

Module 14b – Inference for Regression

[Review Against All Odds: Unit 30 \(Inference for Regression\)](#)

Stat Procedure Diagram – Where we are

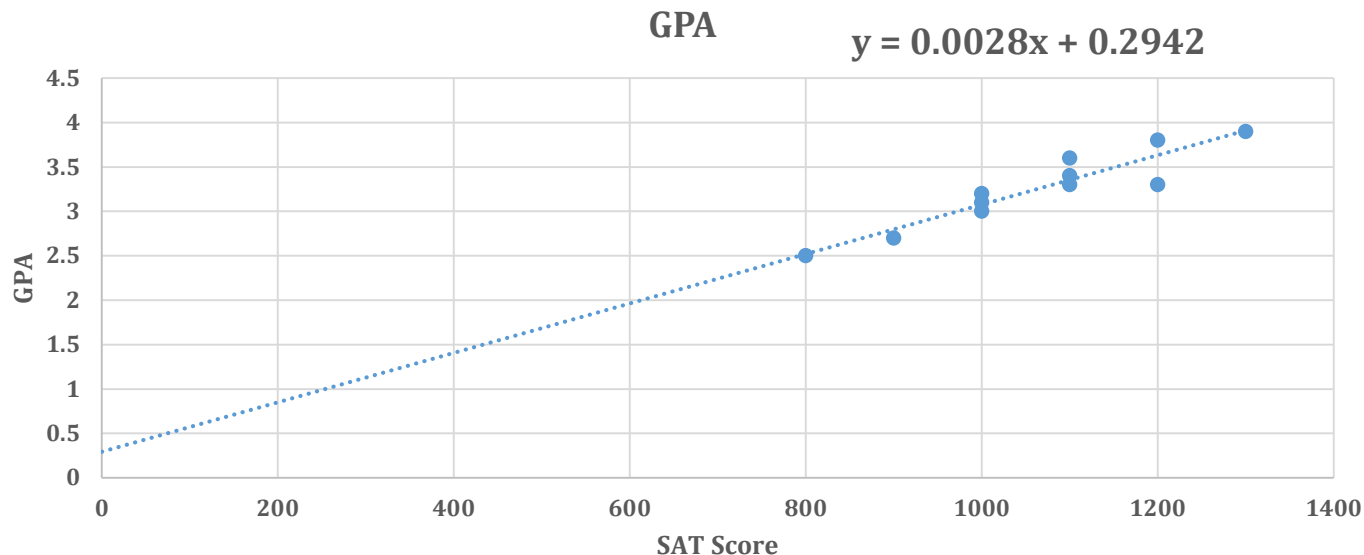
		<u>Descriptive Statistics (Describing Pops or Samples)</u>		<u>Inferential Statistics (from Samples)</u>
	Variable Types	Display	Describe	Estimation
Univariate	categorical (nominal or ordinal)*	Bar Graph/Pie Chart	Counts/Percentages	When binary/dichotomous: Confidence interval for proportions
	quantitative/continuous	Histogram/Stem & Leaf Box Plot	Mean/St Dev (normal) Median/Min, Q1, Q3, Max (skewed)	Confidence interval for means
		Display	Describe	Significance Tests/Hypothesis Tests
Bivariate	2 categorical	Tables or Bar Graphs	Two-way tables/Crosstabulation	Chi-square test (for goodness of fit)
	1 categorical, 1 quant.	Bar Graphs	Comparison of means/averages	T-test (one sample/group, two samples/groups) ANOVA (two or more samples/groups)
	2 quant.	Scatterplot	Correlation Coef. (Coef. of determ)/ Regression Line	T-test for correlation
		Display	Describe	Significance Tests/Hypothesis Tests
Multivariate	Response Variable is Quant.	-	Ordinary Least Squares Regression (OLS)	F-test for overall model T-tests for each explanatory variable
	Response Variable is categorical (dichotomous)	-	Logistic Regression	Chi-square tests of significance

NOTE: Items highlighted in yellow are covered in this course.

*When a categorical variable has two categories, it is called dichotomous.

Regression Line (Review)

- Formula for regression line:
 - $y = a + bx$
 - y = response/dependent variable value
 - x = explanatory/independent variable value
 - b = slope (rise/run)
 - For every one unit increase in x , there is a predicted b change in y
 - a = y -intercept
 - The value of y when $x = 0$ (where the line crosses y -axis)



Regression (Inference)

- How do we know if we can infer/generalize the sample relationship between SAT score and college GPA?
- We test for the statistical significance of “b” (our slope).
- Null Hypothesis: slope (b) = 0
 - (there is no association between SAT score and college GPA)
- Alternative Hypothesis: slope (b) > 0
 - (there is a positive association between SAT score & college GPA)
- OR***
- Alternative Hypothesis: slope (b) < 0
 - (there is a negative association between SAT score & college GPA)
- OR***
- Alternative Hypothesis: slope (b) > or < 0
 - (there is a pos or neg association between SAT score & college GPA)

Regression (Inference)

- Our t-statistic for regression is equal to the slope from our regression line divided by the standard error of the slope:
- $t = b(\text{slope}) / SE(b) \text{ slope}$
- If the resulting t-statistic is greater than our critical value for t, we reject the null hypothesis
- This would mean there IS good evidence that the slope is NOT zero (there is an association between SAT score and GPA).

Regression (Inference)

- A final point about regression. It is not limited to one independent/explanatory variable.
- Multiple regression tests for the impact of multiple independent/explanatory variables simultaneously.
- For example, is there a relationship between ***SAT score, high school GPA, high school extra curricular hours*** AND ***College GPA***.

$$y = a + b_{\text{SAT}} + b_{\text{HS GPA}} + b_{\text{EC hrs}}$$

- This allows us to test the impact of all three variables (while “statistically controlling” for the other two)
 - We have three null and alternative hypothesis (one for each slope) that we test with three t-statistics and critical values.
- We also end up with an R-squared value that tells us the *percentage of the variation in the dependent variable that can be explained by variation in the independent variables*