d. Sometimes measurements come from different groups
   i. Group membership might represent a third variable that might influence mechanism of explanation.
   ii. Use different symbols for group members. See Fig. 17.

Fig. 17: Blood Lead vs. Floor Lead

D&P: 5.1

2. Score strength of relationship: Covariance.
Lecture 5

a. How should score behave?
   i. Want score positive when relationship is strongly positive
   ii. Want score negative when relationship is strongly negative
   iii. Want score near zero when relationship is weak.

b. How to construct?
   i. Calculate means for $X$, $Y$
   ii. Calculate difference between obsn and mean for each observation.
   iii. Multiply deviations for $X$ and $Y$
      - When both observations are well above the mean the contribution is very $+$
      - When both observations are well below the mean the contribution is very $+$
      - When one observation is well above the mean and the other is well below the mean, contribution is very $-$
      - When one observation is very near 0, contribution is near 0.
   iv. Add and divide by number of obsns -1.

c. Example: Figure 18/.

d. Units for covariance is product of units of constituent random
Fig. 18: Example of Covariance Calculation for Owners’ Children’s Log Lead, Intervention Group

<table>
<thead>
<tr>
<th>Log Blood Lead</th>
<th>Log Blood Lead -mean</th>
<th>Log Floor Lead</th>
<th>Log Floor Lead -mean</th>
<th>Product</th>
<th>Covariance</th>
<th>Variance Log Blood Lead</th>
<th>Variance Log Floor Lead</th>
<th>Covariance Log Blood Lead Log Floor Lead</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>0.29</td>
<td>0.52</td>
<td>0.04</td>
<td>0.01</td>
<td>0.22</td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.18</td>
<td>-0.49</td>
<td>0.36</td>
<td>-0.12</td>
<td>0.06</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.78</td>
<td>0.11</td>
<td>0.24</td>
<td>-0.24</td>
<td>-0.03</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.41</td>
<td>-0.25</td>
<td>0.29</td>
<td>-0.18</td>
<td>0.05</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.32</td>
<td>-0.34</td>
<td>-0.08</td>
<td>-0.55</td>
<td>0.19</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.51</td>
<td>-0.16</td>
<td>2.41</td>
<td>1.94</td>
<td>-0.31</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.78</td>
<td>0.11</td>
<td>1.05</td>
<td>0.58</td>
<td>0.07</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.83</td>
<td>0.17</td>
<td>0.08</td>
<td>-0.40</td>
<td>-0.07</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.66</td>
<td>0.00</td>
<td>0.86</td>
<td>0.39</td>
<td>0.00</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>1.06</td>
<td>0.40</td>
<td>1.03</td>
<td>0.56</td>
<td>0.22</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.59</td>
<td>-0.07</td>
<td>-0.05</td>
<td>-0.52</td>
<td>0.04</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.63</td>
<td>-0.03</td>
<td>-0.48</td>
<td>-0.95</td>
<td>0.03</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.34</td>
<td>-0.32</td>
<td>0.23</td>
<td>-0.25</td>
<td>0.08</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.94</td>
<td>0.28</td>
<td>0.08</td>
<td>-0.40</td>
<td>-0.11</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.89</td>
<td>0.22</td>
<td>0.23</td>
<td>-0.25</td>
<td>-0.05</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.80</td>
<td>0.13</td>
<td>0.80</td>
<td>0.33</td>
<td>0.04</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>0.61</td>
<td>-0.05</td>
<td>0.50</td>
<td>0.02</td>
<td>0.00</td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
<tr>
<td>11.28</td>
<td>8.07</td>
<td></td>
<td>0.22</td>
<td></td>
<td></td>
<td>11.28</td>
<td>8.07</td>
<td>0.014</td>
</tr>
</tbody>
</table>
variables

i. Doubling spread of one variable doubles the covariance without
   effecting really how the two variables interact

ii. Remove units and dependence on univariate spreads by
   dividing by sd’s to get correlation

e. Invariance:

i. Correlation is invariant to
   • Interchange of roles of $Y$ and $X$.
   • shift: adding a quantity to one or both variables
   • multiplying one or both quantities by a $+$ constant.

f. Interpretation:

i. $+1$: Perfect linear $+$ association
   • What we get if one is copy of other
   • By invariance, what we get if one is $+$ multiple of another plus
     constant
   • Converse true but harder to see.

ii. $-1$: Perfect linear $-$ association
   • What we get if one is $-$ of other
   • By invariance, what we get if one is $-$ multiple of another plus
Lecture 6

constant

- Converse true but harder to see.

iii. 0: No linear association

g. Examples:

i. Small Correlation: Figure 19/

ii. Moderate Correlation: Figure 20/

iii. Large Correlation: Figure 21/

h. Caveats:

i. Neglects non-linear association as we will see later

ii. Doesn’t imply causation.
Fig. 19: Demonstration of Correlation for Owner’s Children, Intervention Group

Log Floor Lead Level

Log Blood Lead Level

Correlation = 0.082
Fig. 20: Demonstration of Correlation for Owners’ Children, Intervention Group

Correlation = 0.46
Fig. 21: Demonstration of Correlation for All Children

Correlation = 0.975