Unit 7: Statistical Inference and Confidence Intervals

Unit 7 Post Hole:

Interpret a confidence interval from a frequentist perspective and from a Bayesian perspective.

Unit 7 and 8 Technical Memo and School Board Memo:

Using a new data set, and choosing a new continuous outcome, a new continuous predictor and a new dichotomous predictor, fit and discuss two regression models.

Unit 7 (and Units 6 and 8) Reading:

http://onlinestatbook.com/

Chapter 5, Probability	Chapter 6, Normal Distributions
Chapter 7, Sampling Distributions	Chapter 8, Estimation
Chapter 9, Logic Of Hypothesis Testing	Chapter 10, Testing Means
Chapter 11, Power	Chapter 12, Prediction

Unit 7: Technical Memo and School Board Memo

Work Products (Part I of II):

I. Technical Memo: Have one section per biviariate analysis. For each section, follow this outline. (4 Sections)

- A. Introduction
  - i. State a theory (or perhaps hunch) for the relationship—think causally, be creative. (1 Sentence)
  - ii. State a research question for each theory (or hunch)—think correlationally, be formal. Now that you know the statistical machinery that justifies an inference from a sample to a population, begin each research question, “In the population,…” (1 Sentence)
  - iii. List the two variables, and label them “outcome” and “predictor,” respectively.
  - iv. Include your theoretical model.
- B. Univariate Statistics. Describe your variables, using <u>descriptive statistics</u>. What do they represent or measure?
  - i. Describe the data set. (1 Sentence)
  - ii. Describe your variables. (1 Short Paragraph Each)
    - a. Define the variable (parenthetically noting the mean and s.d. as descriptive statistics).
    - b. Interpret the mean and standard deviation in such a way that your audience begins to form a picture of the way the world is. Never lose sight of the substantive meaning of the numbers.
    - c. Polish off the interpretation by discussing whether the mean and standard deviation can be misleading, referencing the median, outliers and/or skew as appropriate.
- C. Correlations. Provide an overview of the relationships between your variables using <u>descriptive statistics</u>.
  - i. Interpret all the correlations with your outcome variable. Compare and contrast the correlations in order to ground your analysis in substance. (1 Paragraph)
  - ii. Interpret the correlations among your predictors. Discuss the implications for your theory. As much as possible, tell a coherent story. (1 Paragraph)
  - iii. As you narrate, note any concerns regarding assumptions (e.g., outliers or non-linearity), and, if a correlation is uninterpretable because of an assumption violation, then do not interpret it.

Unit 7: Technical Memo and School Board Memo

Work Products (Part II of II):

I. Technical Memo (continued)

   D. Regression Analysis. Answer your research question using <u>inferential statistics</u>. (1 Paragraph)

      i. Include your fitted model.

      ii. Use the R^2 statistic to convey the goodness of fit for the model (i.e., strength).

      iii. To determine statistical significance, test the null hypothesis that the magnitude in the population is zero, reject (or not) the null hypothesis, and draw a conclusion (or not) from the sample to the population.

      iv. Describe the direction and magnitude of the relationship in your sample, preferably with illustrative examples. Draw out the substance of your findings through your narrative.

      v. Use confidence intervals to describe the precision of your magnitude estimates so that you can discuss the magnitude in the population.

      vi. If simple linear regression is inappropriate, then say so, briefly explain why, and forego any misleading analysis.

   X. Exploratory Data Analysis. Explore your data using outlier resistant statistics.

      i. For each variable, use a coherent narrative to convey the results of your exploratory univariate analysis of the data. Don't lose sight of the substantive meaning of the numbers. (1 Paragraph Each)

      ii. For the relationship between your outcome and predictor, use a coherent narrative to convey the results of your exploratory bivariate analysis of the data. (1 Paragraph)

II. School Board Memo: Concisely, precisely and plainly convey your key findings to a lay audience. Note that, whereas you are building on the technical memo for most of the semester, your school board memo is fresh each week. (Max 200 Words)

III. Memo Metacognitive

Unit 7: Road Map (VERBAL)

Nationally Representative Sample of 7,800 8th Graders Surveyed in 1988 (NELS 88).
Outcome Variable (aka Dependent Variable):

READING, a continuous variable, test score, mean = 47 and standard deviation = 9

Predictor Variables (aka Independent Variables):

FREELUNCH, a dichotomous variable, 1 = Eligible for Free/Reduced Lunch and 0 = Not
RACE, a polychotomous variable, 1 = Asian, 2 = Latino, 3 = Black and 4 = White

- Unit 1: In our sample, is there a relationship between reading achievement and free lunch?
- Unit 2: In our sample, what does reading achievement look like (from an outlier resistant perspective)?
- Unit 3: In our sample, what does reading achievement look like (from an outlier sensitive perspective)?
- Unit 4: In our sample, how strong is the relationship between reading achievement and free lunch?
- Unit 5: In our sample, free lunch predicts what proportion of variation in reading achievement?
- Unit 6: In the population, is there a relationship between reading achievement and free lunch?
- Unit 7: In the population, what is the magnitude of the relationship between reading and free lunch?
- Unit 8: What assumptions underlie our inference from the sample to the population?
- Unit 9: In the population, is there a relationship between reading and race?
- Unit 10: In the population, is there a relationship between reading and race controlling for free lunch?
- Appendix A: In the population, is there a relationship between race and free lunch?

Unit 7: Roadmap (R Output)

```
> load("E:/User/Folder/RoadmapData.rda")
> library(abind, pos=4)
> numSummary(RoadmapData[,c("FREELUNCH", "READING")],
+   statistics=c("mean", "sd", "quantiles"), quantiles=c(0,.25,.5,.75,1))
               mean       sd    0%   25%   50%   75%  100%    n
FREELUNCH  0.3353846 0.472155  0.00  0.00  0.00  1.00  1.00 7800
READING   47.4940397 8.569440 23.96 41.24 47.43 53.93 63.49 7800
```

Unit 3 (mean, sd) — Unit 2 (quantiles, n)

```
> RegModel.1 <- lm(READING~FREELUNCH, data=RoadmapData)
> summary(RegModel.1, cor=FALSE)
Call:
lm(formula = READING ~ FREELUNCH, data = RoadmapData)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  49.1176     0.1147  428.17   <2e-16 ***
FREELUNCH    -4.8409     0.1981  -24.44   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.26 on 7798 degrees of freedom
Multiple R-squared: 0.07114, Adjusted R-squared: 0.07102
F-statistic: 597.3 on 1 and 7798 DF,  p-value: < 2.2e-16
```

Unit 1 (Estimate) — Unit 8 (Std. Error) — Unit 6 (t value, Pr(>|t|)) — Unit 5 (Adjusted R-squared) — Unit 9 (F-statistic)

```
> library(MASS, pos=4)
> Confint(RegModel.1, level=.95)
             Estimate     2.5 %    97.5 %
(Intercept) 49.117616 48.892742 49.342489
FREELUNCH   -4.840938 -5.229237 -4.452638
```

Unit 7

```
> cor(RoadmapData[,c("FREELUNCH","READING")])
           FREELUNCH    READING
FREELUNCH  1.0000000 -0.2667237
READING   -0.2667237  1.0000000
```

Unit 4

Unit 7: Roadmap (SPSS Output)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.267[a]	.071	.071	8.25952

a. Predictors: (Constant), FREELUNCH

Unit 5

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	40744.322	1	40744.322	597.251	.000[a]
	Residual	531977.541	7798	68.220		
	Total	572721.864	7799			

a. Predictors: (Constant), FREELUNCH

b. Dependent Variable: READING

Unit 9

Statistics

		READING	FREELUNCH
N	Valid	7800	7800
	Missing	0	0
Mean		47.4940	.3354
Std. Deviation		8.56944	.47216
Minimum		23.96	.00
Maximum		63.49	1.00
Percentiles	25	41.2400	.0000
	50	47.4300	.0000
	75	53.9300	1.0000

Unit 3

Unit 2

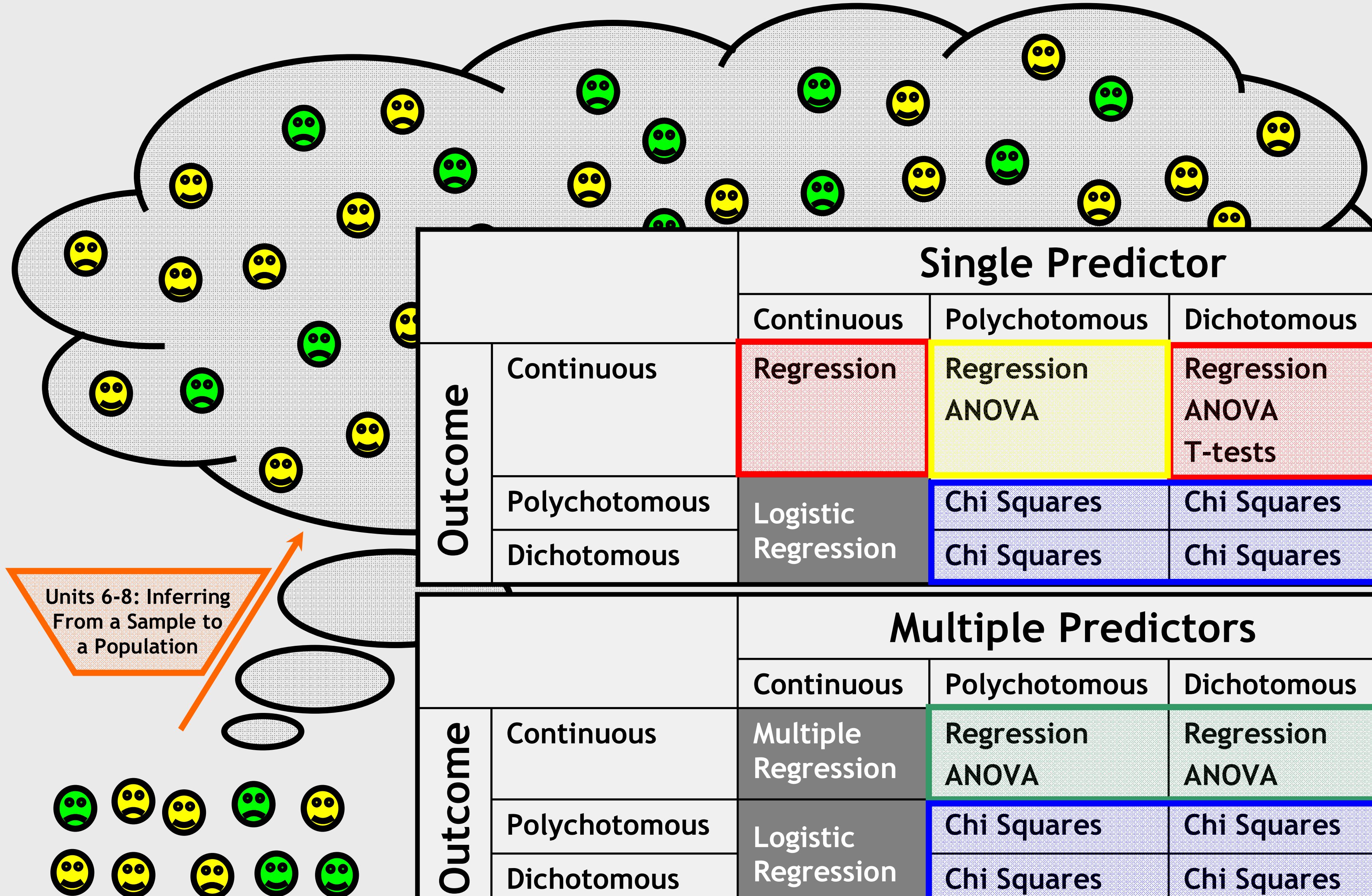
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	49.118	.115		428.169	.000	48.893	49.342
	FREELUNCH	-4.841	.198	-.267	-24.439	.000	-5.229	-4.453

a. Dependent Variable: READING

Unit 1 Unit 8 Unit 4 Unit 6 Unit 7

Unit 7: Road Map (Schematic)



Outcome	Single Predictor: Continuous	Single Predictor: Polychotomous	Single Predictor: Dichotomous
Continuous	Regression	Regression ANOVA	Regression ANOVA T-tests
Polychotomous	Logistic Regression	Chi Squares	Chi Squares
Dichotomous	Logistic Regression	Chi Squares	Chi Squares

Outcome	Multiple Predictors: Continuous	Multiple Predictors: Polychotomous	Multiple Predictors: Dichotomous
Continuous	Multiple Regression	Regression ANOVA	Regression ANOVA
Polychotomous	Logistic Regression	Chi Squares	Chi Squares
Dichotomous	Logistic Regression	Chi Squares	Chi Squares

Epistemological Minute

Epistemic Humility: Whenever we make a decision, we might be wrong. For the sake of simplicity, let's assume the decision is dichotomous: yes or no. If we decide "yes," and we are wrong, that is a false positive. If we decide "no," and we are wrong, that is a false negative. There is only one way to completely avoid false positives: ALWAYS decide "NO"! Likewise, there is only one way to avoid false negatives: ALWAYS decide "YES"! For any meaningful decision, neither strategy (always "no" nor always "yes") is reasonable. Instead, we must strike a careful balance between the risk of false positives and the risk of false negatives. In Unit 6, we learned Null-Hypothesis Significance Testing (NHST) which gives us some control over the risks of false positives (and by extension, false negatives) due to sampling error.

The Choices, by Orin Optiglot (When I get a chance, I will replace this with me **choosing** neckties. Oh, wait, I choose solid-color neckties at random!)

Type I Error is (falsely) rejecting the null *when* the null is true. Alpha level is the probability of Type I Error due to sampling error. By setting our alpha level, we control the probability of Type I Error.

		Our Conclusion	
		Positive	Negative
		We find a relationship in the population, after rejecting the null hypothesis.	Our findings are inconclusive, after failing to reject the null hypothesis.
Our World	There is a relationship in the population.	True Positive	False Negative "Type II Error"
	There is *no* relationship in the population	False Positive "Type I Error"	True Negative

Alpha level is the conditional probability that we will reject the null hypothesis *when* it is true. It is the risk of a false positive, *when* there is no relationship in the populations. Of course, we cannot be sure about the *when*.

Why bother? In the face of unavoidable uncertainty due to sampling error, at least we can quantify our risks and, consequently, control our risks.

IF 100%		
There is *no* relationship in the population	Alpha = .05 False Positive "Type I Error" 5% Chance	True Negative 95% Chance

Dialectic of Statistical Inference (Part I of II)

In my random sample, I found that the intervention group scored 5 points higher on average than the control group ($r = .17$, $p < .05$). Intervention predicted 3% of the score variation.

I suspect that there is no intervention effect, that the relationship you observe in your sample is merely an artifact of sampling error and not reflective of the population.

Well, the $p < .05$ tells us that you might be right (it's not $p = 0$), but if you were right, and thus there were no relationship in the population, we would only observe a relationship so strong (or stronger) less than 5% of the time. That is pretty unlikely.

I see. My suspicion of 0.00000 relationship is not very plausible.

Okay, so we do not want to conclude that the relationship is exactly zero in the population, but that does not mean the relationship is exactly .17 as you observe in your sample.

That's right. We don't want to conclude that the Pearson correlation in the population is exactly .17. Likewise, we do not want to conclude an intervention effect of exactly 5 points. However, they are unbiased estimates of the population values.

How precise are those estimates?

Using the same standard errors that we used to calculate $p < .05$ and reject the null hypothesis, we can construct 95% confidence intervals. Thus, our estimate for the population correlation is $.17 \pm .14$ and, for the intervention effect, 5 ± 4.

Dialectic of Statistical Inference (Part II of II)

Your intervention is boosting children's scores on average. A 1-point boost is your lower-bound estimate for that average, and a 9-point boost is your upper-bound estimate for that average. Should your intervention be an educational funding priority?

That is a difficult question. I can tell you all about ***statistical*** significance, but your question is about ***practical*** significance. To determine whether the intervention is worth implementing, we need to conduct a benefits-costs analysis: Do the benefits of the intervention outweigh its costs? Among the costs, we must consider opportunity costs: How does our intervention stack up against similarly targeted interventions?

My wife is an economist. I'll have her people call your people.

Back to statistical significance—The whole process of inference from a sample to a population is heavily laden with assumptions. If those assumptions do not hold, then it's all lies.

Yes. If my assumptions weren't tenable, my standard errors and, consequently, p-values and confidence intervals would be biased, and I would not be reporting them. Give me a break.

Hey, we're all friends. Were there *any* worrisome assumptions?

Independence, normality, linearity and outliers were okay, but there was a little heteroskedasticity; there was a little less variation in the intervention group than the control group. Heteroskedasticity won't bias our magnitude and strength estimates, but it will bias our precision estimate (i.e., standard error). Next semester, Sean will show me how to fix it.

Unit 7: Research Questions

Theory 1: Since depression leads to introversion, and reading is an introverted activity, depressed children will be stronger readers than non-depressed children.

Research Question 1: In children of immigrants, reading achievement is positively correlated with depression levels.

Theory 2: Since depression conflicts with cognitive functioning, and reading is a cognitively demanding activity, depressed children will be weaker readers than non-depressed children.

Research Question 2: In children of immigrants, reading achievement is negatively correlated with depression levels.

Data Set: ChildrenOfImmigrants.sav

Variables:
- Outcome—Reading Achievement Score (*READING*)
- Predictor—Depression Level (*DEPRESS*)

Model: $READING = \beta_0 + \beta_1 DEPRESS + \varepsilon$

SAT.sav Codebook

Portes, Alejandro and Ruben G. Rumbaut
Children of Immigrants Longitudinal Study
(1992, 1995)

"CILS is a longitudinal study designed to study the adaptation process of the immigrant second generation which is defined broadly as U.S.-born children with at least one foreign-born parent or children born abroad but brought at an early age to the United States. The original survey was conducted with large samples of second-generation children attending the 8th and 9th grades in public and private schools in the metropolitan areas of Miami/Ft. Lauderdale in Florida and San Diego, California" (from the website description of the data set).

Subset of data: Random sample of 880 participants obtained through the website.

Selected references:

Portes, Alejandro, & Ruben G. Rumbaut (2001). *Legacies: The Story of the Immigrant Second Generation*. Berkeley CA: University of California Press.

More information is available at: http://cmd.princeton.edu/ (Center for Migration and Development, Princeton University)

ChildrenOfImmigrants.sav Codebook

Variable Name	Variable Description	Characteristics
ID	Identification #	Integers
Reading	Stanford Reading Achievement Score	Range: 527-830 Mean: 669
FreeLunch	% students in school who are eligible for free lunch program	Range: 0-92.30 Mean: 45.27
Male	Sex dummy variable	1=Male 0=Female
Depress	Depression scale (Higher score means more depressed)	Range: -1.68 - 5.57 Mean: 0.00
SES	Composite family SES score	Range: -1.66 - 2.09 Mean: -0.04

The Children of Immigrants Data Set

ChildrenOfImmigrants.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : ID 4414

	ID	Reading	FreeLunch	Male	Depress	SES
1	4414	558	70.8	0	0.95820	-1.37
2	282	570	27.5	0	0.50933	-0.38
3	3848	580	92.3	1	0.94126	-1.02
4	342	581	27.5	1	-0.72716	0.25
5	3805	584	38.2	1	1.96360	0.07
6	4301	584	82.0	1	0.33707	-0.81
7	3548	586	82.0	0	-0.47796	-1.10
8	2593	589	38.2	0	2.45614	-1.65
9	3545	589	82.0	0	3.14966	-0.35

ChildrenOfImmigrants.sav [DataSet1] - SPSS Data Editor

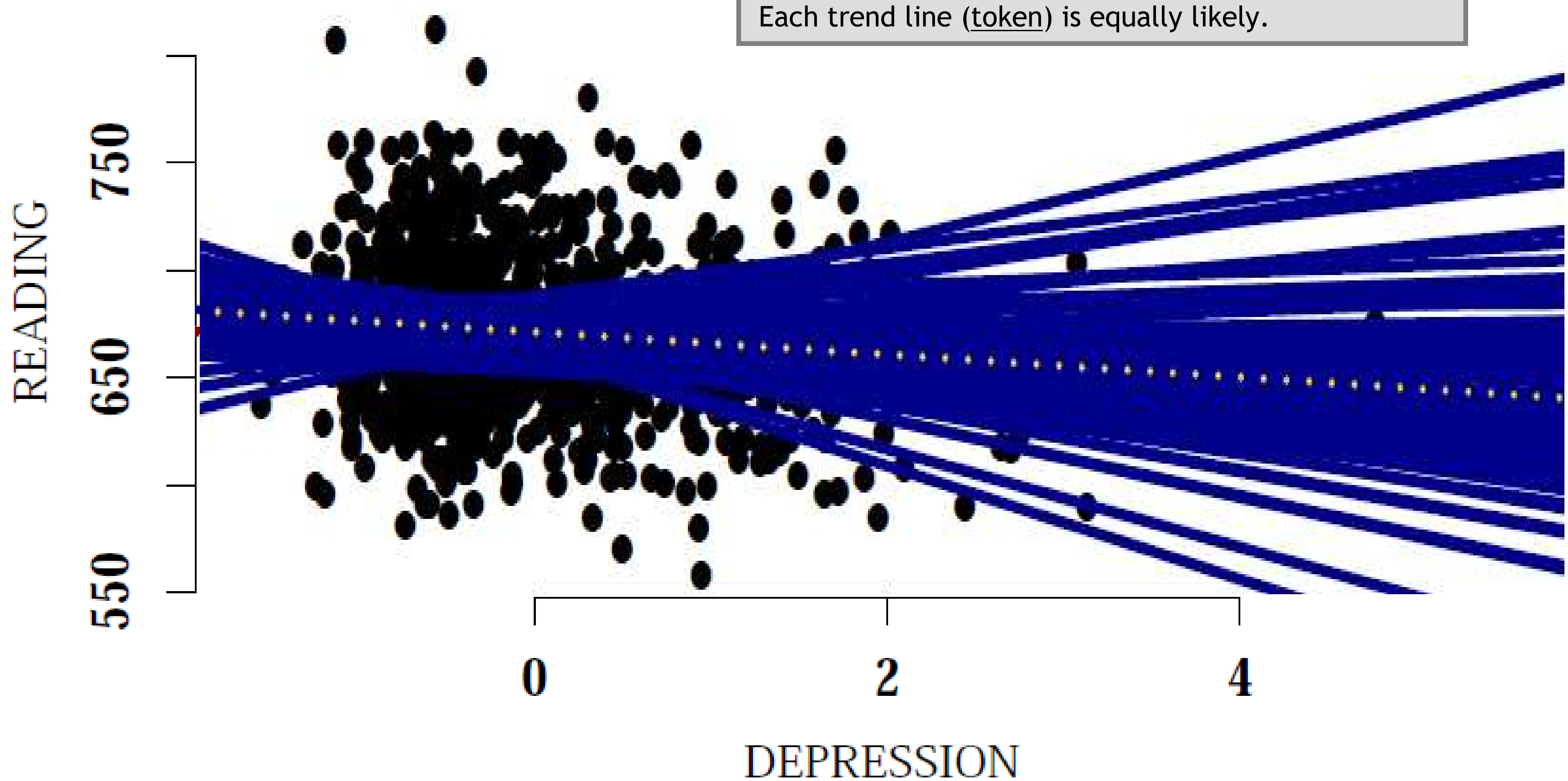
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	ID	Numeric	4	0		None	None	8	Right	Scale
2	Reading	Numeric	3	0	Stanford Readi...	None	None	8	Right	Scale
3	FreeLunch	Numeric	4	1	% of Students i...	None	None	8	Right	Scale
4	Male	Numeric	1	0	Male = 1, Fem...	{0, Female}...	None	8	Right	Nominal
5	Depress	Numeric	8	5	Depression Sc...	None	None	8	Right	Scale
6	SES	Numeric	5	2	Composite Fa...	None	None	8	Right	Scale
7										

The Problem of Sampling Error: Different Samples Yield Different Estimates

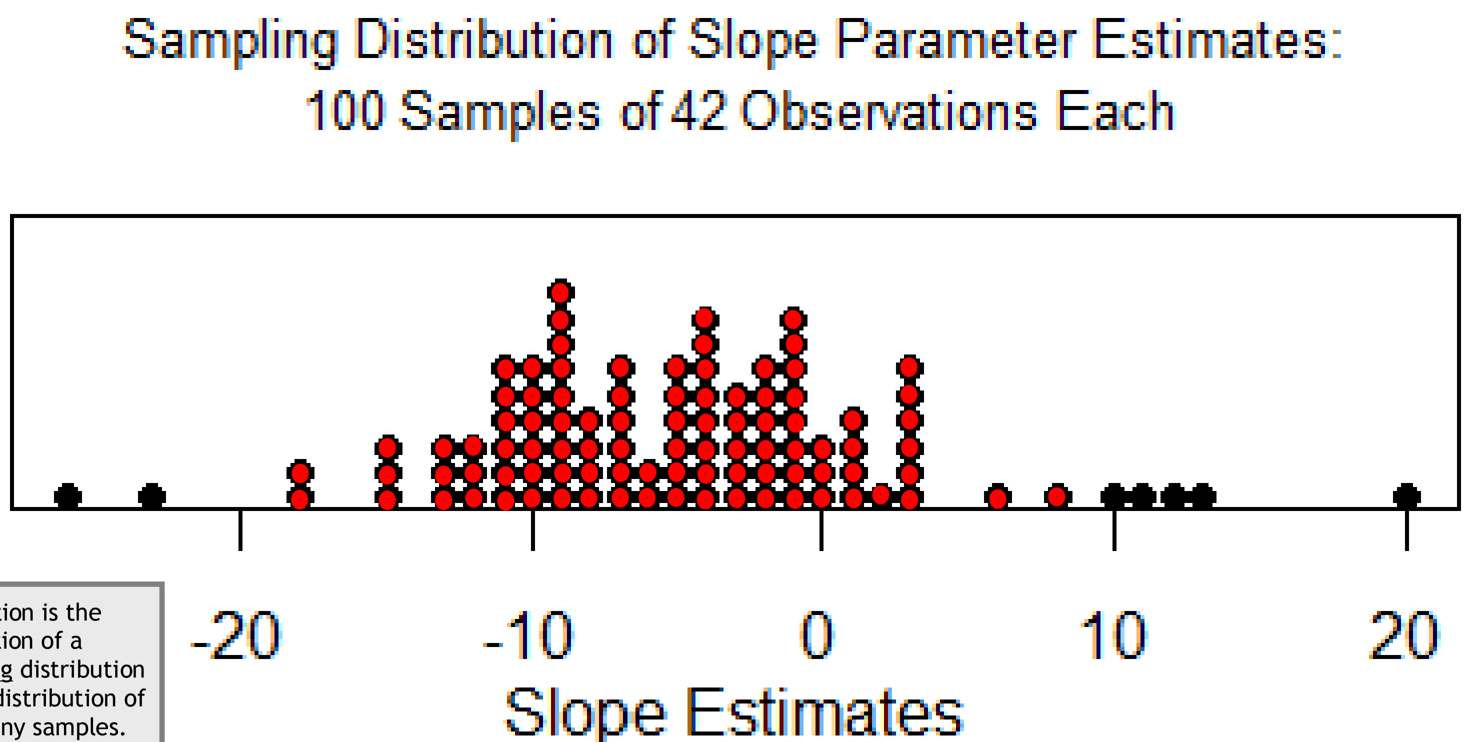
Each of the 100 trend lines comes from a different random sample of 42 children of immigrants.

Notice that every sample gives us a regression line with a different y-intercept and a different slope. Each trend line (<u>token</u>) is equally likely.



Notice the butterfly pattern. This is not an accident. Some trend lines (<u>types</u>) are more likely than others; notice how the butterfly thins out around the edges.

Review: Sampling Distribution In Dot Plot Form



A <u>sample</u> distribution is the observed distribution of a sample. A <u>sampling</u> distribution is a hypothetical distribution of statistics from many samples.

A sampling distribution is a distribution of statistics taken from many (equal sized, random) samples of the same population. Generally, in life, we get one sample. Nevertheless, it's a useful question to ask what would happen if we took many samples.

Thanks to the Central Limit Theorem, we know a lot about the sampling distribution.
SPLaSH: Spread, Location and Shape

The distribution is "fighting" to be normal.

The mean of the distribution is "fighting" to be the true population slope.

The standard deviation of the distribution is also predictably "fighting."

Review: Location Location Location (The Null Hypothesis Approach)

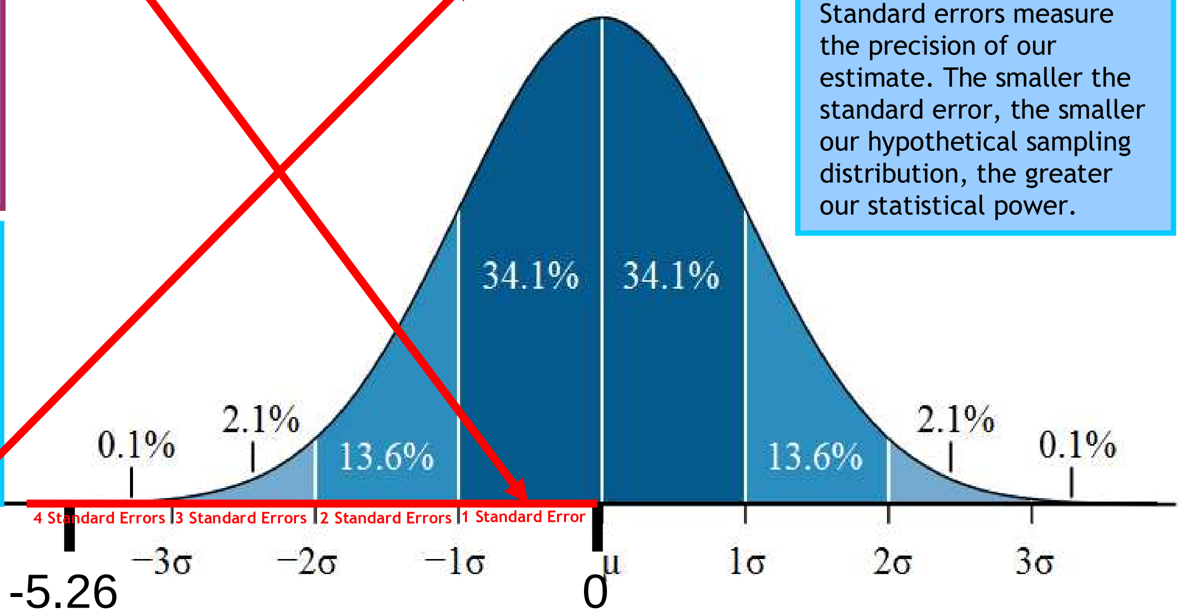
We know that our trend line belongs to a butterfly, and we can guesstimate the spread of the butterfly. We know that our slope belongs to a bell curve, and we can guesstimate the spread of the bell curve. We do not know the location of the butterfly or the bell curve, but we can use our knowledge of the shapes and spreads of the butterfly and bell curve.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	671.607	1.275		526.746	.000	669.105	674.110
	Depress	-5.260	1.429	-.123	-3.680	.000	-8.066	-2.455

There is a statistically significant relationship ($p < 0.05$) between depression levels and reading scores in our sample ($n = 880$) of children of immigrants.

Standard errors measure the precision of our estimate. The smaller the standard error, the smaller our hypothetical sampling distribution, the greater our statistical power.

A t-test is a test to see if our observation is a sufficient number of standard errors away from zero to scare us into rejecting the null hypothesis. A t statistic of $\pm$ 2, indicating that our observation is $\pm$ 2 standard errors from zero, will have a two-tailed significance level (or p value) of about 0.05.



Our observed slope is -3.68 standard errors from zero.

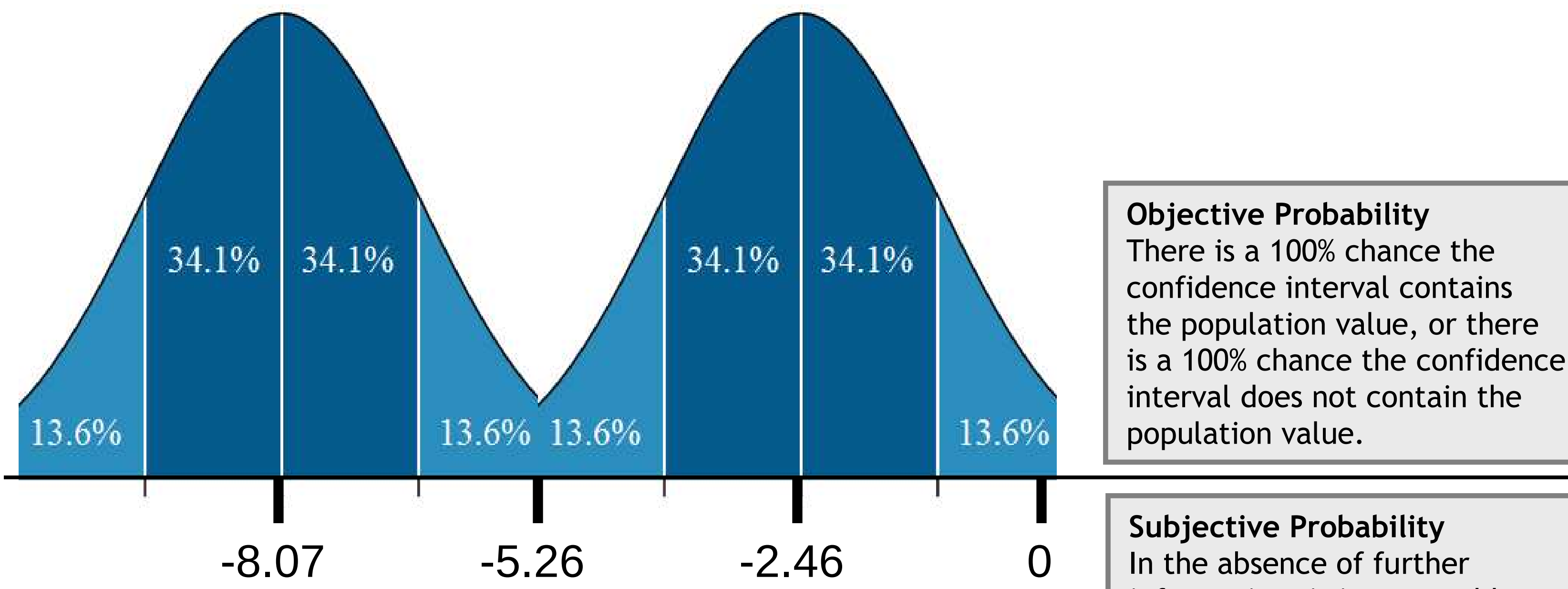
Quiz: What is -5.26 divided by 1.429?

We observe a regression coefficient of -5.26 in our sample. If there were no relationship in the population, we would observe a coefficient this large of larger in less than 0.01% of our samples.

Another (Better) Way to Use Standard Errors!

The purpose of confidence intervals is to contain the true population parameter. Over your lifetime, 95% of (95%) confidence intervals will succeed and 5% will fail. You will not know which are the unlucky 5%! Analogous reasoning holds for alpha level.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	671.607	1.275		526.746	.000	669.105	674.110
	Depress	-5.260	1.429	-.123	-3.680	.000	-8.066	-2.455



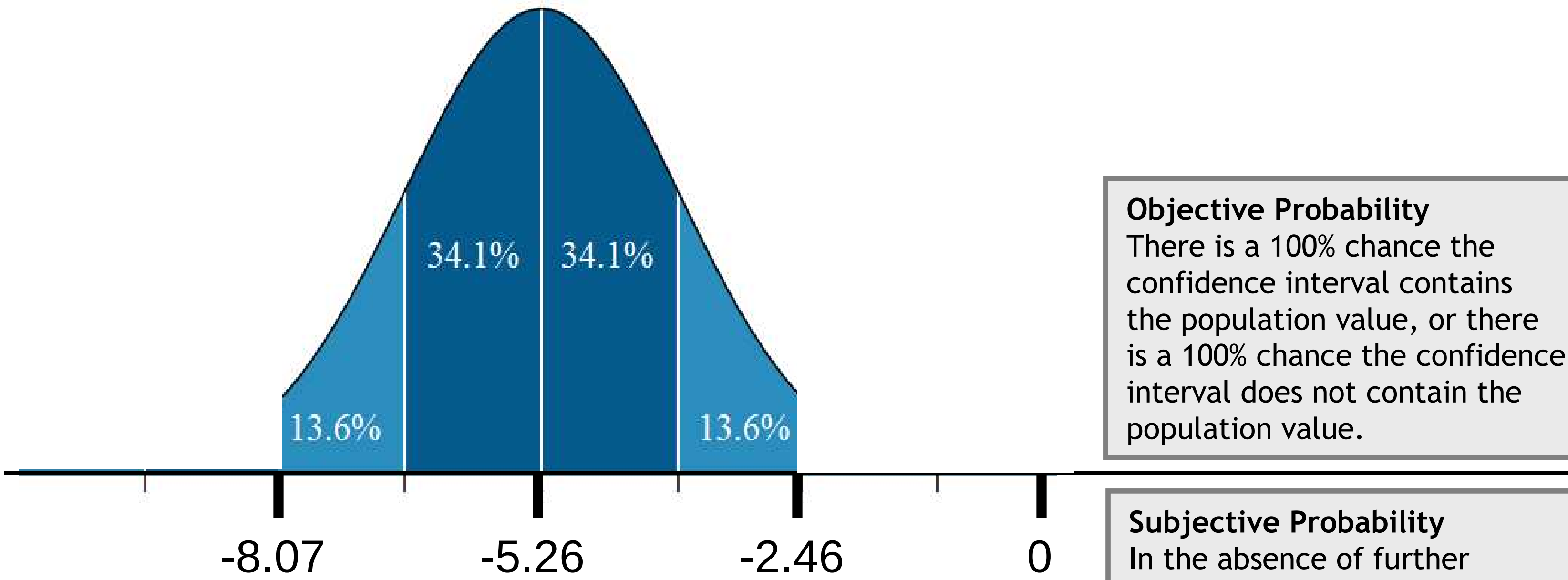
Objective Probability
There is a 100% chance the confidence interval contains the population value, or there is a 100% chance the confidence interval does not contain the population value.

Subjective Probability
In the absence of further information, it is reasonable to conclude that there is a 95% chance the confidence interval contains the population value.

Location Location Location

The purpose of confidence intervals is to contain the true population parameter. Over your lifetime, 95% of (95%) confidence intervals will succeed and 5% will fail. You will not know which are the unlucky 5%! Analogous reasoning holds for alpha level.

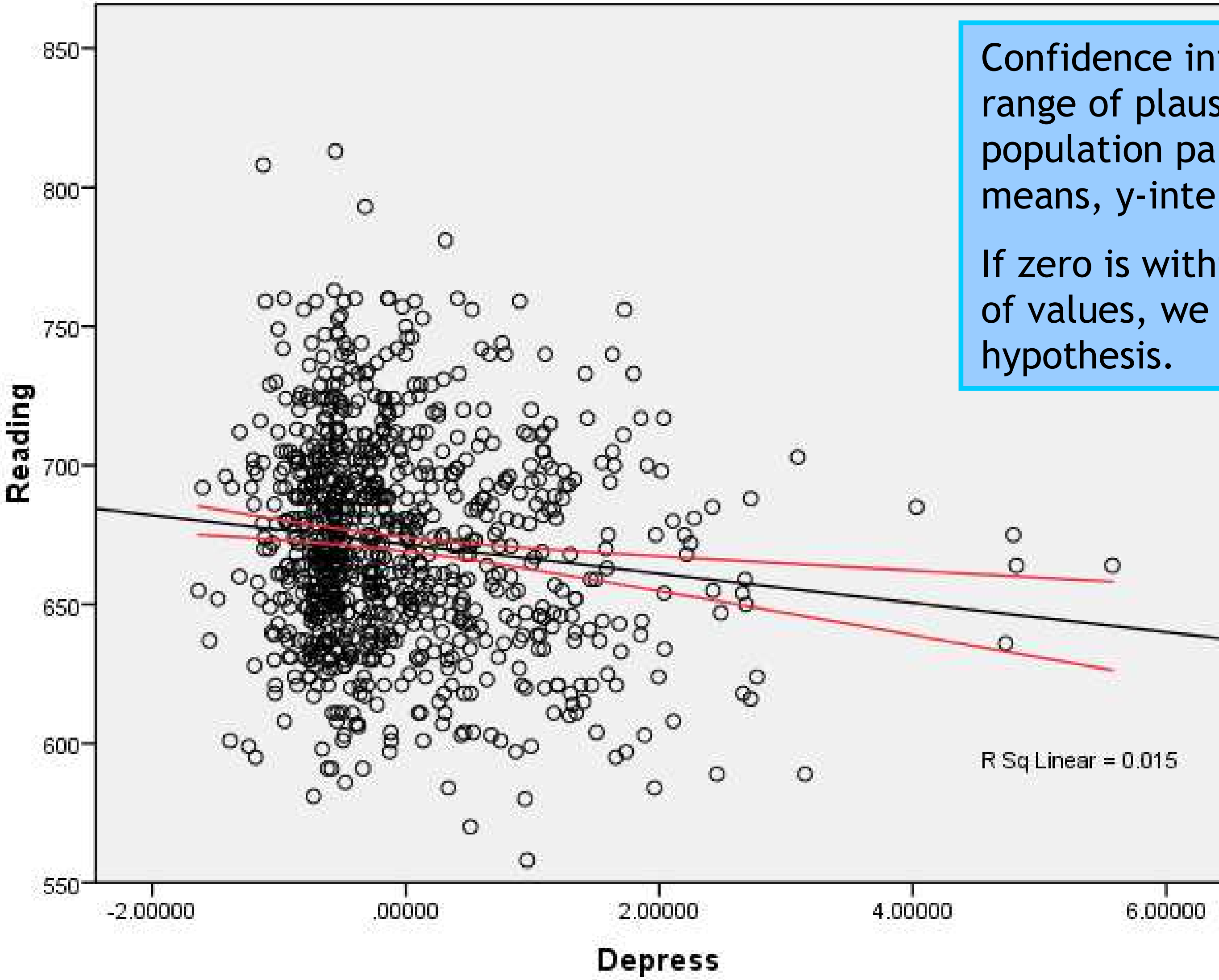
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	671.607	1.275		526.746	.000	669.105	674.110
	Depress	-5.260	1.429	-.123	-3.680	.000	-8.066	-2.455



Objective Probability
There is a 100% chance the confidence interval contains the population value, or there is a 100% chance the confidence interval does not contain the population value.

Subjective Probability
In the absence of further information, it is reasonable to conclude that there is a 95% chance the confidence interval contains the population value.

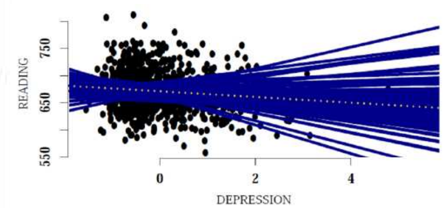
95% Confidence Intervals



Confidence intervals provide a range of plausible values for population parameters (such as means, y-intercepts, and slopes).

If zero is within the plausible range of values, we do not reject the null hypothesis.

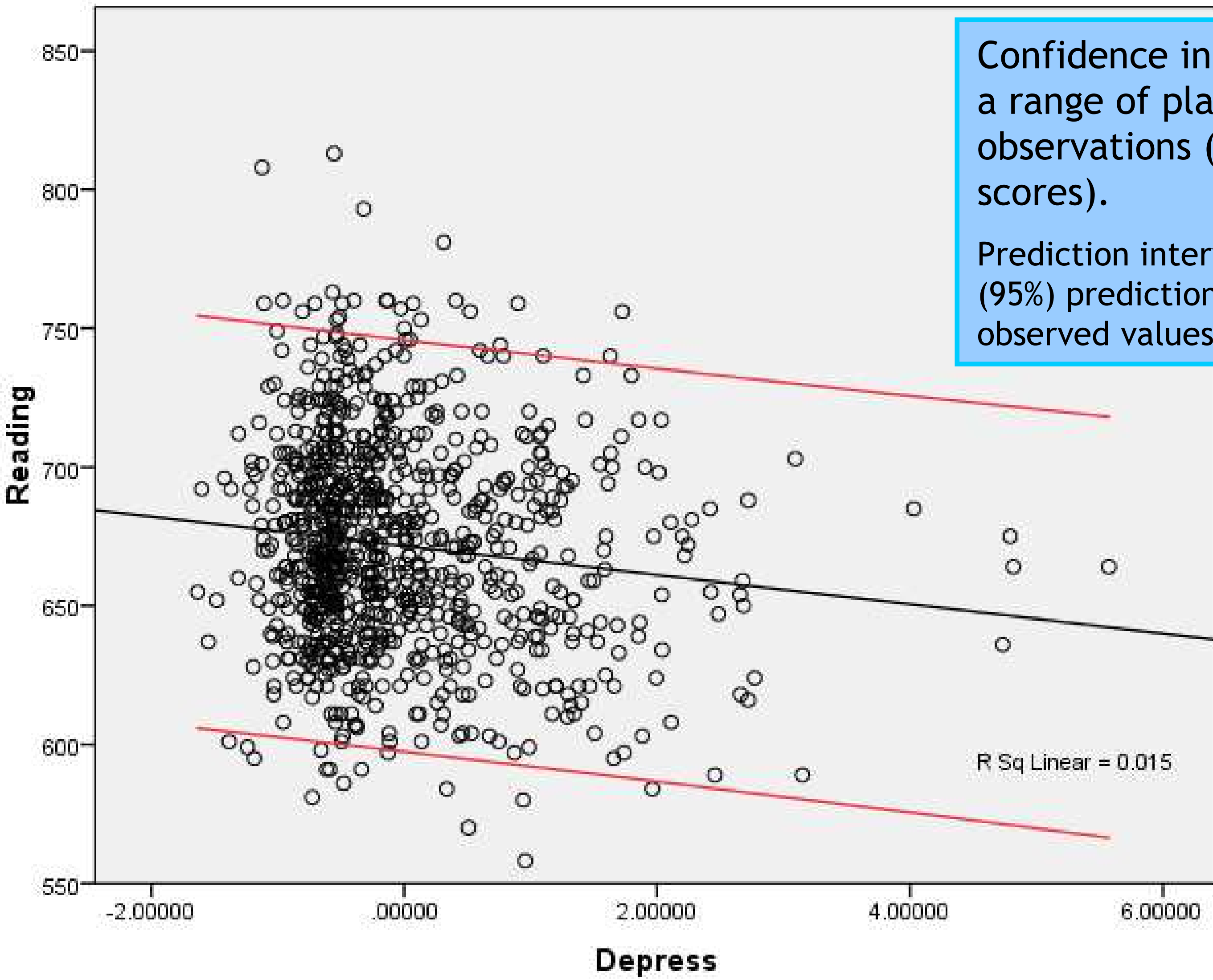
Notice that the confidence interval looks like a butterfly. That is not an accident!



The sampling distribution looks like a butterfly, and the purpose of the confidence interval is to mimic the spread (the middle 95%) of the sampling distribution.

Notice that the red butterfly is skinnier than the blue butterfly. The red butterfly is based on the full sample ($N = 880$). The blue butterfly is based on random subsamples ($n = 42$). The bigger the sample, the skinnier the butterfly.

95% Prediction Intervals



Confidence intervals **do not** provide a range of plausible values for observations (such as individual scores).

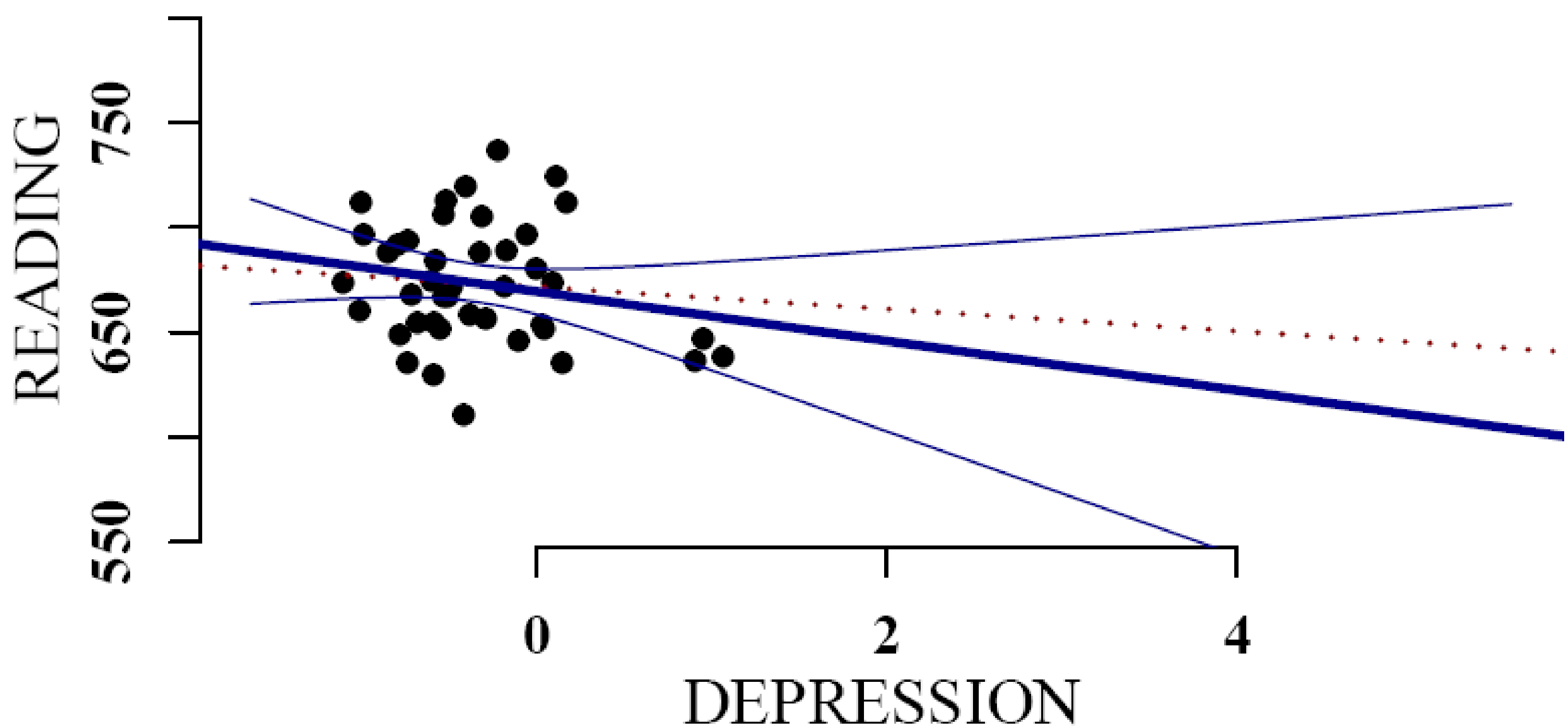
Prediction intervals have that job. 95% of (95%) prediction intervals contain 95% of observed values 95% of the time.

The purpose of regression (and ANOVA) is to predict on average. We almost always do a crummy job of predicting individuals, and prediction intervals highlight that point.

Don't spend any time worrying about prediction intervals. They are virtually useless, except as a good contrast for confidence intervals.

95% Confidence Intervals Work 95% of the Time

Random Sample # 100 : READING vs. DEPRESSION (n = 42)



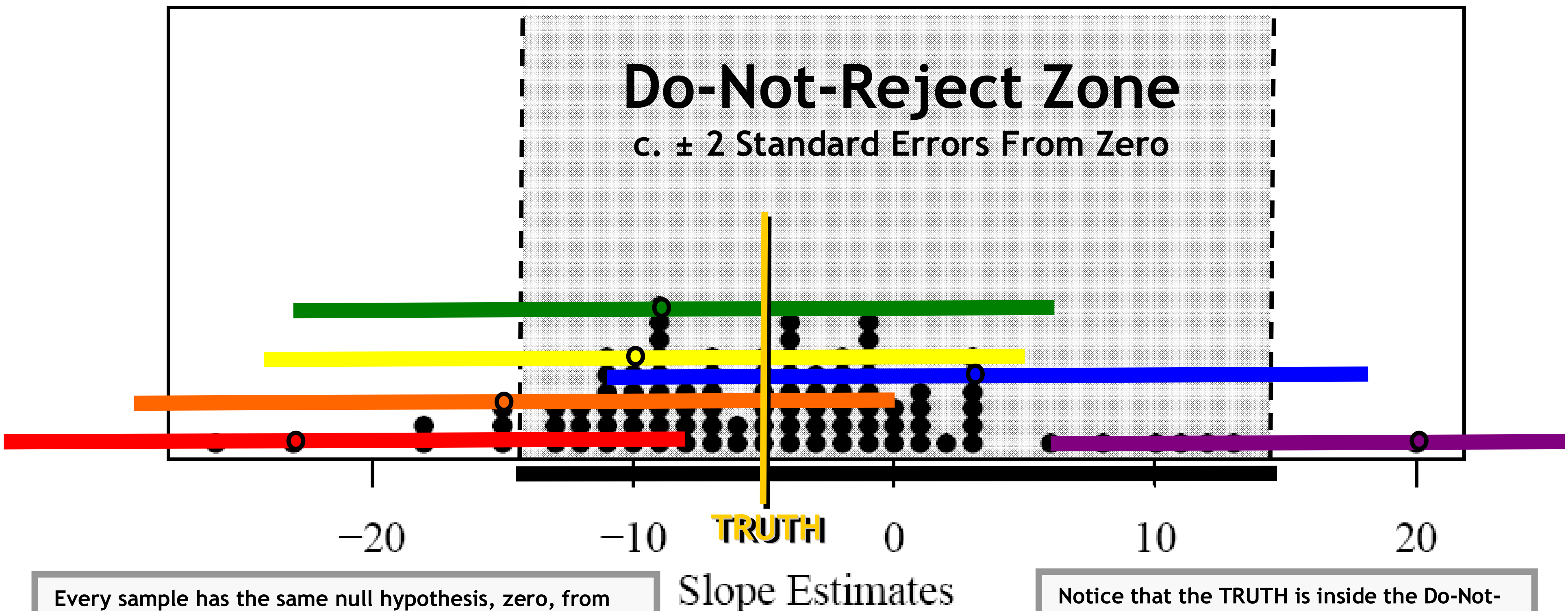
Note that the confidence interval for the slope and the confidence interval for the line may give slightly different results. Different jobs; different results.

Fitted Model: Predicted READING = 669.43 + −11.75 * DEPRESSION   SE For Slope = 8.93   CI For Slope = −29.8 to 6.29   Success!

Two Distinct Approaches to Using Standard Errors: Null and CI

Sampling Distribution of Slope Parameter Estimates:

100 Samples of 42 Observations Each



Every sample has the same null hypothesis, zero, from which we build out a Do-Not-Reject Zone with standard errors. Every sample gets its own confidence interval.

Mean = −4.9 , Compare to the Population Slope of −5.3

Standard Deviation = 7.3 , Compare to the Average Standard Error of 6.9

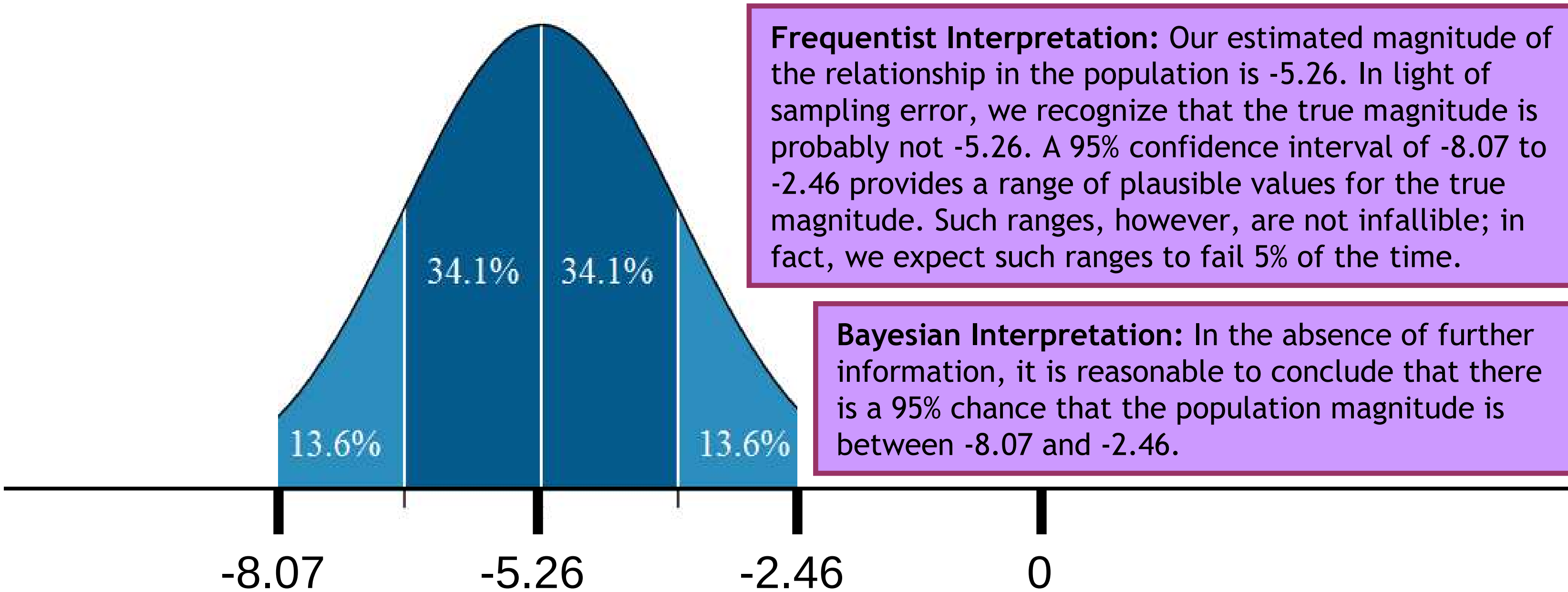
On average, 95% of our confidence intervals will cover the TRUTH. Of our 100 samples, 93% of our CIs cover the TRUTH, but if we had infinite samples, it would be 95%.

Notice that the TRUTH is inside the Do-Not-Reject-Zone. This is bad. It means that if our estimate happens to be the true population value, we still cannot reject the null and , consequently, cannot draw a conclusion to the population. This study is underpowered! The sample size (n = 42) is not large enough to cut down the standard error to a useful length. We need a bigger sample size for the sake of smaller standard errors.

Interpreting Confidence Intervals

The purpose of confidence intervals is to contain the true population parameter. Over your lifetime, 95% of (95%) confidence intervals will succeed and 5% will fail. You will not know which are the unlucky 5%! Analogous reasoning holds for alpha level.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	671.607	1.275		526.746	.000	669.105	674.110
	Depress	-5.260	1.429	-.123	-3.680	.000	-8.066	-2.455



Frequentist Interpretation: Our estimated magnitude of the relationship in the population is -5.26. In light of sampling error, we recognize that the true magnitude is probably not -5.26. A 95% confidence interval of -8.07 to -2.46 provides a range of plausible values for the true magnitude. Such ranges, however, are not infallible; in fact, we expect such ranges to fail 5% of the time.

Bayesian Interpretation: In the absence of further information, it is reasonable to conclude that there is a 95% chance that the population magnitude is between -8.07 and -2.46.

Anatomy of Two Interpretations

How many frequentists does it take to change a light bulb? For 95% of light bulbs, it only takes one.

How many Bayesians does it take to change a light bulb? I'm 95% confident that it only takes one.

I just flipped a (fair) coin. What are the chances that it is tales?

The frequentist/objectivist says it is 100% tales or 100% heads; over repeated flips, however, 50% of the time it will be tales.

The Bayesian/subjectivist says that there is a 50% chance it is tales.

Frequentists take into consideration hypothetical, infinite repeated replications. They reason about the different possible outcomes over the repeated replications, and they draw conclusions about the differing likelihoods of the outcomes based on the differing frequencies.

Bayesians build from the frequentist base. Bayesians take the frequentists' mathematical/objective reasoning into the realm of the psychological/subjective. Bayesians draw conclusions about rational belief and degrees of confidence. They use arguments about rational betting behavior to bridge the gap between the mathematical and the psychological, between the objective and the subjective.*

Notice how the frequentist interpretation focuses on the confidence interval and what happens when we repeatedly use confidence intervals. For the frequentist, the "95%" is about the expected success rate of our confidence intervals over our lifetimes.

Frequentist Interpretation: Our estimated magnitude of the relationship in the population is -5.26. In light of sampling error, we recognize that the true magnitude is probably not -5.26. A 95% confidence interval of -8.07 to -2.46 provides a range of plausible values for the true magnitude. Such ranges, however, are not infallible; in fact, we expect such ranges to fail 5% of the time.

Notice how the Bayesian interpretation focuses on the population magnitude. For the Bayesian, the "95%" is about our degree of confidence that the population magnitude is within our confidence interval.

Bayesian Interpretation: In the absence of further information, it is reasonable to conclude that there is a 95% chance that the population magnitude is between -8.07 and -2.46.

The Bayesian might talk about the probability of the population magnitude falling within the confidence interval. This talk is very subjective, and it raises the hackles of the frequentist/objectivist. Objectively, the population magnitude is what it is. Objectively, the population magnitude does not "fall" inside (or outside) of our confidence interval; rather, our confidence interval succeeds or fails at containing the population magnitude. The population magnitude is NOT a moving target; we're just less-than-perfect at casting our net.

Bayesians allow us to talk about being 95% confident that the true magnitude is between our lower bound and upper bound. This seems natural, but it is deceptively simple. One necessary further consideration, in addition to frequentist/objectivist probability and rational betting behavior, is prior probability. We should only be 95% confident based on a 95% confidence interval if we are analyzing the data in a vacuum. If there is any additional evidence to sway our confidence one way or the other, it is not reasonable to neglect it.

* The game: you win $100 dollars if your 95% confidence interval contains the population magnitude. What is the most you would pay to play the game? $105? $100? $95? $90?

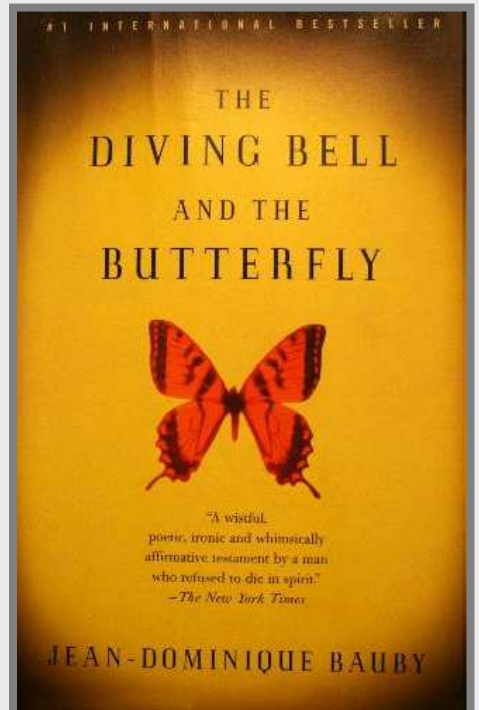
Reporting the Results of Simple Linear Regression

When reporting the results of simple linear regression:

1. Include your fitted model.
2. Use the R^2 statistic to convey the goodness of fit for the model (i.e., strength).
3. To determine statistical significance, test the null hypothesis that the magnitude in the population is zero, reject (or not) the null hypothesis, and draw a conclusion (or not) from the sample to the population.
4. Describe the direction and magnitude of the relationship in your sample, preferably with illustrative examples. Make clear the substance of your findings.
5. Use confidence intervals to describe the magnitude of the relationship in the population
6. If simple linear regression is inappropriate for the data, then say so, briefly explain why, and forego any misleading analysis.

The frequentist interpretation of confidence intervals is about expectations over repeated samples (even though we actually have only one sample). The frequentist interpretation is purely mathematical/statistical. "Over repeated coin flips, we expect 50% heads. Any one coin flip is either 100% heads or 0% heads."

The Bayesian interpretation starts with the frequentist interpretation but then draws conclusions about rational belief. The Bayesian argues that, if confidence intervals work 95% of the time over the course of our lives, then we should be 95% confident in each confidence interval. The bayesian interpretation is a mixture of not only mathematics and statistics but also normative psychology and the logic of belief. "Because, over repeated coin flips, we expect 50% heads, we are 50% confident that any one coin flip is heads."



Never lose sight of the substantive meaning of the numbers.

You have what you need for the Unit 7 post hole. There is practice in back.

Dig the Post Hole

Unit 7 Post Hole:

Interpret a confidence interval from a frequentist perspective and from a Bayesian perspective.

Here is my answer:

FREQ: -2.5 to -8.1 provides a range of plausible values for the true population slope. In my life, 5% of such ranges will fail.

BAYES: In the absence of further information, I am 95% confident that the true population slope is between -2.5 and -8.1.

Evidentiary materials: regression output with CIs (SPSS).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.267[a]	.071	.071	8.25952

a. Predictors: (Constant), FREELUNCH

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	40744.322	1	40744.322	597.251	.000[a]
	Residual	531977.541	7798	68.220		
	Total	572721.864	7799			

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	671.607	1.275		526.746	.000	669.105	674.110
	Depress	-5.260	1.429	-.123	-3.680	.000	-8.066	-2.455

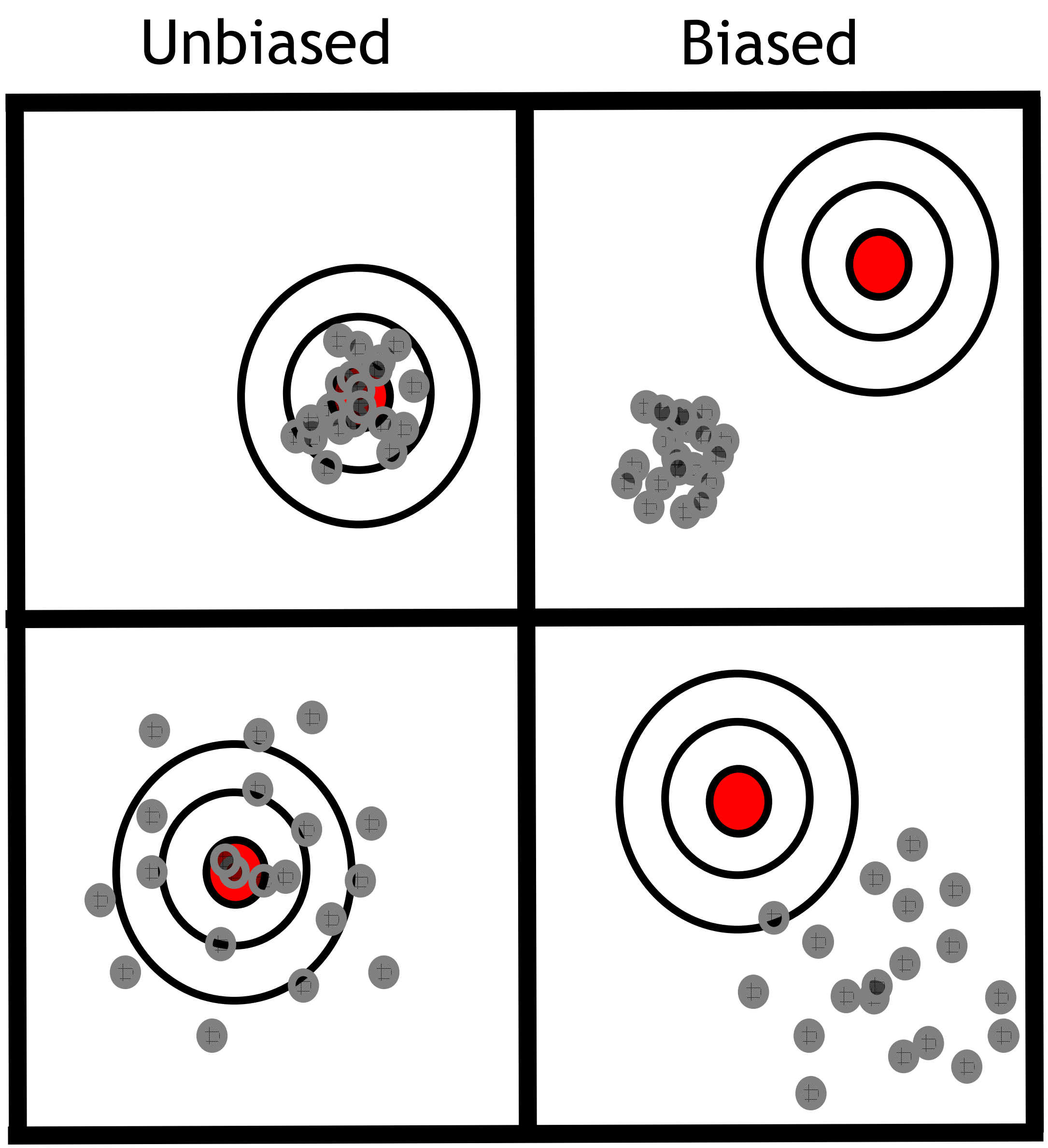
Evidentiary materials: regression output with CIs(R).

```
> model.1 <- lm(Reading~Depress)
> summary(model.1)
Call:
lm(formula = Reading ~ Depress)
Residuals:
    Min      1Q  Median      3Q     Max
-108.57  -26.13   -2.97   25.40  138.48
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  671.607      1.275  526.75  < 2e-16 ***
Depress       -5.260      1.429   -3.68 0.000248 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
Residual standard error: 37.79 on 878 degrees of freedom
Multiple R-squared: 0.01519,    Adjusted R-squared:
0.01407
F-statistic: 13.54 on 1 and 878 DF,  p-value: 0.0002475
> confint(model.1)
                  2.5 %     97.5 %
(Intercept) 669.104755 674.109608
Depress      -8.065788  -2.454707
```

Precision vs. Bias: The Classic Dart Metaphor

Bias is caused by biased (nonrandom) sampling, biased variables that do not measure what they purport to measure, and biased estimation techniques.

Outliers and (unmodeled) non-linearity also cause bias. That's why we look!



Precision and bias apply to almost all statistics, including measurements. Measurements can be imprecise and/or unbiased. Regression is designed to handle imprecise outcomes, but not imprecise predictors (which bias regression estimates).

Standard errors are a measure of imprecision. The best way to decrease standard errors and increase precision is to increase sample size!

In the next room, on the wall, there are painted four bull's eyes, the sort of things at which one throws darts. Imagine that you are tasked with, one at a time, putting your finger on each bull's eye. The trick is that you are blindfolded. You can feel the wall with your hands, but you can't sense the differences in paint. From your perspective, the wall is perfectly smooth other than a few dart holes. Your only guide is those few dart holes, created by four different throwers, each aiming for a different bull's eye.

Anatomy of a Standard Error for the Slope

$$\text{StandardError}_{\beta_{YX}} = \frac{\frac{sd_Y}{sd_X}\sqrt{1-R^2}}{\sqrt{n-2}}$$

Where,

β_{YX} is the slope coefficient from a regression of Y on X

sd_Y and sd_X are standard deviations of Y and X, respectively

R^2 is the proportion of variation in Y predicted by X

n is the sample size

All standard errors have roughly the same form:

$$\text{StandardError} = \frac{\text{Guesstimation of the Population Variation}}{\text{SampleSize}}$$

The standard error for the slope is no exception. In the numerator are the standard deviations (measures of variation) of *Y* and *X* and the R^2 statistic, which is a measure of the variation in Y predicted by X. In the denominator is sample size.

Standard errors measure the <u>imprecision</u> of our estimates, so we want our standard errors to be small. Let's understand the standard error for the slope in terms of our desire for small standard errors.

$$\text{StandardError}_{B_{YX}} = WantSmall = \frac{WantSmall}{WantBig} = \frac{\frac{sd_Y}{sd_X}\sqrt{1-R^2}}{\sqrt{n-2}} = \frac{(IsWhatItIs)\sqrt{(1-WantBig)}}{\sqrt{WantBig-2}}$$

Take-Home Messages:

Bigger R^2 statistics lead to more precise estimates.

Bigger sample sizes lead to more precise estimates.

Because the standard deviations of *Y* and *X* are both 1 for a Pearson product-moment correlation (*r*), we only need the sample size to calculate its standard error. Recall that $R^2 = r*r$ and that the r statistic is just the slope from a regression of standardized Y on standardized X, and, when we standardize our variables, we make their standard deviations equal to 1.

How Big Should I Make My Sample?

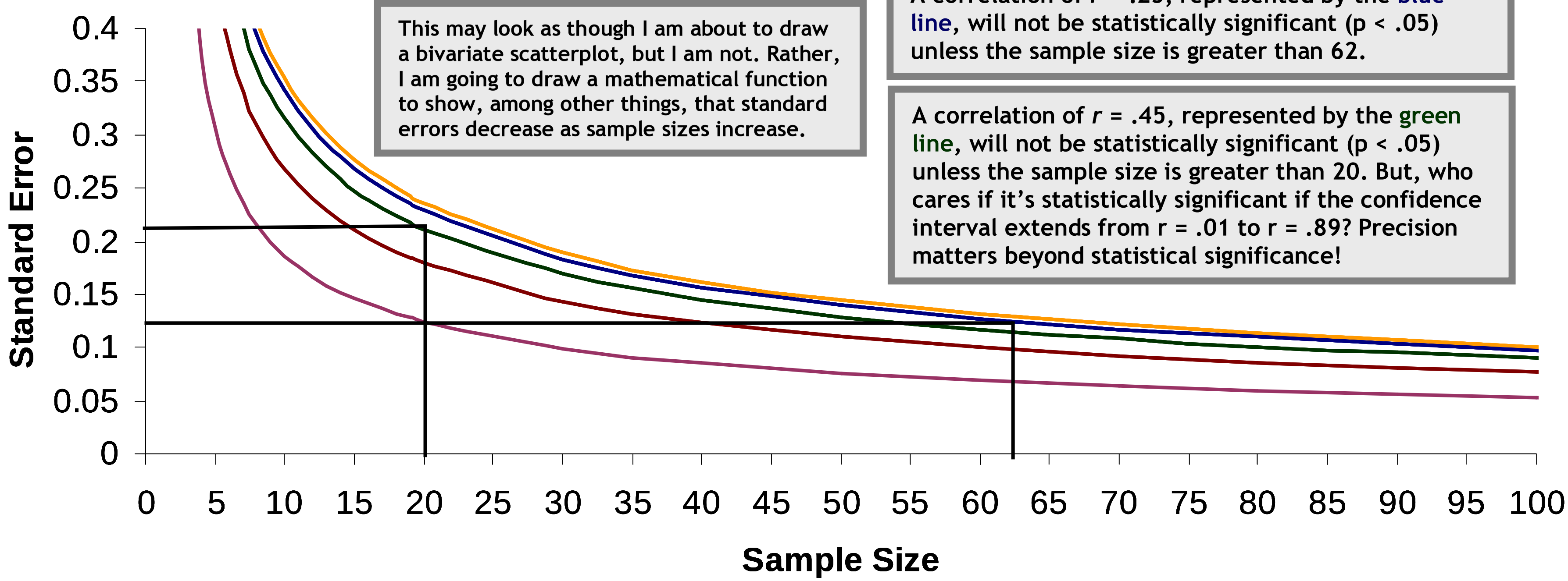
The force at work here is the Central Limit Theorem. Our predictions get more precise with larger sample sizes not because we are sampling a greater proportion of the population. (Remember, the population is infinite!) Rather, our predictions are getting more precise because, the larger our samples, the more opportunity for positive and negative outliers to balance themselves. This math phenomenon is sometimes called "The Law of Large Numbers."

A correlation of r = .05, represented by the orange line, will not be statistically significant ($p < .05$) unless the sample size is greater than 1537. The standard errors have to be smaller than about .025 or the confidence intervals will contain zero!

A correlation of r = .25, represented by the blue line, will not be statistically significant ($p < .05$) unless the sample size is greater than 62.

This may look as though I am about to draw a bivariate scatterplot, but I am not. Rather, I am going to draw a mathematical function to show, among other things, that standard errors decrease as sample sizes increase.

A correlation of r = .45, represented by the green line, will not be statistically significant ($p < .05$) unless the sample size is greater than 20. But, who cares if it's statistically significant if the confidence interval extends from r = .01 to r = .89? Precision matters beyond statistical significance!



Adding a few more subjects buys you a lot when you have a relatively small sample size to begin with, but more subjects buys you much less when you already have many subjects. This is because the *square root* of the sample size is in the denominator:

$$\text{StandardError}_{B_{YX}} = \frac{\frac{sd_Y}{sd_X}\sqrt{1-R^2}}{\sqrt{n-2}}$$

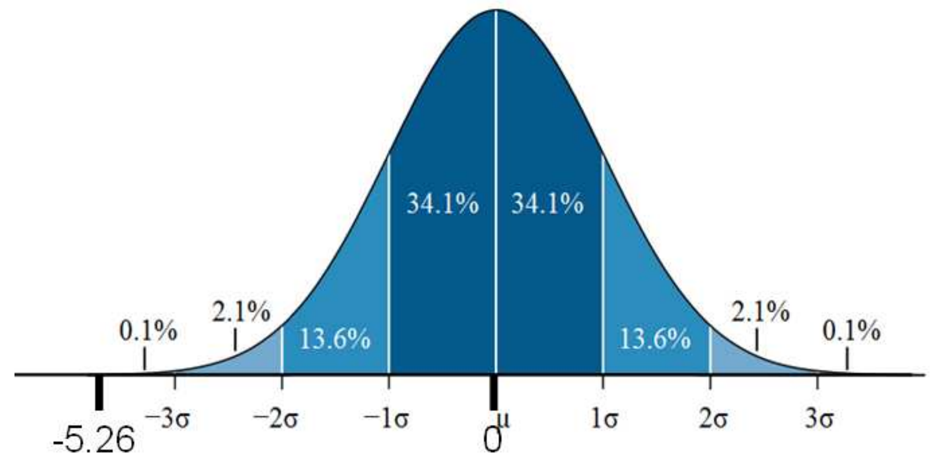
A correlation of r = .65, represented by the red line, will not be statistically significant ($p < .05$) unless the sample size is greater than 8.

A correlation of r = .85, represented by the purple line, will not be statistically significant ($p < .05$) unless the sample size is greater than 6.

Statistical Power (1- ß)

Choose Three!

And the Fourth is Fixed. (In other words, there are only three degrees of freedom! See the Math Appendix for an explanation of degrees of freedom.)



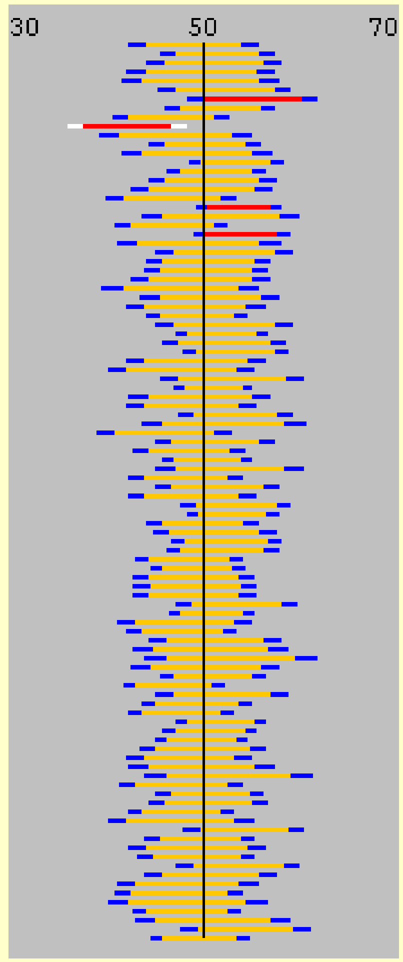
- **Sample Size**
  - ***n* = 880**
- **Effect Size**
  - ***r* = -0.123**
  - A standardized measure of the magnitude of the relationship.
  - Nothing to do with "cause and effect."
  - There are many methods for measuring effect sizes, and correlation (*r*) is one of them.
  - Effect size is largely out of our control, but we can boost it sometimes by creating more focused interventions and more precise outcome measures.
- **Alpha Level**
  - **α** = **0.05** (or, 95% confidence intervals) industry standard (why?)
  - Alpha level is a measure of YOUR tolerance for false positives due to sampling error.
- **Beta Level (not to be confused with any of the other Betas that pop up in stats)**
  - **ß = 0.04** (from http://www.danielsoper.com/statcalc/calc03.aspx)
  - Beta level is a measure of YOUR tolerance for false negatives due to sampling error.
  - We usually worry about **ß** when we are just eking out a sufficient sample size. We guesstimate a reasonable effect size, and we lock in **α** = **0.05** and **ß = 0.20.** Once we've chosen the three, the fourth, sample size, is fixed!
  - ß ≤ 0.20, industry standard (why?)

If we can't be certain, why bother using confidence intervals?

If a 95% confidence ***does not*** fail in 5% of samples, then it is broken! It is designed to fail in exactly 5% of samples, not 4% and not 6%. What would prompt anyone to design such a tool, never mind actually ***use*** the tool?

We use the tool because we realize that there is no way to 100% avoid false positives (Type I Error) without 100% embracing false negatives (Type II Error). It is very easy to produce 100% confidence intervals: all 100% confidence intervals contain all possible values! Now, that's useless. In statistics and in life, we are constantly trading off the risks of false positives and false negatives.

"In the absence of further information…" It is crucial to add this qualifier to our Bayesian interpretation instead of simply saying, "We are 95% confident that…" Based on other samples, we can have other confidence intervals for the population parameter. Those other confidence intervals supply information, and that information would be foolish to ignore. In fact, it would be foolish to ignore any source of information.



Sample size: 20 ▾ Sample Clear

When you click the sample button, 100 samples of the specified sample size (10, 15, or 20) will be taken from a population with a mean of 50 and a standard deviation of 10. The confidence interval on the mean will be computed for each. If the 95% confidence interval contains the population mean of 50 then a line will show the 95% confidence interval in orange and the 99% confidence interval in blue. If the 95% confidence interval does not contain the population mean then it will be shown in red. If the 99% interval does not contain the population mean it will be shown in white.

Cumulative Results:

	99% Conf. Int	95% Conf. Int
Contained 50	9911	9536
Did Not Contain 50	89	464
Proportion Contained	0.991	0.954

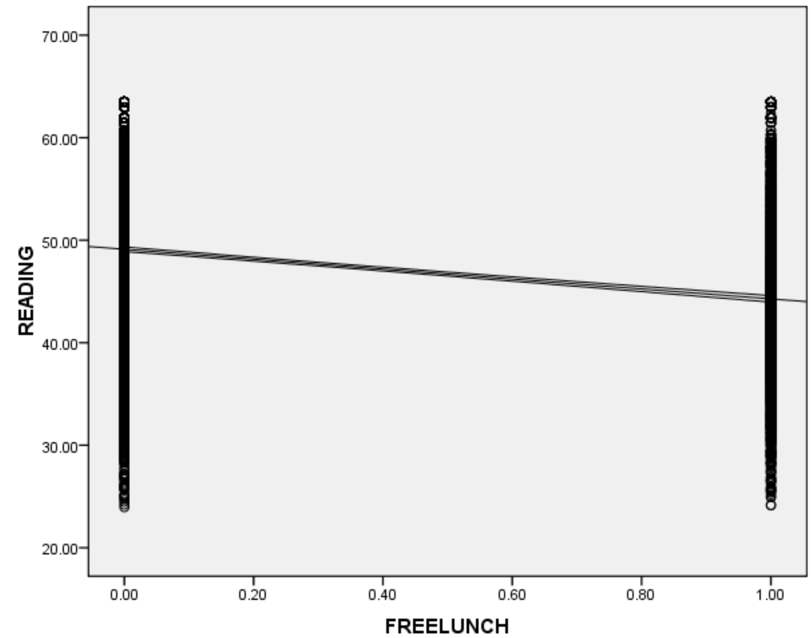
http://onlinestatbook.com/stat_sim/conf_interval/index.html

Answering our Roadmap Question

Unit 7: In the population, what is the magnitude of the relationship between reading and free lunch?

$$\hat{Reading} = 49 - 5FreeLunch$$

There is a statistically significant difference in the reading performance of students who are eligible for free lunch and those who are not ($p < 0.001$). Based on 95% confidence intervals, we estimate that, in the population of U.S. 8^{th} graders (in 1988), the average difference in readings scores is between 5.2 and 4.5 points.



Note that, in our sample of 7,800 students, the difference in averages is 4.841, but what are the chances that the average difference in the population is 4.841? Slim to none. Furthermore, if we take our estimate out to enough decimal places, the chances are simply none.

Be careful! Your audience will be inclined to treat your confidence interval as a prediction interval.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	49.118	.115		428.169	.000	48.893	49.342
	FREELUNCH	-4.841	.198	-.267	-24.439	.000	-5.229	-4.453

a. Dependent Variable: READING

Unit 7 Appendix: Key Concepts

Standard errors are a measure of imprecision. The best way to decrease standard errors and increase precision is to increase sample size!

Our predictions get more precise with larger sample sizes not because we are sampling a greater proportion of the population. (Remember, the population is infinite!) Rather, our predictions are getting more precise because, the larger our samples, the more opportunity for positive and negative outliers to balance themselves. This math phenomenon is sometimes called “The Law of Large Numbers.”

“In the absence of further information…” It is crucial to add this qualifier to our Bayesian interpretation instead of simply saying, “We are 95% confident that…” Based on other samples, we can have other confidence intervals for the population parameter. Those other confidence intervals supply information, and that information would be foolish to ignore. In fact, it would be foolish to ignore any source of information.

Unit 7 Appendix: Key Interpretations

•Confidence Intervals

•**Frequentist Interpretation (Long):** Our estimated magnitude of the relationship in the population is -5.26. In light of sampling error, we recognize that the true magnitude is probably not -5.26. A 95% confidence interval of -8.07 to -2.46 provides a range of plausible values for the true magnitude. Such ranges, however, are not infallible; in fact, we expect such ranges to fail 5% of the time.

•**Frequentist Interpretation (Short):** Our 95% confidence interval of -8.07 to -2.46 provides a range of plausible values for the true magnitude.

•**Bayesian Interpretation:** In the absence of further information, it is reasonable to conclude that there is a 95% chance that the population magnitude is between -8.07 and -2.46.

•Statistical Significance

•There is a statistically significant relationship ($p < 0.05$) between depression levels and reading scores in our sample ($n = 880$) of children of immigrants.

•We observe a regression coefficient of -5.26 in our sample. If there were no relationship in the population, we would observe a coefficient this large of larger in less than 0.01% of our samples.

Unit 7 Appendix: Key Terminology

- Type I Error is (falsely) rejecting the null *when* the null is true. Alpha level is the probability of Type I Error due to sampling error. By setting our alpha level, we control the probability of Type I Error.
- Alpha level is the probability of (falsely) rejecting the null *when* the null is true.
- The purpose of confidence intervals is to contain the true population parameter. Over your lifetime, 95% of (95%) confidence intervals will succeed and 5% will fail. You will not know which are the unlucky 5%! Analogous reasoning holds for alpha level
- Standard errors measure the precision of our estimate. The smaller the standard error, the smaller our hypothetical sampling distribution, the greater our statistical power.
- A t-test is a test to see if our observation is a sufficient number of standard errors away from zero to scare us into rejecting the null hypothesis. A t statistic of $\pm$ 2, indicating that our observation is $\pm$ 2 standard errors from zero, will have a two-tailed significance level (or p value) of about 0.05.
- Confidence intervals provide a range of plausible values for population parameters (such as means, y-intercepts, and slopes).
- If zero is within the plausible range of values, we do not reject the null hypothesis.
- Confidence intervals **do not** provide a range of plausible values for observations (such as individual scores).

Unit 7 Appendix: Math (Degrees of Freedom, Part I of II)

In order to keep our population estimates unbiased, we use degrees of freedom when we might otherwise be inclined to use sample size. You'll find that degrees of freedom are just sample sizes minus little smidges. In our introductory course, the smidges will be either 1 or 2. We've begun to think of certain statistics as population estimates. Your sample mean is an estimate of the population mean. Your sample standard deviation is an estimate of the population standard deviation. Some population estimates are relatively simple, such as the mean. Other population estimates, however, have population estimates already built into them. For example, the standard deviation has the mean built into it. In other words, you must calculate the mean to calculate the standard deviation. As statistics begin to feed on themselves, we invite bias unless we adjust by a smidge. This is easiest to see under the microscope of the smallest of sample sizes, $n = 1$:

Here is the data in our sample of reading scores: 62

The mean of our sample is 62/1 = 62, which is an unbiased estimate of the population mean. (Hey, I said it was unbiased, not precise! By "unbiased," I mean that it is just as likely to be an overestimate as an underestimate.)

Using sample size instead of degrees of freedom, the standard deviation of our sample is 0/1 = 0, which is a biased estimate of the population standard deviation. Zero standard deviation will ALWAYS be an underestimate for a variable! It is biased to be an underestimate. In larger samples, our estimate of the population standard deviation won't always be an underestimate, but it has a slight tendency to be an underestimate. That tendency is bias.

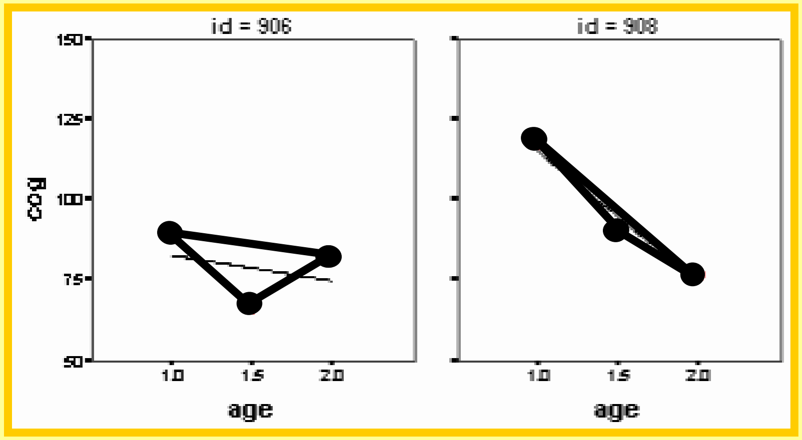
Mad cow disease can come from cows eating cows. Bias can come from statistics eating statistics.

We correct for the bias by subtracting 1 from the sample size for each in-built statistic. The standard deviation has one in-built statistic, the mean, so its degrees of freedom are n-1. The standard error of the slope coefficient has two in-built statistics, the y-intercept and the slope, so its degrees of freedom are n-2.

Unit 7 Appendix: Math (Degrees of Freedom, Part II of II)

The name "degrees of freedom" has to do with the freedom that our statistics have to vary in response to the data. For calculating the standard deviation from a sample size of 1, there are no degrees of freedom, *n* -1 = 1 -1 = 0. Do you see how the standard deviation has zero room to vary in response to the data when the sample size is 1? The standard deviation is constrained to be zero.

If our microscopic, *n* = 1, example seems ludicrous, let me give you a not-so-ludicrous example, n = 3. In longitudinal data analysis (in a couple of semesters), we might fit two levels of regression models, where Level 1 is the individual level, and Level 2 is group level. At Level 1, we might have measures of cognition at ages, 1-year, 1.5-years and 2-years old. For each individual subject, therefore, we have a small sample of reading scores, *n* = 3.



From Singer and Willet's *ALDA*, *Chapter 3*

When *n* = 3, we have 1 degree of freedom to calculate the standard error of the slope, *n*-2=3-2=1. Do you see how, if the sample size were any smaller, the standard error would be constrained to be zero, in other words, it would not be free to vary? A straight line always fits perfectly through two bivariate data points. A straight line has two in-built population estimates, the y-intercept and slope. While the slope estimate is unbiased, the standard error (another population estimate) is biased unless we consider degrees of freedom.

Both the mean and the slope are population estimates, but we estimate them by minimizing the variation in the sample. (Both the mean and the slope are estimated by ordinary least squares methods, or some algebraic equivalent.) Since we designed the parameter estimates to minimize the variation, it is no wonder that our measures of variation from the mean and the slope are biased downward unless unless unless we correct by using degrees of freedom instead of sample size.

Unit 7 Appendix: SPSS Syntax

```
*********************************************************************************.
*To ask for confidence intervals, simply add "CI" to your options line.
*********************************************************************************.

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT SAT
  /METHOD=ENTER PERCENT.

*********************************************************************************.
*Unfortunately, there's no easy way to do this via syntax.
*You must dive into the output.
*********************************************************************************.

GRAPH
  /SCATTERPLOT(BIVAR)=PERCENT WITH SAT
  /MISSING=LISTWISE.
```

Unit 7 Appendix: R Syntax

```
# Load your data
load('F:/CD146 2010/Data Sets/childrenofimmigrants.rda')
# Attach your data
attach(childrenofimmigrants)
# Specify your model and assign it a name
model.1 <- lm(Reading~Depress)
# Summarize your model
summary(model.1)
# Output confidence intervals for your model
confint(model.1)
# So you want to plot the confidence intervals for the line?
# Plot y vs. x
plot(Reading~Depress)
# Add your regression line
abline(model.1, col=“darkblue")
# Generate a dummy dataset, and make predictions for it with CIs
newDepress <- seq(min(Depress),max(Depress), by=.01)
prd <- predict(model.1, newdata=data.frame(Depress=newDepress),
        interval=c("confidence"), level = 0.95, type="response")
# Add your confidence intervals to your plot
lines(newDepress, prd[,2], col="darkblue",lty=2)
lines(newDepress, prd[,3], col="darkblue",lty=2)
```

Perceived Intimacy of Adolescent Girls (Intimacy.sav)

- **Overview: Dataset contains self-ratings of the intimacy that adolescent girls perceive themselves as having with: (a) their mother and (b) their boyfriend.**
- **Source: HGSE thesis by Dr. Linda Kilner entitled Intimacy in Female Adolescent's Relationships with Parents and Friends (1991). Kilner collected the ratings using the Adolescent Intimacy Scale.**
- **Sample: 64 adolescent girls in the sophomore, junior and senior classes of a local suburban public school system.**
- **Variables:**

Self Disclosure to Mother (M_Seldis)
Trusts Mother (M_Trust)
Mutual Caring with Mother (M_Care)
Risk Vulnerability with Mother (M_Vuln)
Physical Affection with Mother (M_Phys)
Resolves Conflicts with Mother (M_Cres)

Self Disclosure to Boyfriend (B_Seldis)
Trusts Boyfriend (B_Trust)
Mutual Caring with Boyfriend (B_Care)
Risk Vulnerability with Boyfriend (B_Vuln)
Physical Affection with Boyfriend (B_Phys)
Resolves Conflicts with Boyfriend (B_Cres)

Perceived Intimacy of Adolescent Girls (Intimacy.sav)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.731[a]	.534	.526	.80682

a. Predictors: (Constant), Self-disclose to boyfriend

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	43.280	1	43.280	66.487	.000[a]
	Residual	37.756	58	.651		
	Total	81.037	59			

a. Predictors: (Constant), Self-disclose to boyfriend

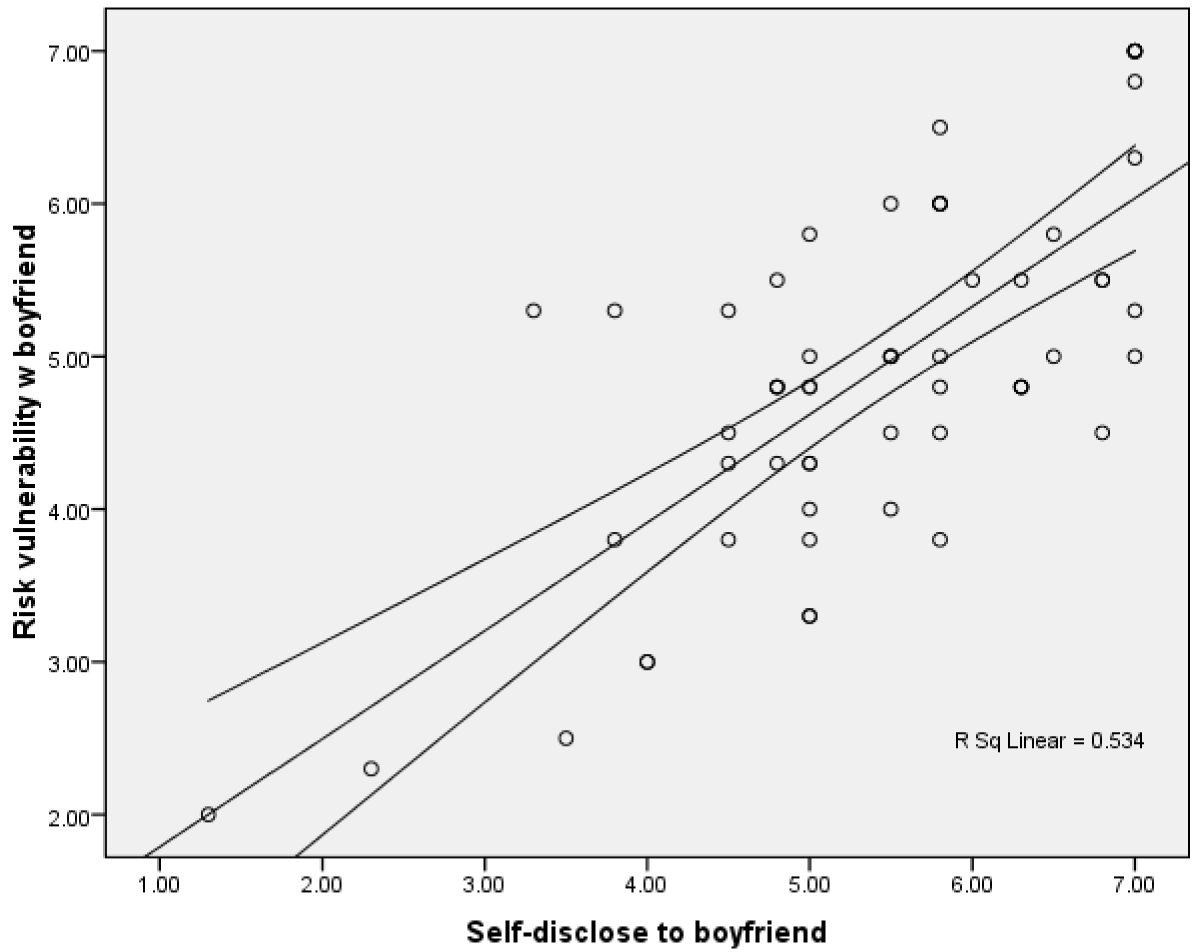
b. Dependent Variable: Risk vulnerability w boyfriend

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1.081	.482		2.244	.029	.117	2.045
	Self-disclose to boyfriend	.708	.087	.731	8.154	.000	.534	.882

a. Dependent Variable: Risk vulnerability w boyfriend

Perceived Intimacy of Adolescent Girls (Intimacy.sav)



Perceived Intimacy of Adolescent Girls (Intimacy.sav)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.002[a]	.000	-.017	1.19785

a. Predictors: (Constant), Self-disclose to mother

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.000	1	.000	.000	.985[a]
	Residual	83.221	58	1.435		
	Total	83.222	59			

a. Predictors: (Constant), Self-disclose to mother

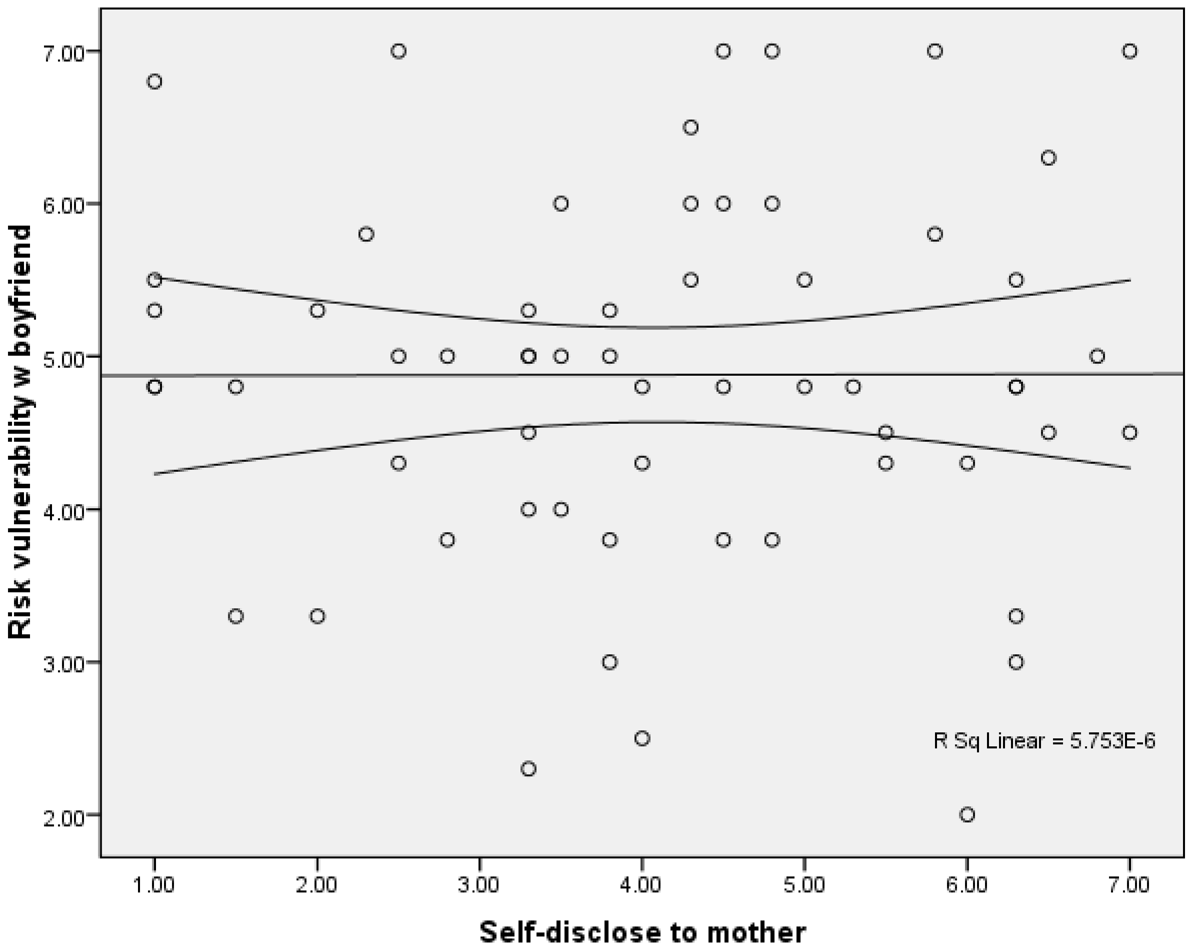
b. Dependent Variable: Risk vulnerability w boyfriend

Coefficients[a]

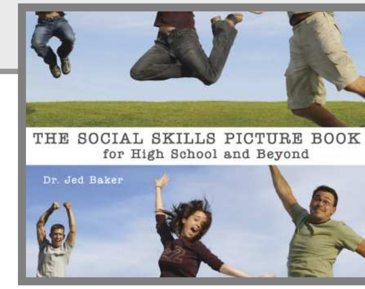
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	4.872	.404		12.050	.000	4.062	5.681
	Self-disclose to mother	.002	.091	.002	.018	.985	-.181	.184

a. Dependent Variable: Risk vulnerability w boyfriend

Perceived Intimacy of Adolescent Girls (Intimacy.sav)



High School and Beyond (HSB.sav)



- **Overview: High School & Beyond – Subset of data focused on selected student and school characteristics as predictors of academic achievement.**
- **Source: Subset of data graciously provided by Valerie Lee, University of Michigan.**
- **Sample: This subsample has 1044 students in 205 schools. Missing data on the outcome test score and family SES were eliminated. In addition, schools with fewer than 3 students included in this subset of data were excluded.**
- **Variables:**

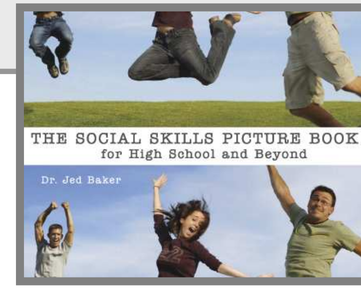
Variables about the student—

(Black) 1=Black, 0=Other
(Latin) 1=Latino/a, 0=Other
(Sex) 1=Female, 0=Male
(BYSES) Base year SES
(GPA80) HS GPA in 1980
(GPS82) HS GPA in 1982
(BYTest) Base year composite of reading and math tests
(BBConc) Base year self concept
(FEConc) First Follow-up self concept

Variables about the student's school—

(PctMin) % HS that is minority students Percentage
(HSSize) HS Size
(PctDrop) % dropouts in HS Percentage
(BYSES_S) Average SES in HS sample
(GPA80_S) Average GPA80 in HS sample
(GPA82_S) Average GPA82 in HS sample
(BYTest_S) Average test score in HS sample
(BBConc_S) Average base year self concept in HS sample
(FEConc_S) Average follow-up self concept in HS sample

High School and Beyond (HSB.sav)



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.440[a]	.193	.192	7.71738

a. Predictors: (Constant), Base Year SES

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14858.061	1	14858.061	249.473	.000[a]
	Residual	62059.321	1042	59.558		
	Total	76917.382	1043			

a. Predictors: (Constant), Base Year SES

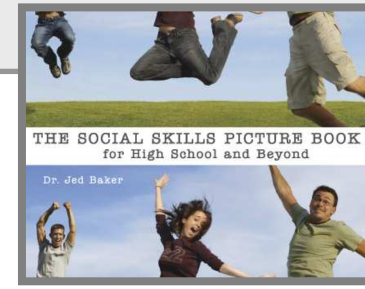
b. Dependent Variable: Base Year Composite Test

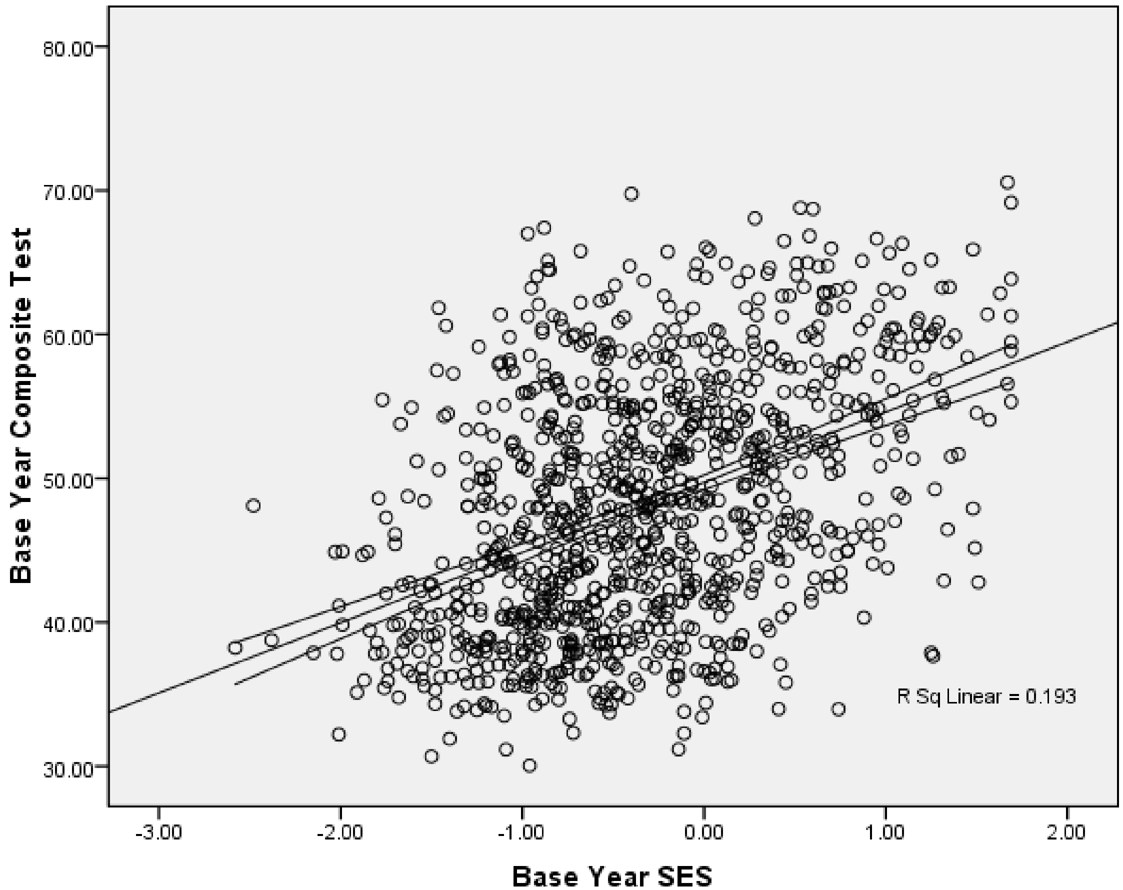
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	49.726	.260		191.448	.000	49.216	50.235
	Base Year SES	4.879	.309	.440	15.795	.000	4.273	5.485

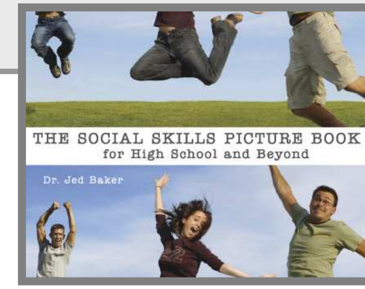
a. Dependent Variable: Base Year Composite Test

High School and Beyond (HSB.sav)





High School and Beyond (HSB.sav)



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.429[a]	.184	.184	7.75965

a. Predictors: (Constant), BY SES, School Avg

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14176.284	1	14176.284	235.439	.000[a]
	Residual	62741.098	1042	60.212		
	Total	76917.382	1043			

a. Predictors: (Constant), BY SES, School Avg

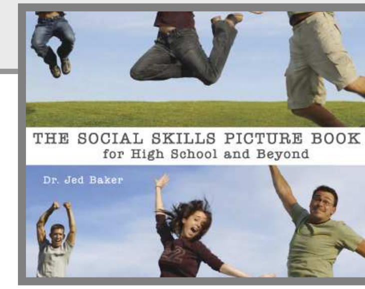
b. Dependent Variable: Base Year Composite Test

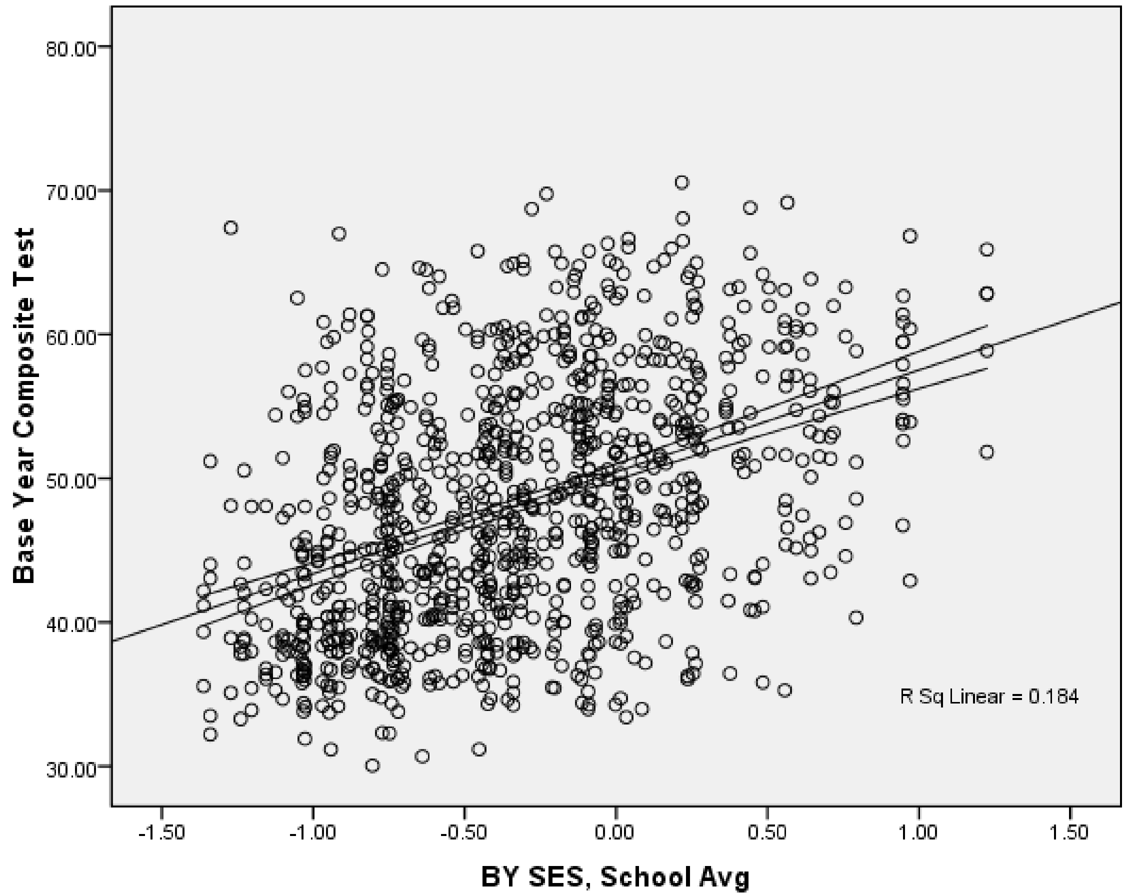
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	50.451	.284		177.397	.000	49.893	51.009
	BY SES, School Avg	7.075	.461	.429	15.344	.000	6.171	7.980

a. Dependent Variable: Base Year Composite Test

High School and Beyond (HSB.sav)





Understanding Causes of Illness (ILLCAUSE.sav)

- **Overview: Data for investigating differences in children's understanding of the causes of illness, by their health status.**
- **Source: Perrin E.C., Sayer A.G., and Willett J.B. (1991).**

  Sticks And Stones May Break My Bones: Reasoning About Illness Causality And Body Functioning In Children Who Have A Chronic Illness, *Pediatrics*, 88(3), 608-19.
- **Sample: 301 children, including a sub-sample of 205 who were described as asthmatic, diabetic,or healthy. After further reductions due to the *list-wise deletion* of cases with missing data on one or more variables, the analytic sub-sample used in class ends up containing: 33 diabetic children, 68 asthmatic children and 93 healthy children.**
- **Variables:**

(ILLCAUSE)	Child's Understanding of Illness Causality
(SES)	Child's SES (Note that a high score means low SES.)
(PPVT)	Child's Score on the Peabody Picture Vocabulary Test
(AGE)	Child's Age, In Months
(GENREAS)	Child's Score on a General Reasoning Test
(ChronicallyIll)	1 = Asthmatic or Diabetic, 0 = Healthy
(Asthmatic)	1 = Asthmatic, 0 = Healthy
(Diabetic)	1 = Diabetic, 0 = Healthy

Understanding Causes of Illness (ILLCAUSE.sav)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.824[a]	.679	.678	.58181

a. Predictors: (Constant), General Reasoning

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	136.226	1	136.226	402.433	.000[a]
	Residual	64.316	190	.339		
	Total	200.542	191			

a. Predictors: (Constant), General Reasoning

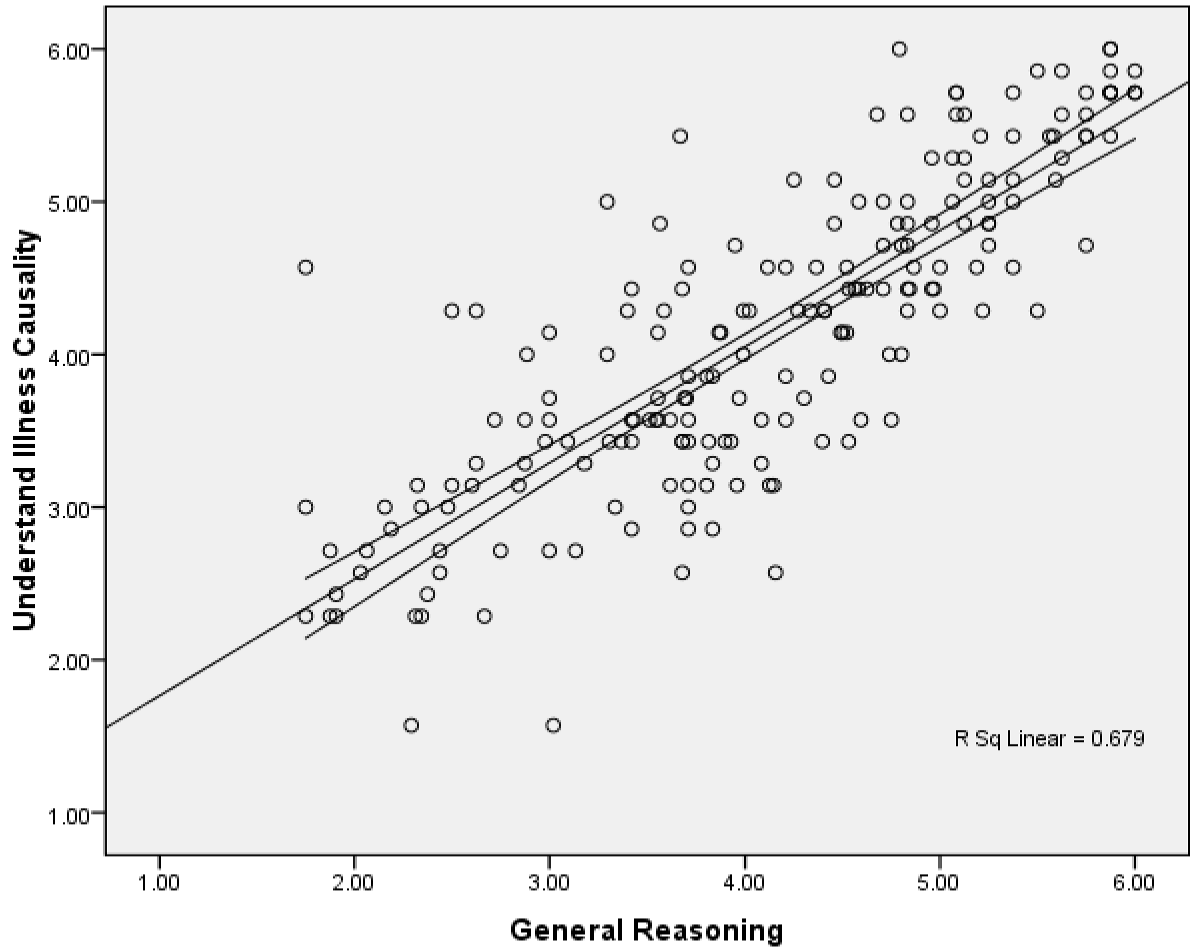
b. Dependent Variable: Understand Illness Causality

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1.004	.162		6.204	.000	.685	1.323
	General Reasoning	.762	.038	.824	20.061	.000	.687	.837

a. Dependent Variable: Understand Illness Causality

Understanding Causes of Illness (ILLCAUSE.sav)



Understanding Causes of Illness (ILLCAUSE.sav)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.440[a]	.194	.189	.94848

a. Predictors: (Constant), 1 = Asthmatic, 0 = Healthy

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	34.383	1	34.383	38.219	.000[a]
	Residual	143.040	159	.900		
	Total	177.423	160			

a. Predictors: (Constant), 1 = Asthmatic, 0 = Healthy

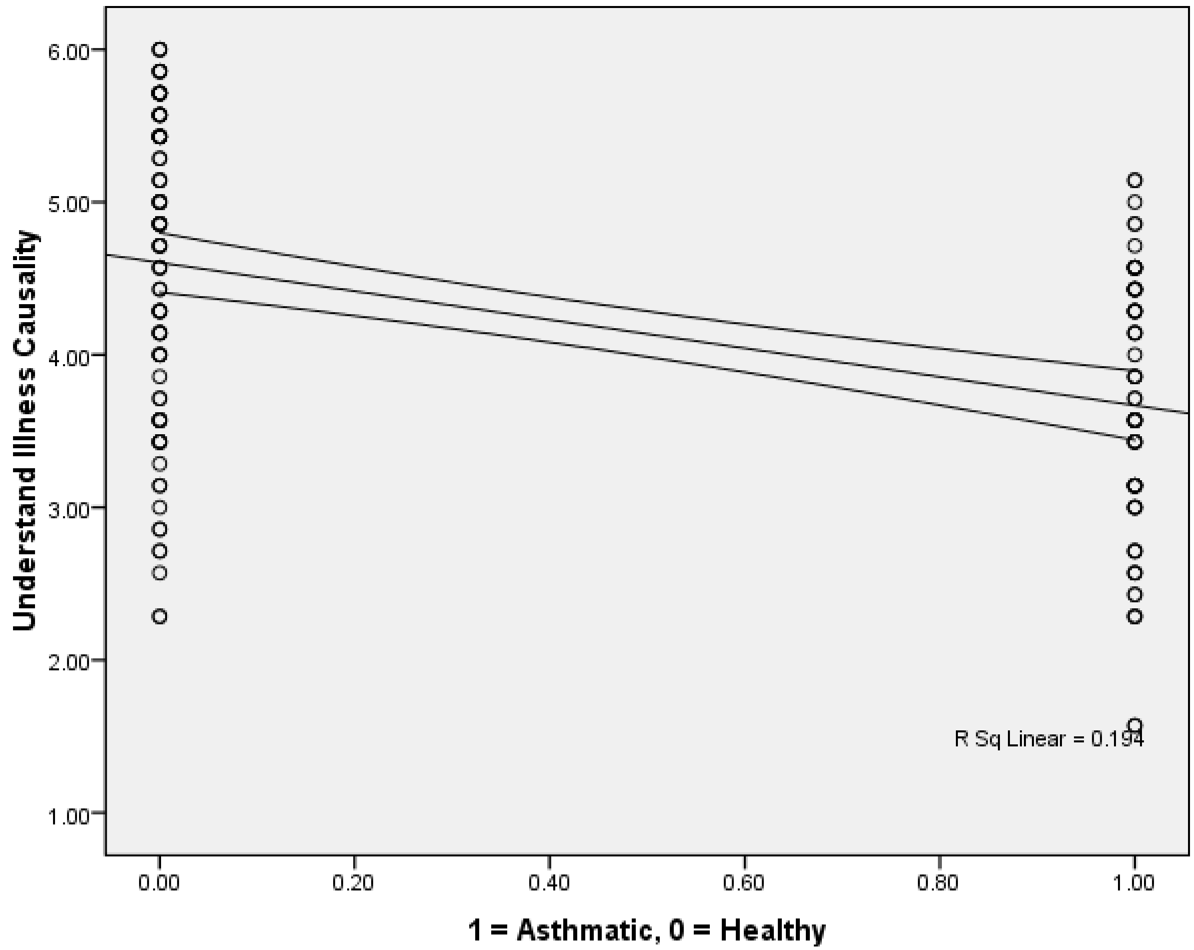
b. Dependent Variable: Understand Illness Causality

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	4.604	.098		46.807	.000	4.409	4.798
	1 = Asthmatic, 0 = Healthy	-.936	.151	-.440	-6.182	.000	-1.234	-.637

a. Dependent Variable: Understand Illness Causality

Understanding Causes of Illness (ILLCAUSE.sav)



Children of Immigrants (ChildrenOfImmigrants.sav)

- **Overview: "CILS is a longitudinal study designed to study the adaptation process of the immigrant second generation which is defined broadly as U.S.-born children with at least one foreign-born parent or children born abroad but brought at an early age to the United States. The original survey was conducted with large samples of second-generation children attending the 8th and 9th grades in public and private schools in the metropolitan areas of Miami/Ft. Lauderdale in Florida and San Diego, California" (from the website description of the data set).**
- **Source: Portes, Alejandro, & Ruben G. Rumbaut (2001). *Legacies: The Story of the Immigrant SecondGeneration*. Berkeley CA: University of California Press.**
- **Sample: Random sample of 880 participants obtained through the website.**
- **Variables:**

(Reading)	Stanford Reading Achievement Score
(Freelunch)	% students in school who are eligible for free lunch program
(Male)	1=Male 0=Female
(Depress)	Depression scale (Higher score means more depressed)
(SES)	Composite family SES score

Children of Immigrants (ChildrenOfImmigrants.sav)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.404[a]	.163	.162	34.837

a. Predictors: (Constant), Composite Family SES Score

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	207358.576	1	207358.576	170.863	.000[a]
	Residual	1065535.601	878	1213.594		
	Total	1272894.177	879			

a. Predictors: (Constant), Composite Family SES Score

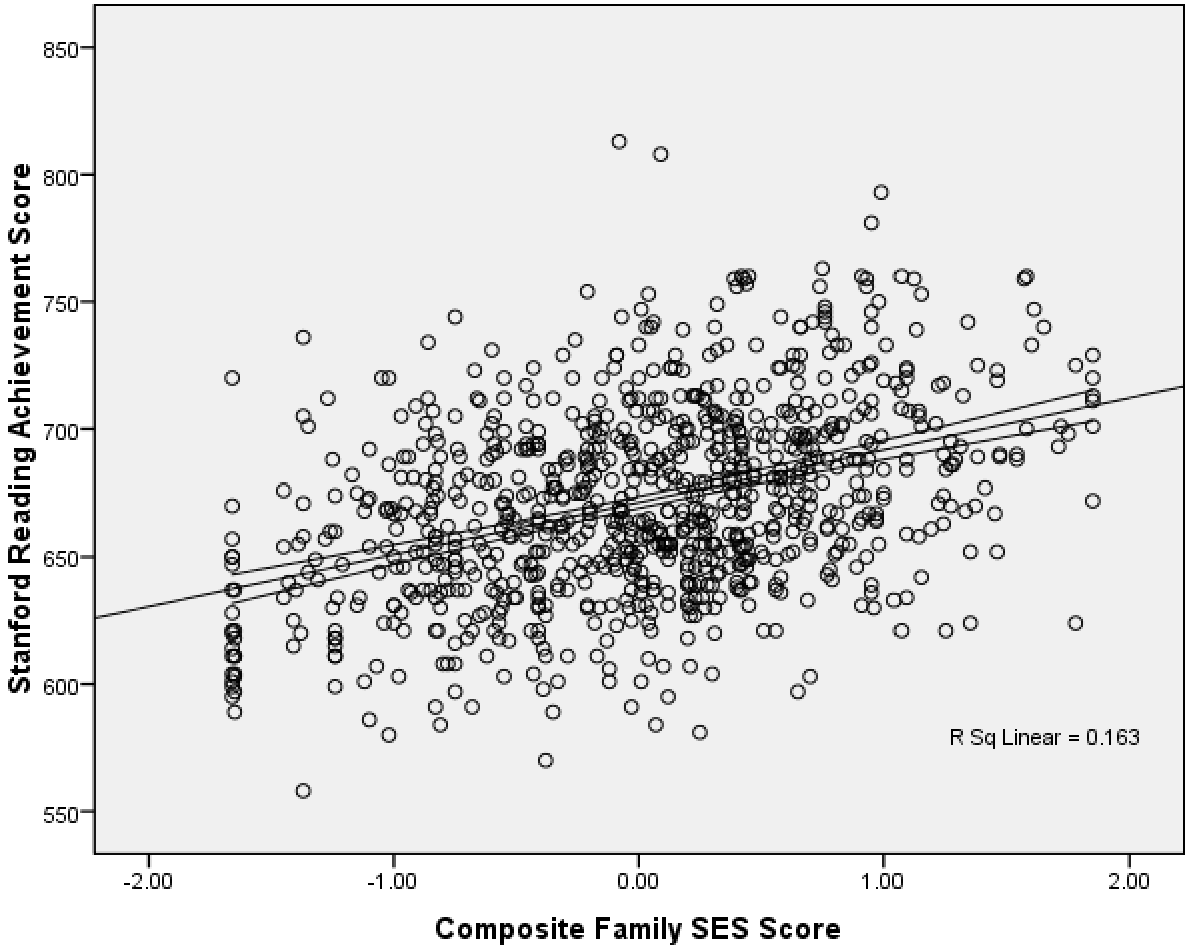
b. Dependent Variable: Stanford Reading Achievement Score

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	671.350	1.175		571.418	.000	669.044	673.656
	Composite Family SES Score	20.418	1.562	.404	13.071	.000	17.352	23.483

a. Dependent Variable: Stanford Reading Achievement Score

Children of Immigrants (ChildrenOfImmigrants.sav)



Children of Immigrants (ChildrenOfImmigrants.sav)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.353[a]	.125	.124	35.624

a. Predictors: (Constant), % of Students in Child's School Eligible for Free Lunch

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	158680.746	1	158680.746	125.040	.000[a]
	Residual	1114213.431	878	1269.036		
	Total	1272894.177	879			

a. Predictors: (Constant), % of Students in Child's School Eligible for Free Lunch

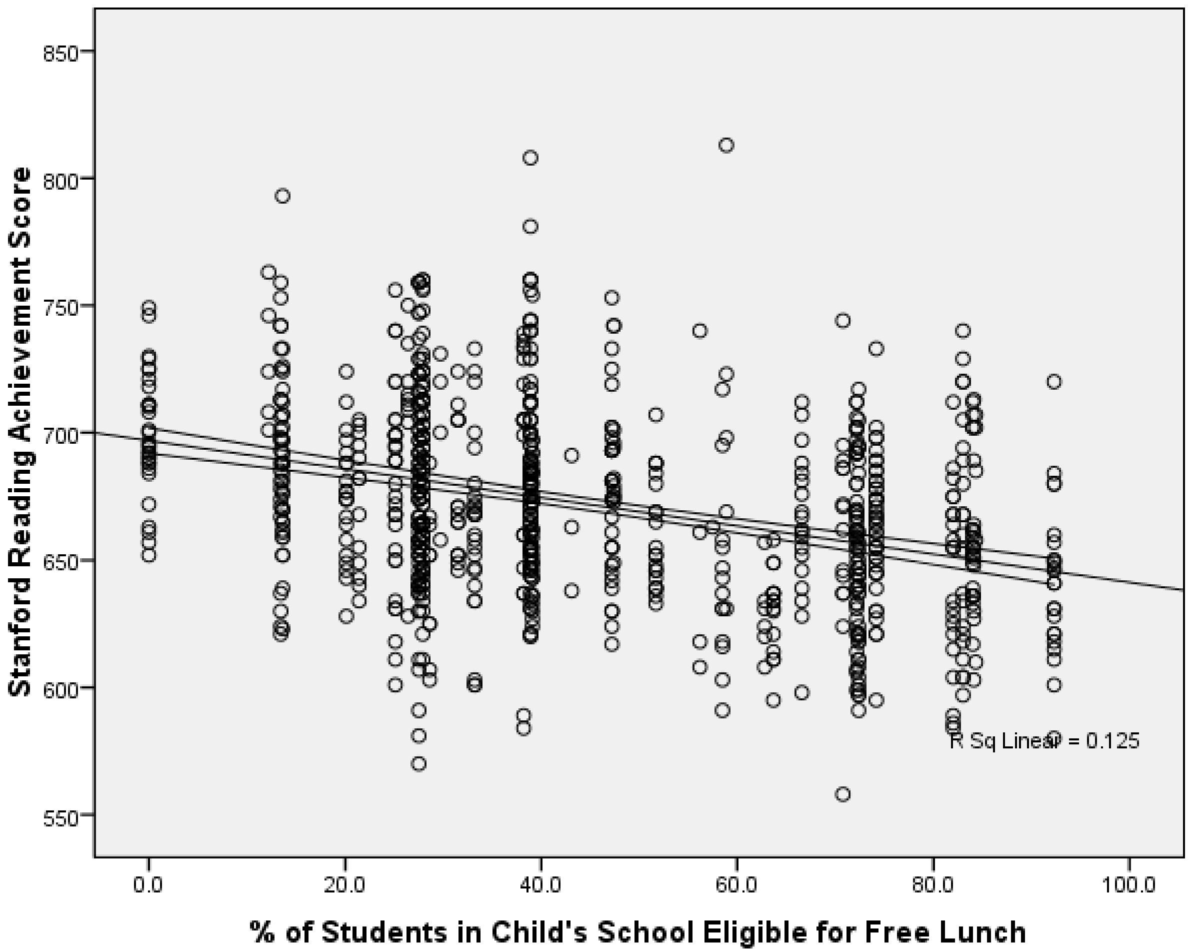
b. Dependent Variable: Stanford Reading Achievement Score

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	696.847	2.540		274.325	.000	691.861	701.832
	% of Students in Child's School Eligible for Free Lunch	-.555	.050	-.353	-11.182	.000	-.653	-.458

a. Dependent Variable: Stanford Reading Achievement Score

Children of Immigrants (ChildrenOfImmigrants.sav)



Human Development in Chicago Neighborhoods (Neighborhoods.sav)

- **These data were collected as part of the Project on Human Development in Chicago Neighborhoods in 1995.**
- **Source: Sampson, R.J., Raudenbush, S.W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277, 918-924.**
- **Sample: The data described here consist of information from 343 Neighborhood Clusters in Chicago Illinois. Some of the variables were obtained by project staff from the 1990 Census and city records. Other variables were obtained through questionnaire interviews with 8782 Chicago residents who were interviewed in their homes.**
- **Variables:**

(Homr90)	Homicide Rate c. 1990
(Murder95)	Homicide Rate 1995
(Disadvan)	Concentrated Disadvantage
(Imm_Conc)	Immigrant
(ResStab)	Residential Stability
(Popul)	Population in 1000s
(CollEff)	Collective Efficacy
(Victim)	% Respondents Who Were Victims of Violence
(PercViol)	% Respondents Who Perceived Violence

Human Development in Chicago Neighborhoods (Neighborhoods.sav)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.382[a]	.146	.143	.91099

a. Predictors: (Constant), Collective efficacy

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	48.191	1	48.191	58.068	.000[a]
	Residual	282.170	340	.830		
	Total	330.361	341			

a. Predictors: (Constant), Collective efficacy

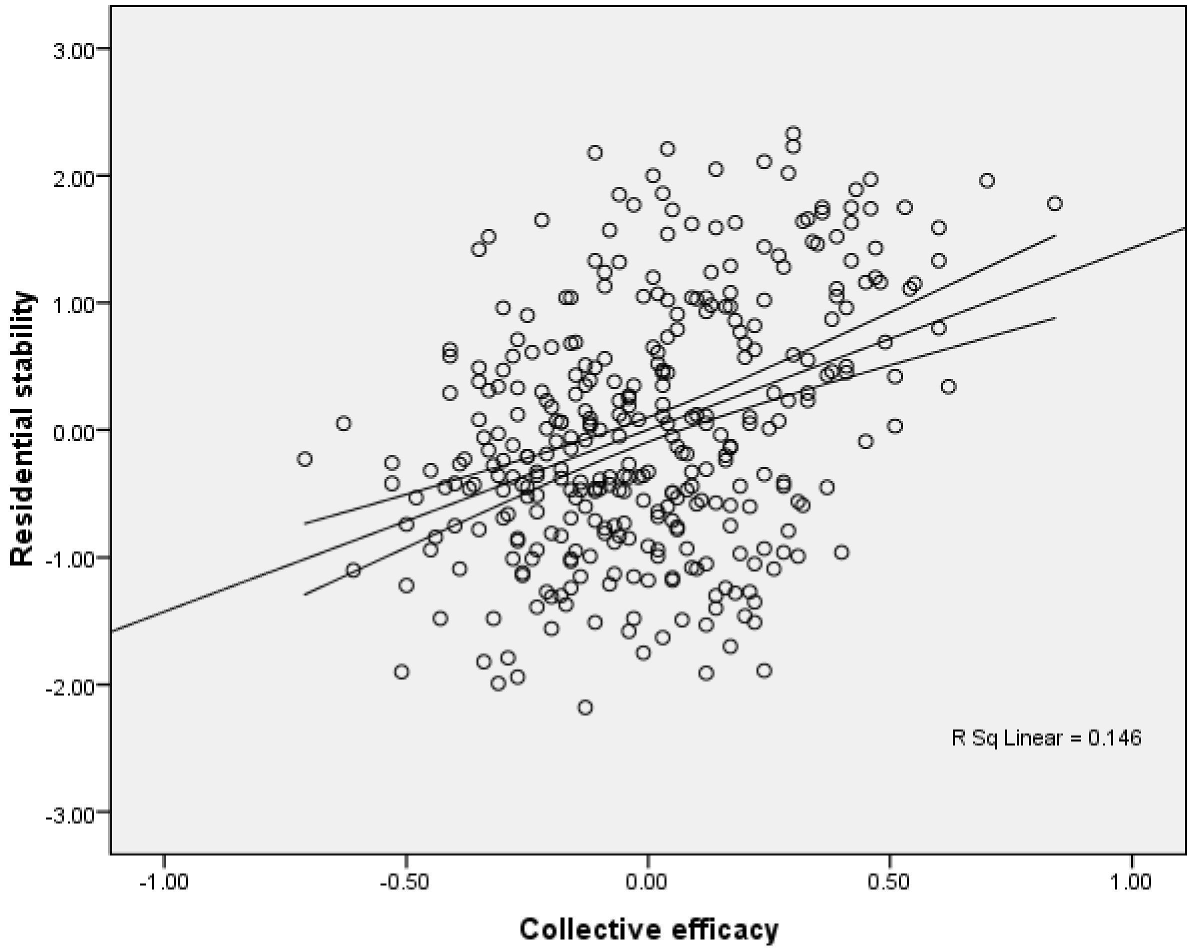
b. Dependent Variable: Residential stability

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	.002	.049		.050	.961	-.094	.099
	Collective efficacy	1.429	.187	.382	7.620	.000	1.060	1.797

a. Dependent Variable: Residential stability

Human Development in Chicago Neighborhoods (Neighborhoods.sav)



Human Development in Chicago Neighborhoods (Neighborhoods.sav)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.147[a]	.022	.019	.97506

a. Predictors: (Constant), Homicide rate 1988-90

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7.112	1	7.112	7.480	.007[a]
	Residual	323.249	340	.951		
	Total	330.361	341			

a. Predictors: (Constant), Homicide rate 1988-90

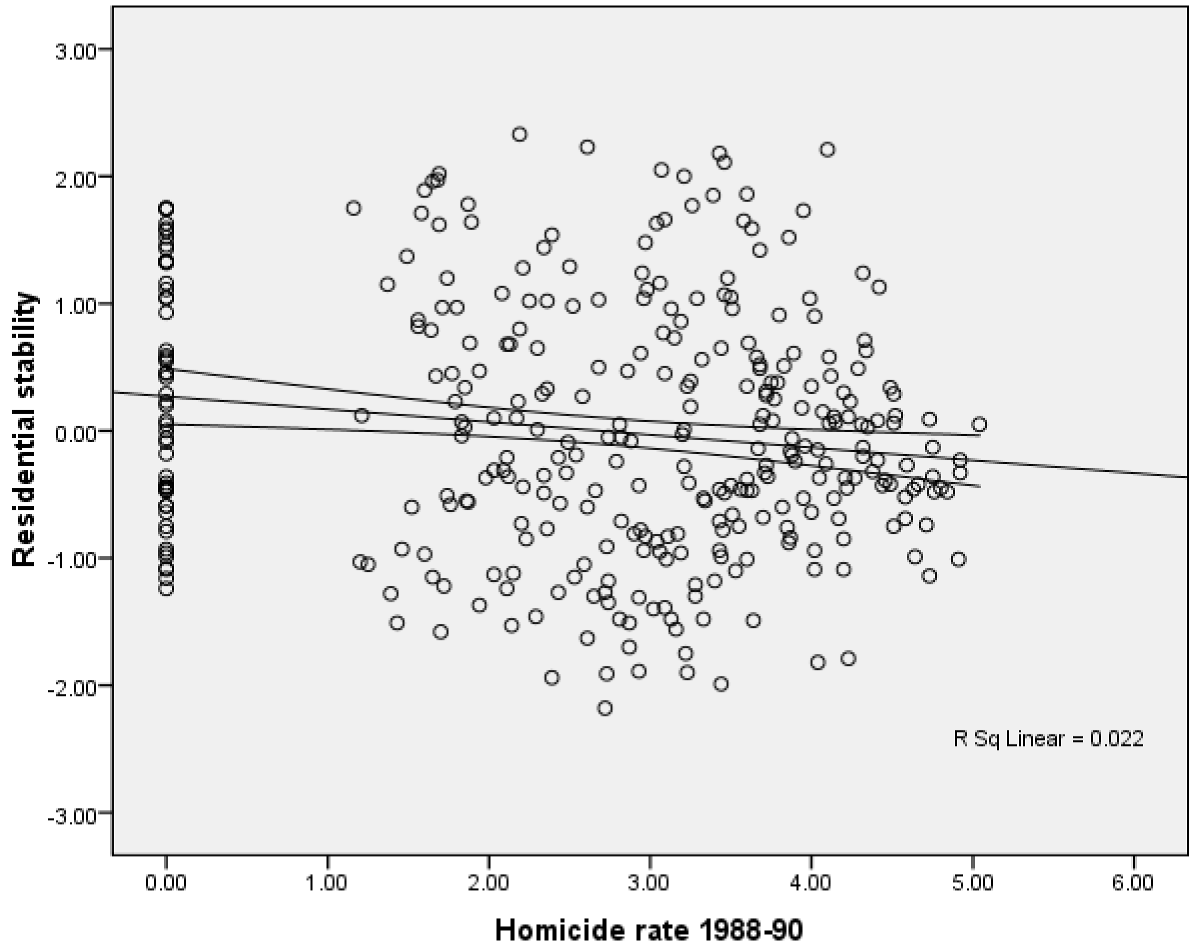
b. Dependent Variable: Residential stability

Coefficients[a]

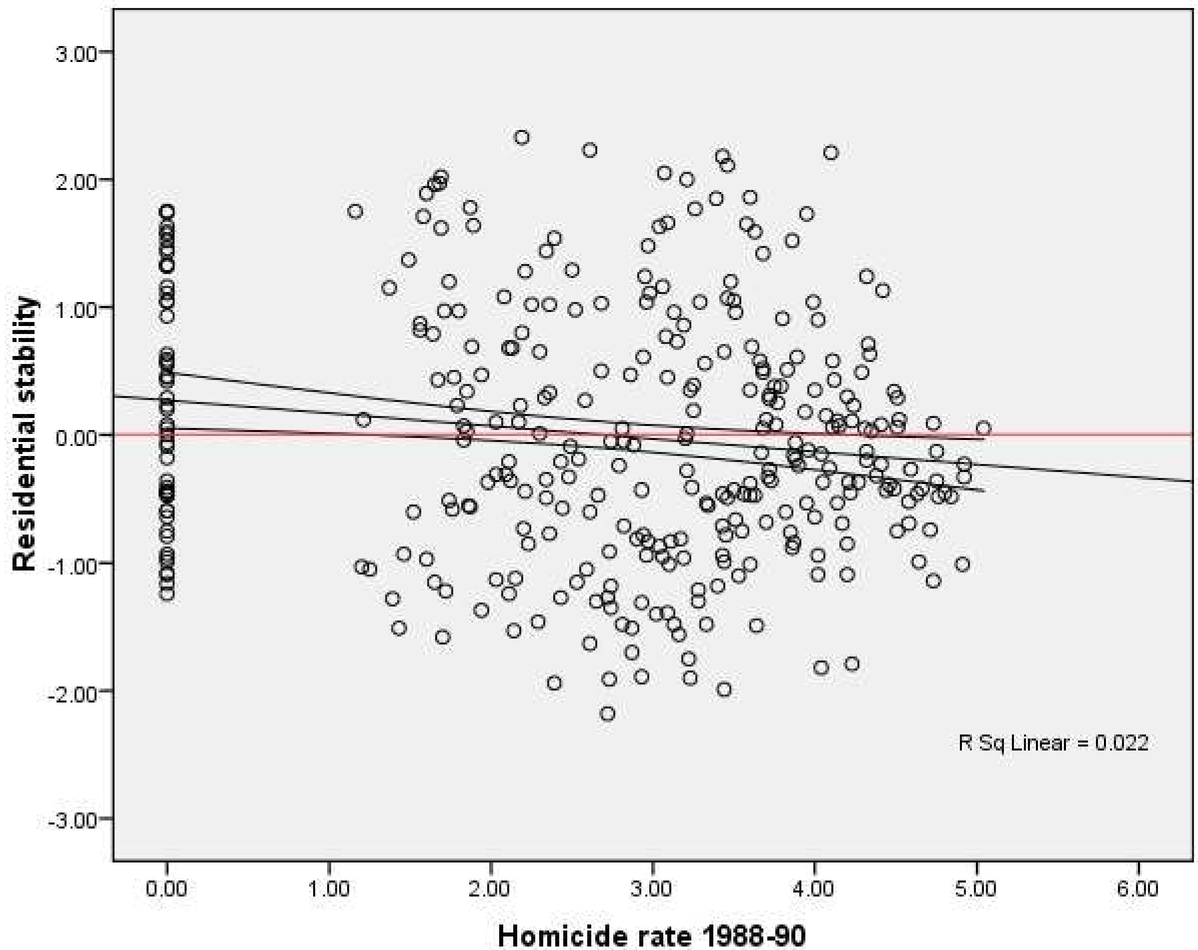
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	.270	.111		2.432	.016	.052	.489
	Homicide rate 1988-90	-.100	.037	-.147	-2.735	.007	-.173	-.028

a. Dependent Variable: Residential stability

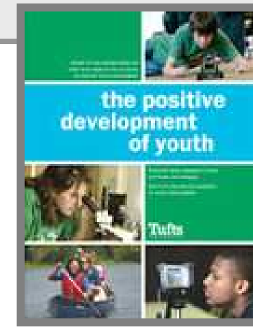
Human Development in Chicago Neighborhoods (Neighborhoods.sav)



Human Development in Chicago Neighborhoods (Neighborhoods.sav)



4-H Study of Positive Youth Development (4H.sav)



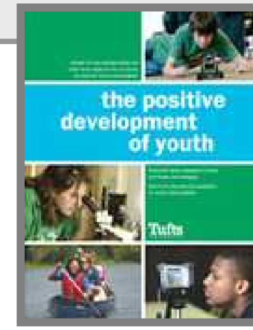
- **4-H Study of Positive Youth Development**
- **Source: Subset of data from IARYD, Tufts University**

- **Sample: These data consist of seventh graders who participated in Wave 3 of the 4-H Study of Positive Youth Development at Tufts University. This subfile is a substantially sampled-down version of the original file, as all the cases with any missing data on these selected variables were eliminated.**
- **Variables:**

(SexFem)	1=Female, 0=Male
(MothEd)	Years of Mother's Education
(Grades)	Self-Reported Grades
(Depression)	Depression (Continuous)
(FrInfl)	Friends' Positive Influences
(PeerSupp)	Peer Support
(Depressed)	0 = (1-15 on Depression) 1 = Yes (16+ on Depression)

(AcadComp)	Self-Perceived Academic Competence
(SocComp)	Self-Perceived Social Competence
(PhysComp)	Self-Perceived Physical Competence
(PhysApp)	Self-Perceived Physical Appearance
(CondBeh)	Self-Perceived Conduct Behavior
(SelfWorth)	Self-Worth

4-H Study of Positive Youth Development (4H.sav)



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.559[a]	.313	.311	.50341

a. Predictors: (Constant), Depression

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	46.912	1	46.912	185.115	.000[a]
	Residual	103.141	407	.253		
	Total	150.053	408			

a. Predictors: (Constant), Depression

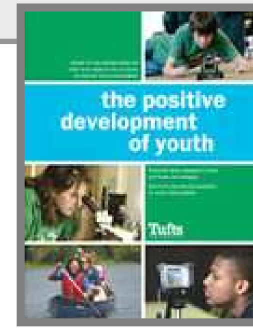
b. Dependent Variable: Self-Worth

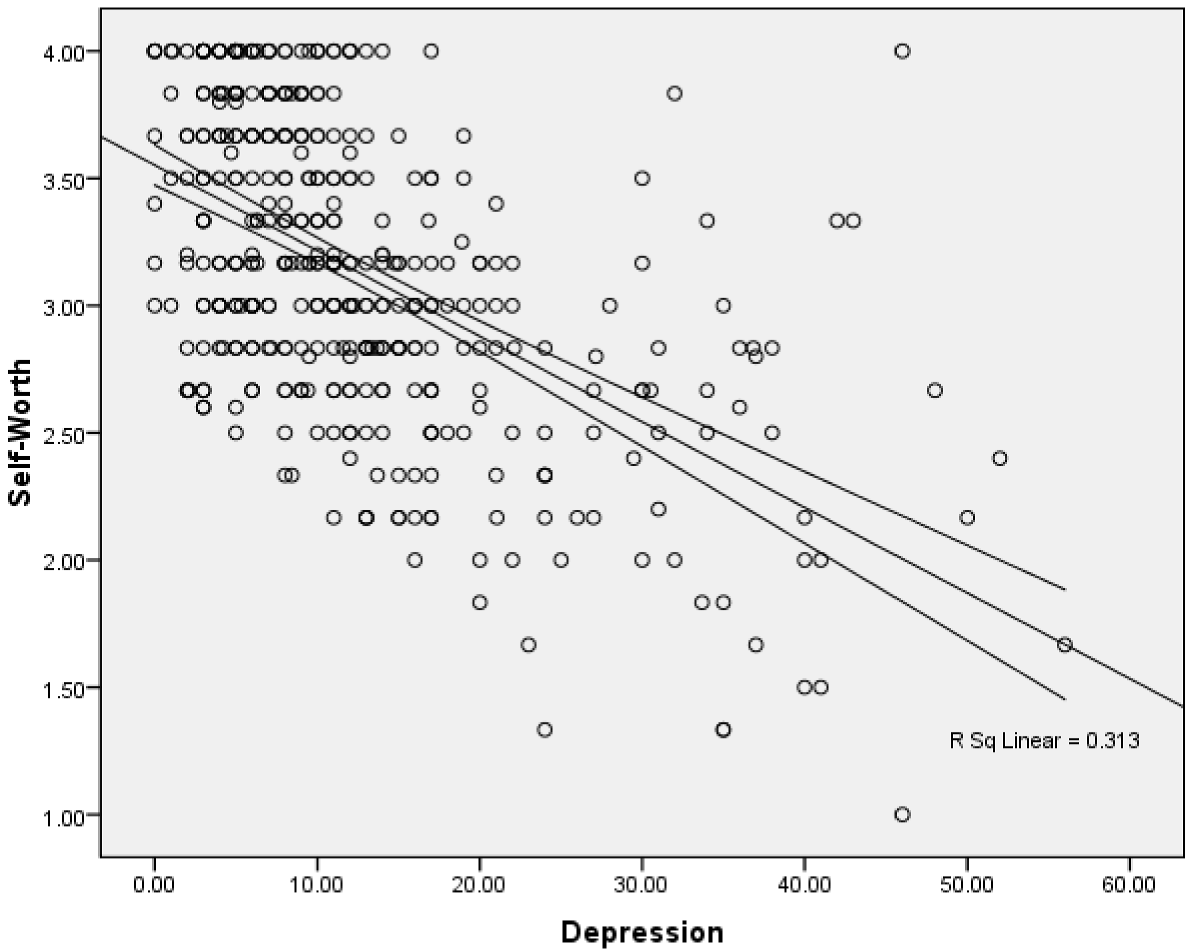
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	3.552	.040		88.146	.000	3.473	3.631
	Depression	-.034	.002	-.559	-13.606	.000	-.038	-.029

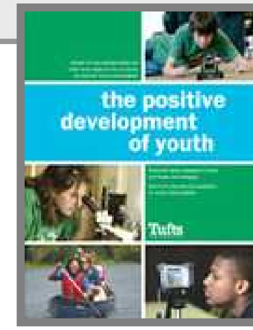
a. Dependent Variable: Self-Worth

4-H Study of Positive Youth Development (4H.sav)





4-H Study of Positive Youth Development (4H.sav)



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.504[a]	.254	.252	.52460

a. Predictors: (Constant), Depressed = 1, Not Depressed = 0

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	38.046	1	38.046	138.247	.000[a]
	Residual	112.007	407	.275		
	Total	150.053	408			

a. Predictors: (Constant), Depressed = 1, Not Depressed = 0

b. Dependent Variable: Self-Worth

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	3.307	.030		108.824	.000	3.247	3.367
	Depressed = 1, Not Depressed = 0	-.686	.058	-.504	-11.758	.000	-.801	-.571

a. Dependent Variable: Self-Worth

4-H Study of Positive Youth Development (4H.sav)