factor design

- Four sets of main effects
- Six two-factor interactions
- Four three-factor interactions
- One four-factor interaction: The nature of the three-factor interaction depends on the value of the 4th factor
- There is an F test for each one
- And so on …

As the number of factors increases

- The higher-way interactions get harder and harder to understand
- All the tests are still tests of sets of contrasts (differences between differences of differences …)
- But it gets harder and harder to write down the contrasts
- Effect coding becomes easier

Effect coding

Like indicator dummy variables with intercept, but put -1 for the last category.

Bact	**B_1**	**B_2**
1	1	0
2	0	1
3	-1	-1

Temperature	**T**
1=Cool	1
2=Warm	-1

$$E(Y|\mathbf{X}=\mathbf{x}) = \beta_0 + \beta_1 B_1 + \beta_2 B_2 + \beta_3 T + \beta_4 B_1 T + \beta_5 B_2 T$$

Interaction effects are products of dummy variables

$$E(Y|\mathbf{X} = \mathbf{x}) = \beta_0 + \beta_1 B_1 + \beta_2 B_2 + \beta_3 T + \beta_4 B_1 T + \beta_5 B_2 T$$

- The A x B interaction: Multiply each dummy variable for A by each dummy variable for B
- Use these products as additional explanatory variables in the multiple regression
- The A x B x C interaction: Multiply each dummy variable for C by each product term from the A x B interaction
- Test the sets of product terms simultaneously

Make a table

$$E(Y|\mathbf{X}=\mathbf{x}) = \beta_0 + \beta_1 B_1 + \beta_2 B_2 + \beta_3 T + \beta_4 B_1 T + \beta_5 B_2 T$$

Bact	Temp	B_1	B_2	T	B_1T	B_2T	$E(Y\|\mathbf{X}=\mathbf{x})$
1	1	1	0	1	1	0	$\beta_0 + \beta_1 + \beta_3 + \beta_4$
1	2	1	0	-1	-1	0	$\beta_0 + \beta_1 - \beta_3 - \beta_4$
2	1	0	1	1	0	1	$\beta_0 + \beta_2 + \beta_3 + \beta_5$
2	2	0	1	-1	0	-1	$\beta_0 + \beta_2 - \beta_3 - \beta_5$
3	1	-1	-1	1	-1	-1	$\beta_0 - \beta_1 - \beta_2 + \beta_3 - \beta_4 - \beta_5$
3	2	-1	-1	-1	1	1	$\beta_0 - \beta_1 - \beta_2 - \beta_3 + \beta_4 + \beta_5$

Cell and Marginal Means

	Bacteria Type			
Tmp	1	2	3	
1=C	$\beta_0 + \beta_1 + \beta_3 + \beta_4$	$\beta_0 + \beta_2 + \beta_3 + \beta_5$	$\beta_0 - \beta_1 - \beta_2$ $+\beta_3 - \beta_4 - \beta_5$	β_0 $+\beta_3$
2=W	$\beta_0 + \beta_1 - \beta_3 - \beta_4$	$\beta_0 + \beta_2 - \beta_3 - \beta_5$	$\beta_0 - \beta_1 - \beta_2$ $-\beta_3 + \beta_4 + \beta_5$	β_0 $-\beta_3$
	$\beta_0 + \beta_1$	$\beta_0 + \beta_2$	$\beta_0 - \beta_1 - \beta_2$	β_0

We see

- Intercept is the grand mean
- Regression coefficients for the dummy variables are deviations of the marginal means from the grand mean
- What about the interactions?

$$E(Y|\mathbf{X} = \mathbf{x}) = \beta_0 + \beta_1 B_1 + \beta_2 B_2 + \beta_3 T + \beta_4 B_1 T + \beta_5 B_2 T$$

A bit of algebra shows

$\mu_{1,1} - \mu_{2,1} = \mu_{1,2} - \mu_{2,2}$ is equivalent to $\beta_4 = \beta_5$

$\mu_{1,2} - \mu_{2,2} = \mu_{1,3} - \mu_{2,3}$ is equivalent to $\beta_4 = -\beta_5$

So $\beta_4 = \beta_5 = 0$

What are “effects?”

$$E(Y|\mathbf{X}=\mathbf{x}) = \beta_0 + \beta_1 B_1 + \beta_2 B_2 + \beta_3 T + \beta_4 B_1 T + \beta_5 B_2 T$$

- **There are 3 main effects for Bacteria** Type: beta1, beta2 and -beta1 -beta2.
- They are deviations of the marginal means from the grand mean.
- **There are 2 main effects for Temperature**: beta3 and - beta3
- They are deviations of the marginal means from the grand mean.
- **There are 6 interaction effects**.
- They are deviations of the cell mean from the grand mean plus the main effects.
- They add to zero across rows and across columns.
- The non-redundant ones are beta4 and beta5.

- This is regression notation. There are ANOVA notations as well.

Factorial ANOVA with effect coding is pretty automatic

- You don’t have to make a table unless asked.
- It always works as you expect it will.
- Hypothesis tests are the same as testing sets of contrasts.
- Covariates present no problem. Main effects and interactions have their usual meanings, “controlling” for the covariates.
- Plot the “least squares means” (Y-hat at x-bar values for covariates).

Again

- Intercept is the grand mean
- Regression coefficients for the dummy variables are deviations of the marginal means from the grand mean
- Test of main effect(s) is test of the dummy variables for a factor.
- Interaction effects are products of dummy variables.

Copyright Information