S1. Linear, Orthogonal, and Orthogonal Polynomial Contrasts (Oh My!)

Readings: Kleinbaum, Kupper, Nizam, and Rosenberg (KKNR): Ch. 15

SAS: PROC REG, PROC LOESS

Supplement to Lectures 23-25 (Tests of General Linear Hypotheses and Polynomial Regression)

Overview

A) Motivation
B) Linear Contrasts and Orthogonal Contrasts
C) Equivalence of Orthogonal Contrasts for Reference Cell and Cell Means Models
D) Orthogonal Polynomial Contrasts
E) Equivalence of Orthogonal Polynomial Contrasts for Reference Cell and Cell Means Models
F) Orthogonal Polynomial Contrasts for Polynomial Regression Models

A. Motivation

In regression modeling and the analysis of variance, we often wish to combine various coefficients or model terms to estimate a combined effect. For example, an example with the following categories:

1. Non-smoker
2. Former smoker
3. Light smoker
4. Heavy smoker

may wish to combine the effects for *current* non-smokers versus *current* smokers (i.e., combining non-smoker and former smoker versus combining light smoker and heavy smoker).

Alternatively, we may wish to combine different variables to estimate an effect or evaluate a model for potential polynomial terms.

There are many approaches to doing this, some of which have special properties. In our lecture slides discussing general linear hypotheses, we saw how we could specify specific tests within the PROC REG statement directly using a TEST statement. The methods described in this supplemental set of slides further discusses special cases of contrasts that may be of benefit in certain circumstances.

B. Linear Contrasts

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

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

Our contrast is estimated from our sample means and we can estimate its variability:

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

Different coding schemes can be used:

- Reference Cell (Dummy codes)
- Cell Means (No Intercept)
- Effect Coding (Design coding)
- Orthogonal Polynomial Coding (Section D in these slides)

Linear contrasts are most often used to test linear combinations of group means in a <u>Cell Means Model</u> (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: $t = \frac{\hat{L}}{SE(\hat{L})}$

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

Independent means:
Covariances are 0

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

$$= c_1^2\, var(\bar{y}_1) + c_2^2\, var(\bar{y}_2) + \ldots + c_k^2\, var(\bar{y}_k) + 2c_1c_2\, cov(\bar{y}_1, \bar{y}_2) + \ldots + 2c_{k-1}c_k\, cov(\bar{y}_{k-1}, \bar{y}_k)$$

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

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

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

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).

Benefits of Orthogonal Contrasts

A priori (pre-planned) orthogonal contrasts are extremely powerful, because they do not need correction for multiple comparisons like *post-hoc* tests do.

This benefit for orthogonal contrasts is because we can partition our model sum of squares into the meaningful components associated with specific comparisons of interest (note, these are subjectively defined by the researchers and may be different for each person).

Assume we have defined *t* pairwise orthogonal contrasts, then our partition of the SS_{Model} is:

$$SS_{Model} = SS(\hat{L}_1) + \cdots SS(\hat{L}_t) + SS_{Remainder}$$

For our categorical variable context, if *t* is equal to the number of groups minus 1, then $SS_{Remainder}$ equals 0. Otherwise, if *t* is less than the number of groups minus 1 (perhaps we don't have more comparisons of interest), then

$$SS_{Remainder} = SS_{Model} - \left[SS(\hat{L}_1) + \cdots SS(\hat{L}_t)\right]$$

However, if the contrasts are not orthogonal, we cannot partition our SS_{Model} correctly and the results will be built on incorrect assumptions and, consequently, incorrect interpretations.

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`, $H_0: \mu_{never} = \mu_{former}$

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`, $H_0: \mu_{never} = \mu_{light}$

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`, $H_0: \mu_{never} = \mu_{heavy}$

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`

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

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`, $H_0: \mu_{never} = \mu_{former}$

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. `light-heavy`, $H_0: \mu_{light} = \mu_{heavy}$

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+former-light-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, light-heavy, never+former-light-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		

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

Brief Interlude: Why Test 4 is identical for the Non-Orthogonal Example Contrast 1 (page 10) and the Orthogonal Example Contrast 2 (page 13).

From page 12, note we can calculate the F-statistic from the matrices directly:

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

Here we define $\hat{\beta} = \begin{pmatrix} 7.44 \\ 7.24 \\ 6.18 \\ 5.96 \end{pmatrix}$, $c_1 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}$, $c_2 \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix}$, $d = \mathbf{0}$, and

$\boldsymbol{\Sigma} = \begin{pmatrix} 0.2148 & 0 & 0 & 0 \\ 0 & 0.2148 & 0 & 0 \\ 0 & 0 & 0.2148 & 0 \\ 0 & 0 & 0 & 0.2148 \end{pmatrix}$ [from COVB specified for cell means model]

For contrast 1: $F = \begin{pmatrix} 0.47 & 0.51 & 2.96 \end{pmatrix} \begin{pmatrix} 0.20 \\ 0.22 \\ 2.54 \end{pmatrix} / 3 = 2.57$

For contrast 2: $F = \begin{pmatrix} 0.47 & -2.44 & 5.91 \end{pmatrix} \begin{pmatrix} 0.20 \\ 0.22 \\ 1.38 \end{pmatrix} / 3 = 2.57$

Why does this happen? Because if we define the maximum number of independent contrasts (in our case this is the number of groups minus 1), any new contrast can be determined as some linear combination of the existing contrasts.

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

$Var(L) = \hat{\sigma}^2_{Y|X} \sum_{i=1}^{k} c_i^2 / n_i$ (from slide 25)

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 directly calculate the F statistic from our formula on slide 27:

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

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

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

$L = (0.5)\times 7.44 + (0.5)\times 7.24 + (-0.5)\times 6.18 + (-0.5)\times 5.96 = 1.27$ lbs

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

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

$p = 0.0145$ (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 (never 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 <u>population</u> are former smokers and 50% of current smokers in the <u>population</u> 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		

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:

$$H_0: \frac{1}{2}\mu_{never} + \frac{1}{2}\mu_{former} - \frac{1}{2}\mu_{light} - \frac{1}{2}\mu_{heavy} = 0$$

versus

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

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

C. 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		

D. 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.

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. Therefore, they avoid the serious collinearity inherent in using natural polynomials.

As the order increases, computational accuracy may decease with the simple polynomial variables due to collinearity. However, the orthogonal polynomial variables are not impacted because they are uncorrelated. ***One of the main motivations for using orthogonal polynomial variables is to avoid the serious collinearity of simple polynomial variables in determining what higher order, if any, is needed.***

Because orthogonal polynomial variables contain the same information as the simple polynomial variables, the overall regression F-test and multiple R^2 values will be identical, even though the β's will be different and have different interpretations.

Because these special contrasts are still orthogonal, we can still partition the Model Sums of Squares into statistically independent sums of squares for each polynomial contrast (linear, quadratic, etc.) and take advantage of more powerful *a priori* tests.

The orthogonal polynomials can also be used to perform linear contrasts in a cell means model by defining the TEST statement using the contrast matrix values.

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

The assumption of equally spaced predictor values may not make intuitive sense in cases with nominal or ordinal groups (e.g., assuming the “space” between smoking statuses is equal). However, in contexts where groups are based on interval values (e.g., different dose levels being study in a trial) this assumption is more straightforward.

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
$\overline{Y}\|X$	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: β_{former}=β_{light}=β_{heavy}=β_{never}=0

or

H_0: μ_{former}=μ_{light}=μ_{heavy}=μ_{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 x 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		

Sum the linear, quadratic, and cubic contrast SS:

7.56250
+0.00050
+0.72250
Σ=8.28550

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		

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
0=Non	-3	1	-1
1=Former	-1	-1	3
2=Light	1	-1	-3
3=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:

11.51558
+0.00144
<u>+0.40199</u>
Σ=11.9190
≠ 11.6727

NOTE: The Contrast SS <u>DO NOT</u> 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 = {0,1,2,3};
    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
0=Non	-.263541	0.174013	-0.07801
1=Former	-.098062	-.214473	0.327640
2=Light	0.067418	-.215651	-.234029
3=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:

11.25561
+0.01510
+0.40198
Σ=11.67269
(matches slide 51)

E. 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

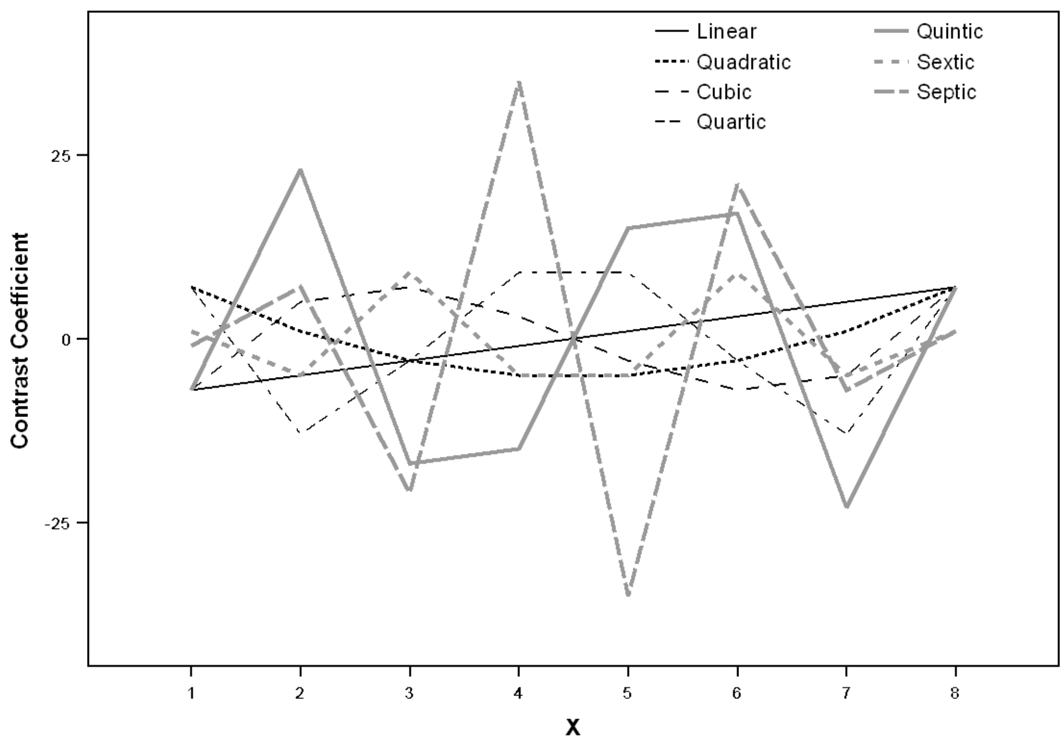
Test OverOrth Results for Dependent Variable birthwt				
Source	**DF**	**Mean Square**	**F Value**	**Pr > F**
Numerator	3	2.76183	2.57	0.0904
Denominator	16	1.07400		

F. Orthogonal Polynomial Contrasts for Polynomial Regression Models

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. *(See SAS code for DATA step to calculate "odose".)*

Coefficients from KKNR Table A7 →

k=8	X								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Σp_i^2**
Linear	-7	-5	-3	-1	1	3	5	7	168
Quadratic	7	1	-3	-5	-5	-3	1	7	168
Cubic	-7	5	7	3	-3	-7	-5	7	264
Quartic	7	-13	-3	9	9	-3	-13	7	616
Quintic	-7	23	-17	-15	15	17	-23	7	2184
Sextic	1	-5	9	-5	-5	9	-5	1	264
Septic	-1	7	-21	35	-35	21	-7	1	3432



Orthogonal Polynomials Example

```
PROC REG DATA=wtgain;
  MODEL wgtgain = odose1 odose2 odose3 odose4 odose5 odose6 odose7 / covb;
  LinearLOF: TEST odose2, odose3, odose4, odose5, odose6, odose7;
  QuadLOF:   TEST odose3, odose4, odose5, odose6, odose7;
RUN;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	169.77958	24.25423	1492.57	<.0001
Error	16	0.26000	0.01625		
Corrected Total	23	170.03958			

ANOVA Table Identical to Reference Cell Model (Lecture 25, Page 13).

	Parameter Estimates					
	Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
	Intercept	1	3.65417	0.02602	140.43	<.0001
Linear	**odose1**	1	**0.55575**	**0.00568**	**97.87**	**<.0001**
Quadratic	**odose2**	1	0.16687	0.00568	29.39	<.0001
Cubic	**odose3**	1	0.00391	0.00453	0.86	0.4003
Quartic	**odose4**	1	-0.00525	0.00297	-1.77	0.0958
Quintic	**odose5**	1	0.00001526	0.00157	0.01	0.9924
Sextic	**odose6**	1	-0.00215	0.00453	-0.47	0.6420
Septic	**odose7**	1	-0.00112	0.00126	-0.89	0.3871

How would we interpret the parameter estimate for ODOSE1?

This is the expected weight gain for a ½ unit increase in dose. To obtain weight gain for a 1-unit increase multiply the parameter estimate by two:
$2\times0.55575 = 1.1115$.

$SE(2\beta_1) = 2SE(\beta_1)$
$= 2\times0.00568 = 0.01136$

Orthogonal Polynomials Example (Linear vs. Quadratic)

Partial Output of Covariance Matrix				
Variable	**Intercept**	**odose1**	**odose2**	**odose3**
Intercept	0.0006770833	0	0	0
odose1	0	0.0000322421	0	0
odose2	0	0	0.0000322421	0
odose3	0	0	0	0.0000205177

0 because of orthogonal contrasts

Test LinearLOF Results for Dependent Variable wgtgain

Source	DF	Mean Square	F Value	Pr > F
Numerator	6	2.35215	144.75	<.0001
Denominator	16	0.01625		

Test QuadLOF Results for Dependent Variable wgtgain

Source	DF	Mean Square	F Value	Pr > F
Numerator	5	0.01591	0.98	0.4602
Denominator	16	0.01625		

Conclusions:

The linear model is **not** adequate.

A quadratic model is adequate.

Orthogonal Polynomials Example (Finding "Pure Error")

NOTE: We need to fit all 7 orthogonal polynomials for the MSE to be equal to our "pure error".

However, if the higher-order polynomials are not necessary, we will approximate the "pure error" with a lower-order polynomial model while using fewer degrees of freedom.

```
PROC REG DATA=wtgain;
    MODEL wgtgain = odose1 odose2;
RUN;
```

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	2	169.70004	84.85002	5247.78	<.0001
Error	**21**	0.33954	**0.01617**		
Corrected Total	23	170.03958			

Recall:
Pure Error = 0.01625

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	3.65417	0.02596	140.78	<.0001
odose1	1	0.55575	0.00566	98.12	<.0001
odose2	1	0.16687	0.00566	29.46	<.0001

ANOVA Table identical to quadratic model using natural polynomials (p.8), but parameter estimate table is different.

Same parameter estimates, slightly different SEs (page 18)

Standardized Orthogonal Polynomials

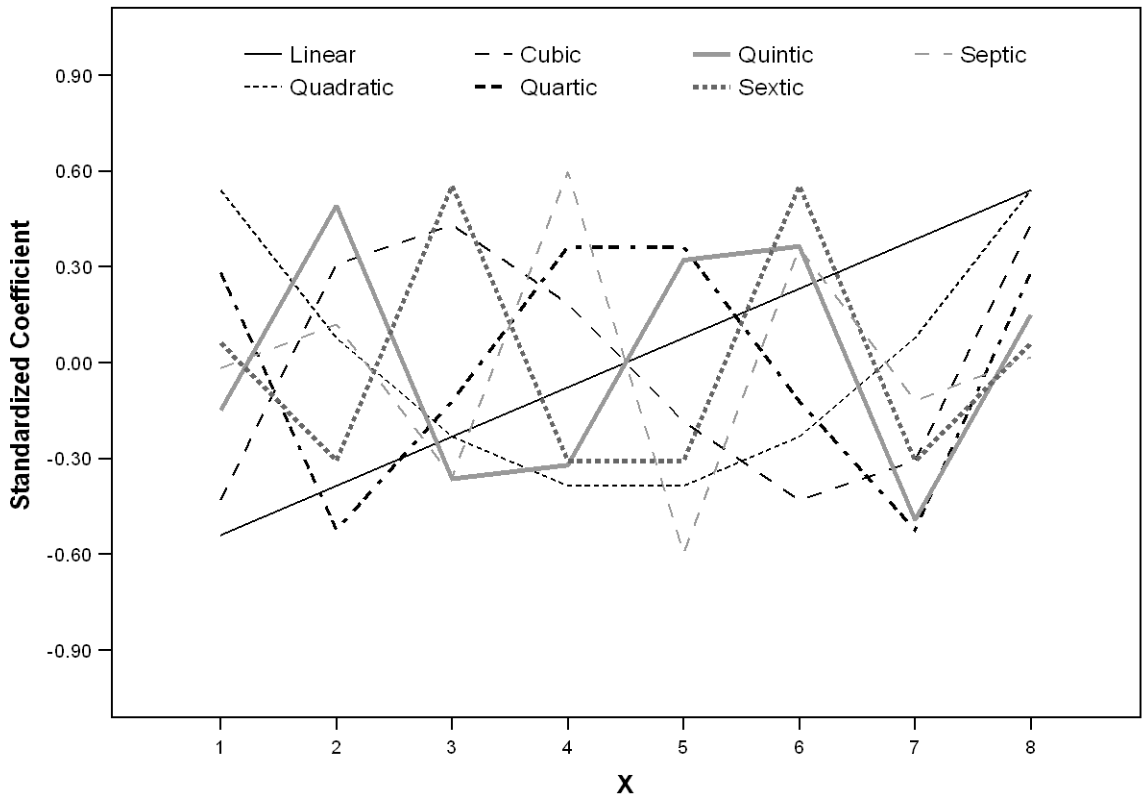
KKNR recommend dividing the orthogonal polynomials by the square root of the sum of squared values of the coefficients (provided in the last column of Table A7).

- The variance of each set of orthogonal polynomial scores is thus equal to 1.
- This improves numerical accuracy by avoiding scaling problems.
- The SEs for all estimated regression coefficients are thus equal, simplifying the task of comparing and interpreting such regression coefficients.

```
data wtgain;
 set wtgain;

 odose1_s = odose1/SQRT(168);
 odose2_s = odose2/SQRT(168);
 odose3_s = odose3/SQRT(264);
 odose4_s = odose4/SQRT(616);
 odose5_s = odose5/SQRT(2184);
 odose6_s = odose6/SQRT(264);
 odose7_s = odose7/SQRT(3432);

RUN;
```



Orthogonal Polynomials Example

```
PROC REG DATA=wtgain;
  MODEL wgtgain = odose1_s odose2_s odose3_s odose4_s odose5_s odose6_s odose7_s;
RUN;
```

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	7	169.77958	24.25423	1492.57	<.0001
Error	16	0.26000	0.01625		
Corrected Total	23	170.03958			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	3.65417	0.02602	140.43	<.0001
odose1_s	1	7.20339	0.07360	97.87	<.0001
odose2_s	1	2.16282	0.07360	29.39	<.0001
odose3_s	1	0.06360	0.07360	0.86	0.4003
odose4_s	1	-0.13027	0.07360	-1.77	0.0958
odose5_s	1	0.00071327	0.07360	0.01	0.9924
odose6_s	1	-0.03488	0.07360	-0.47	0.6420
odose7_s	1	-0.06543	0.07360	-0.89	0.3871

No change in t-values or p-values

SEs now equal

Parameter estimates changed (now standardized)

To test just the linear effect of dose, we could have modeled the data using a cell means model, and then test the linear contrast (or any other contrast of interest):

```
PROC REG DATA=wtgain;
    MODEL wgtgain = dose1 dose2 dose3 dose4 dose5 dose6 dose7 dose8/noint;
    LINEAR: TEST -7*dose1-5*dose2-3*dose3-1*dose4+1*dose5+3*dose6+
    5*dose7+7*dose8;
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	8	490.25000	61.28125	3771.15	<.0001
Error	16	0.26000	0.01625		
Uncorrected Total	24	490.51000			

This is Σy_i^2 (uncorrected sum of squares) in the cell means model

Pure Error

We are still using the pure error estimate of the MSE, but only need to estimate the contrast of interest (linear contrast).

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
dose1	1	0.86667	0.07360	11.78	<.0001
dose2	1	1.13333	0.07360	15.40	<.0001
dose3	1	1.53333	0.07360	20.83	<.0001
dose4	1	2.20000	0.07360	29.89	<.0001
dose5	1	3.36667	0.07360	45.74	<.0001
dose6	1	4.76667	0.07360	64.77	<.0001
dose7	1	6.66667	0.07360	90.58	<.0001
dose8	1	8.70000	0.07360	118.21	<.0001

This is a cell means model, so each of the beta estimates is the observed mean for that dose group.

Test LINEAR Results for Dependent Variable wgtgain

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	155.66669	9579.49	<.0001
Denominator	16	0.01625		

Equivalent to the previous t statistics for the linear effect on pages 11, 18, and 22:

$F = 9579.49$

$t = \sqrt{9579.49} = 97.875$