Introduction:

- **controlled** variables - these are also known as *factors*. For example, we have *treatment* factors and *blocking* factors. These variables are directly controlled by the researcher.

- **uncontrolled** variables - these are also known as *covariates*. The researcher has no control of these.

- A covariate is quantitative in nature.

- Examples of covariates
  1. initial conditions - soil moisture, humidity, temperature, etc.
  2. volunteers - weight, age.
  3. “before” measurements - endurance tests *before* using a training method, or blood cholesterol *before* taking a new drug regime.

- We use **blocking** to control for sources of variability.

- In the presence of uncontrolled sources of variation we can use ANCOVA.

- Uncontrolled variables are sources of variability.

- Uncontrolled sources of variation inflate MSE which reduces power.

**ANCOVA - overview**

- ANCOVA is a technique that combines the ideas of ANOVA and regression.

- ANCOVA can be used with any designed experiment or observational study.

- Main idea - augment the ANOVA model with additional variables in order to reduce the model error variance.

- The model for a one factor study:
  - Notation:
    - $y_{ij}$ - response for treatment $i$, replicate $j$.
    - $x_{ij}$ - covariate for treatment $i$, replicate $j$.
    - $a$ - number of levels of our one Factor A.
    - $n$ - number of replicates.
  - ANCOVA Model:
    $$y_{ij} = \mu + \tau_i + \beta x_{ij} + \epsilon_{ij}$$
– ANOVA Model:

\[ y_{ij} = \mu + \tau_i + \epsilon_{ij} \]

– Model Assumptions
  1. The usual ones on the error terms in the model.
  2. The slope, \( \beta \), does not depend on the treatment.

Choosing Covariates

• First note that some texts call covariates concomitant variables - nonetheless, they are still quantitative variables added to an ANOVA model.

• A covariate needs to be related to the reponse or it will not serve to reduce variability.

• A covariate, however, should not be effected by the treatment.

• It is best to measure covariates prior to the experimental manipulation.

• If measured during the experiment exercise caution to be sure the covariate is not being influenced by the treatment.

• Example - Psychiatric therapies. We may want to test several different psychiatric therapies on teenage depression. The treatment factor is therapy with, say, three levels (three different types of therapies). We may use a prestudy survey to assess each patient’s attitude prior to treatment. A composite score based on the survey could be used as a covariate. We would not want to measure this particular covariate during treatment as a patient’s attitude may change because of the therapy.

• Another nice example is found in the text on p. 1012 and 1013.

Implementation - A Blood Cholesterol Example

An experiment was carried out to determine the efficacy of two drugs (drug A and drug B) for reducing cholesterol of patients showing high cholesterol readings. Twenty volunteers were available. The two drugs were randomly assigned to the volunteers. Blood cholesterol was measured before drug administration, and following 3 months of drug treatment. The SAS program is `ancova_blood.sas` and the data are found below:
To see how a covariate can reduce variability, let’s analyze the data using a one-way ANOVA with and without using the covariate.

Results of using a one-way ANOVA without the covariate.

ANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug</td>
<td>1</td>
<td>6.05</td>
<td>6.06</td>
<td>.01</td>
</tr>
<tr>
<td>Error</td>
<td>18</td>
<td>8960.90</td>
<td>497.83</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>8966.95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:

1. Based on a $p$-value of .9134, one has to conclude that there is no difference between A and B.

2. Mean square error for this model is 497.83. Watch the reduction in MSE when the covariate is added.

Testing Drugs after considering the effects of the covariates - equal slopes.

ANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>1</td>
<td>7453.81</td>
<td>7453.81</td>
<td></td>
</tr>
<tr>
<td>Drug after Before</td>
<td>1</td>
<td>926.64</td>
<td>926.64</td>
<td>26.86</td>
</tr>
<tr>
<td>Error</td>
<td>17</td>
<td>586.50</td>
<td>34.50</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>8966.95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3
Notes:

1. The effect of Drug after adjusting for the covariate is now significant with a $p$-value of .0001.

2. MSE is now 34.50 - a substantial reduction which gained us the power to detect the difference.

3. What is the effect of the covariate after taking into consideration the effects of the Drugs? We will use Type III sums of squares to answer this one.

4. Using Type III sums of squares we see that the covariate is significant after adjusting for the effects of Drugs ($F = 242.74$ w/ df1 = 1, df2 = 17, $p$-value = .0001).

5. These results suggest that we have two lines with the same slope and different $y$-intercepts.

6. Regression line for Drug A: $\hat{y} = -0.985 + 1.0003(Before)$.

7. Regression line for Drug B: $\hat{y} = -15.29 + 1.0003(Before)$.

8. Which Drug is more effective? Drug B - plot the two equations and you will see why.

9. Examine the least square means. These are the means of Drugs A and B after adjusting for the effects of the covariates.

10. Drug A - Sample mean = 245.6, Adjusted mean = 252.2.

11. Drug B - Sample mean = 244.5, Adjusted mean = 237.9.

12. The adjusted means show stronger evidence for the effect of Drug B (notice how much further apart they are).


\[
\begin{array}{c}
\hat{y}_{ij} \\
270 \\
252 \\
238 \\
220 \\
\end{array}
\begin{array}{c}
\bar{x}_i \\
\end{array}
\]

range of covariate $x$
Testing for the effect of Drugs after the effects of covariates - unequal slopes

**ANOVA**

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>1</td>
<td>7453.8</td>
<td>7453.8</td>
<td></td>
</tr>
<tr>
<td>Drug</td>
<td>1</td>
<td>926.6</td>
<td>926.6</td>
<td>26.9</td>
</tr>
<tr>
<td>Interaction</td>
<td>1</td>
<td>6.2</td>
<td>6.2</td>
<td>.17</td>
</tr>
<tr>
<td>Error</td>
<td>17</td>
<td>580.3</td>
<td>36.3</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>19</td>
<td>8967.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:

1. With a \( p \)-value of .6892 we can not reject \( H_0 \). There is no interaction between the covariate and drug. That is, the effect of the drug does not depend on your “Before” blood cholesterol reading.

2. If the slopes were unequal, then the estimated regression line for Drug A is

\[
\hat{y} = 7.66 + 0.97(Before)
\]

3. Similarly, for Drug B we have

\[
\hat{y} = -20.80 + 1.02(Before)
\]