23-24. Categorical Predictors and Testing General Linear Hypotheses

Readings: Kleinbaum, Kupper, Nizam, and Rosenberg (KKNR): Ch. 12

SAS: PROC REG

Homework: Homework 10 due by midnight on December 3
Final Project due by midnight on December 7

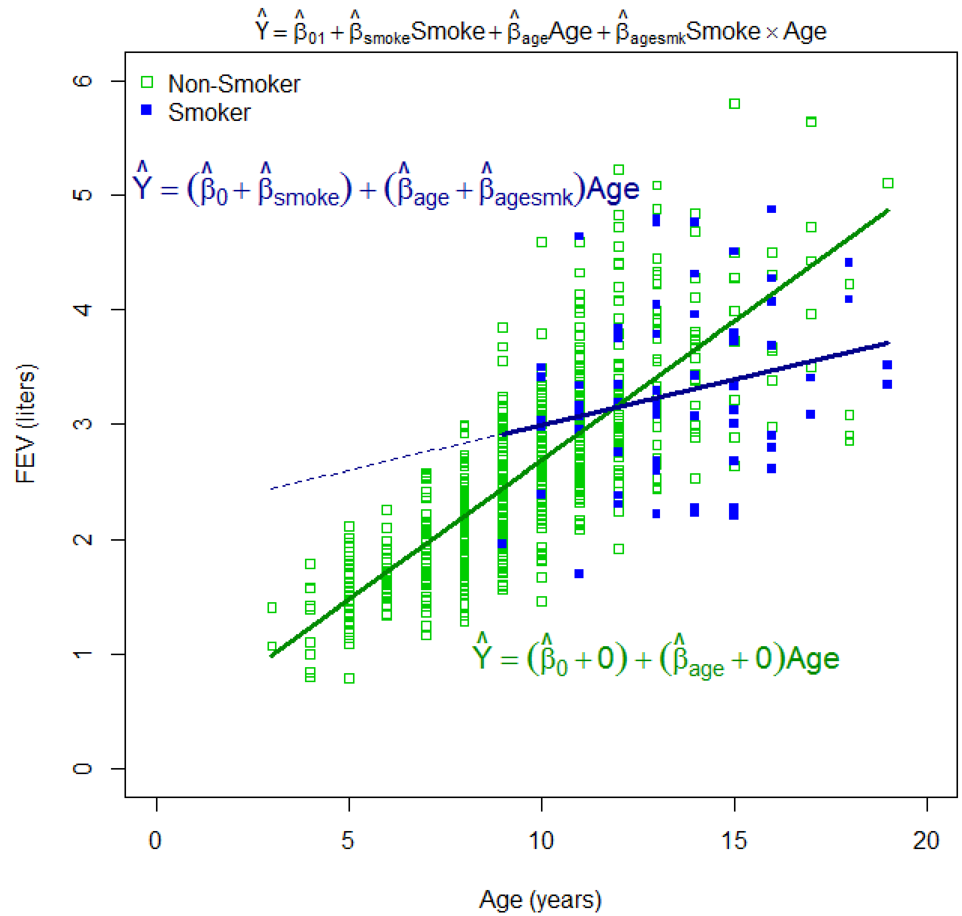
Overview

A) Re/Preview of Topics
B) Categorical Predictors with >2 Categories
C) Tests of General Linear Hypotheses
D) Linear Contrasts
E) Orthogonal Polynomials
F) Equivalence of Reference Cell Coding and Cell Means Coding for Orthogonal Contrasts
G) Equivalence of Reference Cell Coding and Cell Means Coding for Orthogonal Polynomials

A. Review (Lecture 22)/Current (Lecture 23-24)/ Preview (Lecture 25)

Lecture 22:

- Effect Modification (Interactions):
  - $E[FEV_i] = \beta_0 + \beta_{age}Age_i + \beta_{smoke}Smoke_i + \beta_{agesmk}Age_i \times Smoke_i$
  - Allows for different slopes (FEV vs. age) for smokers and non-smokers
- MLE vs LSE (same β's, different variance)



Lectures 23-24:

- Categorical Predicators
  - Indicator variables
- Test of general linear hypothesis
- Linear contrasts
- Orthogonal polynomials

Lecture 25:

- Polynomial Regression: quadratic, cubic, quartic
- Other remedies for non-linearity

B. Categorical Explanatory Variables: More Than 2 Categories

Motivating Example: An investigator is interested in studying the relationship between infant birthweight (pounds) and smoking status of the mother during the first trimester. The investigator chose **five pregnant women from each of four smoking categories** (X) (*never*, *former*, *light*, and *heavy* smokers, with X coded 0, 1, 2, 3) from a larger study. The table below provides the birthweight (Y) of each baby and the average birthweight for each smoking category.

	Never Smokers (X=0)	*Former Smokers (X=1)*	*Light Smokers (X=2)*	*Heavy Smokers (X=3)*
	7.50	5.80	5.90	6.20
	6.20	7.30	6.20	6.80
	6.90	8.20	5.80	5.70
	7.40	7.10	4.70	4.90
	9.20	7.80	8.30	6.20
$\overline{Y}\|X_i$	7.44	7.24	6.18	5.96
$S^2{}_{Y\|X_i}$	1.233	0.833	1.727	0.503

```
proc import
datafile="~/birthweight_smoking_5per
group_dataset.csv"
     out=bwt5 /* name for data set
for SAS to reference */
     dbms=csv /* identify file as
csv */
     replace; /* overwrite BWT if
already present */
     getnames=yes; /* take first row
as column names from data */
run;
```

Potential Scientific Questions:

- Is there an association between smoking status and birthweight?
- Is there a difference in birthweight between never smokers and former smokers?
- Is there a difference in birthweight between non-smokers and current smokers?
- Is there an association between smoking and birthweight adjusting for weight of the mother?

To address these scientific questions in a regression model, you can create a different indicator variable or "dummy variable" for each of the categories:

$$\text{never}=\begin{cases}1 & \text{if smoke}=0.\\ 0 & \text{if smoke}=1,2,3.\end{cases} \qquad \text{former}=\begin{cases}1 & \text{if smoke}=1.\\ 0 & \text{if smoke}=0,2,3.\end{cases}$$

$$\text{light}=\begin{cases}1 & \text{if smoke}=2.\\ 0 & \text{if smoke}=0,1,3.\end{cases} \qquad \text{heavy}=\begin{cases}1 & \text{if smoke}=3.\\ 0 & \text{if smoke}=0,1,2.\end{cases}$$

Any *<u>three</u>* of these indicator variables can be used in the model if an intercept is included.

```
/* create dummy variables */
DATA bwt5;
    set bwt5;

    *** Create dummy variables ****;
    IF momsmoke = 'Never' THEN Never = 1; ELSE Never = 0;
    IF momsmoke = 'Former' THEN Former = 1; ELSE Former = 0;
    IF momsmoke = 'Light' THEN Light = 1; ELSE Light = 0;
    IF momsmoke = 'Heavy' THEN Heavy = 1; ELSE Heavy = 0;

    *** Create variable for two groups with current status ****;
    IF momsmoke = 'Never' THEN group = 0;
    IF momsmoke = 'Former' THEN group = 1;
    IF momsmoke = 'Light' THEN group = 2;
    IF momsmoke = 'Heavy' THEN group = 3;

    non = (group = 0 or group = 1);
    smoke = (group = 2 or group = 3);

RUN;
```

Notes on Using Indicator Variables

The reference category is the category associated with the indicator variable left out of the model (if specifying a model with an intercept).

Using never smoker as the reference category:

$$E[\text{birthweight}] = \beta_0 + \beta_{former} I_{former} + \beta_{light} I_{light} + \beta_{heavy} I_{heavy}$$

From this regression equation we can still determine the estimated mean for each group:

$$
\begin{aligned}
E[\text{birthweight}|\text{never}] &= \beta_0 &&= \mu_{never} \\
E[\text{birthweight}|\text{former}] &= \beta_0 + \beta_{former} &&= \mu_{former} \\
E[\text{birthweight}|\text{light}] &= \beta_0 + \beta_{light} &&= \mu_{light} \\
E[\text{birthweight}|\text{heavy}] &= \beta_0 + \beta_{heavy} &&= \mu_{heavy}
\end{aligned}
$$

The β’s can be used to estimate the difference between the mean of any two groups:

$$
\begin{aligned}
E[\text{birthweight}|\text{light}] - E[\text{birthweight}|\text{never}] &= (\beta_0 + \beta_{light}) - \beta_0 &&= \beta_{light} \\
E[\text{birthweight}|\text{heavy}] - E[\text{birthweight}|\text{never}] &= (\beta_0 + \beta_{heavy}) - \beta_0 &&= \beta_{heavy} \\
E[\text{birthweight}|\text{light}] - E[\text{birthweight}|\text{heavy}] &= (\beta_0 + \beta_{light}) - (\beta_0 + \beta_{heavy}) &&= \beta_{light} - \beta_{heavy}
\end{aligned}
$$

Notes on Using Indicator Variables (cont.)

From our regression equation, we can conduct the **Overall F-test** and make the direct connection to the one-way ANOVA:

$H_0: \beta_{former} = \beta_{light} = \beta_{heavy} = 0$ (Step 1: add β_0 to the H_0)

$H_0: \beta_{former} + \beta_0 = \beta_{light} + \beta_0 = \beta_{heavy} + \beta_0 = \beta_0$ (Step 2: substitute in definition for μ_x)

$H_0: \mu_{former} = \mu_{light} = \mu_{heavy} = \mu_{never}$

The intercept represents the level of the outcome in the reference category:

$$E[\text{birthweight}] = \beta_0 + \beta_{former} I_{former} + \beta_{light} I_{light} + \beta_{heavy} I_{heavy} \Rightarrow E[\text{birthweight} \mid \text{never}] = \beta_0$$

You can choose a different reference category by selecting which indicator variables are included in the model. For example, if we made heavy smoking mothers our reference category:

$$E[\text{birthweight}] = {\beta^*}_0 + \beta^*_{\boldsymbol{never}} I_{\boldsymbol{never}} + \beta^*_{former} I_{former} + \beta^*_{light} I_{light} \Rightarrow E[\text{birthweight} \mid \text{heavy}] = \beta^*_0$$

Notes on Using Indicator Variables – Testing A Category's Coefficient

The test of one category's coefficient is conceptually equivalent a *t*-test of that category against the reference category, *but* it isn't mathematically identical.

- Because we are using a "pooled variance" from all four categories/groups
- Not just the two groups we are comparing

The parameter estimates and some of the p-values for the parameter estimates will change if the reference category is changed.

The *F* test or partial *F* test can be used to test the overall significance of the categorical variable (this does not depend on the reference category).

- The *F* test and partial *F* test will **not** change if the reference category is changed.

Association between smoking and birthweight (reference group: never smokers)

```
/* REFERENCE CELL MODEL (reference group: never smokers) */
PROC REG DATA=bwt5;
    MODEL birthwt = former light heavy;
RUN;
```

$$E[\text{birthweight}] = \beta_0 + \beta_{former} I_{former} + \beta_{light} I_{light} + \beta_{heavy} I_{heavy}$$

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	**8.28550**	2.76183	2.57	0.0904
Error	16	17.18400	1.07400		
Corrected Total	19	25.46950			

SS explained by smoking status.

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	7.44000	0.46347	16.05	<.0001
Former	1	-0.20000	0.65544	-0.31	**0.7642**
Light	1	-1.26000	0.65544	-1.92	**0.0725**
Heavy	1	-1.48000	0.65544	-2.26	**0.0383**

H_0: $\beta_{former}=\beta_{light}=\beta_{heavy}=0$

or

H_0: $\beta_{former}+\beta_0=\beta_{light}+\beta_0=\beta_{heavy}+\beta_0 = \beta_0$

or

H_0: $\mu_{former}=\mu_{light}=\mu_{heavy}=\mu_{never}$

Global Hypotheses vs Multiple Comparisons

If you **reject** the null hypothesis H_0: $\mu_{former}=\mu_{light}=\mu_{heavy}=\mu_{never}$ (which we did **not**, p=0.0904)
AND you want to perform 6 additional tests, *then*:
Correct the alpha level, ***if*** you perform the additional tests (**Review Lectures 14-15**)

1. H_0: $\mu_{former} = \mu_{never}$ 2. H_0: $\mu_{light} = \mu_{never}$ 3. H_0: $\mu_{heavy} = \mu_{never}$	4. H_0: $\mu_{former} = \mu_{light}$ 5. H_0: $\mu_{former} = \mu_{heavy}$ 6. H_0: $\mu_{light} = \mu_{heavy}$

Parameter Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	
Intercept	1	7.44000	0.46347	16.05	<.0001	
Former	1	-0.20000	0.65544	-0.31	**0.7642**	$H_0: \mu_{former} = \mu_{never}$
Light	1	-1.26000	0.65544	-1.92	**0.0725**	$H_0: \mu_{light} = \mu_{never}$
Heavy	1	-1.48000	0.65544	-2.26	**0.0383**	$H_0: \mu_{heavy} = \mu_{never}$

This is the explanation as to why H_0: $\mu_{heavy} = \mu_{never}$ is rejected at alpha=0.05, *but* the overall F-test for H_0: $\mu_{former}=\mu_{light}=\mu_{heavy}=\mu_{never}$ is not significant.

Note: no multiple comparison correction is needed for the null hypothesis of all means are equal (H_0: $\mu_{former}=\mu_{light}=\mu_{heavy}=\mu_{never}$), because it is only 1 test.

Association between smoking and birthweight (reference group: never smokers)

```
PROC REG DATA=bwt5;
    MODEL birthwt = former light heavy / covb;
RUN;
```

$$E[\text{birthweight}] = \beta_0 + \beta_{former} I_{former} + \beta_{light} I_{light} + \beta_{heavy} I_{heavy}$$

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	**8.28550**	2.76183	2.57	0.0904
Error	16	17.18400	1.07400		
Corrected Total	19	25.46950			

SS explained by smoking status.

H_0: $\beta_{former}=\beta_{light}=\beta_{heavy}=0$

or

H_0: $\beta_{former}+\beta_0=\beta_{light}+\beta_0=\beta_{heavy}+\beta_0 = \beta_0$

or

H_0: $\mu_{former}=\mu_{light}=\mu_{heavy}=\mu_{never}$

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	7.44000	0.46347	16.05	<.0001
Former	1	-0.20000	0.65544	-0.31	**0.7642**
Light	1	-1.26000	0.65544	-1.92	**0.0725**
Heavy	1	-1.48000	0.65544	-2.26	**0.0383**

Overall test: Does smoking status (the *entire set* of indicator variables) contribute significantly to the prediction of birthweight?

H_0: $\beta_{former} = \beta_{light} = \beta_{heavy} = 0$; No, because F=2.57, p=0.0904.

REDUCED MODEL for Partial *F*

```
PROC REG DATA=bwt5;
    MODEL birthwt = ;
RUN;
```

$E[\text{birthweight}] = \beta_0 = \bar{Y}$

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	0	0	.	.	.
Error	19	25.46950	1.34050		
Corrected Total	19	25.46950			

Root MSE	1.15780	R-Square	0.0000
Dependent Mean	6.70500	Adj R-Sq	0.0000
Coeff Var	17.26771		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.70500	0.25889	25.90	<.0001

Overall *F* test (using Partial *F* test):

$$\frac{[SS_{model}(full) - SS_{model}(reduced)]/k}{MS_{error}(full)} = \frac{[SS_{model}(full) - 0]/k}{MS_{error}(full)} = \frac{MS_{model}(full)}{MS_{error}(full)} = \frac{2.76183}{1.07400} = \mathbf{2.57}$$

Association between smoking and birthweight (reference group: heavy smokers)

```
PROC REG DATA=bwt5;
    MODEL birthwt = never former light;
RUN;
```

$$E[\text{birthweight}] = \beta_0 + \beta_{never} I_{never} + \beta_{former} I_{former} + \beta_{light} I_{light}$$

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	8.28550	2.76183	2.57	0.0904
Error	16	17.18400	1.07400		
Corrected Total	19	25.46950			

H_0: $\beta_{never}=\beta_{former}=\beta_{light}=0$
or
H_0: $\mu_{former}=\mu_{light}=\mu_{heavy}=\mu_{never}$

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.96000	0.46347	12.86	<.0001
Never	1	1.48000	0.65544	2.26	0.0383
Former	1	1.28000	0.65544	1.95	0.0686
Light	1	0.22000	0.65544	0.34	0.7415

The overall F test does not depend on the choice of reference group used in the model. The parameter estimates table *does* depend on the choice of indicator variables. Note that the ANOVA table is identical to the results on slide 10, but the parameter estimates table has changed.

Tests of Individual Coefficients (reference group: never smokers)

```
PROC REG DATA=bwt5;
    MODEL birthwt = former light heavy / covb;
RUN;
```

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	8.28550	2.76183	2.57	0.0904
Error	16	17.18400	1.07400		
Corrected Total	19	25.46950			

H_0: $\beta_{former}=\beta_{light}=\beta_{heavy}=0$

or

H_0: $\mu_{former}=\mu_{light}=\mu_{heavy}=\mu_{never}$

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	7.44000	0.46347	16.05	<.0001
Former	1	-0.20000	0.65544	-0.31	0.7642
Light	1	-1.26000	0.65544	-1.92	0.0725
Heavy	1	-1.48000	0.65544	-2.26	0.0383

Covariance of Estimates				
Variable	**Intercept**	**Former**	**Light**	**Heavy**
Intercept	0.2148	-0.2148	-0.2148	-0.2148
Former	-0.2148	0.4296	0.2148	0.2148
Light	-0.2148	0.2148	0.4296	0.2148
Heavy	-0.2148	0.2148	0.2148	0.4296

$$\boldsymbol{\Sigma} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2_{Y|X}$$

$$\hat{Y} = 7.44 + (-0.20) \times \text{former} + (-1.26) \times \text{light} + (-1.48) \times \text{heavy}$$

$$\hat{Y} = 7.44 + (-0.20) \times \text{former} + (-1.26) \times \text{light} + (-1.48) \times \text{heavy}$$

What is the interpretation of the intercept?

This is the expected mean birthweight for the reference group (non-smokers) or expected birthweight for an individual baby born to a non-smoking mother.

$\hat{Y}$ = 7.44 + (-0.20)×0 +(-1.26)×0 + (-1.48)×0 = 7.44 lbs

What is the expected birthweight for former smokers?

$\hat{Y}$ = 7.440 + (-0.20)×**1** +(-1.26)×0 + (-1.48)×0 = 7.24 lbs

What is the expected birthweight for heavy smokers?

$\hat{Y}$ = 7.44 + (-0.20)×0 +(-1.26)×0 + (-1.48)×**1** = 5.96 lbs

What is the difference in expected birthweight between heavy smokers and never smokers?

E[birthweight|heavy]-E[birthweight|never] = $\left(\beta_0 + \beta_{heavy}\right) - \beta_0 = \beta_{heavy}$
t = $\hat{\beta}_{heavy}/SE\left(\hat{\beta}_{heavy}\right)$ = -1.48/0.65544 = -2.26, p = 0.0383

Why isn't this mathematically the same as an independent samples t-test?

Because we are using a "pooled variance" from all four smoking categories/groups, not just the two groups we are comparing

Form of the Variance Covariance Matrix

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$$Var(\hat{\beta}) = \hat{\sigma}^2_{Y|X}(X^T X)^{-1}$$

$$X^T X = \begin{bmatrix} 20 & 5 & 5 & 5 \\ 5 & 5 & 0 & 0 \\ 5 & 0 & 5 & 0 \\ 5 & 0 & 0 & 5 \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} 0.2 & -0.2 & -0.2 & -0.2 \\ -0.2 & 0.4 & 0.2 & 0.2 \\ -0.2 & 0.2 & 0.4 & 0.2 \\ -0.2 & 0.2 & 0.2 & 0.4 \end{bmatrix}$$

$$E[\text{birthweight}] = \beta_0 + \beta_{former} I_{former} + \beta_{light} I_{light} + \beta_{heavy} I_{heavy}$$

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	7.44000	0.46347	16.05	<.0001
Former	1	-0.20000	0.65544	-0.31	0.7642
Light	1	-1.26000	0.65544	-1.92	0.0725
Heavy	1	-1.48000	0.65544	-2.26	0.0383

Covariance of Estimates				
Variable	**Intercept**	**Former**	**Light**	**Heavy**
Intercept	0.2148	-0.2148	-0.2148	-0.2148
Former	-0.2148	0.4296	0.2148	0.2148
Light	-0.2148	0.2148	0.4296	0.2148
Heavy	-0.2148	0.2148	0.2148	0.4296

What is the difference in average birthweight between heavy smokers and light smokers?

E[birthweight|heavy]-E[birthweight|light] = $\left(\beta_0 + \beta_{\text{heavy}}\right) - \left(\beta_0 + \beta_{\text{light}}\right) = \beta_{\text{heavy}} - \beta_{\text{light}}$

Then $\hat{\beta}_{\text{heavy}}$-$\hat{\beta}_{\text{light}}$= -1.48 – (-1.26) = -0.22

Is this difference significantly different from zero?

$$t = \frac{\hat{\beta}_{heavy} - \hat{\beta}_{light}}{SE\left(\hat{\beta}_{heavy} - \hat{\beta}_{light}\right)} = \frac{\hat{\beta}_{heavy} - \hat{\beta}_{light}}{\sqrt{Var\left(\hat{\beta}_{heavy}\right) + Var\left(\hat{\beta}_{light}\right) - 2Cov\left(\hat{\beta}_{heavy}, \hat{\beta}_{light}\right)}}$$

$$= \frac{-1.48 - (-1.26)}{\sqrt{\mathbf{0.4296} + \mathbf{0.4296} - 2 * \mathbf{0.2148}}} = \frac{-0.22}{\sqrt{0.4296}} = 0.336 \sim t_{16}; p = 0.742$$

Model treating Smoking Status as a *continuous* variable (no dummy codes):

```
PROC REG DATA=bwt5;
    MODEL birthwt = group;
RUN;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	7.56250	7.56250	7.60	0.0130
Error	18	17.90700	0.99483		
Corrected Total	19	25.46950			

H_0: $\beta_{smkgroup} = 0$

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	7.53000	0.37320	20.18	<.0001
group	1	-0.55000	0.19948	-2.76	0.0130

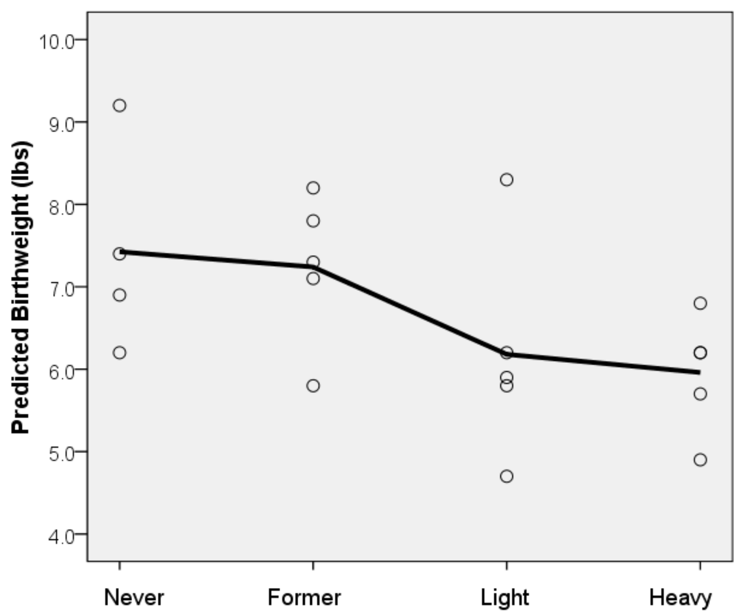
$$\hat{Y} = 7.53 + (-0.55) \times Group$$

$$\hat{Y} = 7.53 + (-0.55) \times Group$$

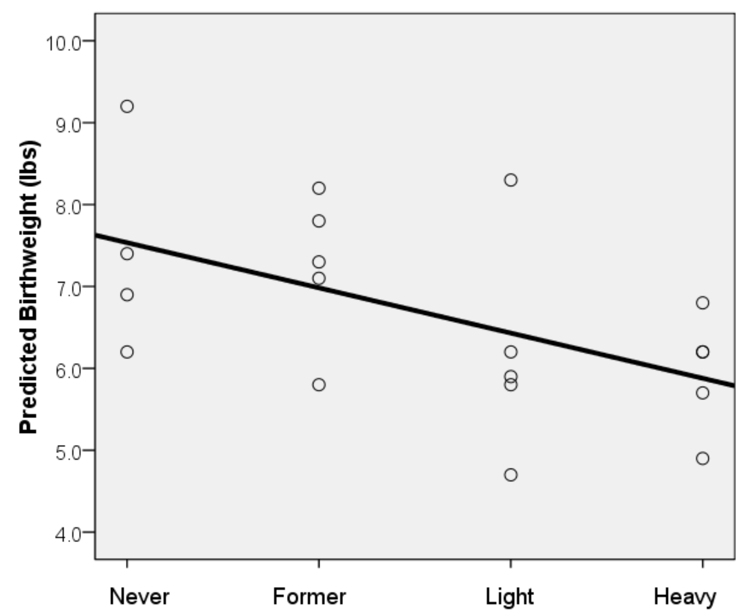
Interpretation of the intercept? Expected birthweight for a non-smoking mother is 7.53 lbs.
$E[\text{birthweight}] = \beta_0 + \beta_{group} Group \Rightarrow E[\text{birthweight} \,|\, \text{never}] = \beta_0$

Interpretation of $\hat{\beta}_1$? On average, birthweight decreases by 0.55 pounds for every category increase in smoking status (assumed to be the same increase between all adjacent categories).

Predicted Model using Dummy Coding (using 4 degrees of freedom)



Predicted Model using Continuous Variable (using 2 degrees of freedom)



C. Tests of General Linear Hypotheses

Tests on individual regression parameters and on subsets of parameters can be put into a more general framework that allows much more flexibility by the use of the general linear hypothesis:

$$H_0: \mathbf{c}\boldsymbol{\beta} = \mathbf{d}$$
$$H_1: \mathbf{c}\boldsymbol{\beta} \neq \mathbf{d}$$

The matrix **c** is an $r \times p^*$ matrix that is of rank r and $r \leq p^*$, where

- $p^* = p$ when an intercept is included in the model
- $p^* = p$-1 for a no intercept model.

In other words, we can postulate $r \leq p^*$ non-redundant and non-contradictory statements about the parameters.

We can use this framework to test a single parameter:

$$H_0: \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = 0 \Rightarrow H_0: \beta_2 = 0$$

We can use this framework to compare two or more parameters:

$$H_0: \begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = 0 \Rightarrow H_0: \beta_1 - \beta_2 = 0 \text{ or } H_0: \beta_1 = \beta_2$$

We can use this framework for simultaneous hypothesis tests:

$$H_0: \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow H_0: \begin{matrix} \beta_2 = 0 \\ \beta_1 = 0 \end{matrix}$$

$$H_0: \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow H_0: \begin{matrix} \beta_1 - \beta_2 = 0 \\ \beta_1 - \beta_3 = 0 \end{matrix} \text{ or } H_0: \begin{matrix} \beta_1 = \beta_2 \\ \beta_1 = \beta_3 \end{matrix}$$

The *F*-test can be used to test our general linear hypotheses:

$$F = \frac{[(\boldsymbol{c}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})'(\boldsymbol{c}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}')^{-1}(\boldsymbol{c}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})/r]}{\hat{\sigma}^2_{Y|X}} \sim F_{r,n-p-1}$$

where *r* is the number of linear combinations of parameters we wish to test (which is equal to the number of rows in **c**).

This reduces to our Partial *F* test for testing a group of variables, since:

$$(\boldsymbol{c}\widehat{\boldsymbol{\beta}})'(\boldsymbol{c}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}')^{-1}(\boldsymbol{c}\widehat{\boldsymbol{\beta}}) = SS_{model}(full) - SS_{model}(reduced).$$

And reduces to our *t* test for a single parameter (or test of a single linear hypothesis):

$$t = \frac{\boldsymbol{c}\widehat{\boldsymbol{\beta}}}{\sqrt{\boldsymbol{c}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}'\hat{\sigma}^2_{Y|X}}}.$$

Recall:
$(X`X)^{-1}\hat{\sigma}^2_{Y|X} = \Sigma$

In terms of the covariance matrix of ***β*, Σ**, the *F* and *t* tests become:

$$F = (\boldsymbol{c}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})'(\boldsymbol{c}\sum\boldsymbol{c}')^{-1}(\boldsymbol{c}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})/r \sim F_{r,n-1-p}$$

$$t = (\boldsymbol{c}\widehat{\boldsymbol{\beta}})\left(\sqrt{\boldsymbol{c}\sum\boldsymbol{c}'}\right)^{-1} \sim t_{n-1-p}$$

Tests of General Linear Hypotheses: Example

$$E[\text{birthweight}] = \beta_0 + \beta_{former} I_{former} + \beta_{light} I_{light} + \beta_{heavy} I_{heavy}$$

We want to test the hypothesis: H_0: $\beta_{heavy} = \beta_{light}$, or equivalently H_0: $\beta_{heavy} - \beta_{light} = 0$

Which can also be written as: $H_0: (0 \quad 0 \quad -1 \quad 1)\begin{pmatrix} \beta_0 \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix} = 0$

$$\boldsymbol{b} = \widehat{\boldsymbol{\beta}} = \begin{pmatrix} 7.440 \\ -0.200 \\ -1.260 \\ -1.480 \end{pmatrix} \quad \boldsymbol{\Sigma} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\hat{\sigma}^2_{Y|X} = \begin{pmatrix} 0.2148 & -0.2148 & -0.2148 & -0.2148 \\ -0.2148 & 0.4296 & 0.2148 & 0.2148 \\ -0.2148 & 0.2148 & 0.4296 & 0.2148 \\ -0.2148 & 0.2148 & 0.2148 & 0.4296 \end{pmatrix}$$

$$t = (\boldsymbol{cb})\left(\sqrt{\boldsymbol{c\Sigma c}'}\right)^{-1} \quad t = (0 \quad 0 \quad -1 \quad 1)\begin{pmatrix} 7.440 \\ -0.200 \\ -1.260 \\ -1.480 \end{pmatrix}\left(\sqrt{(0 \quad 0 \quad -1 \quad 1)\begin{pmatrix} 0.2148 & -0.2148 & -0.2148 & -0.2148 \\ -0.2148 & 0.4296 & 0.2148 & 0.2148 \\ -0.2148 & 0.2148 & 0.4296 & 0.2148 \\ -0.2148 & 0.2148 & 0.2148 & 0.4296 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \end{pmatrix}}\right)^{-1}$$

$$t = \left(-1.260 - (-1.480)\right)\left(\sqrt{(0.4296 + 0.4296 - 2 \times 0.2148)}\right)^{-1} = \frac{0.22}{0.6554} = 0.336 \sim t_{16}$$

Testing Generalized Linear Hypotheses in SAS with Matrices using PROC IML

```
PROC IML;
 beta = {7.44000, -0.20000,  -1.26000,  -1.48000};

 sigma ={0.2148         -0.2148         -0.2148         -0.2148,
        -0.2148          0.4296          0.2148          0.2148,
        -0.2148          0.2148          0.4296          0.2148,
        -0.2148          0.2148          0.2148          0.4296};

PRINT "t statistic for b(heavy)=b(light)";
c = {0 0 -1 1};
t = (c*beta)*INV(SQRT(c*sigma*c`));
PRINT t;

PRINT "F statistic for b(former)=b(heavy)=b(light)= 0";
c = {0 1 0 0, 0 0 1 0, 0 0 0 1};
F = (c*beta)`*INV(c*sigma*c`)*(c*beta)/NROW(c);
PRINT F;
```

t statistic for b(heavy)=b(light)

t
-0.335653

F statistic for b(former)=b(heavy)=b(light)= 0

F
2.5715394

Testing General Linear Hypotheses Directly in SAS

```
PROC REG DATA=bwt5;
    MODEL birthwt = former light heavy;

    /* these 3 statement request the equivalent test */
    TEST light = heavy;
    TEST light-heavy;
    TEST light-heavy=0;

    /* these 3 statement request the equivalent test */
    TEST former=light=heavy=0;
    TEST former,light,heavy;
    test former=0, light=0, heavy=0;
RUN;
```

Test 1 Results for Dependent Variable birthwt

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.12100	0.11	0.7415
Denominator	16	1.07400		

Test 4 Results for Dependent Variable birthwt

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	2.76183	2.57	0.0904
Denominator	16	1.07400		

Note: Tests 2 and 3 results are identical to Test 1 and are not shown here. Similarly, Tests 5 and 6 are identical to Test 4 and are not shown here.

Compare Test 4’s F- and p-value to ANOVA table on Slides 8/10.

D. Linear Contrasts

A linear contrast (L) is any linear combination of the parameters such that the linear coefficients add up to 0. Specifically,

$$L = \sum_{i=1}^{k} c_i \bar{y}_i \quad \text{where} \quad \sum_{i=1}^{k} c_i = 0$$

$$Var(L) = \hat{\sigma}_{Y|X}^2 \sum_{i=1}^{k} c_i^2/n_i$$

Different coding schemes can be used:

- Reference Cell (Dummy codes)
- Cell Means (No Intercept)
- Effect Coding (Design coding)
- Orthogonal Polynomial Coding (Lectures 23-24)

Linear contrasts are most often used to test linear combinations of group means in a Cell Means Model (a model which includes a dummy code for each category/group and specifies no intercept in the model).

A *t*-statistic can be used to test a single linear contrast, and an *F*-statistic can be used for testing several linear contrasts simultaneously: $t = \frac{L}{SE(L)}$

We can show Var(L) from the previous slide applying the various properties we have learned throughout the semester:

Independent means:
Covariances are 0

$$Var(L) = Var\left(\sum_{i=1}^{k} c_i \bar{y}_i\right)$$

$$= c_1^2\, var(\bar{y}_1) + c_2^2\, var(\bar{y}_2) + \ldots + c_k^2\, var(\bar{y}_k) + 2c_1 c_2\, cov(\bar{y}_1, \bar{y}_2) + \ldots + 2c_{k-1} c_k\, cov(\bar{y}_{k-1}, \bar{y}_k)$$

$$= c_1^2\, var(\bar{y}_1) + c_2^2\, var(\bar{y}_2) + \ldots + c_k^2\, var(\bar{y}_k)$$

$$= \frac{c_1^2}{n_1} var(y_1) + \frac{c_2^2}{n_2} var(y_2) + \ldots + \frac{c_k^2}{n_k} var(y_k)$$

$$= \frac{c_1^2}{n_1} var(y|x=1) + \frac{c_2^2}{n_2} var(y|x=2) + \ldots + \frac{c_k^2}{n_k} var(y|x=3)$$

Assume all variances are equal

$$= var(y|x) \sum_{i=1}^{k} \left(c_i^2 / n_i\right)$$

$$= \hat{\sigma}_{Y|X}^2 \sum_{i=1}^{k} \left(c_i^2 / n_i\right)$$

Linear Contrasts (cont.)

<u>Orthogonal contrasts</u>: Two contrasts, L_A and L_B, are orthogonal to one another if:

$$\sum_{i=1}^{k} \frac{c_{Ai}c_{Bi}}{n_i} = 0 \quad \text{or} \quad \sum_{i=1}^{k} c_{Ai}c_{Bi} = 0 \text{ (when the } n_i\text{'s are equal.)}$$

Orthogonality is a desirable property because the Model sums of squares can then be partitioned into statistically independent sums of squares, where the sums of squares for a given contrast, L, is given by:

$$SS(\hat{L}) = \frac{(\hat{L})^2}{\sum_{i=1}^{k} c_i^2/n_i}$$

$$\frac{SS(\hat{L})}{MSE} \sim F_{1,n-k}$$

For a cell means model, the number of orthogonal contrasts cannot exceed the group degrees of freedom (i.e., the number of groups minus 1).

Cell Means Model Example: Mother's Smoking Status and Birthweight

```
PROC REG DATA=bwt5;
    MODEL birthwt = never former light heavy / noint;
RUN;
```

`NOTE: No intercept in model. R-Square is redefined.` It uses the uncorrected sum of squares and is not meaningful to compare to the R^2 from models which include an intercept.

$$E[birthweight] = \beta_{never} I_{never} + \beta_{former} I_{former} + \beta_{light} I_{light} + \beta_{heavy} I_{heavy}$$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	907.42600	226.85650	211.23	<.0001
Error	16	17.18400	1.07400		
Uncorrected Total	20	924.61000			

H_0: $\beta_{never}=\beta_{former}=\beta_{light}=\beta_{heavy}=0$

or

H_0: $\mu_{never}=\mu_{former}=\mu_{light}=\mu_{heavy}=0$

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Never	1	7.44000	0.46347	16.05	<.0001
Former	1	7.24000	0.46347	15.62	<.0001
Light	1	6.18000	0.46347	13.33	<.0001
Heavy	1	5.96000	0.46347	12.86	<.0001

Group Means

Notes on Using Indicator Variables – No Intercept

Model **without** an intercept:

$$E[\text{birthweight}] = \beta_{never} I_{never} + \beta_{former} I_{former} + \beta_{light} I_{light} + \beta_{heavy} I_{heavy}$$

$$E[\text{birthweight} \,|\, \text{never}] = \beta_{never} = \mu_{never}$$
$$E[\text{birthweight} \,|\, \text{former}] = \beta_{former} = \mu_{former}$$
$$E[\text{birthweight} \,|\, \text{light}] = \beta_{light} = \mu_{light}$$
$$E[\text{birthweight} \,|\, \text{heavy}] = \beta_{heavy} = \mu_{heavy}$$

$$E[\text{birthweight} \,|\, \text{former}] - E[\text{birthweight} \,|\, \text{never}] = \beta_{former} - \beta_{never}$$
$$E[\text{birthweight} \,|\, \text{light}] - E[\text{birthweight} \,|\, \text{never}] = \beta_{light} - \beta_{never}$$
$$E[\text{birthweight} \,|\, \text{heavy}] - E[\text{birthweight} \,|\, \text{never}] = \beta_{heavy} - \beta_{never}$$

Overall F-test for the model without an intercept:

$$H_0: \beta_{never} = \beta_{former} = \beta_{light} = \beta_{heavy} = 0 \Rightarrow H_0: \mu_{former} = \mu_{light} = \mu_{heavy} = \mu_{never} = 0$$

Orthogonal Contrasts: Examples

For each set of three linear contrasts, what hypotheses are being tested? Are the contrasts orthogonal?

1. $$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_{never} \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} \beta_{never} - \beta_{former} \\ \beta_{never} - \beta_{light} \\ \beta_{never} - \beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Not Orthogonal [$1 \times 1 + (-1) \times 0 + 0 \times (-1) + 0 \times 0 = 1$ (row 1 and row2)]

2. $$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \beta_{never} \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} \beta_{never} - \beta_{former} \\ \beta_{light} - \beta_{heavy} \\ \beta_{never} + \beta_{former} - \beta_{light} - \beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Orthogonal

2b. $$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0.5 & 0.5 & -0.5 & -0.5 \end{pmatrix} \begin{pmatrix} \beta_{never} \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} \beta_{never} - \beta_{former} \\ \beta_{light} - \beta_{heavy} \\ \frac{\beta_{never} + \beta_{former}}{2} - \frac{\beta_{light} + \beta_{heavy}}{2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Orthogonal

Orthogonal Contrasts: Examples

3. $$\begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}\begin{pmatrix} \beta_{never} \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} \beta_{former} - \beta_{never} \\ \beta_{light} - \beta_{former} \\ \beta_{heavy} - \beta_{light} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Not Orthogonal

4. $$\begin{pmatrix} -3 & 1 & 1 & 1 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix}\begin{pmatrix} \beta_{never} \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} -3\beta_{never} + \beta_{former} + \beta_{light} + \beta_{heavy} \\ -2\beta_{former} + \beta_{light} + \beta_{heavy} \\ -\beta_{light} + \beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Orthogonal

5. $$\begin{pmatrix} -3 & -1 & 1 & 3 \\ 1 & -1 & -1 & 1 \\ -1 & 3 & -3 & 1 \end{pmatrix}\begin{pmatrix} \beta_{never} \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} -3\beta_{never} - 1\beta_{former} + 1\beta_{light} + 3\beta_{heavy} \\ 1\beta_{never} - 1\beta_{former} - 1\beta_{light} + 1\beta_{heavy} \\ -1\beta_{never} + 3\beta_{former} - 3\beta_{light} + 1\beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Orthogonal {Orthogonal Polynomials}

NOTE: These three contrasts are providing the same information as the reference cell model, comparing each smoking group to the never smokers (which we will see again in two slides).

$$1.\ \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_{never} \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Not orthogonal

```
PROC REG DATA=bwt5;
   MODEL birthwt=never former light heavy/noint;
   TEST never-former; * row 1 ;
   TEST never-light;  * row 2 ;
   TEST never-heavy;  * row 3 ;
   TEST never-former, never-light, never-heavy;
RUN;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	907.42600	226.85650	211.23	<.0001
Error	16	17.18400	1.07400		
Uncorrected Total	20	924.61000			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Never	1	7.44000	0.46347	16.05	<.0001
Former	1	7.24000	0.46347	15.62	<.0001
Light	1	6.18000	0.46347	13.33	<.0001
Heavy	1	5.96000	0.46347	12.86	<.0001

ANOVA $H_0: \beta_{heavy} = \beta_{former} = \beta_{light} = \beta_{never} = 0$

$$E[\text{birthweight}] = \beta_{never} I_{never} + \beta_{former} I_{former} + \beta_{light} I_{light} + \beta_{heavy} I_{heavy}$$

Test 1. never-former, $H_0\colon \mu_{never} = \mu_{former}$

Test 1 Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	0.10000	0.09	0.7642
Denominator	16	1.07400		

SS(contrast)
= MS(contrast)*df
= MS(contrast)*1
= MS(contrast)

Test 2. never-light, $H_0\colon \mu_{never} = \mu_{light}$

Test 2 Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	3.96900	3.70	0.0725
Denominator	16	1.07400		

Compare to t tests of betas on next page

Test 3. never-heavy, $H_0\colon \mu_{never} = \mu_{heavy}$

Test 3 Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	5.47600	5.10	0.0383
Denominator	16	1.07400		

$$\sum SS(contrast) = 0.1 \times 1 + 3.969 \times 1 + 5.476 \times 1 = 9.545 \neq 8.2855$$

Test 4. never-former, never-light, never-heavy

$H_0\colon \beta_{never} - \beta_{former} = \beta_{never} - \beta_{light} = \beta_{never} - \beta_{heavy} = 0 \Rightarrow H_0\colon \mu_{never} = \mu_{former} = \mu_{light} = \mu_{heavy}$

Test 4 Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	3	2.76183	2.57	0.0904
Denominator	16	1.07400		

$$\sum SS(contrast) = 2.76183 \times 3 = 8.2855$$

8.2855 is the SS explained by smoking status.

Reference Cell Model Comparison with Cell Means Model for Orthogonal Contrast Example 1

```
PROC REG DATA=bwt5;
 MODEL weight = former light heavy;
RUN;
```

$$E[\text{birthweight}] = \beta_0 + \beta_{former} I_{former} + \beta_{light} I_{light} + \beta_{heavy} I_{heavy}$$

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	8.28550	2.76183	2.57	0.0904
Error	16	17.18400	1.07400		
Corrected Total	19	25.46950			

SS explained by smoking status. Compare to Previous Page

H_0: $\beta_{former}=\beta_{light}=\beta_{heavy}=0$

or

H_0: $\mu_{former}=\mu_{light}=\mu_{heavy}=\mu_{never}$

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	7.44000	0.46347	16.05	<.0001
Former	1	-0.20000	0.65544	-0.31	**0.7642**
Light	1	-1.26000	0.65544	-1.92	**0.0725**
Heavy	1	-1.48000	0.65544	-2.26	**0.0383**

Compare to F tests on previous page

2. $\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \beta_{never} \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ **Orthogonal**

```
PROC REG DATA=bwt5;
    MODEL birthwt=never former light heavy/noint;
    TEST never-former; * row 1 ;
    TEST light-heavy;  * row 2 ;
    TEST never+former-light-heavy;  * row 3 ;
    TEST never-former, light-heavy, never+former-light-heavy;
RUN;
```

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	4	907.42600	226.85650	211.23	<.0001
Error	16	17.18400	1.07400		
Uncorrected Total	20	924.61000			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Never	1	7.44000	0.46347	16.05	<.0001
Former	1	7.24000	0.46347	15.62	<.0001
Light	1	6.18000	0.46347	13.33	<.0001
Heavy	1	5.96000	0.46347	12.86	<.0001

$$E[\text{birthweight}] = \beta_{never} I_{never} + \beta_{former} I_{former} + \beta_{light} I_{light} + \beta_{heavy} I_{heavy}$$

Test 1. never-former, $H_0: \mu_{never} = \mu_{former}$

SS(contrast)
= MS(contrast)*df
= MS(contrast)*1
= MS(contrast)

Test 1 Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.10000	0.09	0.7642
Denominator	16	1.07400		

Compare to t tests of betas on next page

Test 2. never-light, $H_0: \mu_{light} = \mu_{heavy}$

Test 2 Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.12100	0.11	0.7415
Denominator	16	1.07400		

Test 3. never-heavy, $H_0: \mu_{never} + \mu_{former} = \mu_{light} + \mu_{heavy}$

Test 3 Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	8.06450	7.51	0.0145
Denominator	16	1.07400		

$$\sum SS(contrast) = 0.1 \times 1 + 0.121 \times 1 + 8.0645 \times 1 = 8.2855$$

Test 4. never-former, never-light, never-heavy

$H_0: \mu_{never} - \mu_{former} = \mu_{light} - \mu_{heavy} = \mu_{never} + \mu_{former} - \mu_{light} - \mu_{heavy}$

Test 4 Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	3	2.76183	2.57	0.0904
Denominator	16	1.07400		

$$\sum SS(contrast) = 2.76183 \times 3 = 8.2855$$

Calculate the value of a contrast by hand

3rd contrast from example 2: $\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \beta_{never} \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix}$

L = (1)×7.44 + (1)×7.24 + (-1)×6.18 + (-1)×5.96 = 2.540

$$Var(L) = \hat{\sigma}^2_{Y|X} \sum_{i=1}^{k} c_i^2/n_i$$

Var(L) = 1.074 × [$(1)^2$/5 + $(1)^2$/5 +$(-1)^2$/5 +$(-1)^2$/5] = 0.8592

t = 2.540 / √(0.8592) = 2.540/0.92693 = 2.7402 ~ t_{16}

F = 2.7402^2 = 7.509; p = 0.0145

Alternatively, we can also use the estimates provided in our ANOVA table:

$$SS(\hat{L}) = \frac{(\hat{L})^2}{\sum_{i=1}^{k} c_i^2/n_i} = \frac{2.54^2}{[(1)^2/5 + (1)^2/5 + (\text{-}1)^2/5 + (\text{-}1)^2/5]} = \frac{2.54^2}{0.8} = 8.0645$$

$$\frac{SS(\hat{L})}{MSE} = \frac{8.0645}{1.074} = 7.509 \sim F_{1,16}$$

What is the null hypothesis being tested by the third contrast:

$$\begin{pmatrix} 1 & 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \beta_{never} \\ \beta_{former} \\ \beta_{light} \\ \beta_{heavy} \end{pmatrix}$$

1. In terms of the βs?

$\beta_{never} + \beta_{former} - \beta_{light} - \beta_{heavy} = 0$
$\beta_{never} + \beta_{former} = \beta_{light} + \beta_{heavy}$
$\frac{1}{2}(\beta_{never} + \beta_{former}) = \frac{1}{2}(\beta_{light} + \beta_{heavy})$

2. In terms of the 4 population means?

$\mu_{never} + \mu_{former} - \mu_{light} - \mu_{heavy} = 0$
$\frac{1}{2}(\mu_{never} + \mu_{former}) = \frac{1}{2}(\mu_{light} + \mu_{heavy})$

TEST: $\frac{1}{2}(\beta_{never} + \beta_{former}) = \frac{1}{2}(\beta_{light} + \beta_{heavy})$

$L = (0.5)\times7.44 + (0.5)\times7.24 + (-0.5)\times6.18 + (-0.5)\times5.96 = 1.27$ lbs

$\mathrm{Var}(L) = 1.074 \times [(0.5)^2/5 + (0.5)^2/5 + (-0.5)^2/5 + (-0.5)^2/5] = 0.2148$

$t = 1.27 / \sqrt{(0.2148)} = 2.7402 \sim t_{16}$

$p = 0.0145$ (equivalent to previous results)

3. Can the null hypothesis for this contrast be written in terms of 2 population means (non-smokers and current smokers)? What assumptions are being made?

$$\mu_{non} = \mu_{smoker}$$

We are assuming that the sample of non-smokers (current plus former) is representative of the population of non-smokers.

But since we the investigator didn't randomly select non-smokers (the investigator chose 5 never and 5 former smokers or 50% of each in our contrast) the observed average ($\bar{y}_{non}$) for the non-smokers probably isn't equal to the population mean.

Now test the linear contrast, assuming 25% of non-smokers in the population are former smokers and 50% of current smokers in the population are heavy smokers:

$$L = (0.75)\times 7.44 + (0.25)\times 7.24 + (-0.5)\times 6.18 + (-0.5)\times 5.96 = 1.32 \text{ lbs}$$

$$\text{Var}(L) = 1.074 \times [(-0.75)^2/5 + (-0.25)^2/5 + (0.5)^2/5 + (0.5)^2/5] = 1.074*0.225$$

$$t = 1.32 / \sqrt{(0.24165)} = 2.6852 \sim t_{16}$$

$$F = 2.6852^2 = 7.2104 \text{ and } p = 0.0163$$

```
PROC REG DATA=bwt5;
    MODEL birthwt=never former light heavy/noint;
    TEST .75*never + .25*former - .5*light - .5*heavy;
RUN;
```

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	907.42600	226.85650	211.23	<.0001
Error	16	17.18400	1.07400		
Uncorrected Total	20	924.61000			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Never	1	7.44000	0.46347	16.05	<.0001
Former	1	7.24000	0.46347	15.62	<.0001
Light	1	6.18000	0.46347	13.33	<.0001
Heavy	1	5.96000	0.46347	12.86	<.0001

Test 1 Results for Dependent Variable birthwt

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	7.74400	7.21	0.0163
Denominator	16	1.07400		

What is the difference between fitting a cell means model and testing H_0: $\mu_{never} + \mu_{former} - \mu_{light} - \mu_{heavy} = 0$, and testing H_0: $\mu_{non} - \mu_{smoke} = 0$ by estimating the following regression model:

$$Y = \beta_0 + \beta_1 \times non$$

where *non* is coded 1 for non-smokers (never or former) and 0 for current smokers (light or heavy)?

```
PROC REG DATA=bwt5;
    MODEL birthwt=non;
RUN;
```

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	8.06450	8.06450	8.34	0.0098
Error	18	17.40500	0.96694		
Corrected Total	19	25.46950			

NOTE: Different MSE than cell means model which fit all four group means. The SSE must stay the same or increase (it increased slightly in this example) when combining groups, but the degrees of freedom also increase, and thus the MSE in this example is actually smaller than the cell means model. In practice, the MSE usually increases when combining groups.

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	6.07000	0.31096	19.52	<.0001
non	1	1.27000	0.43976	2.89	0.0098

NOTE: Same point estimate as the linear contrast on page 38, but different SE due to the different MSE.

How can we replicate the F-Value and p-value from the contrast statement on slide 37 [(1 1 -1 -1) *replicated below*]?

Test 3 Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	8.06450	7.51	0.0145
Denominator	16	1.07400		

$$SE\left(\hat{\beta}_{non}\right) = \sqrt{\frac{\hat{\sigma}^2_{Y|X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2}} = 0.43976$$

$$t = \frac{\hat{\beta}_{non}}{\sqrt{\frac{MSE(full)}{MSE(reduced)}\left[SE\left(\hat{\beta}_{non}\right)\right]^2}} = \frac{1.27}{\sqrt{\frac{0.96694}{1.074}(0.43976)^2}} = 2.74022$$

$$F = t^2 = 7.5088$$

Unequal Sample Sizes

Finally, note that if we had unequal sample sizes in our groups, then we would also get a different SS(L) and a different mean difference by testing:

$$H_0: 0.5\mu_{never} + 0.5\mu_{former} - 0.5\mu_{light} - 0.5\mu_{heavy} = 0$$

versus

$$H_0: \mu_{non} - \mu_{smoke} = 0$$

i	*Never*	*Former*	*Light*	*Heavy*
1	7.50	5.80	5.90	6.20
2	6.20	7.30	6.20	6.80
3	6.90	8.20	5.80	5.70
4	7.40	7.10	4.70	4.90
5	9.20	7.80	8.30	6.20
6	8.30		7.20	7.10
7	7.60		6.20	5.80
8				5.40
$\bar{Y}_j$=	7.586	7.240	6.329	6.013

What linear contrast would give us the same SS(L) as testing: $H_0: \mu_{non} - \mu_{smoke} = 0$? (Although still with a different MSE).

$$H_0: \frac{7}{12}\mu_{never} + \frac{5}{12}\mu_{former} - \frac{7}{15}\mu_{light} - \frac{8}{15}\mu_{heavy} = 0$$

E. Orthogonal Polynomials

Orthogonal polynomials are a new set of independent variables that are defined in terms of the simple polynomials (e.g., X, X^2, X^3; natural polynomials will be discussed in a future lecture) but have more complicated structures. It avoids the serious collinearity inherent in using natural polynomials since orthogonal polynomials are pairwise uncorrelated.

The orthogonal polynomial variables contain exactly the same information as the simple polynomial variables, but unlike the simple polynomial variables, the orthogonal polynomial variables are uncorrelated with each other.

> Orthogonality is a desirable property because the Model Sums of Squares can then be partitioned into statistically independent sums of squares for each polynomial contrast (linear, quadratic, etc.).

The orthogonal polynomials can also be used to perform linear contrasts in a cell means model.

Table A7 of KKNR provides the orthogonal polynomial coefficients for equally spaced predictor values with the same number of replicates at each value.

Example:

k=4	**X**				
	1	**2**	**3**	**4**	**Σp_i^2**
Linear	-3	-1	1	3	20
Quadratic	1	-1	-1	1	4
Cubic	-1	3	-3	1	20

Example (Orthogonal Polynomial Contrasts, EQUAL N):

	Never Smokers (X=0)	*Former Smokers (X=1)*	*Light Smokers (X=2)*	*Heavy Smokers (X=3)*
	7.50	5.80	5.90	6.20
	6.20	7.30	6.20	6.80
	6.90	8.20	5.80	5.70
	7.40	7.10	4.70	4.90
	9.20	7.80	8.30	6.20
$\bar{Y} \mid X$	7.44	7.24	6.18	5.96

```
PROC REG DATA=bwt5;
    MODEL birthwt=never former light heavy/noint;
    Overall:   TEST never=former=light=heavy;
    Linear:    TEST -3*never -1*former +1*light +3*heavy=0;
    Quadratic: TEST  1*never -1*former -1*light +1*heavy=0;
    Cubic:     TEST -1*never +3*former -3*light +1*heavy=0;
RUN;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	907.42600	226.85650	211.23	<.0001
Error	16	17.18400	1.07400		
Uncorrected Total	20	924.61000			

H_0: $\beta_{former}=\beta_{light}=\beta_{heavy}=\beta_{never}=0$

or

H_0: $\mu_{former}=\mu_{light}=\mu_{heavy}=\mu_{never}=0$

Root MSE	1.03634	R-Square	0.9814
Dependent Mean	6.70500	Adj R-Sq	0.9768
Coeff Var	15.45622		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Never	1	7.44000	0.46347	16.05	<.0001
Former	1	7.24000	0.46347	15.62	<.0001
Light	1	6.18000	0.46347	13.33	<.0001
Heavy	1	5.96000	0.46347	12.86	<.0001

Test Overall Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	3	**2.76183**	2.57	0.0904
Denominator	16	1.07400		

$3 \times 2.76183 = 8.28550$

Sums of Squares
Due to Smoking

H_0: $\mu_{never}=\mu_{former}=\mu_{light}=\mu_{heavy}$

Test Linear Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	**7.56250**	7.04	0.0173
Denominator	16	1.07400		

Test Quadratic Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	**0.00050000**	0.00	0.9831
Denominator	16	1.07400		

Test Cubic Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	**0.72250**	0.67	0.4242
Denominator	16	1.07400		

Sum the linear, quadratic, and cubic contrast SS:

$$
\begin{array}{c}
7.56250 \\
+0.00050 \\
\underline{+0.72250} \\
\Sigma = 8.28550
\end{array}
$$

Example (Orthogonal Polynomial Contrasts Using Data Step):

```
data bwt5;
    set bwt5;
    IF group = 0 THEN DO;
        linear = -3;
        quad   =  1;
        cubic  = -1;
    END;
    IF group = 1 THEN DO;
        linear = -1;
        quad   = -1;
        cubic  =  3;
    END;
    IF group = 2 THEN DO;
        linear =  1;
        quad   = -1;
        cubic  = -3;
    END;
    IF group = 3 THEN DO;
        linear =  3;
        quad   =  1;
        cubic  =  1;
    END;
RUN;

PROC REG data=bwt5;
    MODEL birthwt = linear quad cubic;
RUN;
```

Group	Variable Coding		
	linear	quad	cubic
1=Non	-3	1	-1
2=Former	-1	-1	3
3=Light	1	-1	-3
4=Heavy	3	1	1

PROC REG OUTPUT:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	8.28550	2.76183	2.57	0.0904
Error	16	17.18400	1.07400		
Corrected Total	19	25.46950			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.70500	0.23173	28.93	<.0001
linear	1	-0.27500	0.10363	-2.65	0.0173
quad	1	-0.00500	0.23173	-0.02	0.9831
cubic	1	0.08500	0.10363	0.82	0.4242

Example (Orthogonal Polynomial Contrasts, UNEQUAL N)

Note: KKNR orthogonal polynomial contrasts are for equal N's:

$-3(1) + -1(1) + 1(-1) + 3(1) = 0$, but $-3(1)/7 + -1(1)/5 + 1(-1)/7 + 3(1)/8 \neq 0$

```
PROC REG DATA=bwt;
    MODEL birthwt=never former light heavy/noint;
    Overall:   TEST never=former=light=heavy;
    Linear:    TEST -3*never -1*former +1*light +3*heavy=0;
    Quadratic: TEST  1*never -1*former -1*light +1*heavy=0;
    Cubic:     TEST -1*never +3*former -3*light +1*heavy=0;
RUN;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1234.44639	308.61160	349.60	<.0001
Error	23	20.30361	0.88277		
Uncorrected Total	27	1254.75000			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Never	1	7.58571	0.35512	21.36	<.0001
Former	1	7.24000	0.42018	17.23	<.0001
Light	1	6.32857	0.35512	17.82	<.0001
Heavy	1	6.01250	0.33218	18.10	<.0001

Test Overall Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	3	**3.89090**	4.41	0.0137
Denominator	23	0.88277		

3 x 3.89090 = 11.6727

Sums of Squares
Due to Smoking

Test Linear Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	**11.51558**	13.04	0.0015
Denominator	23	0.88277		

Test Quadratic Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	**0.00144**	0.00	0.9681
Denominator	23	0.88277		

Test Cubic Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	**0.40199**	0.46	0.5065
Denominator	23	0.88277		

Sum the linear, quadratic, and cubic contrast SS:

$$\begin{array}{r} 11.51558 \\ +0.00144 \\ \underline{+0.40199} \\ \Sigma=11.9190 \\ \neq 11.6727 \end{array}$$

NOTE: The Contrast SS DO NOT add up to the Model SS due to unequal N's across groups.

Example (Orthogonal Polynomial Contrasts: Adjusting for unequal n)

```
PROC IML;
    N = {7,5,7,8};
    X = {1,2,3,4};
    op = ORPOL(X,3,N);
PRINT op;

DATA bwt;
    set bwt;
    IF group = 0 THEN DO;
        o1 = -0.263541;
        o2 = 0.1740137;
        o3 = -0.07801;
    END;
    IF group = 1 THEN DO;
        o1 = -0.098062;
        o2 = -0.214473;
        o3 = 0.3276404;
    END;
    IF group = 2 THEN DO;
        o1 = 0.0674175;
        o2 = -0.215651;
        o3 = -0.234029;
    END;
    IF group = 3 THEN DO;
        o1 =  0.2328967;
        o2 = 0.1704784;
        o3 = 0.0682584;
    END;
```

Group	Variable Coding		
	o1	o2	o3
1=Non	-.263541	0.174013	-0.07801
2=Former	-.098062	-.214473	0.327640
3=Light	0.067418	-.215651	-.234029
4=Heavy	0.232897	0.170479	0.068259

```
PROC REG DATA=bwt;
    MODEL birthwt = o1 o2 o3;
    Linear:     TEST o1;
    Quadratic:  TEST o2;
    Cubic:      TEST o3;
RUN;
```

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	11.67269	3.89090	4.41	0.0137
Error	23	20.30361	0.88277		
Corrected Total	26	31.97630			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	6.72963	0.18082	37.22	<.0001
o1	1	-3.35494	0.93956	-3.57	0.0016
o2	1	0.12287	0.93956	0.13	0.8971
o3	1	0.63402	0.93956	0.67	0.5065

Test Linear Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	11.25561	12.75	0.0016
Denominator	23	0.88277		

Test Quadratic Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	0.01510	0.02	0.8971
Denominator	23	0.88277		

Test Cubic Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	0.40198	0.46	0.5065
Denominator	23	0.88277		

Sum the linear, quadratic, and cubic contrast SS:

$$\begin{array}{r} 11.25561 \\ +0.01510 \\ \underline{+0.40198} \\ \Sigma=11.67269 \end{array}$$

(matches slide 51)

F. Equivalence of Orthogonal Contrasts for Reference Cell and Cell Means Models

```
PROC REG DATA=birthsmk2;/* Reference Cell Coding Model*/
 MODEL weight = former light heavy;
/*Algebraic Translation of Orthogonal Contrast Matrix*/
 REFortha: TEST Intercept- Intercept-former = 0,
 Intercept+light - Intercept-heavy = 0,
 Intercept + Intercept+former - Intercept-light –
 Intercept-heavy = 0;
 REForths: TEST -former=0, light-heavy=0, former-light-
 heavy=0; /*Simplified Algebraic*/

 REForth1: TEST -former=0; /* Never vs. Former */
 REForth2: TEST light-heavy=0;
 REForth3: TEST former-light-heavy=0; /* Never+Former - (Light+Heavy) */
RUN;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	8.28550	2.76183	2.57	0.0904
Error	16	17.18400	1.07400		
Corrected Total	19	25.46950			

Root MSE	1.03634	R-Square	0.3253
Dependent Mean	6.70500	Adj R-Sq	0.1988
Coeff Var	15.45622		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	7.44000	0.46347	16.05	<.0001
Former	1	-0.20000	0.65544	-0.31	0.7642
Light	1	-1.26000	0.65544	-1.92	0.0725
Heavy	1	-1.48000	0.65544	-2.26	0.0383

```
PROC REG DATA= birthsmk2; /* Cell Means Coding Model */
 MODEL weight = never former light heavy / noint;

/* Orthogonal Contrast Matrix, 3 rows */
CMortha: TEST never-former=0, light-heavy=0, never+former-light-heavy=0;

 CMorth1: TEST never-former=0;
 CMorth2: TEST light-heavy=0;
 CMorth3: TEST never+former-light-heavy=0;
RUN;
```

NOTE: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	907.42600	226.85650	211.23	<.0001
Error	16	17.18400	1.07400		
Uncorrected Total	20	924.61000			

Root MSE	1.03634	R-Square	0.9814
Dependent Mean	6.70500	Adj R-Sq	0.9768
Coeff Var	15.45622		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Never	1	7.44000	0.46347	16.05	<.0001
Former	1	7.24000	0.46347	15.62	<.0001
Light	1	6.18000	0.46347	13.33	<.0001
Heavy	1	5.96000	0.46347	12.86	<.0001

Test REFortha Results for Dependent Variable weight

Test REFortha Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	3	2.76183	2.57	0.0904
Denominator	16	1.07400		

Test REForths Results for Dependent Variable weight

Test REForths Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	3	2.76183	2.57	0.0904
Denominator	16	1.07400		

Test REForth1 Results for Dependent Variable weight

Test REForth1 Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.10000	0.09	0.7642
Denominator	16	1.07400		

Test CMortha Results for Dependent Variable weight

Test CMortha Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	3	2.76183	2.57	0.0904
Denominator	16	1.07400		

Test CMorth1 Results for Dependent Variable weight

Test CMorth1 Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.10000	0.09	0.7642
Denominator	16	1.07400		

Test REForth2 Results for Dependent Variable weight

Test REForth2 Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	0.12100	0.11	0.7415
Denominator	16	1.07400		

Test REForth3 Results for Dependent Variable weight

Test REForth3 Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	8.06450	7.51	0.0145
Denominator	16	1.07400		

Test CMorth2 Results for Dependent Variable weight

Test CMorth2 Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	0.12100	0.11	0.7415
Denominator	16	1.07400		

Test CMorth3 Results for Dependent Variable weight

Test CMorth3 Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	8.06450	7.51	0.0145
Denominator	16	1.07400		

G. Equivalence of Orthogonal Polynomials for Reference Cell and Cell Means Models

```
PROC REG DATA= birthsmk2;/*Reference Cell Coding Model*/
 MODEL weight = linear quad cubic;
 OverOrth: TEST linear=0, quad=0, cubic=0;
RUN;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	8.28550	2.76183	2.57	0.0904
Error	16	17.18400	1.07400		
Corrected Total	19	25.46950			

Root MSE	1.03634	R-Square	0.3253
Dependent Mean	6.70500	Adj R-Sq	0.1988
Coeff Var	15.45622		

```
PROC REG DATA= birthsmk2; /* Cell Means Coding Model */
 MODEL weight=never former light heavy/noint;
 Overall:   TEST never=former=light=heavy;
 Linear:    TEST -3*never -1*former +1*light +3*heavy=0;
 Quadratic: TEST  1*never -1*former -1*light +1*heavy=0;
 Cubic:     TEST -1*never +3*former -3*light +1*heavy=0;
 OverOrth:  TEST -3*never -1*former +1*light +3*heavy=0,
                  1*never -1*former -1*light +1*heavy=0,
                 -1*never +3*former -3*light +1*heavy=0;
RUN;
```

NOTE: No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	907.42600	226.85650	211.23	<.0001
Error	16	17.18400	1.07400		
Uncorrected Total	20	924.61000			

Root MSE	1.03634	R-Square	0.9814
Dependent Mean	6.70500	Adj R-Sq	0.9768
Coeff Var	15.45622		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Never	1	7.44000	0.46347	16.05	<.0001
Former	1	7.24000	0.46347	15.62	<.0001
Light	1	6.18000	0.46347	13.33	<.0001
Heavy	1	5.96000	0.46347	12.86	<.0001

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	6.70500	0.23173	28.93	<.0001
linear	1	-0.27500	0.10363	-2.65	0.0173
quad	1	-0.00500	0.23173	-0.02	0.9831
cubic	1	0.08500	0.10363	0.82	0.4242

Test Overall Results for Dependent Variable weight

Test Overall Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	3	2.76183	2.57	0.0904
Denominator	16	1.07400		

Test Linear Results for Dependent Variable weight

Test Linear Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	7.56250	7.04	0.0173
Denominator	16	1.07400		

Test Quadratic Results for Dependent Variable weight

Test Quadratic Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	0.00050000	0.00	0.9831
Denominator	16	1.07400		

Test Cubic Results for Dependent Variable weight

Test Cubic Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	1	0.72250	0.67	0.4242
Denominator	16	1.07400		

Test OverOrth Results for Dependent Variable weight

Test OverOrth Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	3	2.76183	2.57	0.0904
Denominator	16	1.07400		

Test OverOrth Results for Dependent Variable weight

Test OverOrth Results for Dependent Variable birthwt				
Source	DF	Mean Square	F Value	Pr > F
Numerator	3	2.76183	2.57	0.0904
Denominator	16	1.07400		