

Regression & Correlation

Regression is the same idea as ANOVA:

- ANOVA is good when treatment effects give discrete separable categories along different dimensions.
- Regression analysis is the same thing:
- Separating out or explaining some of the variation based on a defined cause.
- Regression is used when the magnitude of the cause and the magnitude of the effect are continuous rather than discrete.
- The observed value in an individual can be described as a sum of the explained effects and the "random errors":

$$y = b_0 + b_1 * X_1 + b_2 * X_2 \dots + e$$

where: y = observed value

b₀ = the y intercept

b_i = the regression coefficient that goes with cause X_i

e is the unexplained "random error" that causes individual values of y to deviate from the prediction.

- Significance of the β_i effects are measured by comparison against the size of the error component.
- Correlation coefficient is a measure of how much of the variation is explained by relating the observations to the known cause.
- High correlation does not necessarily imply causality:
 - o toilet flush causes computer reboot.
 - o Vanilla icecream at the convenience store-engine fails to restart other flavor it's okay.

Linear Correlation & Regression

Sir Francis Galton: "regression toward the mean"

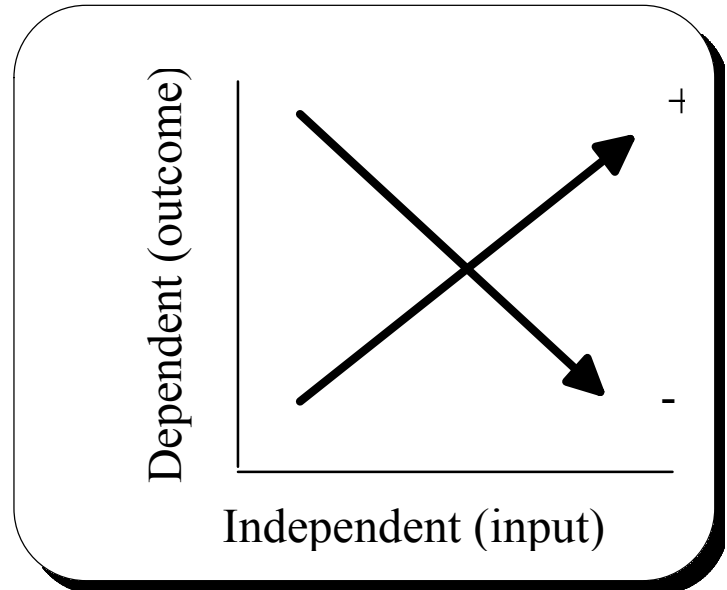
Linear Correlation:

- Strength of Association.
- Prediction equation
- Line of best fit
- "Least Squares" fit to a line

Linear equation $Y = a + b X$
has error term:

Limitations:

- $n < 50$ questionable
- outliers have exaggerated effects
- misleading if nonlinear
- may not be cause -and-effect relationship
- Leaves residual, unexplained, error term



Correlation Coefficient:

Max +1 or -1 depending on slope

Is the Correlation Coefficient Statistically Significant?

Variance around the line:

- Standard error for evaluating whether β is different from zero:
- Confidence interval around estimate of slope:
- t test with d.f.=n-2 test whether slope is "statistically significant"
- d.f.=number of paired observations -2