Effect Size 2

Cohen's *f*

Cohen's *d* can be extended to the multi group situation and is called Cohen's *f*.

$$f = \sqrt{\frac{(\mu_j - \mu)^2}{k\sigma_\epsilon^2}}$$

Cohen's *f* measures the degree of departure from "no effect"

To understand this, consider what "no effect" means

$$\mu_1 = \mu_2 = \ldots \mu_k$$

That would also imply that

$$\mu_j - \mu = 0$$

for all j

$$f = \sqrt{\frac{(\mu_j - \mu)^2}{k\sigma_\epsilon^2}}$$

$$f = \sqrt{\frac{(0)^2}{k\sigma_\epsilon^2}}$$

$$f = 0$$

If there is an effect

$$\mu_j - \mu \neq 0$$

for at least one j

$$f = \sqrt{\frac{(\mu_j - \mu)^2}{k\sigma_\epsilon^2}}$$

That would imply that the numerator is greater than 0.

$$f > 0$$

As the effect gets larger, all things being equal, f would be larger...higher value of f implies more departure from "no effect"

For the Word Recall Data

$$f = \sqrt{\frac{(\mu_j - \mu)^2}{k\sigma_\epsilon^2}}$$

$$\hat{f} = \sqrt{\frac{(6-9)^2 + (10-9)^2 + (11-9)^2}{3(1.574^2)}}$$

$$\hat{f} = 1.37$$

Interpreting the Magnitude of *f*

Cohen provided "rules of thumb" to help applied researchers interpret the magnitude of standardized difference

- $f = 0.10$ corresponds to a small effect size
- $f = 0.25$ corresponds to a medium effect size
- $f = 0.40$ corresponds to a large effect size

The terms "small," "medium," and "large" are relative, not only to each other, but to the area of behavioral science or even more particularly to specific content and research method being employed in any given investigation (Cohen, 1965, p. 23).

Cohen, J. (1965). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

```
# Fit the statistical model using the lm() function
> lm.30 = lm(words ~ 1 + con60 + con180, data = wr)

# Examine the results using the summary() function
> summary(lm.1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.0000     0.5563  10.785 5.08e-10 ***
con60         4.0000     0.7868   5.084 4.92e-05 ***
con180        5.0000     0.7868   6.355 2.67e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.574 on 21 degrees of freedom
Multiple R-squared:  0.6829,   Adjusted R-squared:  0.6527
F-statistic: 22.62 on 2 and 21 DF,  p-value: 5.783e-06
```

To examine whether there is an effect of condition (30s. vs 60s. vs. 180s.) on number of words recalled, a fixed-effects regression model using two dummy variables was fitted to $n = 24$ study participants' data. The results of this analysis were statistically reliable, $F(2, 21) = 22.62$, $p < 0.001$, indicating that there is likely a difference in the mean number of words recalled between the three conditions. Differences in condition accounted for 68.3% of the variation in the number of words recalled.

$$\hat{\eta}^2 = 0.683$$

This is a "variance accounted for" measure of effect rather than a distance measure.

Note that we can also get a "variance accounted for measure" with two groups

```
# Fit the statistical model using the lm() function
> lm.1 = lm(gradRate ~ public, data = mn)

# Examine the results using the summary() function
> summary(lm.1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   63.104      3.446  18.313   <2e-16 ***
public       -12.759      6.147  -2.076   0.0458 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.88 on 33 degrees of freedom
Multiple R-squared:  0.1155,   Adjusted R-squared:  0.08868
F-statistic: 4.309 on 1 and 33 DF,  p-value: 0.04579
```

$$\hat{\eta}^2 = 0.116$$

Cohen's *f* is a function of eta-squared.

$$f = \sqrt{\frac{\eta^2}{1-\eta^2}}$$

$$\hat{f} = \sqrt{\frac{0.116}{1-0.116}} = 1.37$$

Adjusted Eta-Squared

Eta squared tends to overestimate the population "variance accounted for". It is **positively biased**.

$$\text{Adjusted } \eta^2 = 1 - \left[\frac{N-1}{N-p}(1-\eta^2)\right]$$

$$\text{Adjusted } \eta^2 = 1 - \left[\frac{24-1}{24-3}(1-0.6829)\right]$$

$$\text{Adjusted } \eta^2 = 0.6527$$

Eta-squared is adjusted based on the size of the model (p) and the sample size (N)

Omega-Squared

Another measure to account for the overestimation of eta squared is omega-squared.

$$\hat{\omega}^2 = \frac{\mathrm{SS}_{\mathrm{Model}} - (k-1)\hat{\sigma}_\epsilon^2}{\mathrm{SS}_{\mathrm{Total}} + \hat{\sigma}_\epsilon^2}$$

$$\hat{\omega}^2 = \frac{112 - (3-1) \times 2.476}{164 + 2.476}$$

$$\hat{\omega}^2 = 0.64$$

The index omega squared provides is a relative measure of the strength of an independent variable ranging from 0.0 to 1.0.

Cohen provided "rules of thumb" to help applied researchers interpret the magnitude of omega-squared

- $\omega^2 = 0.010$ corresponds to a small effect size
- $\omega^2 = 0.059$ corresponds to a medium effect size
- $\omega^2 = 0.138$ corresponds to a large effect size

It is unlikely that high omega-squared values will be seen, because of the large contribution of error variance in most behavioral research.

It is possible that when computing ω^2, a researcher can end up with a result that is a negative number. If this should happen, the effect size should be set to 0.

Note: This measure of effect size should only be used for fixed-effect models.

```
# Examine the results using the summary() function
> summary(lm.1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.0000     0.5563  10.785 5.08e-10 ***
con60         4.0000     0.7868   5.084 4.92e-05 ***
con180        5.0000     0.7868   6.355 2.67e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.574 on 21 degrees of freedom
Multiple R-squared:  0.6829,	Adjusted R-squared:  0.6527
F-statistic: 22.62 on 2 and 21 DF,  p-value: 5.783e-06
```

$\hat{\eta}^2$, adjusted-$\hat{\eta}^2$, and $\hat{\omega}^2$ are all point estimates for the population η^2

In the regression paradigm, η^2 is usually referred to as R^2

Which estimate of the effect size that you report is up to you. However, they have some strengths and weaknesses depending on the purpose of the study

- $\hat{\eta}^2$ is generally an overestimate of the true η^2. So if the goal is to estimate η^2 then this might not be the best measure of effect to report.
- Adjusted-$\hat{\eta}^2$ and $\hat{\omega}^2$ are adjusted based on sample size and model size. Because of this, they may not be the best measures to report when a goal is to compare different models.

Typically, $\hat{\eta}^2$ (R^2) is always reported and is supplemented by CIs for η^2 or by one of the less biased measures.

Computing Confidence Intervals for Eta-Squared

We will use the `ci.R2()` function from the **MBESS** library to compute the estimate of the standardized effect.

```
# Load MBESS library
> library(MBESS)

# Point estimate for the standardized effect
> ci.R2(R2 = 0.6829, df.1 = 2, df.2 = 21, Random.Predictors = FALSE)

$Lower.Conf.Limit.R2
[1] 0.3703557

$Prob.Less.Lower
[1] 0.025

$Upper.Conf.Limit.R2
[1] 0.7879404

$Prob.Greater.Upper
[1] 0.025
```

The CI for eta-squared is [0.37, 0.79].

Effect Size for Contrasts

We can also compute effect sizes for each of the *post hoc* contrasts that we test. A first step is to compute CIs.

Use the `glht()` function from **multcomp** to set things up

```
# Set up the contrasts to test
> contr = c("`30s` - `60s` = 0",
            "`30s` - `180s` = 0",
            "`60s` - `180s` = 0")

# Fit a model using the grouping variable
> lm.fac = lm(words ~ condition, data = wr)

# Load multcomp library (you may need to install it first)
> library(multcomp)

# Use the ANOVA model to test pairwise contrasts
> glht.1 = glht(lm.fac, linfct = mcp(condition = contr))
```

Simultaneous Intervals

```
# Assign the summary output to an object
> benHoch = summary(glht.1, test = adjusted("BH"))

# Get the Benjamani-Hochberg adjusted confidence intervals
> confint(benHoch)

Linear Hypotheses:
                    Estimate lwr     upr
`30s` - `60s` == 0  -4.0000  -5.9827 -2.0173
`30s` - `180s` == 0 -5.0000  -6.9827 -3.0173
`60s` - `180s` == 0 -1.0000  -2.9827  0.9827
```

The confidence intervals computed have been adjusted (similar to the p-values) so that they are a little wider (more uncertainty in the estimates) because of the number of comparisons. Adjusted interval estimates are referred to as **simultaneous intervals**.

Standardized Effects and Intervals

A standardized effect and interval could be computed for each contrast. Here we show it for the 30s. vs. 60s. contrast

```
# Assign the summary output to an object
> smd(
    Mean.1 = 6,
    s.1 = 1.51,
    n.1 = 8,
    Mean.2 = 10,
    s.2 = 1.41,
    n.2 = 8
    )

[1] -2.738121

> ci.smd(smd = -2.738121, n.1 = 8, n.2 = 8)

$Lower.Conf.Limit.smd
[1] -4.117425

$smd
[1] -2.738121

$Upper.Conf.Limit.smd
[1] -1.311351
```

"Variance Accounted For" Measures

We can also compute R^2 for each of the contrasts

$$R^2 = \frac{\mathrm{SS}_{\mathrm{Model}}}{\mathrm{SS}_{\mathrm{Total}}}$$

Omnibus

$$R^2 = \frac{\mathrm{SS}_{\mathrm{Contrast}}}{\mathrm{SS}_{\mathrm{Total}}}$$

Contrast

To get the SS for the contrast we need to further explore contrast testing.

Contrasts

A contrast is a statistical comparison of two or more group means. More formally, it is a mathematically expressed sum (or difference) of the group means, where each mean is multiplied by some coefficient (or weight). These expressions, are called **linear combinations**, can be symbolically written as

$$\psi = w_1(\mu_1) + w_2(\mu_2) + \ldots + w_k(\mu_k)$$

Consider the hypothesis in which we compare 30s. vs. 60s.

$$H_0 : \mu_{\text{30s.}} - \mu_{\text{60s.}} = 0$$

The contrast is

$$\psi = \mu_{\text{30s.}} - \mu_{\text{60s.}} \longrightarrow H_0 : \psi = 0$$

Which is equivalent to

$$\psi = (1)\mu_{\text{30s.}} - (1)\mu_{\text{60s.}}$$

When using software, we have to include all of the group means for the factor, not only the means of the groups we are interested in. For us, this means also including the mean for the 180s. condition in the contrast.

$$\psi = (1)\mu_{30\text{s.}} + (-1)\mu_{60\text{s.}} + (0)\mu_{180\text{s.}}$$

We can estimate the contrast using the observed data

$$\hat{\psi} = (1)\hat{\mu}_{30\text{s.}} + (-1)\hat{\mu}_{60\text{s.}} + (0)\hat{\mu}_{180\text{s.}}$$

$$\begin{aligned}\hat{\psi} &= (1)(6) + (-1)(10) + (0)(11) \\ &= -4\end{aligned}$$

To test contrasts, we create a vector of the contrast weights (w_j) and then use that in the `linfct=` argument of the `glht()` function.

$$\psi = (1)\mu_{\text{30s.}} + (-1)\mu_{\text{60s.}} + (0)\mu_{\text{180s.}}$$

$$\mathbf{w} = [1, -1, 0]$$

Contrast weights must sum to zero.

$$\text{SS}_\psi = \frac{\hat{\psi}^2}{\sum \frac{w_j^2}{n_j}}$$

$$\text{SS}_\psi = \frac{(-4)^2}{\frac{(1)^2}{8} + \frac{(-1)^2}{8} + \frac{(0)^2}{8}}$$

$$\text{SS}_\psi = 64$$

To estimate the sum of squares for the contrast for 30s. vs. 180s.

$$\psi = (1)\mu_{\text{30s.}} + (0)\mu_{\text{60s.}} + (-1)\mu_{\text{180s.}}$$

$$\mathbf{w} = [1, 0, -1]$$

$$\text{SS}_{\psi} = \frac{\hat{\psi}^2}{\sum \frac{w_j^2}{n_j}} = \frac{(-5)^2}{\frac{(1)^2}{8} + \frac{(0)^2}{8} + \frac{(-1)^2}{8}} = 100$$

To estimate the sum of squares for the contrast for 60s. vs. 180s.

$$\psi = (0)\mu_{\text{30s.}} + (1)\mu_{\text{60s.}} + (-1)\mu_{\text{180s.}}$$

$$\mathbf{w} = [0, 1, -1]$$

$$\text{SS}_{\psi} = \frac{\hat{\psi}^2}{\sum \frac{w_j^2}{n_j}} = \frac{(-1)^2}{\frac{(0)^2}{8} + \frac{(1)^2}{8} + \frac{(-1)^2}{8}} = 4$$

$$R^2 = \frac{SS_{\text{Contrast}}}{SS_{\text{Total}}}$$

$$SS_{\text{Total}} = 164$$

This comes from the anova() output for the omnibus regression or anova model

Contrast	$SS_{Contrast}$	SS_{Total}	R^2
30s. – 60s.	64	164	0.390
30s. – 180s.	100	164	0.610
60s. – 180s.	4	164	0.024

Note that in this case, the $SS_{Contrast}$ values **do not** sum to the SS_{Total} value. This is because the three contrast vectors are not mutually orthogonal to one another.

Orthogonality

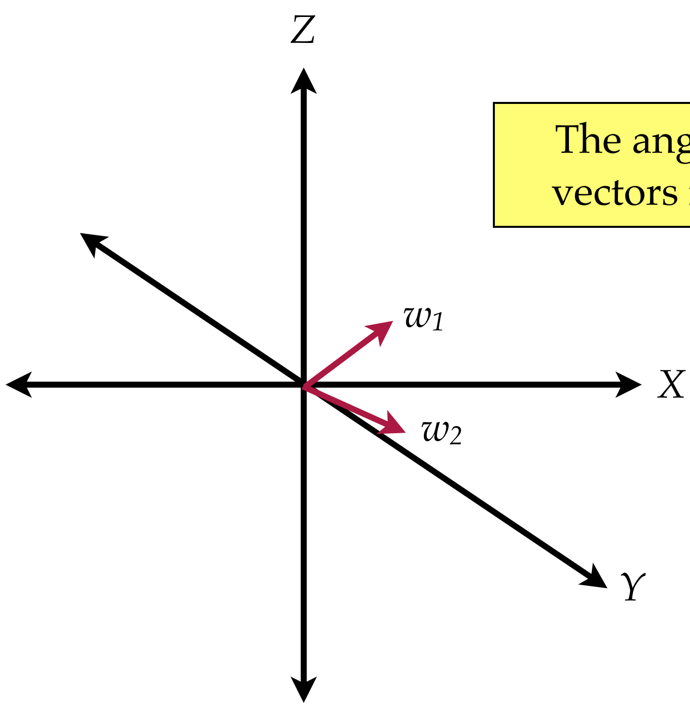
Vectors are orthogonal if the meet at a 90 degree angle.

The three contrast vectors can be represented in 3-dimensional space.

$$\mathbf{w_1} = [1, -1, 0]$$

$$\mathbf{w_2} = [1, 0, -1]$$

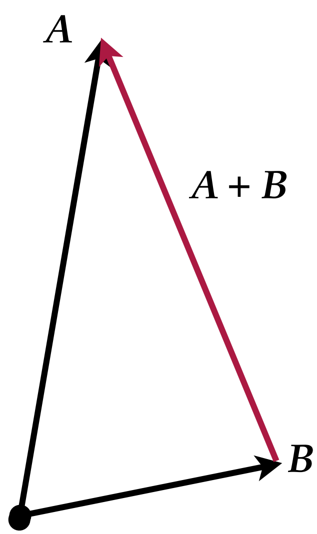
$$\mathbf{w_3} = [0, 1, -1]$$



The angle between these vectors is not 90 degrees!

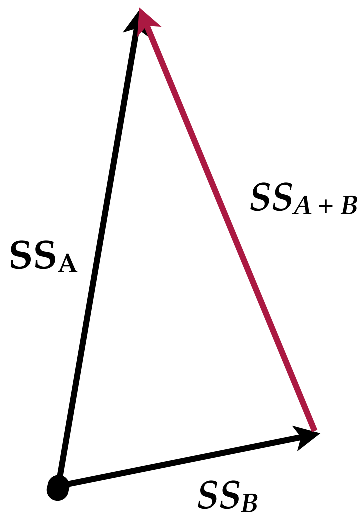
Adding Vectors

We sum vectors by placing them end-to-end and computing the length from tip to tip.



Sum of Squares

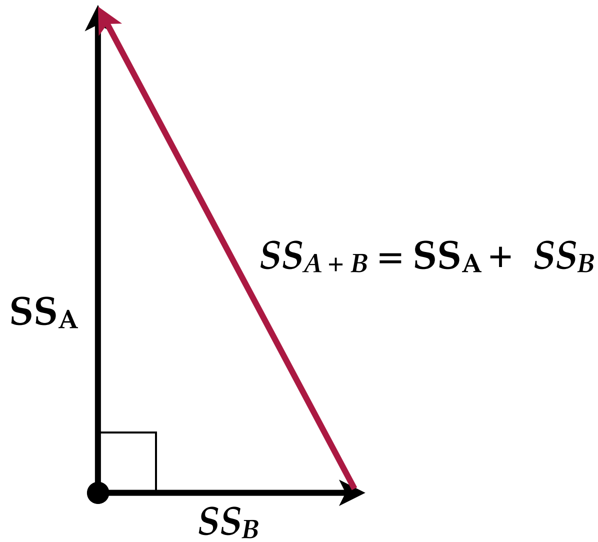
Sums of Squares are lengths of particular vectors.



In this triangle, the side length SS_{A+B} < the side length SS_A + the side length SS_B

$S_A + SS_B > SS_{A+B}$

When would $S_A + SS_B = SS_{A+B}$?



$S_A + SS_B = SS_{A+B}$

when the vectors A and B are orthogonal

$$SS_{\psi_1} + SS_{\psi_2} + SS_{\psi_3} > SS_{\text{Total}}$$

Another term for orthogonal is uncorrelated. The contrasts are correlated with one another because they share information about the group differences.