Experimental Design: Blocking

- A study is conducted to measure the effect of a drug on locomotor activity in hyperactive children.
- Between-subjects design:
  - 3 groups differ in drug dosage: zero, low, & high
- Dependent variable: locomotor activity
  - measured for fixed interval after drug administration
- Before study, measure baseline locomotor activity in each subject
  - baseline measure used as a blocking variable

Randomized Block Design

- 4 blocks of 12 Ss created using baseline locomotor activity measure
- Subjects in each block assigned randomly to drug dose condition

<table>
<thead>
<tr>
<th>Drug Dose (Baseline Locomotor Activity)</th>
<th>zero</th>
<th>low</th>
<th>high</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>medium</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>high</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>very high</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Do blocks differ from each other?

- Blocks are factors that indicate the level of drug received by the subjects. Finally, we create a fake dependent variable `activity` in other words, `activity` is a factor that indicates baseline locomotor activity.

- ANCOVA quantitatively models association between dependent variable and covariate (baseline activity) using each subject’s activity measure rather than dividing subjects into 4 factor groups.

Analysis of Covariance (ANCOVA)

- ANCOVA is a method to remove the differences between groups before comparing the groups themselves. This is done by statistically controlling for the effects of other variables called covariates.

- ANCOVA allows us to examine the relationship between a dependent variable (Y) and an independent variable (X) while controlling for the effects of one or more covariates (Z).

- In the hyperactivity example, Y was linearly related to baseline activity, while controlling for the linear association between Y and baseline activity.

- The residuals `SSresiduals` are calculated after controlling for the linear association between Y and baseline activity.

Effect of blocking factor on `SSresiduals`

```r
> aov.1 <- aov(y ~ drug, data = theData)
> aov.2 <- aov(y ~ block + drug + block:drug, data = theData)
> summary(aov.1)

Df  Sum Sq Mean Sq F value Pr(>F)
block 3  483.1 161.02  106.3 <2e-16 ***
Residuals 45 384.0  8.532

> summary(aov.2)

Df  Sum Sq Mean Sq F value Pr(>F)
block 3  483.1 161.02  106.3 <2e-16 ***
block:drug 6  50.24  8.373  0.03119 *
Residuals 45 383.0  8.567

SSresiduals = 421
```

ANCOVA

```r
> lm.1 <- lm(y ~ activity + drug, data = theData)
> lm.2 <- lm(y ~ activity + drug + activity:drug, data = theData)
> anova(lm.2, lm.1)

Analysis of Variance Table

Model 1: y ~ activity
Model 2: y ~ activity + drug

Df  Sum Sq Mean Sq F value  Pr(>F)
Res.Df RSS  Df  Sum of Sq    F   Pr(>F)
1  44 239.64  5.356
2  43 239.64  5.356

> anova(lm.1)

Analysis of Variance Table

Response: y

Df  Sum Sq Mean Sq F value  Pr(>F)
activity 1 142.116 142.116 18.254 0.0003544 ***
drug 2 39.286 19.643  3.606 0.03544 *
Residuals 43 383.0  8.847
```

Analysis of variance of `SSresiduals` shows `SSresiduals` after controlling for the linear association between Y and baseline activity.

Same result obtained with sequential sums-of-squares ANOVA table for full model.
Order of terms does matter

```r
> lm.1 <- lm(y ~ activity + drug, data = theData)
> anova(lm.1)

Analysis of Variance Table
Df  Sum Sq Mean Sq F value    Pr(>F)
activity   1 142.116 142.116 26.0939 6.739e-06 ***
drug       2  39.286  19.643  3.6066   0.03544 *
Residuals 44 239.638  5.446

> lm.1b <- lm(y ~ drug + activity, data = theData)
> anova(lm.1b)

Analysis of Variance Table
Df  Sum Sq Mean Sq F value   Pr(>F)
drug       2  37.078  18.539  3.4039  0.04222 *
activity   1 144.324 144.324 26.4993 5.91e-06 ***
Residuals 44 239.638  5.446
```

Graphical illustration of ANCOVA

• ANCOVA computes regression lines for each group
  - equal slopes
  - variable intercepts
• group effect ($\alpha_j$) corresponds to shift of regression intercept

Checking linear regression

```r
par(cex = 1.5)
plot(residuals(lm.0) ~ fitted(lm.0), main = "Residuals vs. Fitted Values")
abline(h = 0, lty = 2)
plot(residuals(lm.0) ~ df0$x, main = "Residuals vs. X")
abline(h = 0, lty = 2)
```
Adjusted Group Means

- point where regression line crosses grand mean of covariate
- provides estimates of group differences after removing effects of covariate

\[
Y = \mu + \alpha_1 X + \epsilon_1 \\
Y = \mu + \alpha_2 X + \epsilon_2 \\
Y = \mu + \alpha_3 X + \epsilon_3
\]

Adjusted Group Means

- point where regression line crosses grand mean of covariate
- provides estimates of group differences after removing effects of covariate
- most useful when groups differ considerably in terms of mean covariate
### Computing Adjusted Means

```r
> lm.1 <- lm(y ~ activity + drug, data = theData)
> dummy.coef(lm.1)
```

Full coefficients are:

|          | Estimate | Std. Error | t value | Pr(>|t|) |
|----------|----------|------------|---------|----------|
| (Intercept) | 1.28519  | 0.5124572  | 2.5075  | 0.0171  |
| drug: zero | -0.82161  | 0.4383696  | -1.8884 | 0.0633  |
| drug: low  | -0.62058  | 0.3904978  | -1.6030 | 0.1186  |
| drug: high | 0.10983   | 0.3059997  | 0.3591  | 0.7228  |

dummy coefficients list the parameters for the lines fit to each group

```
> library(effects)
> effect(term = "drug", lm.1)
```

drug effect

- drug
  - zero
  - low
  - high

|          | Estimate | Std. Error | t value | Pr(>|t|) |
|----------|----------|------------|---------|----------|
| (Intercept) | 6.059691 | 0.4383696  | 13.9089 | < 2e-16 |
| drug: zero | 0.1186933| 0.04383696 | 2.6964  | 0.0073  |
| drug: low  | 0.26037  | 0.04383696 | 5.9348  | < 2e-16 |
| drug: high | 1.26037  | 0.04383696 | 28.6955 | < 2e-16 |

### Association strength

I created the fake dependent variable, `activity.c + drug`, and interacted with them:

```
> lm.2 <- lm(y ~ activity, data = theData)
> library(effects)
> effect(term = "drug", lm.2)
```

- `lm.2` uses the `activity` variable to create a model.
- The effect of the `drug` variable is estimated after accounting for `activity`.

```
> lm.3 <- lm(y ~ activity + drug, data = theData)
> library(effects)
> effect(term = "drug", lm.3)
```

- The model `lm.3` includes both `activity` and `drug` as predictors.
- The effect of `drug` is estimated after accounting for both `activity` and `drug`.

```
> lm.4 <- lm(y ~ activity.c + drug + activity.c:drug, data = theData)
> library(effects)
> effect(term = "drug", lm.4)
```

- The model `lm.4` includes the `activity.c` and `drug` variables, as well as their interaction.
- The effect of `drug` is estimated after accounting for both `activity.c` and their interaction.

### Homogeneity of slopes assumption

Check for homogeneity of slopes:

```
> ZvsH        2.082 0.825 44 2.523   0.0153
> ZvS         3.024 0.085  8 3.743   0.0003
```

- The null hypothesis is that the slopes are equal across groups.
- The alternative hypothesis is that the slopes are not equal.
- The test statistic is the Wilcoxon test.
- The p-value is calculated using the null distribution of the Wilcoxon test statistic.

### Linear contrasts on adjusted means

```
> lm.1 <- lm(y ~ activity + drug, data = theData)
> levels(theData$drug)
```

- The model `lm.1` uses the `activity` and `drug` variables to create a model.
- The levels of the `drug` variable are: `zero`, `low`, and `high`.

```
> w1 <- c(-1, 1, 0)
> levels(theData$drug)
```

- The `w1` vector contains the weights for the contrast.
- The levels of the `drug` variable are repeated.

```
> lm.4 <- lm(y ~ activity.c + drug + activity.c:drug, data = theData)
> contrasts(theData$drug)
```

- The model `lm.4` includes both `activity.c` and `drug` as predictors.
- The contrast estimates are calculated using the `contrasts` function.
- The `drug` variable is contrasted against the reference level.

```
> contrasts <- contr.treatment(N, d, code = 1)
```

- The `contr.treatment` function is used to create contrast matrices.
- The `N` argument specifies the number of levels in the factor.
- The `d` argument specifies the desired Type I error rate.
- The `code` argument specifies how to handle ties in the data.

```
> contrast <- t(contrasts[theDrug, theDrug])
```

- The contrast matrix is calculated using the `contr.treatment` function.
- The `theDrug` variable is used to select the contrast matrix.
- The contrast estimates are calculated using the `t` function.
Why not difference scores?

> theData$diff <- theData$y - theData$activity;  
> diff.lm.1 <- lm(diff~drug,data=theData)  
> anova(diff.lm.1);  

Analysis of Variance Table  
Response: diff  

Df  Sum Sq  Mean Sq  F value  Pr(>F)  
drug  2  41.46  20.7318  2.5196  0.09179 .  
Residuals 45 370.27  8.2282

ANOVA on difference scores (activity - baseline.activity) finds drug is not significant. Why?

Why not difference scores?

\[ Y_{ij} - X_{ij} = \mu + \alpha_j + \epsilon_{ij} \]  

ANOVA on difference scores equivalent to ANCOVA with slope fixed at 1.0

\[ Y_{ij} = \mu + \alpha_j + X_{ij} + \epsilon_{ij} \]