ANOVA and Regression Equivalence
Sex and Depression Example

Do older women and men differ on depression? This question could be investigated a number of ways, including through a t-test, a correlation, analysis of variance or a simple regression. Below I illustrate the equivalence of ANOVA and regression in investigating this hypothesis.[1]

SPSS

```
output close   *.
get file='c:\jason\spsswin\mvclass\anova.sav'.

means variables=w1cesd9 by w1sex.

oneway w1cesd9 by w1sex
      /es=overall.
*note that GLM and ANOVA procedures can also be used but do not have effect size option.
*anova variable=w1cesd9 by w1sex(0,1).

regression vars=w1cesd9 w1sex
    /descriptives=mean stdev n sig corr
    /statistics=anova r coeff ses cha
    /dependent=w1cesd9
    /method=enter w1sex.
```

Means

Report

w1cesd9 9-item CES-D

w1sex A1-sex of R	Mean	N	Std. Deviation
.00 male	4.6355	347	4.58732
1.00 female	5.3922	563	5.19943
Total	5.1036	910	4.98591

Oneway

ANOVA

w1cesd9 9-item CES-D

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	122.913	1	122.913	4.966	.026
Within Groups	22474.231	908	24.751		
Total	22597.145	909			

ANOVA Effect Sizes [a,b]

			95% Confidence Interval	
		Point Estimate	Lower	Upper
w1cesd9 9-item CES-D	Eta-squared	.005	.000	.019
	Epsilon-squared	.004	-.001	.018
	Omega-squared Fixed-effect	.004	-.001	.018
	Omega-squared Random-effect	.004	-.001	.018

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

[1] These data come from the Later Life Study of Social Exchanges (LLSSE) Sorkin, D. H., & Rook, K. S. (2004). Interpersonal control strivings and vulnerability to negative social exchanges in later life. Psychology and Aging, 19(4), 555–564. https://10.1037/0882-7974.19.4.555. Newsom, J. T., Rook, K. S., Nishishiba, M., Sorkin, D. H., & Mahan, T. L. (2005). Understanding the relative importance of positive and negative social exchanges: Examining specific domains and appraisals. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 60(6), P304-P312.

Regression

Correlations

		w1cesd9 9-item CES-D	w1sex A1-sex of R
Pearson Correlation	w1cesd9 9-item CES-D	1.000	.074
	w1sex A1-sex of R	.074	1.000
Sig. (1-tailed)	w1cesd9 9-item CES-D	.	.013
	w1sex A1-sex of R	.013	.
N	w1cesd9 9-item CES-D	910	910
	w1sex A1-sex of R	910	910

(1-tailed test reported for correlations descriptive, so need to double p-value)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.074[a]	.005	.004	4.97507

a. Predictors: (Constant), w1sex A1-sex of R

R-square is the same as eta-squared above and Adjusted R square is the same as Omega squared in the ANOVA

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	122.913	1	122.913	4.966	.026[b]
	Residual	22474.231	908	24.751		
	Total	22597.145	909			

a. Dependent Variable: w1cesd9 9-item CES-D

b. Predictors: (Constant), w1sex A1-sex of R

Significance test is the same as correlation (.013x2 = .026) and notice this ANOVA box for the regression is identical for the one from the ANOVA analysis above

Coefficients [a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	Standardized Coefficients Std. Error	t	Sig.
1	(Constant)	4.635	.267			17.356	<.001
	w1sex A1-sex of R	.757	.340	.074	.033	2.228	.026

a. Dependent Variable: w1cesd9 9-item CES-D

.757 is the difference between means reported up above 5.3922 – 4.6355 = .757. Same p-value as R-square test, ANOVA F, and correlation test.

R Example

```
#clear active frame from previous analyses
rm(d)

library(haven)
d = read_sav("c:/jason/spsswin/mvclass/anova.sav")

library(lessR)
#lessR function (brief version of ANOVA function output)
ANOVA(w1cesd9 ~ w1sex, brief=TRUE)


#alternative method
#base R function
#model1 <- aov(w1cesd9 ~ w1sex, data=d)
#summary(model1)

#aov does not give eta-squared, so can use effectize package
#library(effectsize)
#eta_squared(model1, partial = FALSE)
#omega_squared(model1, partial = FALSE)

Data Frame:  d

Response Variable: w1cesd9

Factor Variable: w1sex
  Levels: 0 1
```

```
Number of cases (rows) of data:  916
Number of cases retained for analysis:  910


  DESCRIPTIVE STATISTICS

     n        mean          sd         min          max
0  347   4.6354981   4.5873245   0.0000000   25.0000000
1  563   5.3921594   5.1994327   0.0000000   26.0000000

Grand Mean: 5.1036303


  BASIC ANALYSIS

Summary Table

                df         Sum Sq      Mean Sq    F-value   p-value
w1sex            1    122.9134505  122.9134505 4.9659279    0.0261
Residuals      908  22474.2312530   24.7513560

Association and Effect Size

R Squared: 0.01
R Sq Adjusted: 0.00
Omega Squared: 0.00

Cohen's f: 0.07

  BACKGROUND

Data Frame:  d

Response Variable: w1cesd9
Predictor Variable: w1sex

Number of cases (rows) of data:  916
Number of cases retained for analysis:  910


#lessR function
Regression(w1cesd9 ~ w1sex, brief=TRUE)

  BASIC ANALYSIS

Estimated Model

             Estimate    Std Err  t-value  p-value   Lower 95%   Upper 95%
(Intercept) 4.6354981  0.2670759   17.356    0.000   4.1113402   5.1596561
      w1sex 0.7566612  0.3395481    2.228    0.026   0.0902709   1.4230515


Model Fit

Standard deviation of residuals:  4.9750735 for 908 degrees of freedom

R-squared:  0.005    Adjusted R-squared:  0.004    PRESS R-squared:  0.001

Null hypothesis that all population slope coefficients are 0:
  F-statistic: 4.966     df: 1 and 908     p-value:  0.026


Analysis of Variance

               df         Sum Sq      Mean Sq    F-value   p-value
Model           1    122.9134505  122.9134505  4.9659279     0.026
Residuals     908  22474.2312530   24.7513560
w1cesd9       909  22597.1447035   24.8593451
```

Write-up

With binary predictors, write-up for regression analysis proceeds just as any other regression analysis (as illustrated previously), although the author may wish to emphasize the interpretation of the slope as a difference in means.