Regression Example[1]

Years Since PhD			Number of Publications		
X	$X-\bar{X}$	$\left(X-\bar{X}\right)^2$	Y	$Y-\bar{Y}$	$\left(X-\bar{X}\right)\left(Y-\bar{Y}\right)$
3	-4.67	21.81	**18**	-1.93	9.01
6	-1.67	2.79	**3**	-16.93	28.27
3	-4.67	21.81	**2**	-17.93	83.73
8	0.33	0.11	**17**	-2.93	-0.97
9	1.33	1.77	**11**	-8.93	-11.88
6	-1.67	2.79	**6**	-13.93	23.26
16	8.33	69.39	**38**	18.07	150.52
10	2.33	5.43	**48**	28.07	65.40
2	-5.67	32.15	**9**	-10.93	61.97
5	-2.67	7.13	**22**	2.07	-5.53
5	-2.67	7.13	**30**	10.07	-26.89
6	-1.67	2.79	**21**	1.07	-1.79
7	-0.67	0.45	**10**	-9.93	6.65
11	3.33	11.09	**27**	7.07	23.54
18	10.33	106.71	**37**	17.07	176.33
$\bar{X}=7.67$		$\sum\left(X-\bar{X}\right)^2=293.33$	$\bar{Y}=19.93$		$\sum\left(X-\bar{X}\right)\left(Y-\bar{Y}\right)=581.67$

Unstandardized regression coefficient:

$$B_{YX}=\frac{\sum\left(X-\bar{X}\right)\left(Y-\bar{Y}\right)}{\sum\left(X-\bar{X}\right)^2}=\frac{581.67}{293.33}=1.98$$

Regression line:

$$\hat{Y}=4.74+1.98X$$

Intercept:

$$B_0=\bar{Y}-B_{YX}\bar{X}=19.93-1.98(7.67)=4.74$$

Standardized regression coefficient:

$$sd_X=\sqrt{\frac{\sum\left(X-\bar{X}\right)^2}{n-1}}=\sqrt{\frac{293.33}{15-1}}=4.58$$

$$sd_Y=13.82$$

$$\beta_{YX}=B_{YX}\left(\frac{sd_X}{sd_Y}\right)=1.98\left(\frac{4.58}{13.82}\right)=.66$$

Significance and Confidence Intervals:

$$SE_B=\left(\frac{sd_Y}{sd_X}\right)\sqrt{\frac{1-r^2}{n-2}}=\frac{13.82}{4.58}\left(\sqrt{\frac{1-(.66)^2}{15-2}}\right)=.632$$

$$t=\frac{B_{YX}}{SE_B}=\frac{1.98}{.632}=3.139$$

$$B_{YX}\pm\left(t_{df,\alpha/2}\right)\left(SE_B\right)=1.98\pm(2.16)(.632)=.62,3.35$$

df for the regression coefficient is $n-2$, so the two-tailed α = .05 critical value, $t_{df,\alpha/2}$, from the Table C.3 in the text is 2.16

[1] Numerical example from Cohen, Cohen, West, & Aiken, 2003. The current text uses sd instead of s for standard deviation and SE_B instead of s_b for the standard error. For simple regression, the text uses B_{YX} for the regression slope at the beginning, but B_1 will often be used later and is more common.

Computer Example Simple Regression

SPSS Syntax

```
*correlation test with regression is one-tailed, so request separate correlations procedure.
correlations vars=yrsphd numpubs.

regression vars=yrsphd numpubs
    /descriptives=mean stddev corr sig n
    /statistics=anova coeff ses r ci
    /dependent=numpubs
     /method=enter yrsphd.
```

SPSS Menus

Analyze -> Regression -> Linear, then drag over the dependent and independent variables, then click on the Statistics button and check the Confidence Intervals box.

Note that there are several important statistics that are normally obtained in the output that I have omitted and we are going to skip for now. We will cover all of them in detail evenutally.

Correlations

		yrsphd	numpubs
yrsphd	Pearson Correlation	1	.657
	Sig. (2-tailed)		.008
	N	15	15
numpubs	Pearson Correlation	.657	1
	Sig. (2-tailed)	.008	
	N	15	15

Output from correlation procedure (note that the correlations from the /DESCRIPTIVES subcommand of the REGRESSION procedure gives the 1-tailed significance (for 2-tailed, double the p-value: .004 × 2 = .008.)

Descriptive Statistics

	Mean	Std. Deviation	N
yrsphd	7.6667	4.57738	15
numpubs	19.9333	13.82269	15

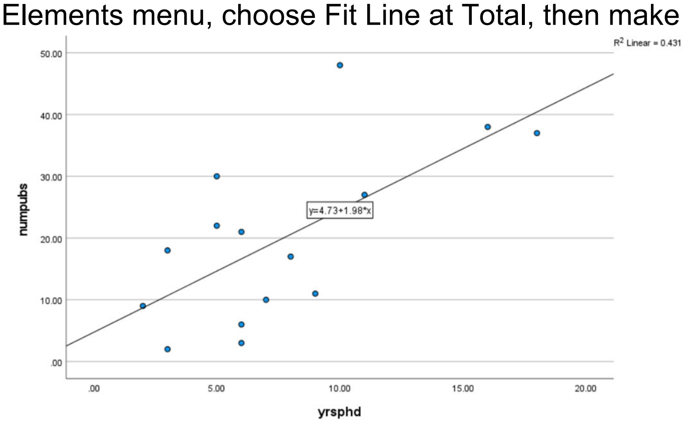
Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients				95.0% Confidence Interval for B	
		B	Std. Error	Beta	Std. Error	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	4.731	5.591			.846	.413	-7.347	16.808
	yrsphd	1.983	.632	.657	.209	3.139	.008	.618	3.348

a. Dependent Variable: numpubs

```
graph
  /scatterplot(bivar)=yrsphd with numpubs
```

For a scatter plot with regression line, obtain the scatterplot, click on the scatterplot chart in the output, go to the Elements menu, choose Fit Line at Total, then make sure the Linear radio button is chosen.



R Code

```
> #clear active frame from previous analyses
> rm(d)
>
> library(haven)
> d = read_sav("c:/jason/spsswin/uvclass/ccwa2_2_2.sav")
>
> library(lessR)
> #lessR
> Regression(NUMPUBS ~ YRSPHD,brief=TRUE)
  #note: we will later use lessR function Regression() without the brief statment, which gives more output
```

Some of the output has been omitted

(Unstandardized)

```
-- Estimated Model for NUMPUBS

             Estimate    Std Err  t-value  p-value   Lower 95%   Upper 95%
(Intercept)     4.731      5.591    0.846    0.413      -7.347      16.808
     YRSPHD     1.983      0.632    3.139    0.008       0.618       3.348

-- Model Fit

Standard deviation of NUMPUBS: 13.823

Standard deviation of residuals:  10.818 for 13 degrees of freedom
95% range of residual variation:  46.744 = 2 * (2.160 * 10.818)

R-squared:  0.431    Adjusted R-squared:  0.387    PRESS R-squared:  0.308

Null hypothesis of all 0 population slope coefficients:
  F-statistic: 9.855     df: 1 and 13     p-value:  0.008
```

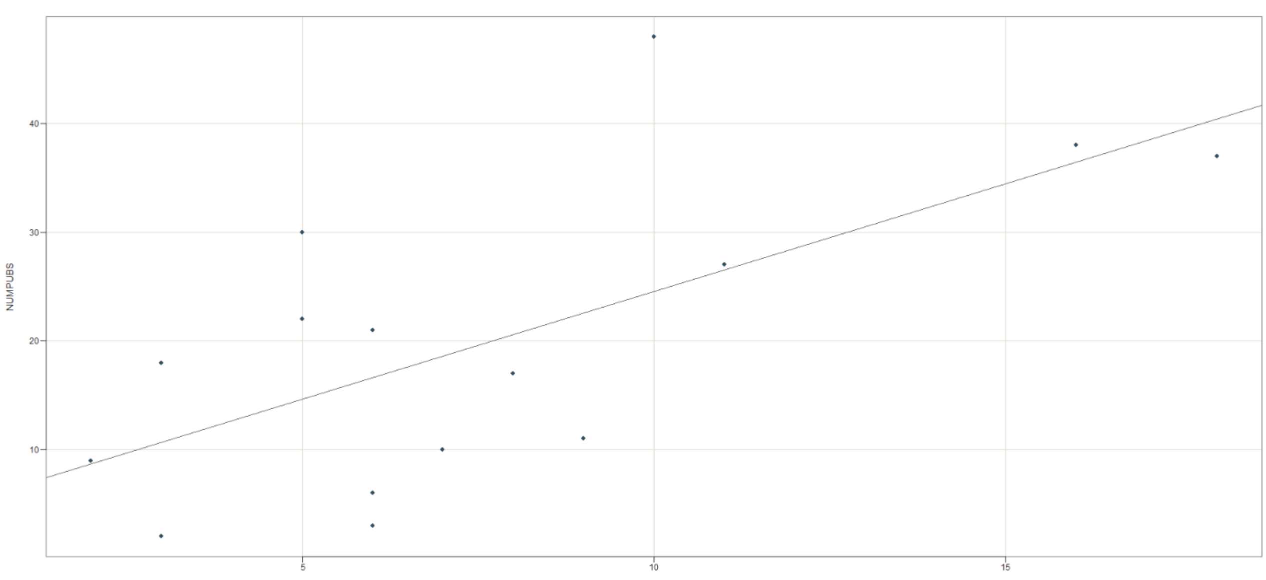
(Standardized)

```
> #In these two lines, I create standardized variables using the rescale command from lessR.
> #do not use the significance tests from this output, just the "Estimate" values.
> d$znumpubs = rescale(NUMPUBS)
> d$zyrsphd = rescale(YRSPHD)
> Regression(znumpubs ~ zyrsphd,brief=TRUE)

-- Estimated Model for znumpubs

             Estimate    Std Err  t-value  p-value   Lower 95%   Upper 95%
(Intercept)    0.0001     0.2020    0.000    1.000     -0.4364     0.4366
    zyrsphd    0.6566     0.2091    3.140    0.008      0.2049     1.1084
```

lessR Scatterplot with Regression Line (produced by default from the unstandardized regression)



Example Write-Up

A simple regression analysis was conducted to examine the relationship between the years of experience of a faculty member and the number of peer-reviewed publications. Results indicated that the years of experience significantly predicted the number of publications, b = 1.98, SE = .632, b^* = .66, p = .01, 95% CIs [.62,3.35].[2] For each additional year of experience, the faculty member published approximately two (1.98) additional publications. Years of experience accounted for a large percentage of variance in the number of publications, R^2 = .44, F(1,13) = 9.86, p = .01.[3]

[2] The APA seventh edition of the publication manual gives conflicting information for the unstandardized and standardized regression coefficients. In Table 6.5, the abbreviations are given as b for unstandardized and b^* for standardized. In the past and in the Table 7.25 example table, B is used or unstandardized and β is used for standardized and B and b seem to still commonly appear in APA journals.

[3] We will begin using R^2, the squared multiple correlation coefficient, instead of r^2, the regular square of the Pearson correlation coefficient, for any regression results.