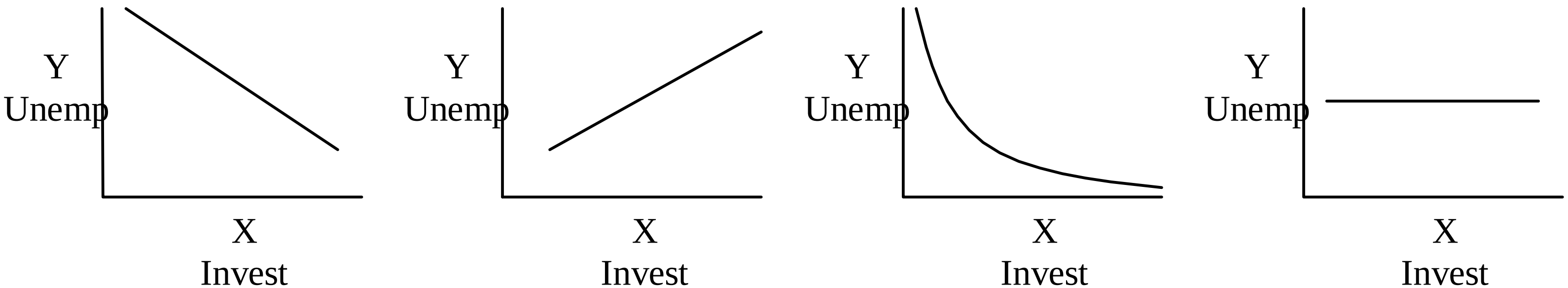
- Interval/Ratio Level Relationships
- Correlation
- Regression
  - ordinary least squares (OLS)

Relationships

- Interval & Ratio level data can co-vary in several ways
- The direction in which variables co-vary determines the type of relationship

Ratio Y Unemp ← Ratio X Invest



Indirect ***Direct*** ***Curvilinear*** ***No Relation***

Slope (b) determines the directionality of relationship

Correlation

- A test to identify if variables co-vary
- The co-variance of two variables has come to be referred to as a statistical relationship
- Pearson's r tests the strength, direction of a relationship, and statistical significance of an observed relationship
  - Values of Pearson's r (-1.0 and 1.0)

Correlation

- Using Pearson's r, a perfect positive relationship would be assigned '1' and indicate a perfect 1-to-1 relationship between variables

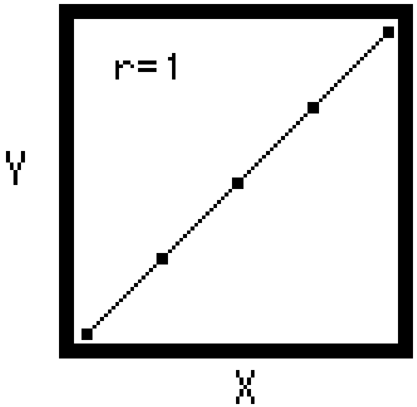
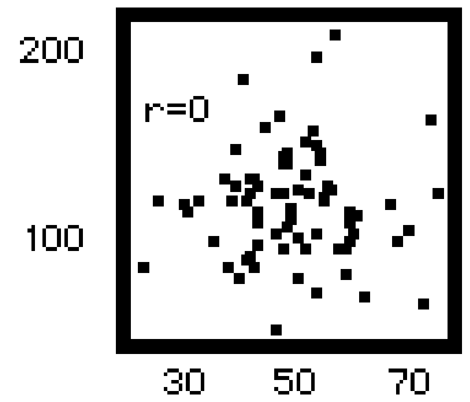


Image From Hyperstat On-Line

Correlation

- Most relationships are not perfect
- Relative Strength of Correlation

  – 0 No relationship



  – .5-.7 Moderate/Strong

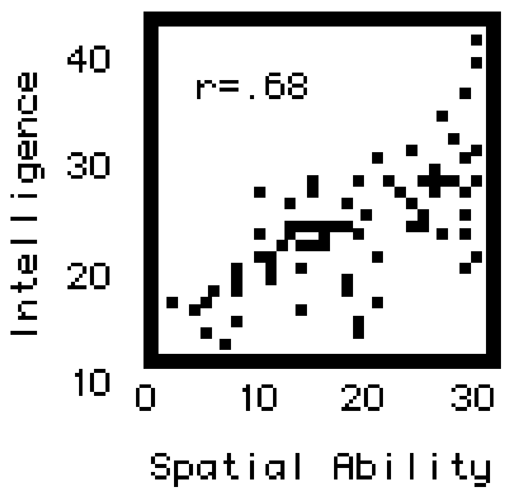


Image From Hyperstat On-Line

Correlation

$$r = \frac{\Sigma z_x z_y}{N}$$

Pearson's r where "Z" represent standardized variables

Image From Hyperstat On-Line

Regression

- Regression (r-square) tests whether (or not) the variance of one variable can ‘statistically explain’ the variance of another and the percent of explanation
  - assumes a linear relationship
  - assumes a specific ‘directionality’ exists, but not causation

Y=Dependent Variable X=Independent Variable

X → Y

Regression

- Regression is about identifying "The Best Fit Line"

$$\hat{Y}=a+bX$$

- "The Best Fit" line (a sophisticated 'mean' of sorts) is the line that minimizes variation between observed cases
- The exact equation is based on (1) calculating the specific deviation of each case (called a residual) and (2) minimizing the sum of the residuals squared

$$\Sigma(Y_i-\hat{Y})^2$$

Y_i=Observed Value $\hat{Y}$=Predicted Value

Other Components of Regression

- Obtain 'a' (intercept) $a = \bar{Y} - b\bar{X}$
- Obtaining 'b' (slope)

$$b = \frac{\Sigma (X_i - \bar{X})(Y_i - \bar{Y}) \quad \text{[CoVariation in X \& Y]}}{\Sigma (X_i - \bar{X}) \quad \text{[Variation in X]}}$$

- Testing the "Line"

Assume $\beta = 0$

$$t = \frac{b}{\hat{\sigma}_b}$$

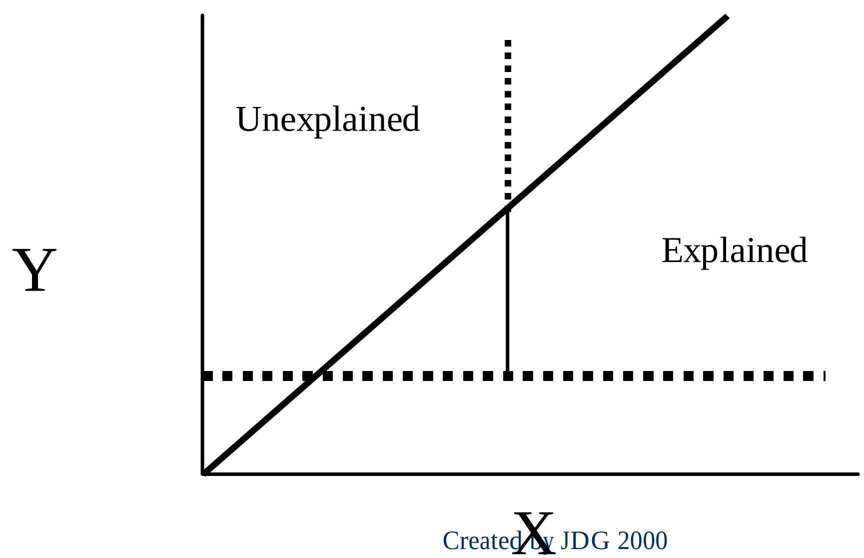
Regression

- Explaining Variance (R-square)

Explained Variance

$$R^2 = \frac{\text{Explained Variance}}{\text{Total Variance}} = \frac{[\Sigma(Y-Y)^2]/n}{[\Sigma(Y-Y)^2]/n} = \frac{\Sigma(Y-Y)^2}{\Sigma(Y-Y)^2}$$

= % of Variance by “X”



Regression

- Is explained variance statistically significant?

$$F = \frac{\text{Explained}}{\text{Unexplained}}$$

$$\text{Unexplained} = 1 - \text{Explained}$$

For critical values see **<u>Appendix C</u>** tables in reserve notes