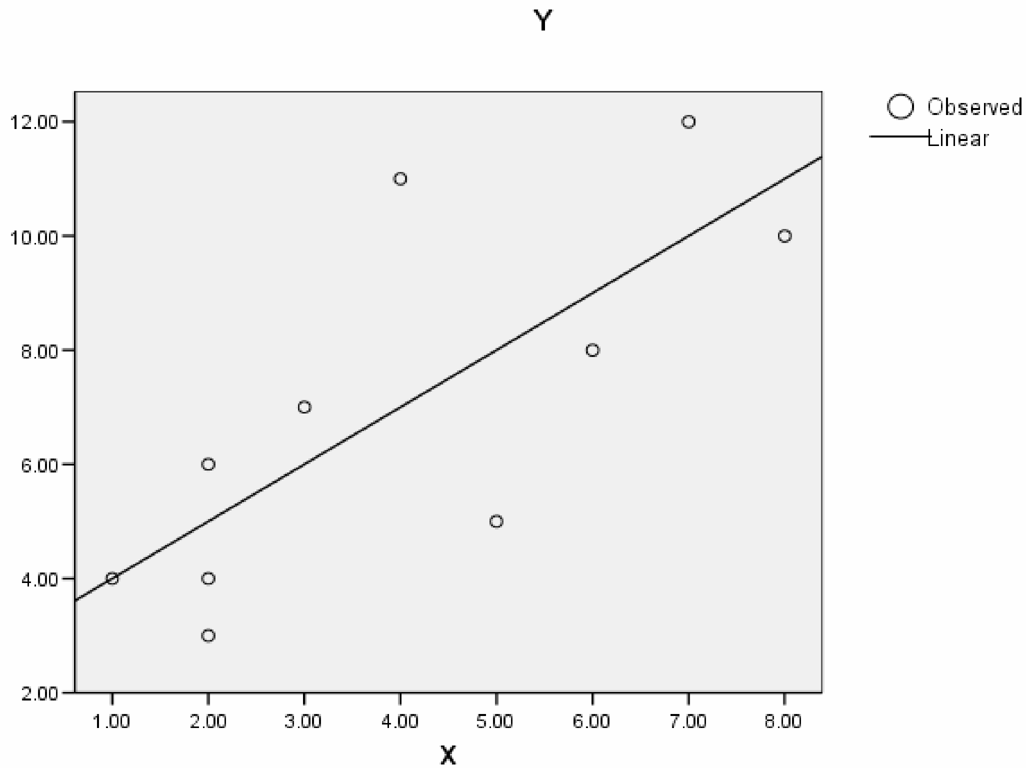
Linear Regression
Significance of Regression Equation

Course: Statistics 1
Lecturer: Dr. Courtney Pindling

Regression Model

Y = 1.368X + 3.35



X ***Independent***	**Y** ***Dependent***
4	11
3	7
1	4
7	12
2	6

Significance of Regression Equation

- The significance of Pearson correlation can be used to test the significance of the regression equation (when single X)
- *H_0:* There is no relationship between X and Y ***or***
- *H_0:* The regression equation does not account for a significant portion of the variance of Y scores
- Process of testing significance of a regression equation is called analysis of regression and similar to analysis of variance (ANOVA)

Sums of Squares of Regression

$$\text{std error of est} = \sqrt{\frac{SS_{Res}}{df}} = \sqrt{\frac{6.33}{3}} = 1.45$$

$$SS_{Reg} = r^2 SS_Y = (0.929)^2(46) = 39.70$$

$$SS_{Res} = (1 - r^2)SS_Y = [1 - (0.929)^2](46) = 6.30$$

$$r = \frac{SP}{\sqrt{SS_X SS_Y}} = \frac{29}{\sqrt{21.2(46)}} = 0.929 \text{ and } r^2 = 0.863$$

Analysis of Regression

- The regression analysis uses an *F*-ratio to determine the amount of variance accounted for by the regression equation
- *F*-ratio: *MSR/MSE* (MS is mean square)
  - MSR is MS_{Reg} or variance predicted by regression equation
  - MSE is MS_{Res} or unpredicted variance due to chance or other than regression equation

Analysis of Regression Table

Source	**SS**	***df***	***MS***	***F***
Regression	SSReg	1	MSR = SSReg/1	MSR/MSE
Residual	SSRes	n – 2	MSE = SSRes/(n-2)	
Total	Sum SS	Sum *df*		

Mean Square Calculations

$$MS_{\text{Reg}} = \frac{SS_{regression}}{df_{reg}} = \frac{39.67}{1} = 39.67$$

$$MS_{\text{Res}} = \frac{SS_{residual}}{df_{res}} = \frac{SS_{residual}}{n-2} = \frac{6.33}{3} = 2.11$$

$$F-ratio = F = \frac{MS_{regression}}{MS_{residual}} = \frac{39.67}{2.11} = 18.80$$

ANOVA: Regression Table

Source	**SS**	***df***	***MS***	***F***
Regression	39.67	1	MSR = 39.67	MSR/MSE = **18.80**
Residual	6.33	3	MSE = 2.11	
Total	46	4		

Conclusion

- F_{cv} = **10.13**
  df = 1, 3, and a = 0.05
- $F_{stat} > F_{cv}$ or 18.80 > 10.13
- Reject H_0 that the regression equation does not accounts for a significant portion of the variability for the Y scores
- **Conclude** that regression equation is a **good model**
- Pearson ***r*** is 0.93, high

df2	df = 1	df = 2
1	161.45	199.50
2	18.51	19.00
3	10.13	9.55
4	7.71	6.94
5	6.61	5.79

SPSS Output: ANOVA

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	39.670	1	39.670	18.800	.023[a]
	Residual	6.330	3	2.110		
	Total	46.000	4			

a. Predictors: (Constant), X

b. Dependent Variable: Y

SPSS Output: Regression Model

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.349	1.254		2.671	.076
	X	1.368	.315	.929	4.336	.023

a. Dependent Variable: Y