Effect Size: Comparing Cohen d and Pearson r

Notes:

(1) This PDF is part of YouTube tutorials (https://youtu.be/AdIwx9j0ARM). This PDF is for individual, personal usage only.
(2) The author accepts no responsibility for the topicality, correctness, completeness or quality of the information provided.

Effect size quantifies the strength of the relationship between variables or the magnitude of the difference between groups.

Effect size - Cohen's *d*

Cohen d is used to quantify the size of the difference between two groups, taking into account the variability within each group.

$$d = \frac{m_2 - m_1}{s_{pooled}} = \frac{m_2 - m_1}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

s_1 is the sample standard deviation for group 1, whereas s_2 is the sample standard deviation for group 2.

Effect Size - Pearson Product Moment *r*

The Pearson product-moment correlation coefficient, often referred to as the Pearson correlation coefficient, quantifies the strength of a linear relationship between two variables.

$$r_{xy} = \frac{cov(x, y)}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Cohen's *d* and Pearson's *r*

The following effect size numbers are based on Jacob Cohen. (*reference is included at the end of this tutorial.*)

Effect Size	d	r
Small	0.2	0.1
Medium	0.5	0.3
Large	0.8	0.5

Comments

Cohen's d can assume values ranging from 0 to infinity (i.e., can be greater than 1), whereas Pearson's r is confined to a range between -1 and 1.

Note that, based on my best understanding (not sure), Cohen's d can be negative as well, depending on how to define m_1 and m_2 and which order to put them.

Convert r to d, and vice versa

While Cohen's d looks different from Pearson's r, there is internal connection between these two.

Group 1 and group 2 may be considered to be a dichotomy or a two point scale.

Thus, you can convert r to d, and vice versa.

Convert r to d

$$d = \frac{2r}{\sqrt{1-r^2}}$$

Convert d to r

$$d^2 = \frac{4r^2}{1-r^2}$$

$$d^2 = 4r^2 + r^2d^2 = r^2(4+d^2)$$

$$r^2 = \frac{d^2}{4+d^2}$$

$$r = \frac{d}{\sqrt{4+d^2}}$$

According to Cohen, sometimes, people might want to think of effect sizes for mean differences d in terms of r. Thus, you can convert d to r if needed.

For instance, group 1 and group 2 are two experimental manipulations. The effect size typically is d. But you can convert it into r.

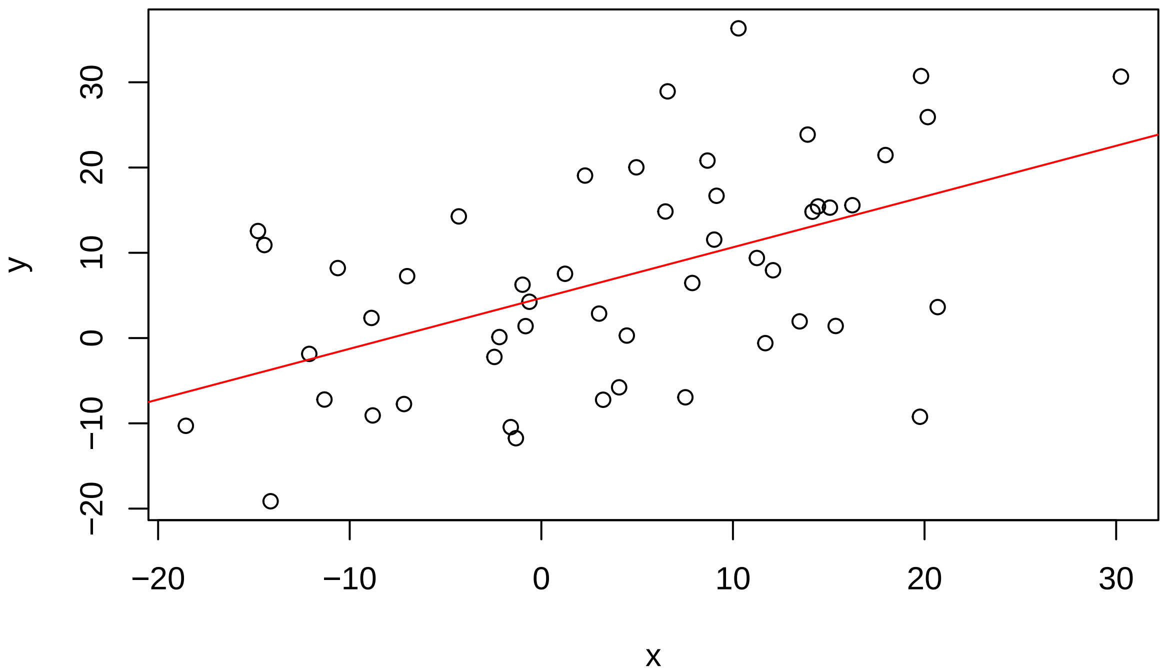
I will provide data to demonstrate the connection between d and r.

Simulate both X and Y that are numerical

```r
# set the seed for reproducibility
set.seed(123)

# specify marginal means and standard deviations
mean_x <- 5
mean_y <- 6
sd_x <- 12
sd_y<- 14

## specify coefficient, which must be between -1 and 1
corcoef <- 0.5

## covariance between two variables
covariance <- corcoef * sd_x * sd_y

## variance-covariance matrix
sigma <- matrix(c(sd_x^2, covariance, covariance, sd_y^2), nrow = 2)

xy <- MASS::mvrnorm(n = 50, mu = c(mean_x, mean_y), Sigma = sigma)
colnames(xy) <- c("x", "y")
df <- data.frame(xy)
head(df)
```

```
##           x          y
## 1 -2.188776  0.1167428
## 2  3.014416  2.8711033
## 3 20.168055 25.9184510
## 4 -4.307292 14.2689957
## 5  7.878421  6.4575243
## 6 10.288163 36.3225845
```

```r
# plot it
plot(df)
abline(lm(y ~ x,data=df), col = "red")
```




```
cor.test(df$x,df$y)
```

```
##
##  Pearson's product-moment correlation
##
## data:  df$x and df$y
## t = 4.2497, df = 48, p-value = 9.8e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2861663 0.6994209
## sample estimates:
##       cor
## 0.5228658
```

r vs. d: Difference and Connection

Effect Size	d	r
Small	0.2	0.1
Medium	0.5	0.3
Large	0.8	0.5

```r
# dichotomize x
df$dichotomized_x<- ifelse(df$x >mean(df$x), 1, 0)
head(df)
```

```
##           x          y dichotomized_x
## 1 -2.188776  0.1167428              0
## 2  3.014416  2.8711033              0
## 3 20.168055 25.9184510              1
## 4 -4.307292 14.2689957              0
## 5  7.878421  6.4575243              1
## 6 10.288163 36.3225845              1
```

Method 1: Direct Calculating d

Since we got a column of dichotomized x, we can actually use the formula mentioned to calculate Cohen's d. That is, calculate the means for groups 1 and 2. Then, calculate the sample standard deviations.

$$d = \frac{m_2 - m_1}{s_{pooled}} = \frac{m_2 - m_1}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

```r
# calculate group means
means<-aggregate(df$y, list(df$dichotomized_x), FUN=mean)

# calculate sample standard deviation - pooled
variances<- aggregate(df$y, list(df$dichotomized_x),FUN=var)
s_pooled=sqrt((variances$x[1]+variances$x[2])/2)

d_calculated<-(means$x[2]-means$x[1])/s_pooled
print(d_calculated)
```

```
## [1] 1.252203
```

Method 2: Converting r to d

$$d = \frac{2r}{\sqrt{1 - r^2}} = \frac{2 * 0.52}{\sqrt{1 - 0.52^2}} = 1.2$$

```r
# check the correlation coefficient
cor_result<-cor.test(df$x,df$y)
correlation_coefficient<-cor_result$estimate
d_converted<-2*correlation_coefficient/sqrt(1-(correlation_coefficient^2))
print(d_converted)
```

```
##      cor
## 1.226787
```

References

1. Jacob Cohen, 1988, Statistical Power Analysis for the Behavioral Sciences, Second Edition
2. What is the relation between the effect size and correlation?

https://stats.stackexchange.com/questions/412590/what-is-the-relation-between-the-effect-size-and-correlation

3. What is Effect Size and Why Does It Matter? (Examples)

https://www.scribbr.com/statistics/effect-size/

4. How to simulate a strong correlation of data with R

https://stackoverflow.com/questions/72894192/how-to-simulate-a-strong-correlation-of-data-with-r

5. FAQ How is effect size used in power analysis?

https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/effect-size-power/faqhow-is-effect-size-used-in-power-analysis/