Tutorial Concept Sheet 6
Psychology 2101
Fall, 2008

Continuous vs. Bivariate Data

Scatter Plots

Covariance between X and Y (two formulas)

$$S_{X,Y} = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}_\bullet\right)\left(Y_i - \bar{Y}_\bullet\right)$$

$$S_{X,Y} = \frac{1}{N-1}\left(\sum_{i=1}^{N} X_i Y_i - \frac{\sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i}{N}\right)$$

Problems with covariance
- Not scale free

Correlation between X and Y

$$r_{X,Y} = \frac{1}{N-1}\sum_{i=1}^{N} Zx_i Zy_i$$
$$= \frac{1}{N-1}\sum_{i=1}^{N}\frac{\left(X_i - \bar{X}_\bullet\right)}{S_X}\frac{\left(Y_i - \bar{Y}_\bullet\right)}{S_Y}$$
$$= \frac{S_{X,Y}}{S_X S_Y} = \frac{S_{X,Y}}{\sqrt{S_X^2 S_Y^2}}$$

Interpreting a Correlation Coefficient
- Sign
- Absolute Value

Interpreting a Scatter Plot

The Spearman Rank-Order Correlation

$$r_s = 1 - \frac{6\sum_{i=1}^{N} D_i^2}{N(N^2 - 1)}$$

Linear Regression

$$Y_i = \hat{Y}_i + E_i$$

$$\hat{Y}_i = bX_i + a$$

The "Least Squares Criterion"

Minimize $\sum_{i=1}^{N} E_i^2$

The "Least Squares Solution" Use

$$b = \frac{r_{YX} S_Y}{S_X}, \quad a = \bar{Y}_\bullet - b\bar{X}_\bullet$$

Linear Regression on *Z*-scores

$$\hat{Y}_i = r_{YX} X_i$$

Two Uses of Regression

- Description
- Prediction

Identifying Univariate and Bivariate (Regression) Outliers

Correlational Problems and Fallacies

- Interpreting a correlation
- Correlation and causality
- Perfect correlation and equivalence
- No Correlation vs. No Relation
- Combining Populations, and Ignoring Explanatory Variables
- Restriction of Range