Quantitative Methods (GEO 441)
SPSS Lab 4: Correlation
Dr. Paul Marr

Please copy the file **S:\GEO\Marr\Quantitative Methods\SPSS Example Data\Pearsons Correlation.sav**, **Spearmans Rank Correlation.sav**, and **Cramers Phi.sav** to your portable media.

- Start SPSS.

Parametric Bivariate Correlation

1. Open **Pearsons Correlation.sav.**
2. Analyze > Descriptive Statistics > Explore.
    a. Run normality test on the variables **Male Standing Height**, **Male Leg Length**, and **Male Arm Length**.
        i. Are these data normally distributed?
3. Analyze > Correlate > Bivariate.
    a. Move **Male Standing Height**, **Male Leg Length**, and **Male Arm Length** to the *Variables* list.
    b. Check the **Pearson** box under *Correlation Coefficients*.
    c. Make sure the **Flag significant correlations** box is checked.
    d. Click **Ok**.
        i. Which of these 3 variables are correlated?
        ii. Is the correlation significant? What are the probabilities?
        iii. What are the signs of the correlation coefficients?
4. Graphs > Legacy Dialog > Scatter/Dot
    a. Choose **Matrix Scatter**.
    b. Click **Define**.
    c. Move **Male Standing Height**, **Male Leg Length**, and **Male Arm Length** to the *Matrix Variables* list.
    d. Click **Ok**.
        i. The scatterplots make the relationship between the variables clear: in all cases as one variable increases the other increases. This is termed a ***positive*** relationship and the sign of the correlation coefficient is positive. If one variable increase while the other decreased, this would be termed a ***negative*** relationship and the sign of the correlation coefficient is negative.

Partial Correlation

If it is suspected that another variable(s) may be influencing the correlation between two variables, the influence of this 'confounding variable' can be controlled using *Partial Correlation*. In our example, it is thought that the influence of age may be diminishing the correlation, in that age influenced body morphology.

1. Analyze > Correlate > Partial
    a. Move **Male Standing Height**, **Male Leg Length**, and **Male Arm Length** to the *Variables* list.
    b. Move **Mean Age** to the *Controlling for* list.
    c. Click on **Options**.
        i. Check the **Zero-order correlations** box.

ii. Click **Continue**.

d. Click **Ok**.
   i. The *upper* part of the table displays the results **without** controlling for Mean Age.
   ii. Note the correlation coefficients.
   iii. The *lower* part of the table displays the results **while** controlling for Mean Age.
   iv. Note that the coefficient for Male Arm Length*Male Standing Height increased while the coefficient for Male Leg Length*Male Standing Height decreased.
   v. This indicates that age is having a slightly positive influence on leg length to height, while a slightly negative influence on arm length to height.

<u>Spearman's Non Parametric Correlation</u>

1. Open **Spearmans Rank Correlation.sav**.
2. Analyze > Descriptive Statistics > Explore.
   a. Run normality test on the variables **Crude Rate per 100000** and **Average PM10 Level**.
      i. Are these data normally distributed?
3. Graphs > Legacy Dialog > Scatter/Dot
   a. Choose Simple **Scatter**.
   b. Click **Define**.
   c. Move **Crude Rate per 100000** to the *Y-Axis* list and **Average PM10 Level** to the *X-Axis* list.
   d. Click **Ok**.
      i. Do these variables appear to be linearly correlated? Are they correlated at all?
      ii. There appears to be a correlation but it is not linear. We can address this with a log transformation.
4. Transform> Compute Variable
   a. Type in **LNCrude** in the *Target Variable* field.
   b. Under *Function Group* click on **Arithmetic**.
   c. Under the *Functions and Special Variable* list double click on **Ln**.
   d. Under the *Type and Label* list double click on **Crude Rate per 100000**.
   e. Click **Ok**.
      i. This creates a new variable called LNCrude that has the logged values of the crude asthma rate per 100000.
5. Graphs > Legacy Dialog > Scatter/Dot
   a. Choose Simple **Scatter**.
   b. Click **Define**.
   c. Move **LNCrude** to the *Y-Axis* list and **Average PM10 Level** to the *X-Axis* list.
   d. Click **Ok**.
      i. Do these variables appear to be linearly correlated?
6. Analyze > Correlate > Bivariate.
   a. Move **LNCrude** and **Average PM10 Level** to the *Variables* list.
   b. Uncheck the Pearson box and check the **Spearman** box.
   c. Click **Ok**.
      i. Are these variables correlated?
      ii. Is the correlation significant? What are the probabilities?
      iii. What are the signs of the correlation coefficients?

Spearman's Partial Correlation

We are interested in whether the number of young and old residents is influencing our correlation results. Therefore we need to perform a Spearman's Partial Correlation… unfortunately SPSS makes this a little challenging. We have to use a *Syntax script*.

1. From the Data Editor window, File > New > Syntax.
    a. Enter the commands below into the Syntax window exactly as they appear:

```
NONPAR CORR LNCrude AvgPM10 Children Elderly
/PRINT=NOSIG
/MISSING = LISTWISE
/MATRIX OUT(*).
RECODE rowtype_ ('RHO'='CORR') .
PARTIAL CORR LNCrude AvgPM10 BY Children Elderly
/SIGNIFICANCE=ONETAIL
/STATISTICS=CORR
/MISSING = LISTWISE
/MATRIX IN(*).
```

    Note: The above syntax performs a Spearman's correlation and writes the correlation matrix output to a data file, recodes the RHO variable to be CORR (SPSS will only run partial correlation on a variable called CORR), then runs a Pearson's partial correlation on the Spearman results. The (*) command makes the output matrix the active worksheet.
    b. In the Syntax Editor window, Run > All.
        i. What influence are the Children and Elderly variables having on the correlation between logged asthma rates and average PM10?

Cramer's Correlation

Cramer's correlation is used to determine if dichotomous variables (1,0) are correlated.

1. Open Cramers Phi.sav.
2. Analyze > Descriptive Statistics > Crosstabs.
    a. Move **RoadPath** to the *Row(s)* list and **Crafts** to the *Column(s)* list.
    b. Click on **Statistics**.
        i. Check the **Phi and Cramer's V** box.
        ii. Click **Continue**.
    c. Click **Ok**.
        i. What is the Cramer's correlation coefficient?
        ii. Is it significant?
3. Analyze > Descriptive Statistics > Crosstabs.
    a. Move **LakeMnts** to the *Row(s)* list and **Employment** to the *Column(s)* list.
    b. Click on **Statistics**.
        i. Check the **Phi and Cramer's V** box.
        ii. Click **Continue**.
    c. Click **Ok**.
        i. What is the Cramer's correlation coefficient?
        ii. Is it significant?
        iii. What does this mean?