BUS 735: Business Decision Making and Research
Instructor: Dr. James Murray
SPSS Handout: Correlation, Comovement, and Regression Analysis

1 Correlation

Example: Public Expenditure

- Data from 1960 about public expenditures per capita, and variables that may influence it:
  - Economic Ability Index
  - Percentage of people living in metropolitan areas.
  - Percentage growth rate of population from 1950-1960.
  - Percentage of population between the ages of 5-19.
  - Percentage of population over the age of 65.
  - Dummy variable: Western state (1) or not (0).
- Is there a statistically significant linear correlation between the percentage of the population who is young and the public expenditure per capita?
- Is there a statistically significant linear correlation between the public expenditure per capita and whether or not the state is a western state?

1. Open the dataset *publicexp.xls* in SPSS.
2. For a parametric test (Pearson correlation):
3. Select `Analyze` menu, select `Correlate`, then select `Bivariate`.
4. Select at least two variables (it will do all pairwise comparisons) on the left and click right arrow button.
5. Select check-box for `Pearson` and/or `Spearman`.
6. Click OK!

2 Chi-Squared Test for Independence

Unsatisfied Customers: Reason for Hotel Guests' Stay vs. Reasons They will Not Return

	Reason for Not Returning		
Reason for Stay	Price	Location	Amenities
Personal/Vacation	56	49	0
Business	20	47	27

Using SPSS:

- Dataset: `hotel.sav`.
- First column, `ReasonStay`: 0=Personal/Vacation, 1=Business.
- Second column, `NoReturn`: 0=Price, 1=Location, 2=Amenities.
- Go to `Analyze`, `Descriptive Statistics`, `Crosstabs`.
- Put one of the variables in the `Row(s)` box.
- Put the other variable in the `Column(s)` box.
- Click `Statistics` button.
- Check the box for `Chi-square`.
- Click OK!

3 Regression Analysis

Example: Public Expenditure

- Data from 1960 about public expenditures per capita, and variables that may influence it.
- Select your independent (aka explanatory) variables. These are the variables that you think can explain the dependent variable. I suggest you select these:
  - ECAB: Economic Ability
  - MET: Metropolitan
  - GROW: Growth rate of population
  - WEST: Western state = 1.

Conduct Regression Analysis to answer the following questions:

1. If the percentage of the population living in metropolitan areas in expected to increase by 1%, what change should we expect in public expenditure?
2. Is this change statistically significantly different from zero?
3. Accounting for economic ability, metropolitan population, and population growth, how much more to Western states spend on public expenditure per capita?

Using SPSS

1. Open *publicexp.sav* in SPSS.
2. Select from menu: `Analyze`, `Regression`, then `Linear`.

3. Move EX to the Dependent variable list.
4. Move ECAB, MET, GROW, and WEST to your Independent variable list.
5. Click OK!

Regression output shows:

- Coefficient of Determination (aka R^2) (more on this ahead...)
- Analysis of Variance Table (more on this ahead...)
- Coefficient Estimates, including standard errors, t-statistics, p-values.

Using SPSS to examine violations of assumptions

- To examine multicollinearity possibilities:
  - Check standard errors / significance levels of your coefficients - if variables that could be related are insignificant (have a large standard error), then there may be a problem.
  - Compute Pearson correlation coefficients for potential problematic variables.
    1. Select from menu: `Analyze`, `Correlate`, then `Bivariate`.
    2. Move all your *Explanatory variables* to the `Variables` box.
    3. Select checkbox for `Pearson`.
    4. Click OK.
  - Do you find any variables highly correlated with one another?
- To examine normality of error term:
  - Check to see if the residuals are normally distributed.
    1. Set up regression dialog as before.
    2. Click `Plots`
    3. Select checkbox for `Normal Probability Plot`.
    4. Select checkbox for `Histogram`.
    5. Click `Continue`
    6. Click OK.
  - The histogram of standardized residuals should appear bell-shaped.
  - The Normal Probability Plot should contain datapoints close to the line, with no discernible pattern.
  - Do the residuals appear to be approximately normally distributed?
- To examine homoscedasticity / linearity issues
  - Compute standardized residuals.

1. Set up regression dialog as before.
2. Click `Save`
3. Under *Residuals*, select checkbox for `Standardized`.
4. Click `Continue`
5. Click OK.

– Plot residuals against one of the explanatory variable to look for a pattern (there shouldn't be any).

1. Select menu item `Graphs`, `Legacy Dialogs`, `Scatter/Dot`
2. Select `Simple Scatter` and click `Define`
3. Move standardized residuals to the Y-Axis, move one of the continuous explanatory variables to the X-Axis.
4. Click OK.

– Things to look for:

* These plots should have residuals randomly above and below zero with no discernible pattern (violation may imply a non-linear relationship).
* Variability of residuals (how spread out they are) should not change as explanatory variable changes (violation implies heteroscedasticity).