Valeria Salinas-Lopez

Fall 2018

Nov 16, 2018

SOC 3305

Lab 6

(Due in by 11:59 pm on Nov 29, 2018)

Please save your final work as lastname_lab6.doc (put your name on the title page) and email it to instructor. Please explain your results with your own words. (You must use Stata program to do this lab). Provide your Stata command for each question.

1. a. Please explain with your OWN words: F statistics, t statistics, and linear regression.

F-statistic: This is the statistic that you get once you conduct an F-test. An F-test tests the null hypothesis that all the coefficients of the independent variables in a regression model are equal to zero, meaning they have no effect on the dependent variable. If the F-statistic is less than or equal to .05, then you reject the null hypothesis that all the independent variable's coefficients are equal to zero, meaning they do have an impact on the dependent variable.
t-statistic: This is the statistic you get once you conduct a t-test. A t-test tests the null hypothesis that a particular independent variable's coefficient is equal to zero, meaning it does not have an impact on the dependent variable. If the t-statistic is less than or equal to .05 then you reject the null hypothesis, which means that the independent variable does have an impact on the dependent variable.
Linear Regression: This is done whenever you want to see the impact of one or more independent variables on a dependent variable. This model fives you both the F-statistic and the t-statistics of each independent variable.

b. Please explain the difference between F-statistics and t-statistics. Can both statistics be ever same? Why? (10)

The main difference between the two statistics is that the F-statistic tests all of the coefficients of the independent variables, with exemption of the constant, where as the t-statistic only tests one independent variable. There is only one way that both the F-statistic and the t-statistic can be the same. This is when you run a linear regression model with only one independent variable. This happens because the F-statistic tests all independent variables, and since there is only one independent variable, you will get the same result in the t-statistic because that is where you test only one independent variable.

2. Use Nations2.dta. See how life expectancy predicts average school years, GDP, child mortality and adolescence fertility to the model by using linear regression.

Command: regress life school gdp chldmort adfert

Source	SS	df	MS	Number of obs	=	178
				F(4, 173)	=	337.70
Model	15635.6868	4	3908.92171	Prob > F	=	0.0000
Residual	2002.50296	173	11.5751616	R-squared	=	0.8865
				Adj R-squared	=	0.8838
Total	17638.1898	177	99.6507898	Root MSE	=	3.4022

life	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
school	-.1442219	.1438604	-1.00	0.317	-.4281693	.1397256
gdp	.0001338	.000023	5.81	0.000	.0000884	.0001793
chldmort	-.1652258	.0091747	-18.01	0.000	-.1833346	-.147117
adfert	.0041719	.0095542	0.44	0.663	-.0146858	.0230297
_cons	75.70404	1.399551	54.09	0.000	72.94164	78.46643

a. Report correlation and pairwise correlation (with .05 level of significance) tables and explain the difference if any.

Command for Correlation: correlate life school gdp chldmort adfert

	life	school	gdp	chldmort	adfert
life	1.0000				
school	0.7313	1.0000			
gdp	0.6062	0.5717	1.0000		
chldmort	-0.9294	-0.7724	-0.5160	1.0000	
adfert	-0.7424	-0.6798	-0.5121	0.7888	1.0000

Command for Pairwise Correlation: pwcorr life school gdp chldmort adfert

	life	school	gdp	chldmort	adfert
life	1.0000				
school	0.7252	1.0000			
gdp	0.6112	0.5733	1.0000		
chldmort	-0.9236	-0.7727	-0.5160	1.0000	
adfert	-0.7318	-0.6752	-0.5171	0.7774	1.0000

Interpretation: There are a few changes in some of the correlation coefficients, however they are not great differences that change the strength of the correlation.

b. Write formula and explain your result (F-statistics, t-statistics and significance levels).

(Predicted Y)Yi=β0+β1x1+β2x2+β3x3+β4x4+e

Predicted Life= 75.7 + (-.144)(school) + (.0001)(gdp) + (-.165)(chldmort) + (.004)(adfert) + e

F-statistic: Since the F-statistic is less than 0.05, the null hypothesis is rejected, which means that all of the coefficients of the independent variables are not equal to zero, which then means they impact the dependent variable.

t-statistic: GDP and child mortality are the only independent variables that have a significant impact on the dependent variable, because the t-statistic is less than 0.05 for both. The other independent variables are greater than 0.05 so they have no significant impact on the dependent variable.

c. Transform gpd with natural logarithm ("ln" function) and run the regression with it.

Command to Transform: generate loggdp= ln(gdp)

Command for Regression: regress life school gdp chldmort adfert loggdp

Source	SS	df	MS	Number of obs	=	178
				F(5, 172)	=	269.62
Model	15642.4196	5	3128.48391	Prob > F	=	0.0000
Residual	1995.77024	172	11.6033153	R-squared	=	0.8868
				Adj R-squared	=	0.8836
Total	17638.1898	177	99.6507898	Root MSE	=	3.4064

life	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
school	-.1847652	.1535547	-1.20	0.231	-.4878595	.118329
gdp	.0001106	.0000382	2.89	0.004	.0000352	.000186
chldmort	-.1616093	.0103403	-15.63	0.000	-.1820195	-.1411991
adfert	.0036516	.0095901	0.38	0.704	-.0152779	.0225811
loggdp	.4367957	.5734215	0.76	0.447	-.6950536	1.568645
_cons	72.34118	4.631775	15.62	0.000	63.19874	81.48362

d. Come up with the reduced model to predict schooling and report the formula with coefficients. (30 pts)

3. **Use shuttle.dta and generate new variable "dist" that shows the distress level and make "distress" dummy (binary, dichotomous) variable. Then run logit regression and report coefficients as well as likelihood of having any distress as increase on date by a day. Would you let Challenger take off? Why? (20)**
   Commands to Make it Binary: generate dist=distress and **replace dist=1 if distress==2**
   Command to Run Regression: logit dist date

```
Logistic regression                               Number of obs     =         23
                                                  LR chi2(1)        =       4.81
                                                  Prob > chi2       =     0.0283
Log likelihood = -12.991096                       Pseudo R2         =     0.1561
```

dist	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
date	.0020907	.0010703	1.95	0.051	-6.94e-06	.0041884
_cons	-18.13116	9.517253	-1.91	0.057	-36.78463	.5223142

Command to Determine Distress: display exp(.002907) and **display 1.0029112-1**

These give me the number .0029*100= 0.29%

Interpretation: As a day increases, there is a 0.29% increase in distress, therefore, it is relatively safe to send the Challenger off because there is little distress added each day.

4. a. With your OWN words please explain logistic regression and multinomial logistic regression.

A **logistic regression** is the same as a linear regression, with a small difference. A logistic regression model shows the impact of one or more independent variables on one dependent variable, which is considered to be binary.

A **Multinomial Logistic Regression** is when you do a logistic regression, but when there are more than 2 options. So if you want to run a model to figure out how political affiliation impacts belief in global warming, you would create dummy variables, which there would be more than 2.

b. What is the difference between "logit" and "logistic" Stata command? (10)

In Stata, the **logit** command runs a regression for binary variables, and reports the independent variables as odd-coefficients whereas the **logistic** command reports the independent variables as odd-ratios.

5. **Use Granit2011_6.dta. Generate new dummy variable called "lessice" from "warmice". 0 for the ones who believe less ice on the Artic compared to last 30 years and 1 for all others.**

Commands to Create Dummy Variables: generate less=0 and **replace less=1 if warmice!=1**

a. **Run a logistic regression by using these variables: age, sex, educ, party and warmop2.**

Command: logit less age sex educ party warmop2

```
Iteration 0:   log likelihood = -286.18868
Iteration 1:   log likelihood = -259.15094
Iteration 2:   log likelihood =  -258.3257
Iteration 3:   log likelihood =   -258.323
Iteration 4:   log likelihood =   -258.323

Logistic regression                             Number of obs     =        486
                                                LR chi2(5)        =      55.73
                                                Prob > chi2       =     0.0000
Log likelihood =   -258.323                     Pseudo R2         =     0.0974
```

less	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0002321	.0067182	0.03	0.972	-.0129354	.0133995
sex	-.1238605	.2199695	-0.56	0.573	-.5549927	.3072717
educ	-.2587727	.1037827	-2.49	0.013	-.4621831	-.0553622
party	.355585	.1311894	2.71	0.007	.0984586	.6127114
warmop2	-.9980062	.2375625	-4.20	0.000	-1.46362	-.5323923
_cons	-.5108925	.5897664	-0.87	0.386	-1.666813	.6450284

b. **Report your codes, coefficients and also your likelihood percentages.**
   - **Age codes: display exp(.0002321)** and **display (1.0002-1)*100 = 0.02%**
   - **Sex codes: display exp(-.1238605)** and **display (0.8835-1)*100 = -11.65%**
   - **Education codes: display exp(-.25877)** and **display (0.772-1)*100 = -22.8%**
   - **Party codes: display exp(.3556)** and **display (1.427-1)*100 = 42.7%**
   - **Warmop2 codes: display exp(-.998)** and **display (0.3686-1)*100 = -63.14%**

c. **Explain your findings and age, sex, educ, party and warmop2 changes likelihood in one's belief? (30)**
   - **Age:** As age goes up by one unit, the belief in less ice increases by 0.02%.
   - **Sex:** As sex changes by one unit, the belief in less ice decreases by 11.65%.
   - **Education:** As education increases one unit, the belief in less ice decreases by 22.8%.
   - **Party:** As party increases by unit, the belief in less ice increases by 42.7%.
   - **Warmop2:** As the unit increases on the belief that global warming is human caused, the belief in less ice decreases by 63.13%.