Analysis of Covariance

Background

- Consider factor $x$ which is correlated with $y$
- Can measure $x$ but can’t control it (block)
- Nuisance factor $x$ called a covariate
- ANCOVA adjusts $y$ for effect of covariate $x$
- Combination of regression and analysis of variance
- Without adjustment, effects of $x$ may inflate $\sigma^2$
  alter treatment comparisons

Examples

- Pretest/Posttest score analysis: The gain in score $y$ may be associated with the pretest score $x$. Analysis of covariance provides a way to “handicap” each student. That way, one does not need to find a group of students with similar pretest scores and randomly assign them to a control and treatment group. Similar to analyzing difference in scores.

- Weight gain experiments in animals: If wishing to compare different feeds, the weight gain $y$ may be associated with the original weight of the animal. Analysis of covariance provides a way to use a herd and adjust for the varying original weights.

- Comparing competing drug products: The effect of the drug $y$ after two hours (measured on a scale from 1 to 10) may be associated with the initial mental and physical shape of the subject. Variables describing the initial mental and physical shape may be used as covariates.

Model Description

- Consider single covariate in CRD
- Statistical model is
  \[
  y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}) + \epsilon_{ij} \quad \left\{ \begin{array}{l} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n_i \end{array} \right. 
  \]
  - Additional assumptions
    \[
    x_{ij} \text{ not affected by treatment} \\
    x \text{ and } y \text{ are linearly related} \\
    \text{Constant slope}
    \]
  - General Procedure:
    Fit one-way model ($y = \text{trt}$)
    Fit one-way model ($x = \text{trt}$)
    Regress residuals (residuals1 = residuals2)
    Model estimates are
    \[
    \bar{\mu} = \bar{y} \\
    \bar{\beta} = \frac{\sum \sum (y_{ij} - \bar{y})(x_{ij} - \bar{x})}{\sum \sum (x_{ij} - \bar{x})^2} \\
    \bar{\tau}_i = \bar{y}_i - \bar{y} - \bar{\beta}(\bar{x}_i - \bar{x})
    \]

10-1

10-2

10-3
Analysis of Covariance

- Test $H_0: \tau_1 = \tau_2 = \ldots = \tau_k = 0$
  - Compare treatment means after adjusting for differences among treatments due to differences in covariate levels
  - Trt and covariate not orthogonal (order of fit matters)
    \[
    F_0 = \frac{SS(trt \mid x)/a - 1}{SS_E/(N-a-1)}
    \]

- Adjusted treatment means
  - Estimate $\bar{\mu} = \bar{\mu} + \bar{\tau} = \bar{y}_i - \hat{\beta}(\bar{x}_i - \bar{x})$
  - Variance: $\hat{\sigma}^2 (1/n + (\bar{x}_i - \bar{x})^2/\sum\sum (x_{ij} - \bar{x}_i)^2)$

- Test: $\beta = 0$
  - Sum of Squares regression ($SS_x$): $\hat{\beta}^2 \sum\sum (x_{ij} - \bar{x}_i)^2$
    \[
    F_0 = \frac{SS_x/1}{SS_E/(N-a-1)}
    \]

Two Examples

1. No treatment differences
   - Positive linear relationship
   - Covariate larger in each group
   - Thus, appears to be treatment difference

2. Treatment differences exist
   - Positive linear relationship
   - Covariate larger in each group
   - Thus, no apparent treatment difference

Using SAS

```
options nocenter ls=80;
data example1;
  input trt x y @@;
cards;
1 1.2 7 1 1.9 13 1 3.4 16
2 4.0 20 2 5.2 22 2 5.8 32
3 7.7 31 3 8.3 45 3 8.9 42
;
proc sort; by trt;
symbol1 v=circle i= c=black;
symbol2 v=square i= c=black;
symbol3 v=triangle i= c=black;
proc gplot;
  plot y*x=trt;
run;
proc glm; class trt;
  model y=trt;
  output out=resid r=resy;
proc glm; class trt;
  model x=trt;
  output out=resid1 r=resx;
proc glm; model resy=resx;
symbol1 v=circle i=r;
proc gplot; plot resy=resx;
run;
proc glm data=example1;
  class trt; model y=trt x / solution;
  means trt /lines lsd;
  lsmeans trt / tdiff adjust=t;
run;
```
The GLM Procedure

### Dependent Variable: resy

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sums of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>138.2699594</td>
<td>138.2699594</td>
<td>10.18</td>
<td>0.0153</td>
</tr>
<tr>
<td>Error</td>
<td>7</td>
<td>95.0633749</td>
<td>13.5804820</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>8</td>
<td>233.3333333</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square 0.592586
Coeff Var 1.03728E17
Root MSE resy Mean 3.685171

---

### Source DF Type I Sums of Squares Mean Square F Value Pr > F

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I Sums of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>resx</td>
<td>1</td>
<td>138.2699594</td>
<td>138.2699594</td>
<td>10.18</td>
<td>0.0153</td>
</tr>
</tbody>
</table>

---

### Standard Parameter Estimate Error t Value Pr > |t|

| Interceptor | 0.000000000 | 1.22839018 | 0.00 | 1.0000 |

---

### t Tests (LSD) for resy

<table>
<thead>
<tr>
<th>Alpha</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Degrees of Freedom</td>
<td>8</td>
</tr>
<tr>
<td>Error Mean Square</td>
<td>19.01267</td>
</tr>
<tr>
<td>Critical Value of t</td>
<td>2.57058</td>
</tr>
<tr>
<td>Least Significant Difference</td>
<td>9.1518</td>
</tr>
</tbody>
</table>

Means with the same letter are not significantly different.

---

The GLM Procedure

### Dependent Variable: y

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sums of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3</td>
<td>1260.936626</td>
<td>420.312209</td>
<td>22.11</td>
<td>0.0026</td>
</tr>
<tr>
<td>Error</td>
<td>5</td>
<td>95.0633749</td>
<td>19.012675</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>8</td>
<td>1356.000000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square 0.592586
Coeff Var 1.03728E17
Root MSE y Mean 3.685171

---

### Source DF Type I Sums of Squares Mean Square F Value Pr > F

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I Sums of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>2</td>
<td>1122.666667</td>
<td>561.3333333</td>
<td>29.52</td>
<td>0.0017</td>
</tr>
<tr>
<td>x</td>
<td>1</td>
<td>138.269959</td>
<td>138.269959</td>
<td>7.27</td>
<td>0.0430</td>
</tr>
</tbody>
</table>

---

### Standard Parameter Estimate Error t Value Pr > |t|

| Interceptor | -4.637573297 | B | 16.49828508 | -0.28 | 0.7899 |
| trt 1       | 5.159224177  | B | 12.56372646 | 0.41  | 0.6983 |
| trt 2       | 2.8157419944 | B | 7.39601943  | 0.38  | 0.7191 |
| x            | 5.297699594  | B | 1.96446828  | 2.70  | 0.0430 |

---

The GLM Procedure

### Least Squares Means

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>LSMEAN</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27.8342365</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25.4907533</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>22.6750113</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

---

### Least Squares Means for Effect trt

<table>
<thead>
<tr>
<th>i/j</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.354685</td>
<td>0.410644</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7373</td>
<td>0.6983</td>
<td>0.38071</td>
</tr>
<tr>
<td>3</td>
<td>0.6983</td>
<td>0.7191</td>
<td>0.38071</td>
</tr>
</tbody>
</table>

Means with the same letter are not significantly different.

---

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.
**Regression Approach to ANCOVA**

- Consider the following model \((a = 3)\)
  
  \[ y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_3 X_{3j} + \epsilon_j \]
  
  \(j = 1, 2, \ldots, N\)

  \(X_{1j} = 1\) if Trt 1 and \(X_{1j} = -1\) if Trt 3
  
  \(X_{2j} = 1\) if Trt 2 and \(X_{2j} = -1\) if Trt 3
  
  \(X_{3j} = (x_j - \bar{x})\) \(x\) is the covariate

- Trt 1: \(y_j = \beta_0 + \beta_1 + \beta_3(x_j - \bar{x}) + \epsilon_j\)

- Trt 2: \(y_j = \beta_0 + \beta_2 + \beta_3(x_j - \bar{x}) + \epsilon_j\)

- Trt 3: \(y_j = \beta_0 - \beta_1 - \beta_2 + \beta_3(x_j - \bar{x}) + \epsilon_j\)

- Results in estimates
  
  \(\hat{\mu} = \beta_0\)
  
  \(\hat{\beta}_1 = \beta_1\)

  \(\hat{\beta}_2 = \beta_2\)

  \(\hat{\beta}_3 = \beta_3\)

**Nonconstant Slope in ANCOVA**

- Statistical model for constant slope is
  
  \[ y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}) + \epsilon_{ij} \]

- Can allow for different slope by including interaction
  
  \[ y_{ij} = \mu + \tau_i + (\beta + (\beta\tau_i))(x_{ij} - \bar{x}) + \epsilon_{ij} \]

- In SAS, simply add interaction term into model

- Provides test for nonconstant slope
Using SAS

```
options nocenter ls=75;

data example1;
  input trt x y @@;
  cards;
1 1.2 7
1 1.9 13
1 3.4 16
2 4.0 20
2 5.2 22
2 5.8 32
3 7.7 31
3 8.3 45
3 8.9 42
;
proc sort; by trt;
symbol1 v=circle i= c=black;
symbol2 v=square i= c=black;
symbol3 v=triangle i= c=black;
proc gplot;
plot y*x=trt;
run;
proc glm;
class trt;
model y=trt x / solution;
lsmeans trt / tdiff;
run;
proc glm;
class trt;
model y=trt x trt*x / solution;
lsmeans trt / tdiff;
run;
```

**Analysis of Covariance**

- Can incorporate covariate into any model
- For two factor model
  
  \[ y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \beta(x_{ijk} - \bar{x}) + \epsilon_{ijk} \]

  - Assume constant slope for each \( ij \) combination
  - Can include interaction terms to vary slope
  - Plot \( y \) vs \( x \) for each combination

---

```
10-15
LSMEAN
trt y LSMEAN Number
1 27.8342355 1
2 25.4907533 2
3 22.6750113 3

Least Squares Means for Effect trt
t for H0: LSmear(i)=LSMean(j) / Pr > |t|
Dependent Variable: y
i/j 1 2 3
1 0.354685 0.410644 0.7373
2 -0.35468 0.38071 0.7191
3 -0.41064 -0.38071 0.6983

10-16
LSMEAN
trt y LSMEAN Number
1 23.2379068 1
2 25.5925926 2
3 10.5092593 3

Least Squares Means for Effect trt
t for H0: LSmear(i)=LSMean(j) / Pr > |t|
Dependent Variable: y
i/j 1 2 3
1 -0.22548 0.591 0.8361
2 0.225476 0.781205 0.8361
3 -0.591 -0.78121 0.5961
10-17
LSMEAN
trt y LSMEAN Number
1 23.2379068 1
2 25.5925926 2
3 10.5092593 3

Least Squares Means for Effect trt
t for H0: LSmear(i)=LSMean(j) / Pr > |t|
Dependent Variable: y
i/j 1 2 3
1 -0.22548 0.591 0.8361
2 0.225476 0.781205 0.8361
3 -0.591 -0.78121 0.5961
10-18
```