# **Lecture 7 - Regression**

Regression allows you to predict variables based on another variable. Let's begin with the example used in the text in which mental health symptoms are predicted from stress.

- ✓ **Open** symptoms and stress.sav.
- ✓ Select Analyze/Regression/Linear.

🔛 Linear Regression		×
🛷 id 🖋 stress	Dependent:     Symptoms  Block 1 of 1	Statistics
	Previous Next Independent(s)	Ogtions
	Mathod: Enter	
	Sglection Veriable:           Sglection Veriable:           Rule           Case Labels:	
	VLS Weight	
ОК	Baxle Rexet Cancel Help	]

Figure 1

✓ Select symptoms as the **Dependent** variable and stress as the **Independent** variable. Then, click on **Statistics** to explore our options. The following dialog box will appear.

E	Linear Regression: Sta	tistics X
	Regression Coefficient	Model fit
	✓ Estimates	R squared change
	Confidence intervals	Descriptives
	Level(%): 95	Part and partial correlations
	Covariance matrix	Collinearity diagnostics
	Residuals	
	Durbin-Watson	
	Casewise diagnostics	
	Outliers outside:	3 standard deviations
	◯ <u>A</u> ll cases	
	Continue	ancel Help

Figure 2

✓ As you can see there are many options. Estimates and Model Fit are selected by default. Leave them that way. Then select Descriptives and Part and partial correlations. SPSS will then calculate the mean and standard deviation for each variable in the equation and the correlation between the two variables. Then, click **Continue**.

- Linear Regression: Plots Continue DEPENDINT \*ZPRED Scatter 1 of 1 Net Cancel \*ZRESID \*DRESID Help DEPENDNT ADJFRED SRESID SDRESID × ADJPRED Standardized Residual Plots Produce all partial plots Histogram Normal probability plot
- ✓ At the main dialog box, click on **Plots** so we can see our options.



✓ It looks like we can create scatterplots here. Click **Help** to see what the abbreviations represent. I'd like to plot the Dependent variable against the predicted values to see how close they are. Select **Dependnt** for **Y** and **Adjpred** for **X**. Adjpred is the adjusted prediction. Used **Help/Topics/Index** to find out what this means for yourself. Then, click **Continue**.

Linear Regression: Save 🛛 🛛 🔀								
Predicted Values  Unstandardized  Standardized  Adjusted  S.E. of mean predictions  Distances  Mehalanobix Cock's  Leverage values  Prediction Intervals  Mean I Individual Confidence Interval: 55 %  Save to New File Coefficient statistics File  Export model information to XML	Residuals  Unstandardized  Studentized  Deleted  Studentized deleted  Influence Statistice  DfBeta(s)  Standardized DtBeta(s)  DfFit  Standardized DtFit  Covariance ratio  file  Browse	Continue Cancel Help						



52

✓ In the main dialog box, click Save, and the dialog box to the left will appear. For Predicted Values, select Unstandardized and Standardized. For Residuals, also select Unstandardized and Standardized. Now, SPSS will save the predicted values of symptoms based on the regression equation and the residual or difference between the predicted values and actual values of symptoms in the data file. This is a nice feature. Remember, the standardized values are based on z score transformations of the data whereas the unstandardized values are based on the raw data. Click Continue.

✓ Finally, click on **Options**.



✓ **Including a constant in the equation** is selected by default. This simply means that you want both a slope and an intercept (the constant). That's good. We will always leave this checked. Excluding cases listwise is also fine. We do not have any missing cases in this example.

### **Descriptive Statistics**

	Mean	Std. Deviation	N
SYMPTOMS	90.70	20.27	107
STRESS	21.47	13.10	107

### Correlations

		SYMPTOMS	STRESS
Pearson Correlation	SYMPTOMS	1.000	.506
	STRESS	.506	1.000
Sig. (1-tailed)	SYMPTOMS		.000
	STRESS	.000	
Ν	SYMPTOMS	107	107
	STRESS	107	107

#### Coefficients<sup>a</sup>

		Unstandardized Coefficients		Standardi zed Coefficien ts				Correlations	
Model		В	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part
1	(Constant)	73.890	3.271		22.587	.000			
	STRESS	.783	.130	.506	6.012	.000	.506	.506	.506

a. Dependent Variable: SYMPTOMS

## Dependent Variable: symptoms



✓ Take a moment to identify all of the key pieces of information. Find the regression coefficients used to calculate the regression equation. One difference is that the text did not include the scatterplot. What do you think of the scatterplot? Does it help you see that predicting symptoms based on stress is a pretty good estimate

✓ Now, click **Window/Symptoms and stress.sav** and look at the new data (residuals and predicted values) in your file. A small sample is below. Note how they are named and labeled.

	id	stress	symptoms	pre_1	res_1	zpr_1	zre_1
1	1	30	99	97.3830 <mark>U</mark>	nstandardize	d Predicted V	alue 09207
2	2	27	94	95.03368	-1.03368	.42248	05886
3	3	9	80	80.93762	93762	95202	05339
4	4	20	70	89.55188	-19.5519	11204	-1.11328

54

Let's use what we know about the regression equation to check the accuracy of the scores created by SPSS. We will focus on the unstandardized predicted and residual values. This is also a great opportunity to learn how to use the Transform menus to perform calculations based on existing data.

We know from the regression equation that:

Symptoms Predicted or = 73.890 + .783\* Stress. Y<sup>^</sup>

We also know that the residual can be computed as follows:

Residual = Y-or Symptoms – Symptoms Predicted Values. Y<sup>^</sup>

We'll use SPSS to calculate these values and then compare them to the values computed by SPSS.

✓ In the Data Editor window, select **Transform/Compute**.

Imagin Famession:         sympred         Type & Label         id         id         stress         sympred         Instancerdized Predict         Unstancerdized Predict         Adjusted Predicted I         Adjusted Predicted I         Standardized Predicted I         Standardized Predicted I         Standardized Residual (         Standardized Residual (S         Standardized Residual (S         Imagin Prediction Residual Re	Compute Variable			
[T] (optional case selection condition)	Compute Variable Sympred Type & Label  f id f stress f symptoms f Unstandardized Predict f Unstandardized Residu f Adjusted Predicted Usl f Standardized Residual [ f Studentized Residual [S f Studentized Residual [S] f Studen	-	Numeric Expression: $73890+873$ *stress         +       >       7       8         -       <=	Function group: Al Arithmetic CDF & Noncentral CDF Conversion Current Date/Time Date Arthmetic
[T] (optional case selection condition)			* () Delete	
	[t] (optional case selecti	ion conc	ition)	

 $\checkmark$  Check the Data Editor to see if your new variable is there, and compare it to pre\_1. Are they the same? The only difference I see is that our variable is only expressed to 2 decimal places. But, the values agree.

✓ Follow similar steps to calculate the residual. Click on **Transform/Compute**. Name your **Target Variable** sympres and **Label** it symptoms residual. Put the formula symptoms-sympred in the **Numeric Expression** box by double clicking the two pre-existing variables and typing a minus sign between them. Then, click **Ok**.

✓ Compare these values to res\_1. Again they agree. A portion of the new data file is below.

	id	stress	symptoms	pre_1	res_1	zpr_1	zre_1	sympred	sympres	
1	1	30	99	97.38302	1.61698	.65157	.09207	g <mark>Sym</mark> p	toms Predicted	d
2	2	27	94	95.03368	-1.03368	.42248	05886	95.03	-1.03	_
3	3	9	80	80.93762	93762	95202	05339	80.94	94	
4	4	20	70	89.55188	-19.5519	11204	-1.11328	89.55	-19.55	_
-	-	<u>0</u>	100	70.00000	00 20102	1 11010	1.05005	70.01		

Now that you are confident that the predicted and residual values computed by SPSS are exactly what you intended, you won't ever need to calculate them yourself again. You can simply rely on the values computed by SPSS through the Save command.