HYPOTHESIS TESTING: CONFIDENCE INTERVALS, T-TESTS, ANOVAS, AND REGRESSION

Lecture Background

- This is a lightning speed summary of introductory statistical methods for senior undergraduate students in the honors program. They had designed their own research studies and had collected quantitative data.
- The previous lecture had covered data entry, data cleaning, and univariate and bivariate analyses.
- The lecture was occurring in a computer lab. The class had access to a common dataset collected from undergraduate students. After I introduced each new analytic method, we took time to do the procedure and explore the output thoroughly.
- Lecture time is four hours.
- Students were not expected to be experts on statistical analysis themselves.
 Rather, they should have been able to know possible avenues of analysis and begin to analyze their data with significant support

Inferential Statistics

- The only way to know the true mean (or any other parameter) of a population is by surveying the entire population
- When we take only a sample of the population, we are inferring or generalizing known characteristics of the sample to the population
- Unfortunately, there is always uncertainty about whether the characteristics of our sample reflect the characteristics of the population
 - This is true even if we have a representative sample and followed the best sampling methods available

Confidence Intervals

Confidence Intervals

The confidence interval is an expression of that uncertainty, expressing an area around the sample mean we think the population mean is likely to fall

- The larger our sample, the smaller the confidence interval
 - The larger the sample, the more sure we are that the sample mean approximates the population mean

Confidence Intervals

- There are different confidence intervals, depending on how sure we want to be that the population mean falls within the CI
 - The more sure we want to be, the wider the range of the confidence interval
- \square CI = $\dot{X} \pm \frac{1-\alpha/2}{2} Z(\sigma/\sqrt{n})$
 - □ 90% CI: $\dot{X} \pm 1.645 (\sigma/\sqrt{n})$
 - □ 95% CI: $\dot{X} \pm 1.960(\sigma/\sqrt{n})$
 - □ 99% CI: $\dot{X} \pm 2.576(\sigma/\sqrt{n})$

Confidence Interval

- □ Interpretation: if we took the a sample from the population repeatedly, in [90, 95, 99, or whatever confidence level we specify] percent of the samples the true population parameter will fall within the range generated by the CI equation
- Or, we are X% sure that the population mean falls within the range of the Cl

CAUTION

Before we go on to examine hypothesis testing, there are several operative assumptions:

Your data is properly entered and cleaned

 You know the univariate and bivariate distributions of your data

T-tests

T-tests

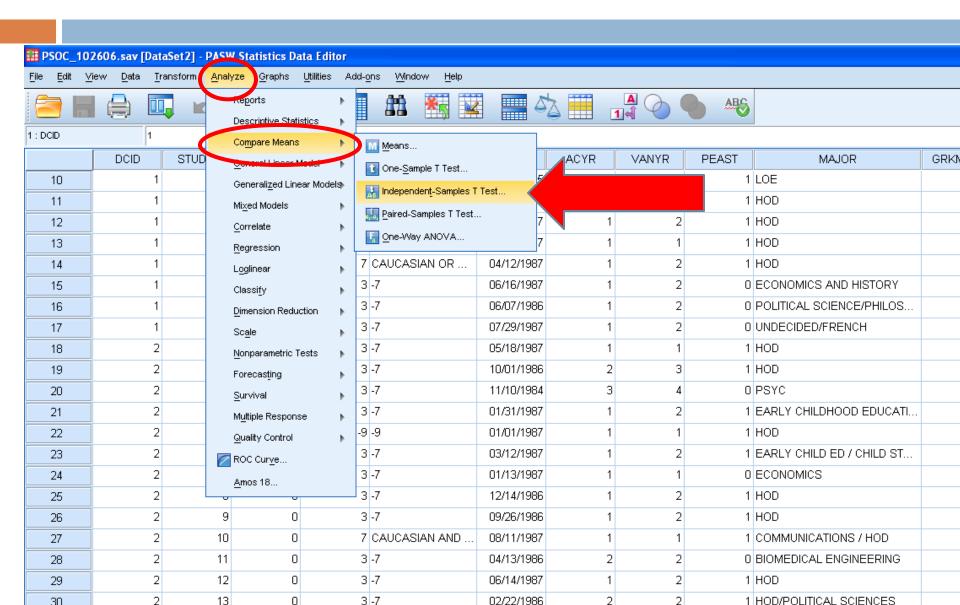
- When we have a continuous variable and we want to know if its mean differs in value between <u>two</u> groups, we use a *t*-test
 - It is extremely rare that the groups will have exactly the same mean, but the t-test compensates for the uncertainty involved in measuring a population using a sample—basically, it measures whether or not the group means are far enough apart that we can be confident that they come from different distributions
 - Example: Do male and female undergraduate students at Vanderbilt have different scores on the Satisfaction with Life scale?

T-test assumptions

- The dependent—continuous—variable
 (Satisfaction with Life) is normally distributed when
 it's examined on both categories
- 2. The dependent variable has a more-or-less equal variance / standard deviation when broken up into the two categories
- 3. The cases are independent from each other

T-tests are, however, reasonably robust even if these assumptions are violated

T-test in SPSS



TUDID	MALE	RACE	RACEOTH	BIRTHDT	ACYR	VANYR	PEAST	MAJOR	GRKMEM
10	0	3	-7	11/11/1985	2	3	•	1 LOE	0
11	1	3	3 CD	01/05/1987	1	2	•	1 HOD	1
12	1-1	est ³	in SP	74/6 71987	1	2	Г	1.00	0
13	0		-7	01/19/1987	1	1		I. Continuous	0
14	0	7	CAUCASIAN OR	04/12/1987	1	2	,	dependent)	0
15	1	3	-7	06/16/1987	1	2	V	/ariable(s) goes	1
16	1	3	-7	06/07/1986	1	2	_	nere	1
17	0	3	-7	07/29/1987	Т	2		ONDEOIDED/I NEMOIT	1
1	0	3	- Independent-S	amples T Test			X	HOD	1
2	1	3	-		<u>T</u> est ∀ariable	(s):	Options	HOD	1
3	0	3	- Data Collector ID	(D 📥	Ø SWLTOT		Options	PSYC	Ω
4	0	3	- Student ID [STUI					EARLY CHILDHO NB: NO	need to
5	0	-9	Race/ethnicity [F					HOD change	default
6	0	3	RaceOther [R/					EARLY CHILD ED options	
7	1	3	- Academic Year					ECONOMICS	1
8	0	3	- Number of years		Grouping Vari	iable.	-	HOD	0
9	0	3	Student at Peab	ody	MALE(??)			HOD	1
10	0	7	Member of Gree	k or	Define Group			COMMUNICATIONS / HOD	0
11	0	3			Panet Carry	el Hele		BIOMEDICAL ENGINEERING	0
12	0	3		OK Paste	Reset Cance	Help		HOD	0
13		3		02/22/1986	2	2	·	HOD/POLITICAL SCIENCES	0
14	2. C	Categorie	cal	03/03/1987	1			1 HOD	0
15	(ind	epende	nt)	01/29/1987	1	2	3	Click on define g	roups
16			es here	08/29/1986	1	2		I HOD	- I
1				09/20/1987	1	2	(Political Science	1
2	0	3	-7	06/20/1986	2	3	•	1 Human and Organizational Dev	1
3	0	3	-7	12/03/1986	1	1	•	1 Human and Organizational Dev	1
4	1	3	-7	01/23/1987	1	1	•	1 Economics	1

DCID	STUDID	MALE	RACE	RACI	EOTH	BIRTHDT	ACYR	VANYR		DEAGT	MAJOR	1
1	10	0	3	-7		11/11/1985	2		1.	Type	in values of	
1	11	1	3			01/05/1987	1		gro	oups y	you want to	
1	T_#	est	in	7 -		04/03/1987	1		tes	st (cor	nsult your	
1	13		3	-7		01/19/1987	1			•	ok if need	
1	14	0	7	CAUCASI	AN OR	04/12/1987	1		be			
1	15	1	3	-7		06/16/1987	1		DC	: /		TORY
1	16	1	3	-7		06/07/1986	1		2	0	POLITICAL SCIENCE/	PHILOS
1	17	0	3	-7		07/29/1987	1		2	0	UNDECIDED/FRENCH	ł
2	1	0	3	Inde	pendent-Sa	amples T Test				×	HOD	
2	2	1	3	-			<u>T</u> est Vari <mark>a</mark> ble(s):		ptions	HOD	
2	3	0	3	Dat	a Collector ID	[D 📥 🚃	C C			puoris	PSYC	
2	4	0	3		dent ID [STUD	נטות	fine Crut pa		X		EARLY CHILDHOOD I	EDUCATI
2	5	0	-9	-	ce/ethnicity [R ceOther [RA	(U); U	se specified val	ıes			HOD	
2	6	0	3		e of Birth [BIF		Group <u>1</u> : 1				EARLY CHILD ED / CI	HILD ST
2	7	1	3		ademic Year		Group 2: 0		/		ECONOMICS	
2	8	0	3		mber of years	(C) C	at point:		1		HOD	
2	9	0	3		ident at Peabo ior [MAJOR]	ody					HOD	
2	10	0	7		mber of Greel	k or	tinue Cance	Help			COMMUNICATIONS /	HOD
2	11	0	3	-			Parat Cara	L Llala	_		BIOMEDICAL ENGINE	ERING
2	12	0	3	-		OK Paste	Reset Cance	Help			HOD	
2	13	0	3	-7		02/22/1986	2		2	1	HOD/POLITICAL SCIE	NCES
2	14	1	1	-7		03/03/1987	1		1	1	HOD	
2	15	2 Proc	c Contir	2110		01/29/1987	1		2	1	HOD	
2	18	2. Pres		iue,		08/29/1986	1		2	1	HOD	
3	1	then Ok	•			09/20/1987	1		2	0	Political Science	
3	2	0	3	-7		06/20/1986	2		3	1	Human and Organizati	onal Dev
3	3	0	3	-7		12/03/1986	1		1	1	Human and Organizati	onal Dev
3	4	1	3	-7		01/23/1987	1		1	1	Economics	

T-test: Interpreting the output

Group Statistics

	Male gender	N	Mean	Std. Deviation	Std. Error Mean
SWLTOT	Male	46	3.8087	.59473	.08769
	Female	79	3.9456	.61092	.06873

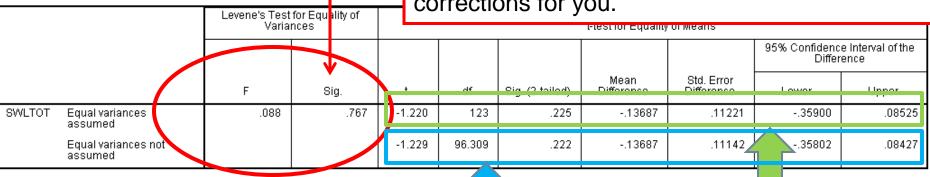
Independent Samples Test

		Levene's Test Varia	for Equality of nces	t-test for Equality of Means							
				95% Confidence Interval Difference							
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper	
SWLTOT	Equal variances assumed	.088	.767	-1.220	123	.225	13687	.11221	35900	.08525	
	Equal variances not assumed			-1.229	96.309	.222	13687	.11142	35802	.08427	

T-test: Interpreting the output

Group Statistics Male gender N Mean Std. Deviation SWLTOT Male 46 3.8087 .59473 Female 79 3.9456 .61092

This information tells you whether or not the second assumption of t-tests, equality of variances, has been violated. If the number under Sig is less than 0.05, there is no violation. If the Sig < .05, then there is a problem—BUT SPSS has automatically made corrections for you.



If the variances are not equal, use the second line If the variances are equal, use this first line

T-test: Interpreting the output

Group Statistics

	Male gender	N	Mean	Std. Deviation	Std. Error Mean
SWLTOT	Male	46	3.8087	.59473	.08769
	Female	79	3.9456	.61092	.06873

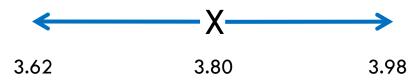
Independent Samples Test

		Levene's Test Varia	for Equality of nces	t-test for Equality of Means								
									95% Confidenc Differ			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Lower Upper		
SWLTOT	Equal variances assumed	.088	.767	-1.220	123	.225	13687	.11221	35900	.08525		
	Equal variances not assumed			-1.229	96.309	.222	13687	.11142	35802	.08427		

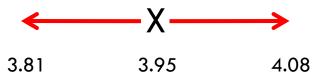
This significance result tells you the probability that we'd be making an error if we said there was a difference in the scores between male and female students. By convention, there needs to be less than a 5% (.05) chance of error if we are to declare that there is a difference. Alternately, we need to be 95% sure that there is a difference in the population means based on our sample means.

Understanding t-tests

Male



Female



T-tests and confidence intervals

If the value of one of the group means falls within the confidence interval (degree of uncertainty) of the other, then the risk of making an error when saying the two have different means is unacceptably high

- SPSS reports the confidence intervals of the mean difference (mean of group one minus mean of group two)
 - If there is no significant difference, the confidence interval will contain zero

T-tests: Reporting results

 Unless you have a whole bunch of t-tests, there is no need for a table

The means & standard deviations for both groups, t value, degrees of freedom, and p value all need to be reported.

Text: There was no difference in satisfaction with life scores in male and female students ($M_{male} = 3.81$, $SD_{male} = .59$, $M_{female} = 3.95$, $SD_{female} = .61$; t(123) = -1.22, p = .23)

ANOVA

ANOVA

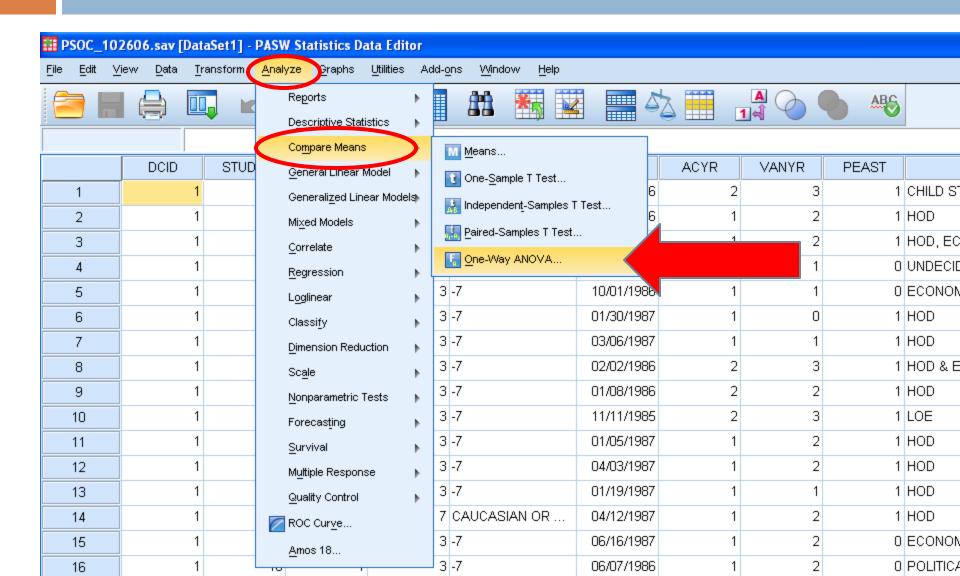
- When we have a continuous variable and we want to know if it differs in value between three or more groups, we use ANOVA (ANalysis Of VAriance)
- ANOVA asks whether the mean of all groups are equal to each other or not
 - Example: Do students at different racial groups at Vanderbilt University differ in Sense of Community (Emotional Connection)?

ANOVA assumptions

- The dependent variable (Satisfaction with Life) is normally distributed when it's examined individually for all categories
- The errors are normally distributed
- □ The cases are independent from each other

Like the t-test, ANOVA is robust in the face of relatively minor violations of these assumptions

ANOVA in SPSS



DCID	310010	MUNICL	ACL INCLO	III DIKI	1101 AC	/II VA	1411	LAUI	1917	20 N
1	1	0	3 -7	08/23	3/1986	2	3	1	CHILD STUDIES	S & HOD
1	2	0	7 MULTIRACIA	L 10/13	3/1986	1	2	1	HOD	
1	3		3-7	20/11	1/1986	1	2	1	HOD, ECONON	1ICS
1	A		4 In 31	PS 5/04	1/1987	1	1	4	2 (
1	5	1	3 -7	10/01	1/1986	1	1		Continuo	
1	6	0	3 -7	01/30	0/1987	1	0		pendent	
1	7	0	3 -7	03/08	6/1987	1	1	var	iable(s)	goes
1	8	1	3 -7	02/02	2/1986	2	3	her		
1	9	0	3 🚣	04.00	24000	2	2		пов	
1	10	0	3 III One-Way	y ANOVA				×	LOE	
1	11	1	3		Deper	ndent List:	Con	ntrasts	HOD	
1	12	1	3 ⊘ 17. If I c	ould live my	Ø F	PSOC Emotional C	onne		HOD	
1	13	0	3						HOD	
1	14	0	/ * .	leeds Fullfill	*			ptions	HOD	
1	15	1	2 *	Froup Memb Influence Su					ECONOMICS A	ND HISTORY
1	16	1	3 ⊘ PSOCTO						POLITICAL SCI	ENCE/PHILOS
1	17	0	3 🔗 UNIQUE		Factor	r:			UNDECIDED/FI	RENCH
2	1	0	3 SWLTO	T 💆	Ø F	Race/ethnicity [R/	ACE]		HOD	
2	2	1	3	OK E	este <u>R</u> eset	Cancel He	ql		HOD	
2	3	0	3						PSYC	
2	4	0) /	1 11 1	01/31	1/1987	1	2	1	EARLY CHILD	OOD EDUCATI
2	5	2. Variak		01/01	1/1987	1	1	3 CI	ick on op	ntions
2	6	defines t	he groups	U3/12	2/1987	1	2	J. OI	- The office	
2	7	goes her	e	01/13	3/1987	1	1	0	ECONOMICS	
2	8	0	3 -7	12/14	1/1986	1	2	1	HOD	
2	9	0	3 -7	09/28	6/1986	1	2	1	HOD	
2	10	0	7 CAUCASIAN	AND 08/11	1/1987	1	1	1	COMMUNICATI	ONS / HOD
2	11	0	3 -7	04/13	3/1986	2	2	0	BIOMEDICAL E	NGINEERING
	40		0.7	00.4	14007					

	STUDID	MALE	RACE	RACEOTH	BIRTHDT	ACYR	VANYR	PEAST	MAJOR	GR
1	1	0	3 -	7	08/23/1986	2	3	1	CHILD STUDIES & HOD	
1	3		\	MULTIRACIA	ns/1 <i>an</i> 986	1	2	1	HOD	
1	3	1110) V A	7 III 3 F	10/11/1386	1	2	1	HOD, ECONOMICS	
1	4	1	3 -	7	05/04/1987	1	1	0	UNDECIDED	
1	5	1	3 -	7	10/01/1986	1	1	0	ECONOMICS	
1	6	0	3 -	7	01/30/1987	1	0	1	HOD	
1	7	0	3 -	7	03/06/1987	1	1	1	HOD	
1	8	1	3 -	7	One-Way AN	OVA: Options	3	1	HOD & EAST ASIAN STUDIES	
1	9	0	3	One-Way ANO	-Statistics				HOD	
1	10	0	3	One-way ANO	Descriptive			×	LOE	
1	11	1	3	7	Fixeuranuma	ndom errects		Contrasts	HOD	
1	1. Ensi	ure that	3	77. If I could live		of variance test	e	Post Hoc	HOD	
1	these t	hree op	tions 🤘	AGE	Brown-rors	yırıe		Options	HOD	
1	are sel	•	7	PSOC Needs F	<u>W</u> elch				HOD	
1	are ser	Coloa	В	PSOC Influence					ECONOMICS AND HISTORY	
1	16	1	3		✓ Means plot		ノ		POLITICAL SCIENCE/PHILOS	
1	17	0	3		-Missing Values-				UNDECIDED/FRENCH	
2	1	0	3		Exclude case	es analysis by ar	nalysis		HOD	
2	2	1	3	1	© Exclude case	es listwise			HOD	
2	3	0	3		Continue	Cancel Hel	,		PSYC	
2		الاماء ٥٠٠	. 4!	7 L		, 101	2	1	EARLY CHILDHOOD EDUCATI	
2		lick Cor	•		01/01/1987	1	1	1	HOD	
2	then	OK on t	the		03/12/1987	1	2	1	EARLY CHILD ED / CHILD ST	
2	mair	NOVA	4	_	01/13/1987	1	1	0	ECONOMICS	
2	scre	en			12/14/1986	1	2	1	HOD	
2	9	U	J -	4	09/26/1986	1	2	1	HOD	
2	10	0	7 (CAUCASIAN AND	08/11/1987	1	1	1	COMMUNICATIONS / HOD	

ANOVA: Interpreting the Output

Descriptives

PSOC Emotional Connection Subscale

Descriptive statistics, including the confidence interval for each mean

					95% Co			
					Me	an		
	N	Mean	Std. Deviation	Std. Error	Lower Bound	Upper Bound	Minimum	Maximum
Africa-American, Black	3	4.4444	.19245	.11111	3.9664	4.9225	4.33	4.67
Asian-American, Asian, including Indian Subcontinent	3	3.0000	1.52753	.88192	7946	6.7946	1.67	4.67
Caucasian or white	110	4.2091	.43275	.04126	4.1273	4.2909	2.67	5.00
Mexican-American, Chicao/a	2	4.3333	.00000	.00000	4.3333	4.3333	4.33	4.33
Other (describe)	6	3.8333	.34960	.14272	3.4664	4.2002	3.33	4.33
Total	124	4.1694	.50202	.04508	4.0801	4.2586	1.67	5.00

Test of Homogeneity of Variances

PSOC Emotional Connection Subscale

Levene Statistic	df1	df2	Sig.	
7.873	4	119	.000	

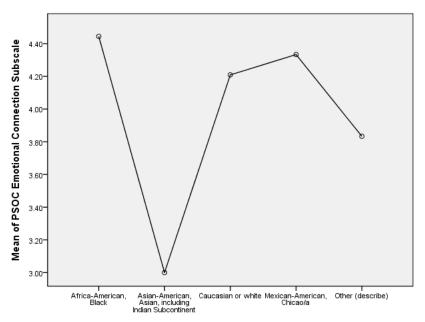
This information tells you whether or not the second assumption of ANOVA, equality of variances, has been violated. If the number under Sig is more than 0.05, there is no violation. If the Sig < .05, then there is a problem, especially if the number of cases in each group is unbalanced.

ANOVA: Interpreting the Output

ANOVA

PSOC Emotional Connection Subscale

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5.234	4	1.309	6.044	.000
Within Groups	25.765	119	.217		
Total	30.999	123			



This significance result tells you the probability that we'd be making an error if we said there was a difference in the mean scores of the different groups. Again, there (usually) needs to be less than a 5% (.05) chance of error if we are to declare that there is a difference.

ANOVA: A caution

- A positive result in an ANOVA (p < .05) only indicates that one or more of the means is not equal to the other means in the group. It does not tell us which group(s) is the one that doesn't belong.</p>
 - Sometimes the probable group can be identified just by looking at the descriptive statistics or means plot
 - Otherwise, there are post-hoc tests to help identify the group(s)

Writing up the results

Report means and standard deviations for all groups, as well as the F value, degrees of freedom, and p value.

□ Text for this example \rightarrow Analysis of variance indicated that the different racial groups present in our sample report unequal levels of Sense of Community: F(4, 119) = 6.04, p < .001.

Regression

Caution: If you're planning to use a regression in your analysis, Josh will help you through the process. It's a lot more complicated than we're showing here.

Regression

- Regression involves using multiple known characteristics of a dataset to try and build a mathematical model which predicts an outcome variable (dependent) using predictor (independent) variables
- The resulting model is then tested against the real observations, and the degree to which the predicted values of the dependent variable matches the real observations is called the model fit
- The advantage of regression is that it allows us to measure the relative, independent contributions of multiple predictors to the outcome

Regression Types

- The type of regression one uses depends on the nature of the outcome variable:
 - Ordinary Least Squares (OLS): when predicting a continuous, normally distributed outcome variable
 - Logistic: when predicting a dichotomous outcome
 - Poisson/Negative Binomial: when predicting a count variable
 - Many, many more

Regression: Predictor variables

- Continuous variables should be normally distributed
 - Significantly skewed variables need to be transformed
- Dichotomous categorical variables can be entered straight into the model, as long as it is coded zero and one; other codings need to be transformed

 Categorical variables with more than two values need to be transformed into a series of dummy variables

Regression assumptions

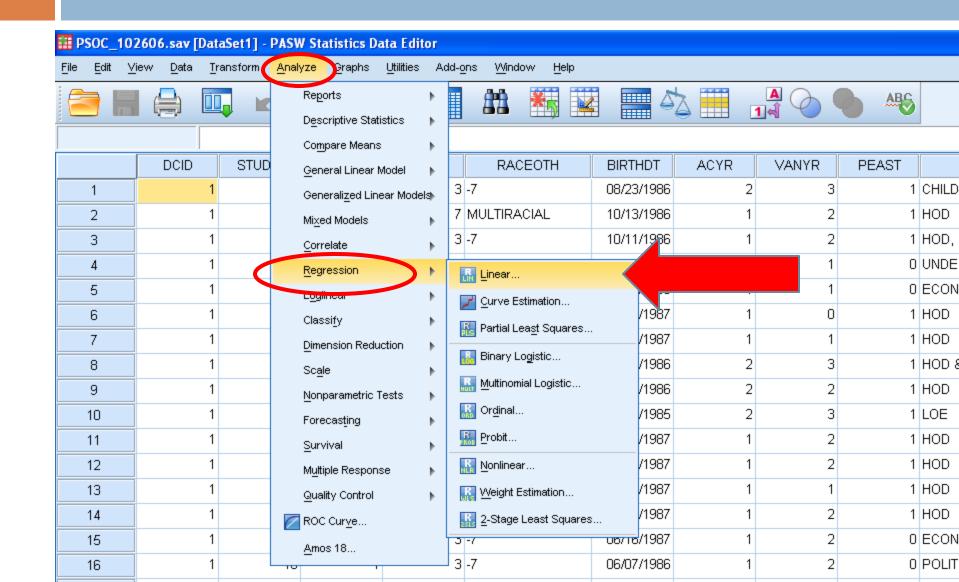
 You've chosen the right kind of regression to suit the dependent variable

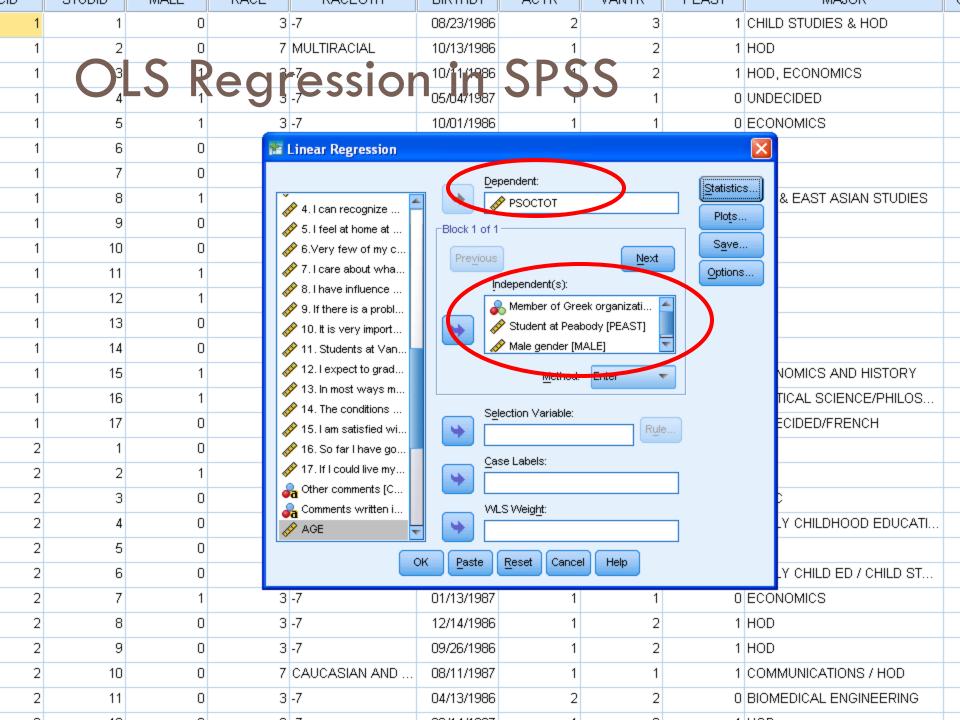
 There is a more-or-less linear relationship between each predictor and the outcome variable

 None of the predictor variables are too highly correlated with each other

■ Many, many more

OLS Regression in SPSS





Interpreting the output

Coefficients^a

		Unstandardize	d Coefficients	Standardized Coefficients			95.0% Confider	ice Interval for B
Model		В	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	3.614	.097		37.358	.000	3.423	3.806
	Member of Greek organization at Vanderpilt	.223	.072	.269	3.09	.002	.080	.365
	Student at Peabody	.053	.084	.058	.631	.529	113	.218
	Male gender	076	.075	093	-1.007	.316	224	.073

a. Dependent Variable: PSO<mark>CTOT</mark>

The constant or intercept represents the value of the outcome if all the predictor variables had zero values (which may or may not be meaningful).

The significance level indicates whether that variable is or is not a statistically significant predictor of the outcome. Again, we generally need to be 95% certain (p < .05) that we're not making an error when we declare something to be significant.

Interpreting the output

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	
1	.300ª	.090	.068	.38145	

a. Predictors: (Constant), Male gender, Member of Greek organization at Vanderbilt, Student at Peabody

b. Dependent Variable: PSOCTOT

Coefficients^a

		Upstanus rdized Coefficients		Standardized Coefficients				95.0% Confidence Interval for B		
Mode	el		В	Std. Error		Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)		3.614	.097	7		37.358	.000	3.423	3.806
	Member of Greek organization at Vanderbilt	4	.223	.072		.269	3.096	.002	.080	.365
	Student at Peabody	\ /	.053	.084	\	.058	.631	.529	113	.218
	Male gender	V	076	.075	\	093	-1.097	.316	224	.073

a. Dependent Variable: PSOCTO

The unstandardized beta represents the coefficients of our regression equation which uses the independent variables to predict the dependent variable

The standardized beta represents the relative contribution of each predictor variable in influencing the outcome

R-squared is the model fit

index, which ranges from 0 to

1, with values closer to one

indicating a better fit.

Understanding what it's all about

The results tell us that the best linear model for predicting PSOC is as follows:

$$PSOC = 3.61 + .22(GRK) + .05(PBDY)^* - .09(MALE)^*$$

 \square But the model fit is quite poor (R² = .09)

* Non-significant coefficients and variables can be deleted from the final model

Interpreting the results

If all the variables were significant, a table like this is one way to make the results more meaningful:

Greek member	N	lo	Yes		
Peabody student	No	Yes	No	Yes	
Female	3.61	3.66	3.83	3.92	
Male	3.52	3.57	3.74	3.79	

Writing it up

- Tables are very appropriate when presenting regression results:
 - http://owl.english.purdue.edu/owl/resource/560/19/
- Don't include any equations
- Include at minimum the intercept, all unstandardized betas, an indication of whether each beta was significant, and the model fit

Table One Student PSOC Regression Results

	В	SE(B)	β	B CI
Intercept	3.61**	.10		[3.42, 3.81]
Greek member	.22**	.07	.27	[.08, .37]
Peabody student	.05	.08	.06	[11, .22]
Male	08	.08	09	[22, .08]

^{*}Significant at p < .05 **Significant at p < .01