Chapter 6
Compare Two Groups

Section 6.3: Standardized Mean Difference, *d*

David W. Gerbing

The School of Business
Portland State University

- Standardized Mean Difference, *d*
  - Definition and Meaning
  - Separation of Distributions
  - Application

6.3a
Definition and Meaning

Statistical Significance

An hypothesis test is the search for statistical significance

- Does a mean difference exist, that is, does $\mu_1 - \mu_2 \neq 0$?
- Regarding the relationship of grouping variable X with response variable Y, an hypothesis test results in either

  "Yes, one true mean is likely larger than the other mean"
  or
  "No difference between true means detected"

- The *p*-value from the hypothesis test provides evidence as to the existence of a mean difference, but . . .
- **Key Concept**: If significant, less than $\alpha = 0.05$, the *p*-value by itself provides *no* information regarding the size of the difference between μ_1 and μ_2

Effect Size

Strength of a relationship

- **Effect Size**: The magnitude of an existing difference
- The effect size is the separation between the group means, expressed, for example, directly as the mean difference
  - sample: $m_1 - m_2$
  - population: $\mu_1 - \mu_2$, estimated with the confidence interval
- The effect size answers the question:

  If the group means are not equal, $\mu_1 - \mu_2 \neq 0$, how large is is the difference between them?

- That is, the difference between μ_1 and μ_2 is real, but . . .
  - Is $\mu_1 - \mu_2$ small and more likely not of interest even though it is nonzero?
  - Is $\mu_1 - \mu_2$ large and more likely of interest?

Practical Importance

Ultimate focus for the decision maker

- **Practical Importance**: An assessment as to whether the effect size is large enough to be useful
- Consider the effect of remodeling a retail outlet: compare average sales on a daily basis before and after the remodeling
  - If average sales per day increases, but only by $0.01, the investment in remodeling is not justified
  - An average increase of $148.34 per day may have justified the investment, so the effect would be of practical importance
- The effect size provides more information than provided by just the *p*-value
  - if *p*-value $> \alpha = 0.05$, then conclude a relationship exits
  - effect size indicates the *magnitude* of this relationship

Standardized Mean Difference

Traditionally called Cohen's *d*, after Jacob Cohen

- Express the mean difference in the sample as $m_1 - m_2$, and in the population as $\mu_1 - \mu_2$
- One assessment of effect size is the mean difference itself, expressed with units such as $ USD, length, and hours
- Effect size can also be expressed in the universal standardized metric, as a *standardized* mean difference, based on the standard deviation of the data
- There are two samples of data, with two standard deviations, s_1 and s_2, so standardize with the within-group standard deviation, s_w, calculated as an average from both samples
- **Standardized mean difference**: Number of standard deviations of the data that separate the group means

Sample: $d = \dfrac{m_1 - m_2}{s_w}$ Population: $\delta = \dfrac{\mu_1 - \mu_2}{\sigma_w}$

Standardized Mean Difference is Unitless

d and measurement scales

- The size of the expressed mean difference depends, in part, on the scale of measurement
- For example, the same physical distance is expressed with larger numbers if in inches instead of centimeters
- True of all standardized values, *d* is unitless
- **Unitless Statistic**: The value of a statistic does not depend on the measurement scale in which the variable is expressed
- For *d*, this unitless property means that, for example, in the analysis of two different average lengths, the same effect size is obtained if length is measured in inches or centimeters
- This unitless property facilitates the comparison of the strength of an effect across different studies of variable Y in which Y is measured with different measurement scales

Qualitative Description of Effect Size

Some guidelines that often, but not always, apply

- In any one situation the following guidelines may not apply, but they provide a useful context in the interpretation of *d* as long as flexibility in their interpretation is maintained
- Cohen studied many published scientific studies and, when possible from the available information, calculated and classified the resulting values of *d*
- From this analysis, he formulated some general guidelines
  - Trivial Effect: $d < .2$
  - Small Effect: $d > .2$ and $d < .6$
  - Medium Effect: $d > .6$ and $d < .8$
  - Large Effect: $d > .8$

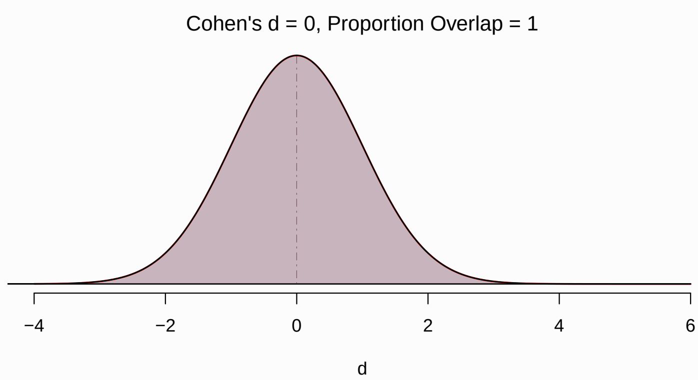
6.3b
Separation of Distributions

Express *d* in Terms of the Overlap Between Distributions

Meaning of the Standardized Mean Difference, *d*

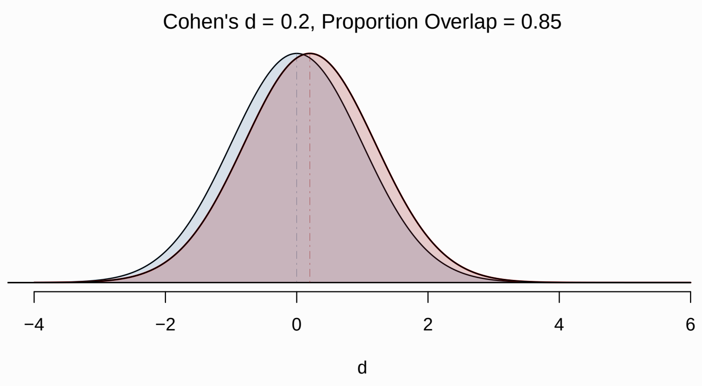
- A focus on the size of the mean difference naturally leads to a consideration of how much the two distributions of Y, one for each group, overlap
- Consider the amount of overlap when Y is normally distributed for each group
- Express the distributions as standardized normal distributions, that is, expressed in terms of standard or *z*-values
- The following examples illustrate the relation of the mean difference to the differences of the *entire* two standardized normal distributions
- The standardized mean difference, *d*, calibrates the separation of distributions, best illustrated for normal distributions
- Question: How distinct are the two distributions depending on the value of *d*

Meaning of $d = 0.0$ in Terms of Distribution Overlap



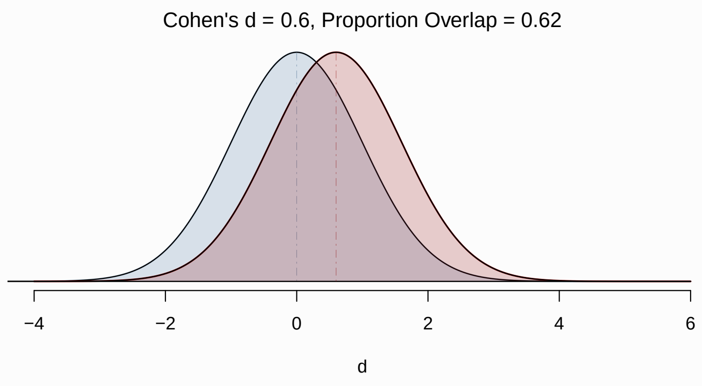
- Null hypothesis is true, $\mu_1 - \mu_2 = 0$, complete overlap

Meaning of $d = 0.2$ in Terms of Distribution Overlap



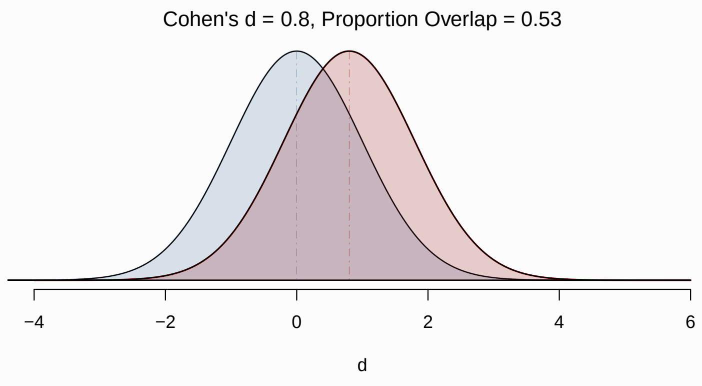
- $\mu_1 - \mu_2 \neq 0$, Effect is small, distributions mostly overlapped

Meaning of $d = 0.6$ in Terms of Distribution Overlap



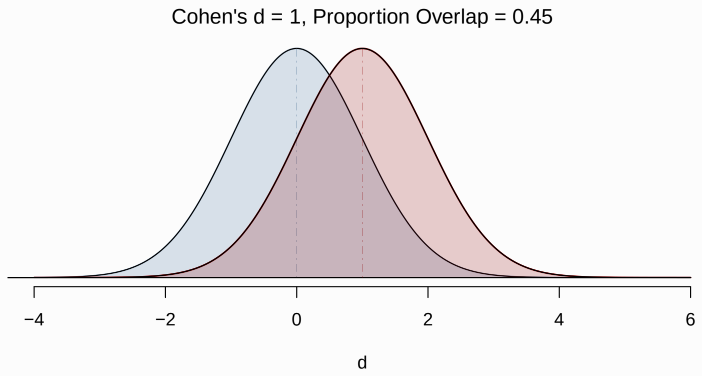
- $\mu_1 - \mu_2 \neq 0$, Effect is moderate, overlap pronounced

Meaning of $d = 0.8$ in Terms of Distribution Overlap



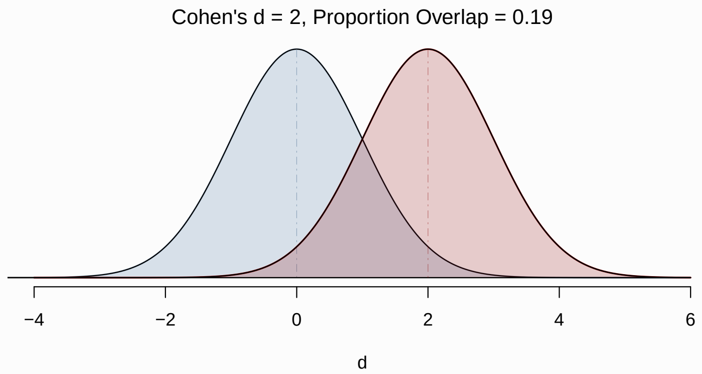
- $\mu_1 - \mu_2 \neq 0$, Effect is large, still more than 50% overlap

Meaning of $d = 1.00$ in Terms of Distribution Overlap



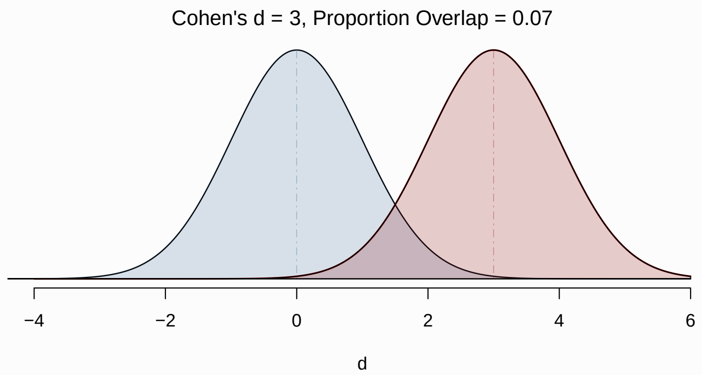
- $\mu_1 - \mu_2 \neq 0$, Effect is quite large, overlap still substantial

Meaning of $d = 2.00$ in Terms of Distribution Overlap



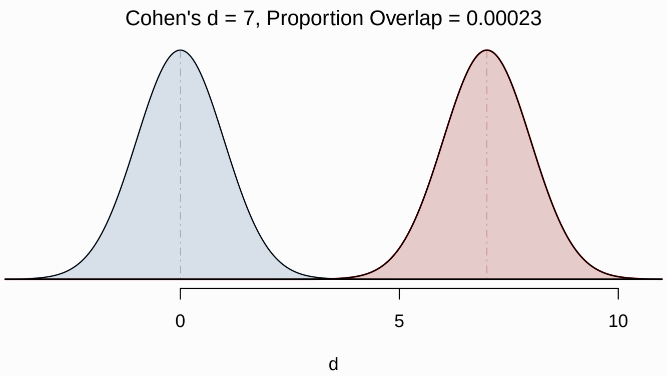
- $\mu_1 - \mu_2 \neq 0$, Effect is very large, overlap still considerable

Meaning of $d = 3.00$ in Terms of Distribution Overlap



- $\mu_1 - \mu_2 \neq 0$, Effect is huge, still 7% percentage overlap

Meaning of $d = 7.00$ in Terms of Distribution Overlap



- 7 standard deviations of separation of means finally results in mostly distinct distributions, that is, with very little overlap

A More Complete Consideration of Differences

The mean difference is one part of a more complete scenario

- The difference between means is just one characteristic of the distinction between two distributions
- Effect size in terms of d can be expressed as the amount of overlap between the distributions of the response variable Y for each group
- Focus on the standardized mean difference, d, changes the emphasis from a summary index, the mean difference, to $\rightarrow$

  a more complete consideration of the relation between two distributions

- Relate the two distributions plotted on the same graph to more completely reveal their differences
- Even a relatively large mean difference can reflect a substantial overlap between the distributions

6.3c
Effect Size and Significance

t-statistic Depends on Size of Effect, d, and Sample Size

d and t can be expressed in terms of each other

- Divide the sample mean difference, $m_1 - m_2$, by either ...
  - standard deviation of the data, to obtain d
  - standard deviation of the mean difference, to obtain t
-
$$t_{m_1 - m_2} = \frac{m_1 - m_2}{s_{m_1 - m_2}} = \frac{m_1 - m_2}{s_w} \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = d \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$
- **Key Concept**: The assessment of statistical significance, assessed with the t-statistic, confounds the magnitude of the relationship, assessed with d, with the sample sizes, n_1 and n_2
- For any given value of effect size d, a significant t-value can be obtained by increasing the sample sizes of the two groups

Statistical Significance vs. Practical Importance

A significant p-value does not imply an important result

- The t-statistic from an hypothesis test, and its associated p-value, do *not* estimate the size of the population mean difference
- Even for a very small separation between the distributions of response variable Y for each group, d slightly larger than 0, large sample sizes, n_1 and n_2, result in statistical significance
- When d is close to but still larger than zero, a difference from the null exists so that the null hypothesis is false, but the result may be of little or no practical importance
- Even with a result too small to be of practical importance, with a large enough sample, statistical significance is *always* obtained

Meaning of t vs. d

Move beyond just significance testing

- t and associated p-value: statistical significance, does a difference between population means exist?
- Mean difference and standardized mean difference, d: size of the difference as it relates to practical importance
- A statistically significant difference between means can *always* be found for sufficiently large sample sizes, no matter how small the difference between population means
- Accordingly, analysis of a mean difference should generally also include a consideration of the minimum mean difference of practical importance
- Express this minimum effect size in terms of the mean difference, in original units of measurement, `mmd`, and/or as the minimum standardized mean difference, `msmd`

6.3d
Application

Begin the Analysis

R instructions

- Invoke the `lessR` t-test function, `ttest()`, or `tt_brief()`, with response variable Time as a function of the grouping variable Supplier

```
> tt_brief(Time ~ Supplier)
```

- This application extends the previous analyses to a consideration of effect size and practical importance
  - Section 6.1: Statement of the problem, data, descriptive statistics, evaluation of stable process and normality assumptions, inferential analysis, conclusion
  - Section 6.2: Evaluation of homogeneity of variance assumption, standard error, inferential analysis not assuming homogeneity of variance

Effect Size

```
--- Effect Size ---

-- Assume equal population variances of Time
   for each Supplier

Sample Mean Difference of Time:  1.49

Standardized Mean Difference of Time, Cohen's d:  0.98

95% Confidence Interval for smd:  0.26 to 1.70
```

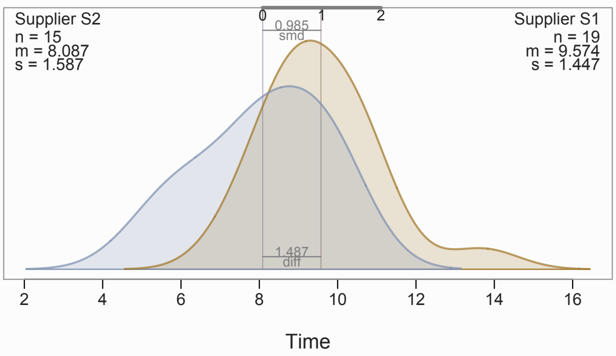
Graphical Display of Effect Size

Compare more than just the means

- Display the mean difference in the metric of
  - unit of measurement of the response variable, Y
  - standardized units of the response variable, Cohen's *d*
- Directly display the overlapping distributions, expressed as smoothed histograms, called density functions
- Observe the difference between the entire distributions
- An additional parameter for invoking `ttest()` specifies the minimum effect size, displayed in the resulting analysis
  - `mmd` for the minimum mean difference, or
  - `msmd` for the minimum standardized mean difference
- For example,

```
> tt_brief(Time ~ Supplier, mmd=0.75)
```


Overlapping Density Curves



Practical Importance

- As is true of any statistic, including the standardized mean difference, *d*, a confidence interval provides the range of plausible values for the corresponding population value, here δ

```
--- Practical Importance ---

Minimum Mean Difference of practical importance: mmd
Compare mmd = 0.75  to the obtained value of md =  1.49
Compare mmd to the confidence interval for md:
   0.42  to  2.55

Minimum Standardized Mean Difference of
   practical importance: msmd
Compare msmd =  0.5  to the obtained value of
   smd =  0.98
Compare msmd to the confidence interval for smd:
   0.26  to  1.7
```

Extent of Graphics Smoothing

```
--- Graphics Smoothing Parameter ---

Density bandwidth for Supplier S1:  0.91
Density bandwidth for Supplier S2:  1.05
```

- Increase the default bandwidth for a smoother density curve
- Decrease the default bandwidth for a density curve with more fluctuations
- Change the density bandwidth from the default with `bw1` and `bw2` options, for the first and second groups, respectively

Index

Subtract 2 from each listed value to get the Slide #

- The End