

An Overview of Effect Size Statistics and Why They Are so Important

Jeff Meyer

THE ANALYSIS
FACTOR

The “d-Family” Effect Size

Difference in means example 1:

- A drug company is developing two new drugs.
- Both new drugs have a difference in means between treatment and control groups of 0.50.
- Both new drugs have two-tailed p-values of 0.000
- The drug company has limited funding and can only develop one drug.
- Which drug does it promote?

Treatment Group 1 versus Control Group 1

Two-sample t test with equal variances

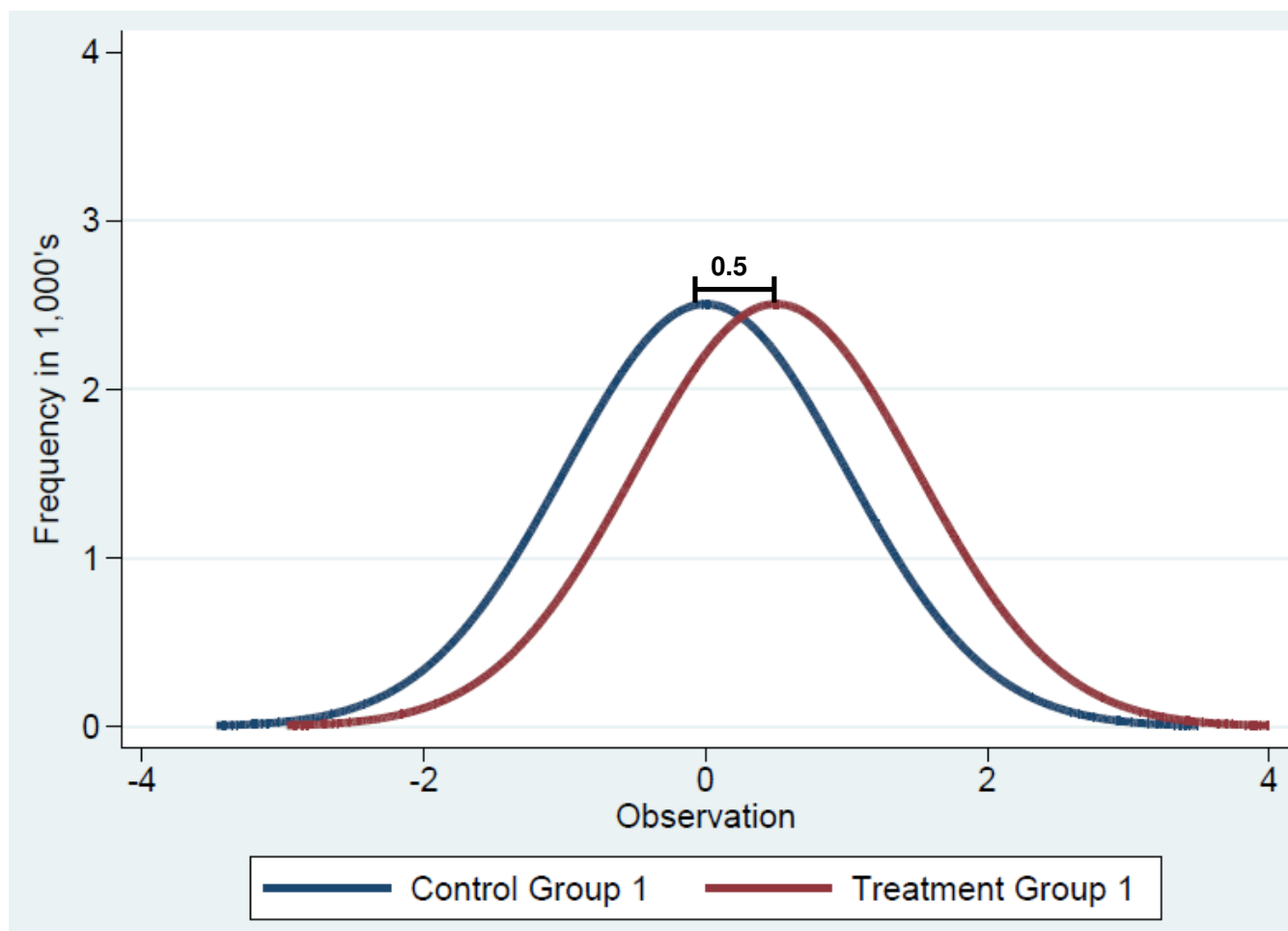
Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
control1	1000	.0852022	.0638943	2.020515	-.0401802	.2105847
treatm~1	1000	.5852022	.0638943	2.020515	.4598198	.7105847
combined	2000	.3352022	.0455136	2.035429	.2459432	.4244612
diff		-.5	.0903602		-.6772101	-.3227899

diff = mean(control1) - mean(treatment1) t = -5.5334
Ho: diff = 0 degrees of freedom = 1998

Ha: diff < 0
Pr(T < t) = 0.0000

Ha: diff != 0
Pr(|T| > |t|) = 0.0000

Ha: diff > 0
Pr(T > t) = 1.0000



Treatment Group 2 versus Control Group 2

Two-sample t test with equal variances

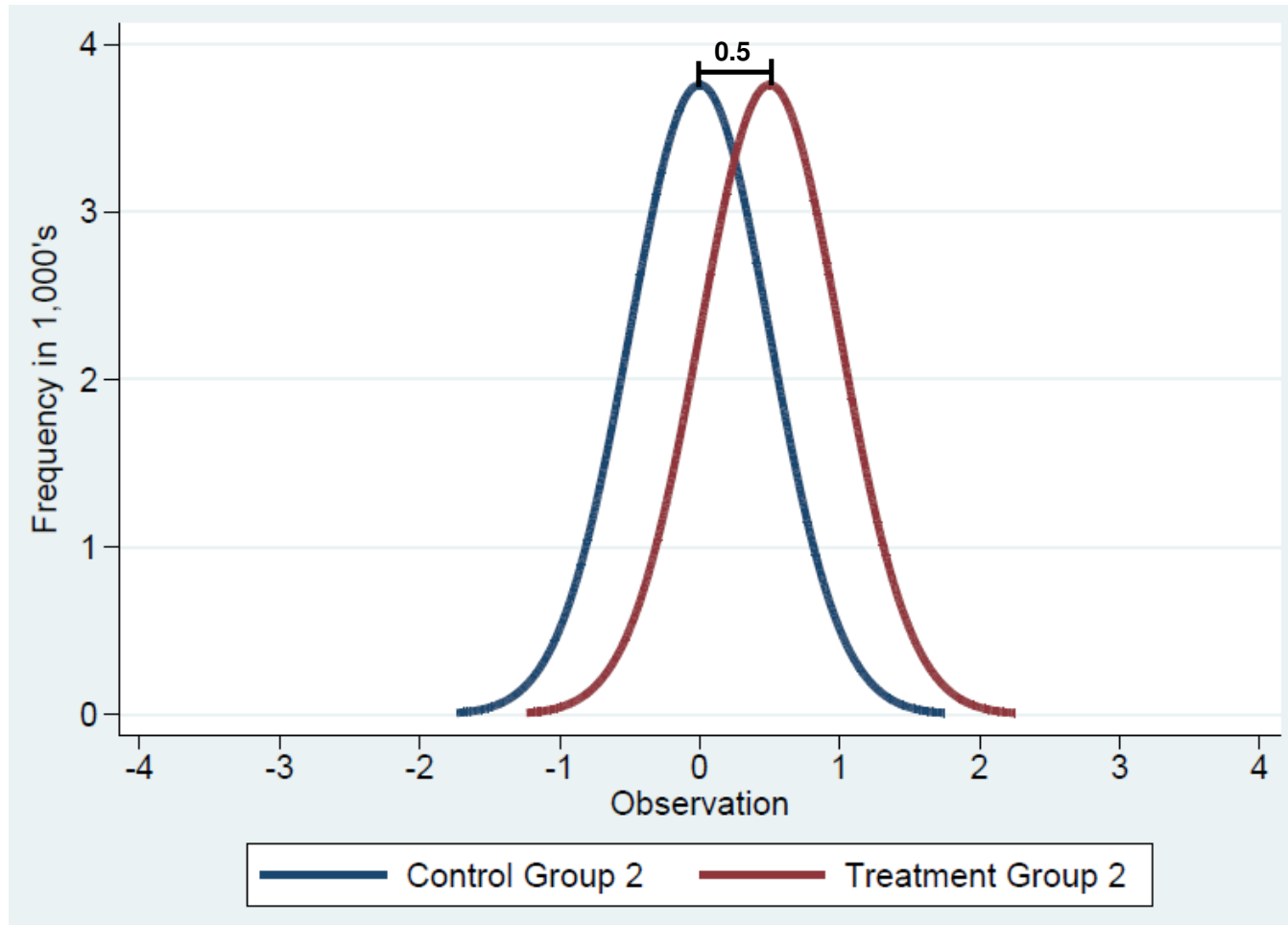
Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
control2	1000	.0426011	.0319472	1.010258	-.0200901	.1052923
treatm~2	1000	.5426011	.0319472	1.010258	.4799099	.6052923
combined	2000	.2926011	.0232663	1.040501	.2469724	.3382298
diff		-.5	.0451801		-.588605	-.411395

diff = mean(control2) - mean(treatment2) t = -11.0668
Ho: diff = 0 degrees of freedom = 1998

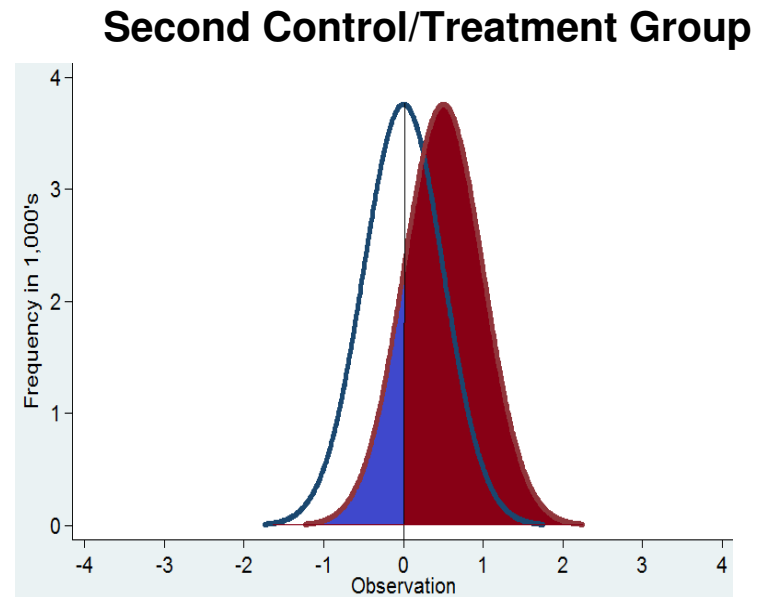
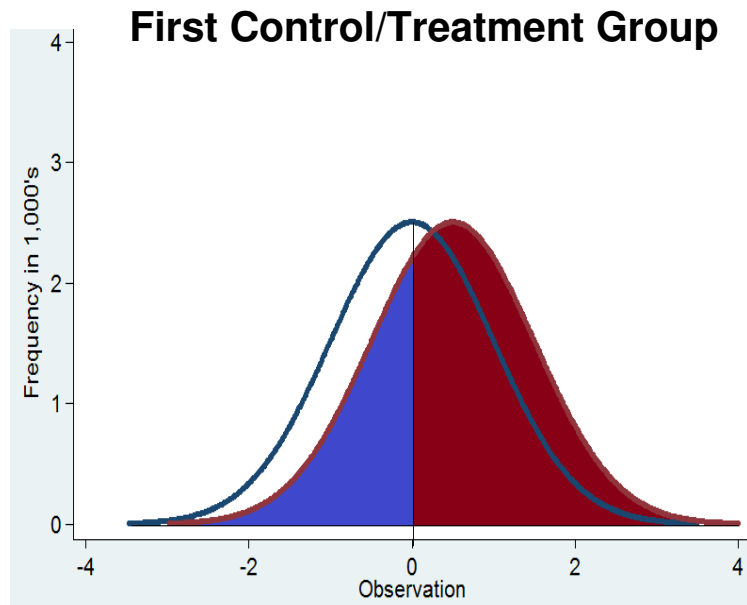
Ha: diff < 0
Pr(T < t) = 0.0000

Ha: diff != 0
Pr(|T| > |t|) = 0.0000

Ha: diff > 0
Pr(T > t) = 1.0000



Same p-values and Same Difference Between the Means



Red shaded: treatment observations greater than the mean of the control group.

Blue shaded: treatment observations less than the mean of the control group.

The “d-Family” Effect Size

Difference in means example 2:

One control group, three treatment groups

- All three treatment groups have two-tailed p-values of 0.000
- Treatment two costs more to produce than treatment one. Treatment three costs more to produce than treatment two.
- The drug company needs to determine which drug has the best cost/benefit ratio.

Treatment 1 vs Control Group

Two-sample t test with equal variances

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
control treat1	1000	-.0883709	.0627212	1.983417	-.2114513	.0347094
	1000	.4116291	.0627212	1.983417	.2885487	.5347094
combined	2000	.1616291	.0446906	1.998626	.0739839	.2492742
diff		- .5	.0887011		-.6739564	-.3260436

```
diff = mean(control) - mean(treat1)          t = -5.6369  
Ho: diff = 0                                degrees of freedom = 1998
```

$$H_a: \text{diff} < 0$$
$$\Pr(T < t) = 0.0000$$

Ha: $\text{diff} \neq 0$
 $\Pr(|T| > |t|) = 0.0000$

$$H_a: \text{diff} > 0$$
$$\Pr(T > t) = 1.0000$$

Treatment 2 vs Control Group

Two-sample t test with equal variances

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
control	1000	-.0883709	.0627212	1.983417	-.2114513	.0347094
treat2	1000	.9116291	.0627212	1.983417	.7885487	1.034709
combined	2000	.4116291	.045728	2.045019	.3219495	.5013086
diff		-1	.0887011		-1.173956	-.8260436

diff = mean(control) - mean(treat2) t = -11.2738
Ho: diff = 0 degrees of freedom = 1998

Ha: diff < 0
Pr(T < t) = 0.0000

Ha: diff != 0
Pr(|T| > |t|) = 0.0000

Ha: diff > 0
Pr(T > t) = 1.0000

Treatment 3 vs Control Group

Two-sample t test with equal variances

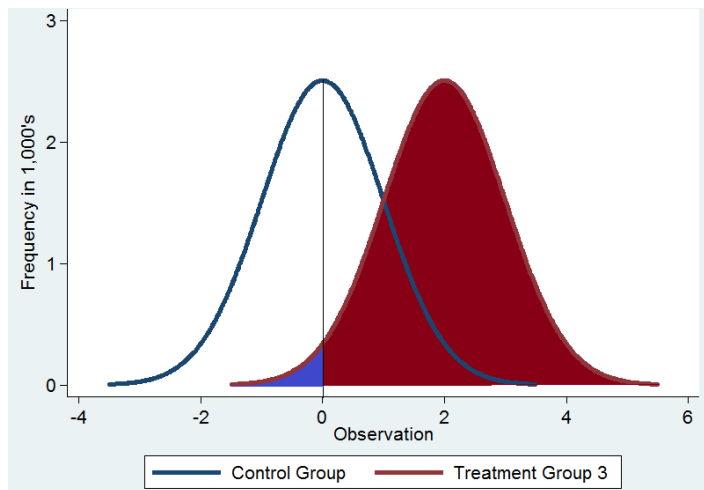
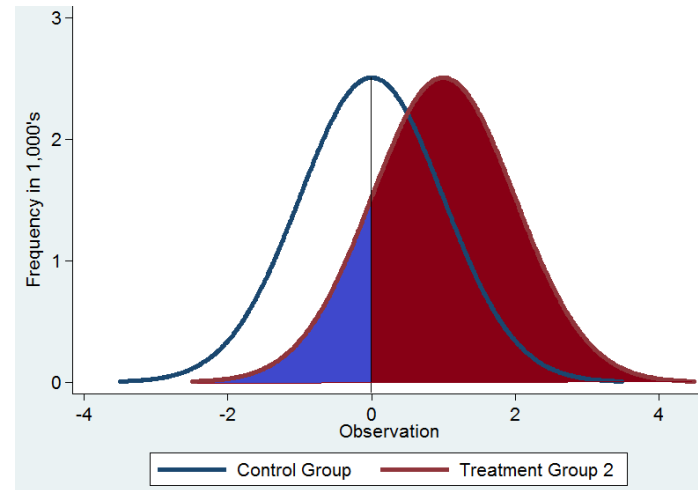
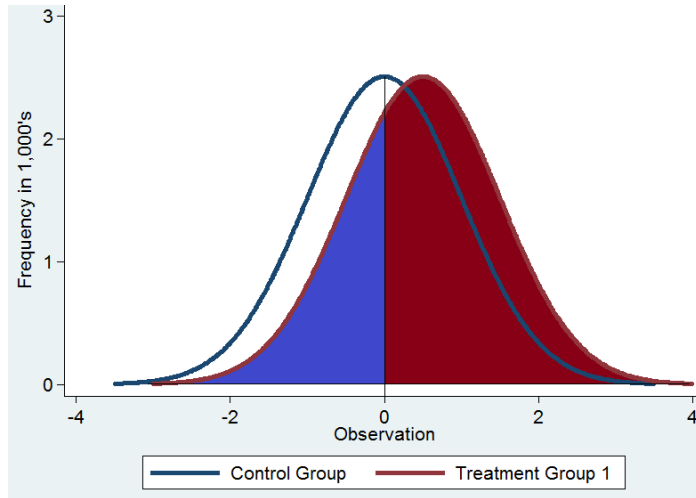
Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
control	1000	-.0883709	.0627212	1.983417	-.2114513	.0347094
treat3	1000	1.911629	.0627212	1.983417	1.788549	2.034709
combined	2000	.9116291	.0496612	2.220918	.8142359	1.009022
diff		-2	.0887011		-2.173956	-1.826044

```
diff = mean(control) - mean(treat3)          t = -22.5476
Ho: diff = 0                                degrees of freedom = 1998
```

$$H_a: \text{diff} < 0$$
$$\Pr(T < t) = 0.0000$$

Ha: diff != 0
Pr(|T| > |t|) = 0.0000

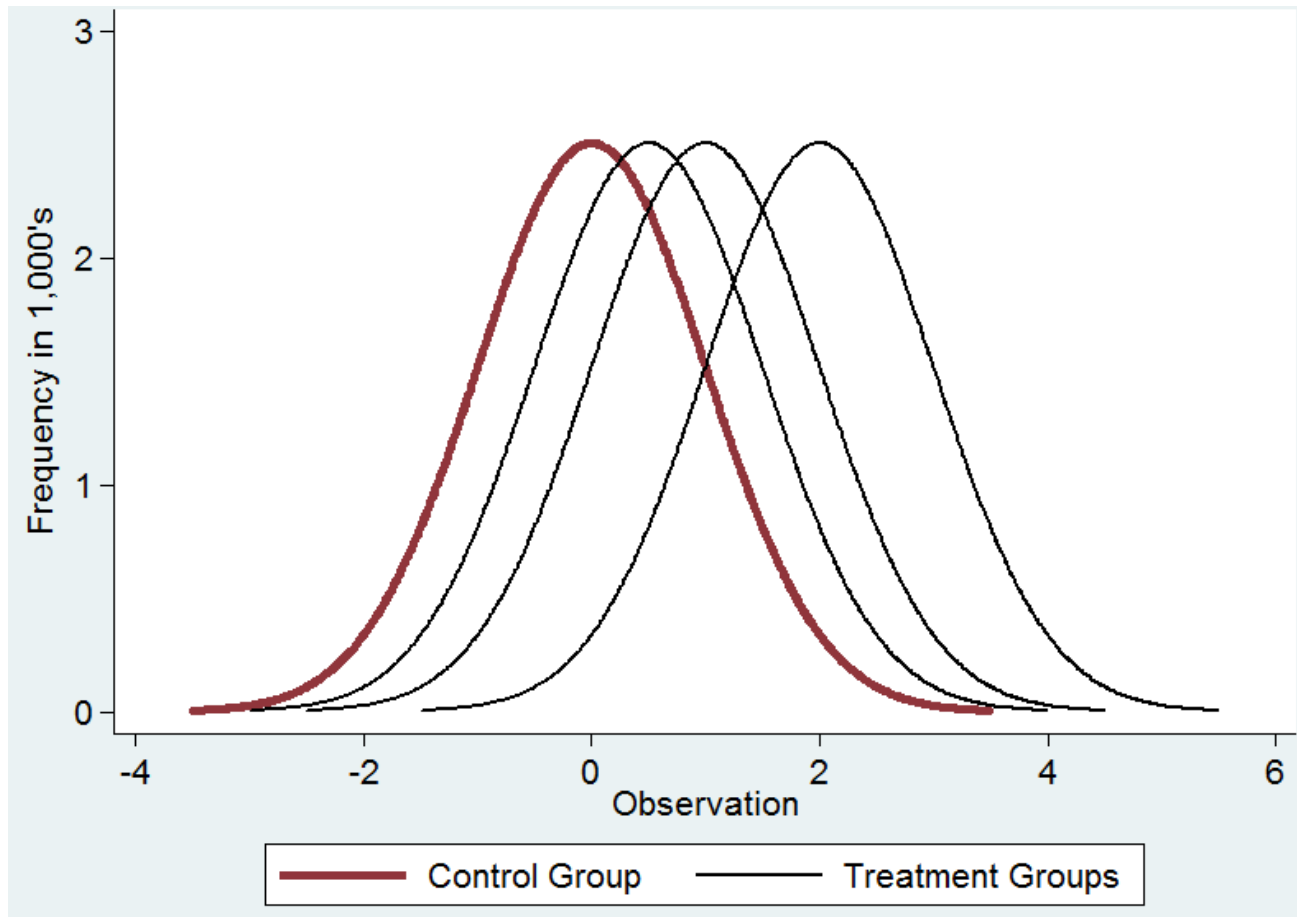
$$\begin{aligned} H_a: \text{diff} &> 0 \\ \Pr(T > t) &= 1.0000 \end{aligned}$$



Red shaded:
treatment observations greater than
the mean of the control group.

Blue shaded:
treatment observations less than the
mean of the control group.

How do We Measure the Differences



The Equation for the “d” (difference) Family Effect Size

(Mean of experimental group) - (Mean of control group)
Standard Deviation

Standard deviation using Hedge's g : $s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

Standard deviation using Cohen's d : $s = \sqrt{\frac{n_1s_1^2 + n_2s_2^2}{n_1+n_2-2}}$

95% confidence interval: $EffectSize \pm 1.96*SE$

$$SE = \sqrt{\frac{n_1+n_2}{n_1n_2} + \frac{EffectSize^2}{2(n_1+n_2)}}$$

Have you seen this formula before?

Have you seen this formula before?

Standardizing a variable:
$$\frac{\text{variable's mean} - \text{observation}}{\text{Standard Deviation}}$$

A standardized variable has a mean of zero and a standard deviation of 1.

The only difference with “effect size” is we use the “sample’s mean” rather than the “observation”.

Difference in Means Example 1

The difference in means standardized

```
. esize unpaired control1 == treatment1
```

Effect size based on mean comparison

Number of obs = 2000

Effect Size	Estimate	[95% Conf. Interval]	
Cohen's <i>d</i>	-.2474616	-.3354182	-.1594435
Hedges's <i>g</i>	-.2473687	-.3352923	-.1593836

```
.  
. esize unpaired control2 == treatment2
```

Effect size based on mean comparison

Number of obs = 2000

Effect Size	Estimate	[95% Conf. Interval]	
Cohen's <i>d</i>	-.4949232	-.5838482	-.4058778
Hedges's <i>g</i>	-.4947374	-.583629	-.4057255

Difference in Means Example 2

The difference in means standardized

```
. esize unpaired control == treat1
```

Effect size based on mean comparison

Number of obs = 2000

Effect Size	Estimate	[95% Conf. Interval]	
Cohen's <i>d</i>	-.2520902	-.3400589	-.1640588
Hedges's <i>g</i>	-.2519955	-.3399312	-.1639973

```
. esize unpaired control == treat2
```

Effect size based on mean comparison

Number of obs = 2000

Effect Size	Estimate	[95% Conf. Interval]	
Cohen's <i>d</i>	-.5041803	-.5931542	-.415084
Hedges's <i>g</i>	-.503991	-.5929315	-.4149281

```
. esize unpaired control == treat3
```

Effect size based on mean comparison

Number of obs = 2000

Effect Size	Estimate	[95% Conf. Interval]	
Cohen's <i>d</i>	-1.008361	-1.101309	-.9151874
Hedges's <i>g</i>	-1.007982	-1.100895	-.9148438

How do We Interpret the *d-family* Effect Size Estimate?

The *d-family* effect size estimate represents the number of standard deviations the means of the two groups are from each other.

For example, an effect size score of 1.2 tells us that the mean of the experimental group is 1.2 standard deviations greater than the mean of the control group.

Interpreting the Effect Size Estimates?

We can use a standard normal distribution z-table

Z	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

Data Analysis Brown Bag: July 2015
An Overview of Effect Size Statistics

Variables		Effect Size (Stand. Dev.)	% of treat group greater than Mean of Control Group
Control mean = Treatment mean		0.00	50.0%
Control 1	Treatment 1	0.247	59.9%
Control 2	Treatment 2	0.495	69.2%
Control	Treat 1	0.252	59.9%
Control	Treat 2	0.504	69.2%
Control	Treat 3	1.010	84.4%

Z	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830

Effect Size Table

Effect Size	Percentile below	Percentile above	Percent of non-overlap
0.0	50%	50%	0.0%
0.1	46%	54%	7.7%
0.2	42%	58%	14.7%
0.3	38%	62%	21.3%
0.4	34%	66%	27.4%
0.5	31%	69%	33.0%
0.6	27%	73%	38.2%
0.7	24%	76%	43.0%
0.8	21%	79%	47.4%
0.9	18%	82%	51.6%
1.0	16%	84%	55.4%
1.1	14%	86%	58.9%
1.2	12%	88%	62.2%
1.3	10%	90%	65.3%
1.4	8.1%	91.9%	68.1%
1.5	6.7%	93.3%	70.7%
1.6	5.5%	94.5%	73.1%
1.7	4.5%	95.5%	75.4%
1.8	3.6%	96.4%	77.4%
1.9	2.3%	97.7%	79.4%
2.0	2.3%	97.7%	81.1%

The *r-Family* Effect Size

- proportion of variance explained.
- the ratio of variance attributed to an effect versus the total variance.

The formula, which is commonly known as eta-squared:

$$\eta^2 = \frac{SS_{model}}{SS_{total}}$$

An Example of the *r*-Family Effect Size

From a research study on wages, we regress the worker's gender on wages and calculate eta-squared:

Source	SS	df	MS	Number of obs = 534		
Model	593.713779	1	593.713779	F(1, 532) = 23.43		
Residual	13482.9849	532	25.3439566	Prob > F = 0.0000		
Total	14076.6987	533	26.4103165	R-squared = 0.0422		
				Adj R-squared = 0.0404		
				Root MSE = 5.0343		

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sex	-2.116056	.4371957	-4.84	0.000	-2.974898	-1.257215
_cons	9.994913	.296134	33.75	0.000	9.413178	10.57665

Note that Eta-squared for the model and for the variable "sex" equal R-squared.

Effect sizes for linear models

Source	Eta-Squared	df	[95% Conf. Interval]	
Model	.0421771	1	.0150494	.0801068
sex	.0421771	1	.0150494	.0801068

An Example of the *r*-Family Effect Size (cont.)

Now we want to add the indicator variable for union membership

Source	SS	df	MS	Number of obs = 534		
Model	835.81628	2	417.90814	F(2, 531) = 16.76		
Residual	13240.8824	531	24.9357484	Prob > F = 0.0000		
Total	14076.6987	533	26.4103165	R-squared = 0.0594		
				Adj R-squared = 0.0558		
				Root MSE = 4.9936		

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sex	-1.901207	.439108	-4.33	0.000	-2.763809	-1.038605
union	1.775493	.5698107	3.12	0.002	.6561335	2.894853
_cons	9.57715	.3228908	29.66	0.000	8.94285	10.21145

Now that we have added an additional variable eta-squared for the model equals R-squared and the eta-squared for the variables do not. “sex” explains 3.4% of “wages” variance while union only explains about half of it. Note that both were statistically significant at 0.05

Effect size measures magnitude of relationship

Source	Eta-Squared	df	[95% Conf. Interval]	
Model	.0593759	2	.0248309	.1000036
sex	.0341	1	.0102594	.0694044
union	.0179562	1	.0024553	.0462399

An Example of the *r-Family* Effect Size (cont.)

Now lets add the indicator variable for marital status

Source	SS	df	MS	Number of obs = 534		
Model	951.629532	3	317.209844	F(3, 530) = 12.81		
Residual	13125.0692	530	24.7642814	Prob > F = 0.0000		
				R-squared = 0.0676		
				Adj R-squared = 0.0623		
				Root MSE = 4.9764		
Total	14076.6987	533	26.4103165			

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sex	-1.926098	.437747	-4.40	0.000	-2.78603	-1.066166
union	1.656903	.57049	2.90	0.004	.5362036	2.777602
marr	.9845776	.4552854	2.16	0.031	.0901923	1.878963
_cons	8.964568	.4286988	20.91	0.000	8.12241	9.806725

All three variables are statistically significant at 0.05. But notice that marital status amounts to less than 1% of wages variance.

Source	Eta-Squared	df	[95% Conf. Interval]	
Model	.0676032	3	.0290796	.1082288
sex	.0352414	1	.0108903	.0709863
union	.0156662	1	.0016375	.0426903
marr	.0087466	1	.	.0310048

An Example of the *r-Family* Effect Size (cont.)

Using “anova” for our partial sum of squares

```
. anova wage sex union marr
```

Number of obs = 534 R-squared = 0.0676
Root MSE = 4.97637 Adj R-squared = 0.0623

Source	Partial SS	df	MS	F	Prob > F
Model	951.629532	3	317.209844	12.81	0.0000
sex	479.441928	1	479.441928	19.36	0.0000
union	208.892904	1	208.892904	8.44	0.0038
marr	115.813252	1	115.813252	4.68	0.0310
Residual	13125.0692	530	24.7642814		
Total	14076.6987	533	26.4103165		

Source	Eta-Squared	df	[95% Conf. Interval]	
Model	.0676032	3	.0290796	.1082288
sex	.0352414	1	.0108903	.0709863
union	.0156662	1	.0016375	.0426903
marr	.0087466	1	.	.0310048

$$\eta^2 = \frac{SS_{model}}{SS_{total}} = \frac{951.6}{14,076.7} = 0.676$$

$$\eta^2 = \frac{SS_{sex}}{SS_{total}} = \frac{479.4}{14,076.7} = 0.034$$

$$\eta^2 = \frac{SS_{union}}{SS_{total}} = \frac{208.9}{14,076.7} = 0.014$$

$$\eta^2 = \frac{SS_{marr}}{SS_{total}} = \frac{115.8}{14,076.7} = 0.008$$

Effect Size indexes

d-family: continuous outcomes

Cohen's d : standardized mean difference between two groups based on the pooled standard deviation (formula shown on slide 14)

Hedges' g : standardized mean difference between two groups based on the pooled, weighted standard deviation (formula shown on slide 14)

Glass's d : standardized mean difference between two groups based on the standard deviation of the control group

Probability of superiority: the probability that a random value from one group will be greater than a random value drawn from another

The Essential Guide to Effect Sizes, Table 1.1

Effect Size indexes

d-family: dichotomous outcomes

Risk difference in probabilities: difference between the probability of an event or outcome occurring in two groups.

Risk or rate ratio: compares the probability of an event or outcome happening in one group with the probability of it occurring in another

Odds ratio: compares the odds of an event or outcome occurring in one group with the odds of it occurring in another

The Essential Guide to Effect Sizes, Table 1.1

Effect Size indexes

r-family: proportion of variance indexes

The coefficient of determination (r^2): used in bivariate regression analysis.

R squared (R^2): used in multiple regression analysis

Adjusted R squared ($_{adj}R^2$): adjusted for sample size and the number of predictor variables

Eta squared (η^2) : used in ANOVA and regression analysis, uses sample variance.

Partial Eta-squared (η_p^2): the proportion of the total variability attributable to a given factor after excluding variance explained by other predictor variables.

Epsilon squared (\mathcal{E}^2): an unbiased alternative to η^2

Omega squared (ω^2): an unbiased alternative to η^2 . Uses population variance.

Cohen's f : quantifies the dispersion of means in three or more groups

Cohen's f^2 : an alternative to R^2 in regression analysis and ΔR^2 in hierarchical regression analysis. Represents the proportion of explained variance over unexplained variance.

The Essential Guide to Effect Sizes, Table 1.1

Effect Size indexes

r-family: correlation indexes

Pearson product moment correlation coefficient(r): used when both variables are measured on an interval or ratio scale.

Spearman's rho(ρ): used when both variables are measured on an ordinal or ranked scale

Kendall's tau(t): used when both variables are measured on an ordinal or ranked scale;
tau-b is used for square-shaped tables: tau-c for rectangular tables

Point-biserial correlation coefficient: used when one variable is measured on a binary scale and the other variable is continuous.

Phi coefficient (ψ): used when variables and effects can be arranged in a 2 x 2 contingency table

Pearson's contingency coefficient (C): used when variables and effects can be arranged in a contingency table of any size

Cramer's V : adjusted version of phi that can be used with categorical variables with more than two categories.

Goodman & Kruskal's lambda(λ): used when both variables are measured on nominal scales

The Essential Guide to Effect Sizes, Table 1.1

Warning!

Jacob Cohen developed the following effect size gauges in his book *Statistical Power Analysis for the Behavioral Sciences*:

Test	Relevant effect size	Effect size classes		
		Small	Medium	Large
Comparison of independent means	d, Δ , Hedges' g	0.20	0.50	0.80
Comparison of two correlations	q	0.10	0.30	0.50
Difference between proportions	Cohen's g	0.05	0.15	0.25
Correlation	ζ_w	0.10	0.30	0.50
	r^2	0.01	0.09	0.25
Crosstabulation	ϕ, V, C	0.10	0.30	0.50
ANOVA	f	0.10	0.25	0.10
	η^2	0.01	0.06	0.14
Multiple Regression	R^2	0.02	0.13	0.26
	f^2	0.02	0.15	0.35

It is tempting to use the above guidelines when reporting effect size but definitely not advisable. Every field of research has different standards. Discussion should be focused within the context of previously reported effect size for similar research.

Careful Which Effect Size You Report

1988 Physicians' Health Study Research Group Analysis on the Effect of Aspirin on Reducing Heart Attacks

Treatment Group = 11,000 doctors taking one low-dose aspirin daily over a five year period
Control Group = 11,000 doctors taking one placebo daily over a five year period

Proportion-of-variance accounted for, r^2 , was extremely small (.001)

Study was discontinued due to unfairness to those taking the placebo, they had a significantly greater chance of having a heart attack.

There was a 44% reduction in the risk of myocardial infarction in the aspirin group:
254.8 per 100,000 per year vs 439.7 in the placebo group.
Relative risk 0.56, 95% CI 0.45 to 0.70

References:

Breaugh, James. "Effect Size Estimation: Factors to Consider and Mistakes to Avoid", Journal of Management, 2003 29(1), 79-97

Coe, Robert. (2002) "It's the Effect Size, Stupid. What effect size is and why it is important" Paper presented at the Annual Conference of the British Educational Research Association.
<http://www.leeds.ac.uk/educol/documents/00002182.htm>

Durlak, Joseph. "How to Select, Calculate, and Interpret Effect Sizes" Journal of Pediatric Psychology 34(9) pp. 917-928, 2009 <http://jpepsy.oxfordjournals.org/>

Ellis, Paul (2010) The Essential Guide to Effect Sizes, Cambridge University Press. Kindle Edition

Kirk, Roger. "Practical Significance: A Concept Whose Time Has Come" Educational and Psychological Measurement, 1996 56:746 <http://epm.sagepub.com/content/56/5/746.refs.html>

Kline, Rex (2004) Beyond Significance Testing, American Psychological Association, Washington D.C.

Lakens, Daniel. "Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs", Frontiers in Psychology, November, 2013
<http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00863/abstract>

Thompson, Bruce. "Effect Sizes, Confidence Intervals, and Confidence Intervals for Effect Sizes", Psychology in the Schools, Vol 44(5), 2007
<http://onlinelibrary.wiley.com/doi/10.1002/pits.20234/abstract>

Wuensch, Karl. "Standardized Effect Size Estimation: Why and How" (2015) copy and paste following into your browser: core.ecu.edu/psyc/wuenschk/StatHelp/Effect%20Size%20Estimation.pdf
http://www.psychometrica.de/effect_size.html