OLS Regression and ANOVA in Excel 2013

Note: Other versions of excel generally have these same functions, but the menus/paths may differ

Note: Regression and ANOVA is appropriate if you are working with two continuous variables (e.g., 1.1, 1.2, 1.4, 2.7). Discrete numerical categories (e.g., 1, 2, 3, 4) can sometimes be treated as continuous. ANOVA is also appropriate if you have a categorical independent (e.g., yes/no coded as 1, 2) and a continuous dependent variable. Neither is appropriate if you the independent variable is continuous.

(1) **Input data into excel.** Make sure data are formatted as numbers with no additional information in the cells:

F1	Ste En Copy + Ste Formal F Clipboard	B I	<u>u</u> - ⊞ - . Font	<mark>⊘-∆</mark> -	
F1	~ 0.	XV	f.		
4			Ja		
	A	В	C	U	
1	Independant Variable	Dependant Variable 1	Dependant Variable 2		
2	2	10	164.5		
3	1.L	15	87.8		
4	2.3	11	39.7		
5	5.2	15.3	183.7		
6	2.9	11.6	58.7		
7	5.4	15.8	174.1		
8	3.5	12.4	78		
9	4.6	15.8	20.5		
10	3.6	13	97.6		
11	3.9	13.3	105.9		
12	4	14	153.7		
13	3.2	12	125.1		
14	4.2	14.8	135.3		
15	3.4	12.7	144.5		
16	4.8	15.6	30.1		
17	5	16	68.2		
18	2.2	10.5	49.2		
19	2.6	11.5	115.5		

(2) Make sure you have the Data Analysis Toolpak. Go to the DATA tab, and select the Data Analysis Toolpak (on the right). If you do not see that option, do a google search for "Data Analysis Toolpak in Excel" to find the instructions on how to add it (for free)

E	B 🗄 🤊 🕯	e - 🕯 - +						Dook1	- Lacoel								? 🖂 🗕 🗆
4	FILE HOME	INSERT.	PAGE LAYOUT	FORMUL	AS 1	DATA REV	IEW VEW /	DD-INS NI	TRC PRO								👍 Marisse Rams er 🕤
	à From Access From Web From Text s Get D	rom Other Sources * Con dernal Data	assting Refresh nections AI *	Connect Propert	tions irs ks	21 ZZ ZI Sort	Thiter	Lect to I Columns	lash terr Fil Dupli	ove Der cotes Validat D	z Conso ion * atz Tools	loste Wha Analy	₹ t-l+ lelation sis*	nships Cro	aup Ungro	up Subtotal	+2 Cola Analysis
-	15 -	: × ~	fx c	a	F	F	5 н	ſ	1	к	E E	м	М	0	P	Data Ana Loois for analysis.	hysis Tools financial and scientific data
1	Independant Variable	Dependant Variable 1	Dependant Variable 2													FUNK Tell o	CRES ne more
2	41	10	04.5			-					-			-			
4	7.8	11	39.7														
5	5.2	15.3	183.7														
8	2.9	11.0	58.7														
7	5.4	15.8	174.1														
8	3.5	12.4	78														
9	1.5	15.8	3 20.5														
14	9 <u>1.</u> 0	1:	97.6														
11	1.7	11.1	106.9														
12	4	14	153.7														
13	3.2	. 13	126.1														
14	4.2	. 14.:	135.3														
1.	3.4	12.5	144.5								1				-		
14	4.8	15.4	30.1											-	-		

Regression & ANOVA In Excel

(3) In the Data Analysis popup window, scroll down to select "Regression" and then "OK"

KE - 5- 1	er 📲 🔹 🕫							Boo	k1 Excel	
FILE HOME	INSERT	PAGE LAYOUT	FORMUL	AS DA	TA R	EVIEW	VIEW AD	D-INS	NITRO PRO	
☐ From Access ☐ From Web ☐ From Text Get Er	orn Other E Sources - Con	kisting nections Refresh All -	Connections	ies 4	Sort	Filter	T _{ic} Clear T _{ic} Reapply Ty Advanced	Text to Columns	Flash Rem Fill Duplie	ove D ates Valid
K16 *	: × 🗸	fx .								
A	В	С	D	E	F	1.0	G H	1	J	К
Independant 1 Variable	Dependant Variable 1	Dependant Variable 2								
2 2	10	154.5				1	Data Analysis		? ×	
3 4.1	15	87.8		<u>Analysis</u>	Tools					
4 2.3	11	39.7		Histogr	om				OK	
5 5.2	16.3	183.7		Moving	Average	Concenti			Cancel	
6 2.9	11.6	58.7		Refik dr	id Percent	ile	011		Unin	
7 5.4	16.8	174.1		Regress	ion				liep	
0 3.5	12.4	78		t-lest: H	aired Iwo	Sample	or Means			
9 4.6	15.8	20.5		t Test: T	wo Sampl	e Assum	ng Equal Variant	es		
10 3.6	13	97.6		7-Test T	wo Sampl	e for Me	ans	v V		
11 3.9	13.3	106.9								
12 4	14	153.7								
12 20	12	126.1								

(4) Put your mouse in the box for "Input Y Range" and then drag a box around one of your dependent variables. Then put your mouse in "Input X Range" and drag a box around your independent variable. Then click "OK"

	HOME	INSERT	PAGE LAYOUT	FORM	IULAS DATA	REVIEW	VIEW AD	D-INS	N TRO PRO		
	From Access From Web From Text 5 Get Ex	om Other E Sources - Con ternal Data	xisting nections All -	S Con Prop E Fdit	nertions 21 perties Links A4	Sort & Filter	K Clear Reapply Advanced	Text to Columns	Flash Remove	Data S Validation Data	Conso id Tools
AZ	•	: X 🗸	f_x				Regression		? ×	1	
il	A	В	с	D	Input					K	L
1	Independant Variable	Dependant Variable 1	Dependant Variable 2		Input <u>Y</u> Rang	je:	\$E\$2:\$E\$19		Cancel		
2	2	10	164.5			je:	\$4\$2:\$4\$19	1.81			
3	4.1	15	87.8		Labels] Constant is <u>Z</u> ero		<u>II</u> elp		
4	2.3	11	. 39.7		Confider	ice Level: 93	5 %				
5	5.2	16.3	183.7		Output optio						
6	2.9	11.6	58.7			113	-	FK			
7	5.4	16.8	174.1		Output	ange:	-	Hita			
8	3.5	12.1	78		New Wo	rksheet <u>P</u> ly:	1				
9	4.6	15.8	20.5		O New Wo	rkbock					
10	3.6	13	97.6		Residuals		Residual	Plots		-	
П	3.9	13.3	106.9		Slandard	ized Residuals	Line Fil Pl	ots			-
12	4	14	153.7		Duran d Prak						
13	3.2	12	126.1		Normal P	robability Plots					
14	4.2	14.8	135.3								
15	3.4	12.7	144.5								-
16	4.8	15.0	30,1								
16	2	10	08.2								
18	2.2	10.5	49.2								
14	7.h	11.5	6.011						-		
20											

(5) You should see the output of the regression analysis pop up in a new worksheet. It will look like the output below. Expand the width of column A if needed so you can see the text. (note: I added the blue highlighting)

X		≠ ≑ ERI PAGELA	VOUT FORML	JEAS D7	VIA RE	VIEW VIEW	NDD- NS	BookT - Excel NITRO PRO		
1	the cut	Calibri	- 11 - A A	; = =	- %-	'∰ Wrap Text	Ge	eneral	-	
Fa	iste • 🎺 Format Painter	в <u>т</u> т я	🗉 • 🔷 • 🔺	• = =		= 🖽 Merge & (Center 🔹 S	- % • 50	Condition Formatting	al Formata Table -
	Clipboard I#	F	ont	8	All	gnment	lat .	Number	2	Styles
G	21 ~ ! 🔉	$\langle \sqrt{f_x} \rangle$								
	A	Б	С	D	E	F	G	н	1	J
1	SUMMARY OUTPUT				1.1					
2										
3	Regression St	atistics								
1	Multiple R	0.985395735								
5	R Square	0.971004755								
6	Adjusted R Square	0.969192552								
7	Standard Error	0.376793048								
З	Observations	18	ļ							
9										
10	ANOVA									
11		df	55	M.S	F	Significance F				
12	Regression	1	75.07120976	76.07121	535.8146	9.93995E-14				
13	Residual	16	2.271568017	0.141973						
14	Lotal	1/	/8.342////8							
15	82									
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
17	Intercept	5.969885774	0.335592351	17.7891	5.78E-12	5.25846177	6.53130977	7 5.25846177	6.681309777	
18	X Variable 1	2.015576324	0.087074596	23.14767	9.94E-14	1.830986214	2.20016643	4 1.830986214	2.200166434	

(6) We are particularly interested in five items on the output (highlighted above):

- a. The <u>R Square</u>. This is how linear the relationship between the variables are. The closer to 1, the more linear the relationship, and the more variation in Y (dependent variable) that is explained by X (independent variable). In the example below, the R Square is 0.97 (round to 2 decimal places). This means the relationship is highly linear ("a good fit"), and 97% of variation in X is explained by Y. Remember that this does NOT reflect causation (we can't know from this analysis if one variable causes variation in the other), but instead just shows there is a relationship.
- b. Excel runs an "Analysis of Variance" (ANOVA) along with the OLS regression.
 - i. The F is a measure of the strength of the relationship. In the example, the F is 535.81, which is good. Don't worry too much about the F, just write it down.
 - ii. The Significance F is the P-Value. This tells us how significant the relationship is. In the example, the P-Value is very low, 9.39 E-14, which is excel's way of using scientific notation. This means 9.39 x 10⁻¹⁴, or 0.000000000000939. It is excessive to report all this though. The convention is that the relationship is "statistically significant" if P<0.05, and "highly statistically significant" if P<0.01. So here you can simply record P<0.01. If you want to show it is very very good, just report one more zero (P<0.001).
- c. The coefficients give you the formula of your regression line. The formula of a line is: y = mx+b, where m is the slope and b is the y-intercept. In the excel output, the slope is the <u>X Variable 1 coefficient</u>, and the y-intercept is the <u>Intercept Coefficient</u>. So for the example, y = 2.0156x + 5.9699. I have rounded to 4 places, but 2 is usually OK also. This is our predictor equation. So, if we wanted to predict variable Y with a value of X, we would simply insert the X value into the equation and get Y.

- (7) This is how you would state your results in a paper. There was a statistically significant relationship between (insert name of one variable) and (insert name of the other variable) (R²=0.97, ANOVA F=535.81, P<0.001).</p>
- (8) Now, let's represent the data graphically. Go back to your data, and click the INSERT tab up top, them charts>scatterchart

X	E HOME	c≥ - 🔏 + = INSERT	PAGE LA	AYOU⊤ F	ORMJ	LAS D.	A⊤A REVI	EW	VIEW AD	B D INS	ook I - Exce NITRO PI	20		
Fiv	otTable Recomm Pivo.T Tables	nendec Table	Pictures	Online Pictures	• •	Store My Apps	<mark>▶</mark> Bing M ▼ 📑 People Add-ins	aps Graph	Recommend Charls	led 📣 🗸	Lev Scatter	FivotChart	Line	Colui
C	art 1 T	: X	fr								U0	See 9		
	A	. в	JA (c	D	E	F	G	н	1.1		<u>M</u>	0	L
1	Independant Variable	Dependant Variable 1	Depend Variabl	dant le 2							1255	X	S	
2	2	10		164.5							DALL			
3	4.1	15		87.8							Bubble			
4	2.3	11		39.7			C					0.	- 3	
5	5.2	16.3		183.7							Qõ	00		
b	2.9	11.6		58.7							11. 1.1.	Contras Chara		
7	5.4	16.8		174.1								e scatter Chan	Esar	
8	3.5	12.4		78										
y	4.6	15.8		20.5										
10	3.6	13		97.6										

(9) Right click on the chart area and "Select Data"



(10)Put your mouse in "Chart Data Range" and then draw a box around both columns of data (they will need to be next to each other).



(11) Your data should now be in the chart:





(12) Adjust your chart axes – double clock on the numbers in the axes, and a dialog box should pop up on the right. Change the min and max values to bracket your data, then close the dialog box.

(13) Right click on the data points (the actual dots in the chart) and "Add Trendline"



(14)In the trendline dialog box that pops up on the right, select linear (if your data are linear), and scroll down to check the boxes for "Display equation..." and "Display R-squared...", then close the dialog box.



- (15)The equation in the trendline should match the equation achieved with your OLS. If not, check to make sure you have your columns organizes so that the independent variable is to the left of the dependent, then redo the chart. Note: If you change any of the data, you will need to delete and re-add the trendline, as the equation of the line and the R² will not automatically update.
- (16) Now, move your columns around so that your next dependent variable is next to the independent variable, and repeat all steps. For this analysis (see below), the relationship was not significant (R2=0.02, ANOVA F=0.36, P=0.56). Note that since the P was not <0.01, I am going to report the actual value, which you would do for any value >0.01, such as 0.02, 0.04, 0.12...).

S 8 5 C-	Q = <i>q</i>					Excel Basi	cs - Excel		
HLE HOME	INSERT PAG	ELAYOUT FOR	MULAS DA	IA REVIEW	VIEW A	DD INS NIL	ROTRO		
Paste	Calibri B I U	- 1 - A - - - -		● »· ;	ि Wrap Text च Merge & Cente	General r = \$ = 9	- • • • 0.00 • • 00	Conditional Fo	imat a
Cinhoard		Foot	5	Alicinine	et	rs M	umber 5	Formatting *	ab e 🗸
Cipicouru		Tenc	2(4)	Angeline		G B	amoer 3	54	103
E6 T	$\times \checkmark f$	r i							
A 0	В	с	D	E	F	G	н	1	J
1 SUMMARY OUTPU	л								
2									
3 Regression	Statistics								
4 Multiple R	0.14849239								
5 R Square	0.02204999				L				
6 Adjusted R Square	e -0.039071886								
7 Standard Error	52.13994822								
8 Observations	18								
9									
10 ANOVA									
11	df	55	MS	F	Significance F				
12 Regression	1	980.7377957	980.7377957	0.36075447	0.55650073				
13 Residual	16	43497.1872	2718.5742						
14 Total	17	44477.925							
15									
16	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
17 Intercept	75.05205162	46.43866944	1.615154221	0.125504522	-23.3935298	173.4976331	-23.3935298	173.4976331	
10 X Variable 1	7.237116155	12.04924076	0.600628396	0.55550073	-18.30613319	32.7803655	-13.30613319	32.7303655	
19									

X	5.	et e 🔒 e 💷				Excel Bas	sics Excel				CHAI	el o or tr			
F	LE HOME	INSERT	PAGE LAYOUT	FORMU	LAS	DATA	REVIEW	VIEW ADI	D-INS INI	TRC PRO	DESIGN	FORMA	т		
	From Access From Web From Test 5 Get Ea	ion Conce – Bit ion Conce – Bit ion Conce ternal Data	sting ections All 7	Le Conn E Prope	ections mics nics	21 ZIA 21 Sort	Filter	To Clear To Reapply To Advanced er	Text to F	ash Remov	e Dat tes Validat Da	a Conse ion - rta Tools	⊧≡ Ilidate	What f	Rc atio
ch	iart 1 🔹	$\times \times \vee$	fx												
4	A Independant	8 Dependant	C Dependant	D	E	F	G	Н	1	1	K	L	A	A	N
1	Variable	Variable 2	Variable 1												
2	2	164.5	10												
3	4.1	87.8	15	¢					10						
4	2.3	39.7	11					Ch	art Title					+	
5	5.2	183.7	16.3		200										
6	2.9	58.7	11.5		205									1	
1	5.4	1 /4.1	16.4		185									-	
8	3.5	78	12.4		168									11	
9	4.6	20.5	15.8		148					1 100					
10	i.h	97.5	1.1		128			100							
11	3.9	105.9	13.3		108					· · · · ·		02		IL	
12	4	153.7	14		22										
13	1.7	125.1	17		CD										
14	4.2	135.3	14.3		100									L	
15	3.4	144.5	12.7		48										
In I	4.8	30.1	15.6		28										
10		68.2	10		8	5913		900 30	1	10.2	100	-	- 1		
18	2.2	19.2	10.5		1	1.5	2	2.5 3	3.5	4 4.5	5	5.5	G		
1	×.0	110.5	11.5												

(17)There are various other things that can happen with your data. For example, there may be an exponential relationship between the data (common in growth studies), in which case an exponential trend line would need to be fit:





Chart Title

y = 0.1146e^{2.19006}

R2 - 0.008

9

(18) You cannot run an OLS with exponential data and get a correct P value for the relationship, unless you first log transform the data. For example:



1.71+09

1E+09

800000008

600000008

100000 11.5129255

5000000 15,4249485

5000000 17.72/5336

10000000 18.4206807

15000000 18.3261459

30000000 19.519293

50000000 20.0301187

100000000 20.7232658

4

5

6

7

8

9

10

12

5

7.5

8.5

9.25

9.5

9.75

10

9

10

11

X		°- % -∓								Freel	Rasirs-	Fxcel				
	HOME	INSERT	PAGE LAYOUT	FORMUL	AS DAT	A	REVIEW	VIEW	ADD-1	NS	NITRO	PEO				
	From Access From Web From Text Set D	rom Other Exis sources Conne cernal Data	sting ections All -	Connections	tions 4↓ ies 2↓	Z A A Z Sort	Filter	Clear No Reapply No Advance ther	ed Co	Fext to olumns	Flash Fill	Remove Duplicates	Data Validation Cata	Consoli • Fools	idate Wł Ana	at-If Relati
02	4		J.	D	F					Regre	ssion		?	x	м	N
1	Independant Variable	Dependant Variable 1	Dependant Variable 1				Input Input <u>Y</u> R	ange:		SC\$2	SCSTT	1	O Car	K		i v
2	4	1000	6.90775528			1	Input <u>X</u> P	lange:		SA\$2	SAS11	1				
3	5	10000	9.21034037				<u>L</u> abel	s		Constar	it is <u>Z</u> ero		H	lp		
4	75	500000	15 4246495				Con <u>f</u>	idence Level:	95	%						
6	2.5	5000000	17 7275226		1.2E+09	6 - -	Output or	tions						2		
7	9	100000000	18 4206807				Outo	ut Ranner				1				
8	9.25	150000000	18.8261459		1E+09	(<u>-</u>	() New)	Worksheet Pl	by:							
9	9.5	300000000	19.519293				O Naw 1	Varkback	<i>.</i>							
10	9.75	500000000	20.0301187		80000008		Residual	(1		
11	10	1000000000	20.7232658				Resid	uals			Residual	Plots				
12				-	60000008		Stanc	lardized Resi	duals		ine Fit P	lots				
13							Normal P	robability								
14					40000008		Norm	al Probability	Plots							
15					25222222									_		
16					ZURIOORR									-		
17													least T			
18						3	4	5	6		7	в	9	10	11	
19																
20																
21																

Select the Ln (Natural log) transformed data for your Y variable:

Now, you can use the OLS output R Square and ANOVA F and P (significance), but use the equation of the line from the scatterplot of the non-log transformed data.

FLE	HOME IN:	¥ ∓ SERT PAGE	LAYOUT FORM	IJLAS DATA	REVIEW	VIEW ADD	Excel Basics - -INS NITRC	Excel PRO	
-	K Cut	Calibri	- 11 - A	A' = = =	87 - E	Wrap Text	General	-	
Paste	· 『의 Copy マ 《 Format Painter Clipboard 대	B <i>I</i> U -	🖂 + 🏠 + 🛕 Fort	• = = = 5	Alignment	Merge & Center	- \$ - %	+.0 .00 nn →n Fer ⊑	Conditional Format Formatting * Table Styles
E7	- 1 2	< 🗸 fx							
4 1 SU	A JMMARY OUTPUT	В	С	D	E	F	G	н	I
2									
3	Regression St	atistics							
4 M	lultiple R	0.999011882							
5 R 3	Square	0.998024741							
6 Ac	djusted R Square	0.997777833							
7 St	andard Error	0.231317858							
8 01	bservations	10							
9									
10 AI	NCVA								
11		df	<i>SS</i>	MS	Г	Significance F			
12 Re	egression	1	216.2845599	216.2845599	4042.10129	4.1658E-12			
13 Re	esidual	8	0.428063612	0.053507952					
14 TO	otal	9	216.7126235						
15									
16		Coefficients	Standard Error	lStat	P-value	Lower 95%	Upper 95%	Luwer 95.0%	Upper 95.0%
17 In	tercept	-2.166590655	0.292368672	-7.410474737	7.54359E-05	-2.840794021	-1.492387239	-2.84079402	1 -1.492387289
18 X	Variable 1	2.292597628	0.036059877	63.5775219	4.1658E-12	2.209443403	2.375751854	2.20044340	3 2.375751854
19									
20									

Regression & ANOVA In Excel

by M. Ramsier, HSU, 4-14-