Effect Sizes and Power Analyses

Nathaniel E. Helwig

Assistant Professor of Psychology and Statistics
University of Minnesota (Twin Cities)

Updated 04-Jan-2017

Copyright

Outline of Notes

1) Effect Sizes:

- Definition and Overview
- Correlation ES Family
- Some Examples
- Difference ES Family
- Some Examples

2) Power Analyses:

- Definition and Overview
- One sample t test
- Two sample t test
- One-Way ANOVA
- Multiple regression

Effect Sizes

What is an Effect Size?

An effect size (ES) measures the strength of some phenomenon:

- Correlation coefficient
- Regression slope coefficient
- Difference between means

ES are related to statistical tests, and are crucial for

- Power analyses (see later slides)
- Sample size planning (needed for grants)
- Meta-analyses (which combine ES from many studies)

Population versus Sample Effect Sizes

Like many other concepts in statistics, we distinguish between ES in the population versus ES in a given sample of data:

- Correlation: ρ versus r
- Regression: β versus $\hat{\beta}$
- Mean Difference: $(\mu_1 - \mu_2)$ versus $(\bar{x}_1 - \bar{x}_2)$

Typically reserve Greek letter for population parameters (ES) and Roman letter (or Greek-hat) to denote sample estimates.

Effect Sizes versus Test Statistics

Sample ES measures are related to (but distinct from) test statistics.

- ES measures strength of relationship
- TS provides evidence against H_0

Unlike test statistics, measures of ES are not directly related to significance (α) levels or null hypotheses.

Standardized versus Unstandardized Effect Sizes

Standardized ES are unit free

- Correlation coefficient
- Standardized regression coefficient
- Cohen's *d*

Unstandardized ES depend on unit of measurement

- Covariance
- Regression coefficient (unstandardized)
- Mean difference

Overview of Correlation Effect Size Family

Measures of ES having to do with how much variation can be explained in a response variable Y by a predictor variable X.

Some examples of correlation ES include:

- Correlation coefficient
- R^2 and Adjusted R^2
- η^2 and ω^2 (friends of R^2 and R_a^2)
- Cohen's f^2

Correlation Coefficient

Given a sample of observations (x_i, y_i) for $i \in \{1, \ldots, n\}$, Pearson's product-moment correlation coefficient is defined as

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

where

- $s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$ is sample covariance between x_i and y_i
- $s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$ is sample variance of x_i
- $s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$ is sample variance of y_i

Measures strength of linear relationship between X and Y.

Coefficient of Multiple Determination

The coefficient of multiple determination is defined as

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

and gives the amount of variation in Y that is explained by $X_1, \ldots, X_p$

When interpreting R^2 values, note that...

- $0 \leq R^2 \leq 1$ so contains no directional information
- Larger R^2 values imply stronger relationship in given sample

Adjusted Coefficient of Multiple Determination (R_a^2)

The adjusted R^2 is a relative measure of fit:

$$R_a^2 = 1 - \frac{SSE/df_E}{SST/df_T} = 1 - \frac{\hat{\sigma}^2}{s_Y^2}$$

where $s_Y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$ is the sample estimate of the variance of Y.

Note that $R^2 = 1 - \tilde{\sigma}^2/\tilde{s}_Y^2$ where

- $\tilde{\sigma}^2 = SSE/n$ is MLE of error variance
- $\tilde{s}_Y^2 = SST/n$ is MLE of variance of Y

so R_a^2 replaces the biased estimates $\tilde{\sigma}^2$ and $\tilde{s}_Y^2$ with the unbiased estimates $\hat{\sigma}^2$ and s_Y^2 in definition of R^2.

ANOVA Coefficient of Determination (η^2 and η_k^2)

In the ANOVA literature, R^2 is typically denoted using

$$\eta^2 = \frac{SSR}{SST}$$

which is the amount of variation in Y attributable to group membership.

Also could consider partial η^2 for k-th factor

$$\eta_k^2 = \frac{SSR_k}{SST}$$

which is the proportion of variance in Y that can be explained by the k-th factor after controlling for the remaining factors.

Calculating η_k^2 in R (Balanced ANOVAs)

R's `aov` function does not calculate this, but you can (easily) write your own function for this using output of `anova` function:

```
eta.sq <- function(mod,k=NULL){
  atab = anova(mod)
  if(is.null(k)){ k = 1:(nrow(atab)-1) }
  sum(atab[k,2]) / sum(atab[,2])
}
```

This function is only appropriate for balanced multiway ANOVAs.

Adjusted ANOVA Coefficient of Determination (ω^2)

Note that η^2 suffers from the same over-fitting issues as R^2:

- If you add more groups, you will have higher η^2

For a one-way ANOVA we could adjust η^2 as follows

$$\omega^2 = \frac{SSB - df_B SSW/df_W}{SST + SSW/df_W}$$

where *SSB* and *SSW* are the SS Between and Within groups.

- Note that ω^2 is less biased estimate of population η^2

Calculating ω^2 in R (One-Way ANOVA)

R's `aov` function does not calculate this, but you can (easily) write your own function for this using output of `anova` function:

```
omega.sq <- function(mod){
  atab = anova(mod)
  ssb = atab[["Sum Sq"]][1]
  ssw = atab[["Sum Sq"]][2]
  dfb = atab[["Df"]][1]
  msw = atab[["Mean Sq"]][2]
  (ssb - dfb*msw) / (ssb + ssw + msw)
}
```

Cohen's f^2 Measure

Jacob Cohen's f^2 measure is defined as

$$f^2 = \frac{X^2}{1 - X^2}$$

where X^2 is some R^2-like measure.

Can define f^2 using any measure we've discussed so far:

- Regression: $f^2 = \frac{R^2}{1-R^2}$
- ANOVA: $f^2 = \frac{\eta^2}{1-\eta^2}$

Note that f^2 increases as R^2 (or η^2) increases.

Cohen's f^2 Measure for "Hierarchical" Regression[1]

Suppose we have a regression model with two sets of predictors:

- A: contains predictors we want to control for (i.e., condition on)
- B: contains predictors we want to test for

Suppose there are q predictors in set A and $p - q$ predictors in set B.

- Model A: $y_i = b_0 + \sum_{j=1}^{q} b_j x_{ij} + e_i$
- Model AB: $y_i = b_0 + \sum_{j=1}^{p} b_j x_{ij} + e_i$

Can use a version of Cohen's f^2 to examine contribution of B given A:

$$f_{B|A}^2 = \frac{R_{AB}^2 - R_A^2}{1 - R_{AB}^2}$$

[1] Note that this has nothing to do with hierarchical linear models (multilevel models).

Example 1: One-Way ANOVA

```
> set.seed(1)
> g = factor(sample(c(1,2,3),100,replace=TRUE))
> e = rnorm(100)
> mu = rbind(c(0,0.05,0.1),c(0,0.5,1),c(0,5,10))
> eta = omega = rep(NA,3)
> for(k in 1:3){
+     y = 2 + mu[k,g] + e
+     mod = lm(y~g)
+     eta[k] = summary(mod)$r.squared
+     omega[k] = omega.sq(mod)
+ }
> eta
[1] 0.03222293 0.22131646 0.94945042
> omega
[1] 0.01214756 0.20362648 0.94791418
```

Example 2: Two-Way ANOVA

```
> A = factor(rep(c("male","female"),each=12))
> B = factor(rep(c("a","b","c"),8))
> set.seed(1)
> e = rnorm(24)
> muA = c(0,2)
> muB = c(0,1,2)
> y = 2 + muA[A] + muB[B] + e
> mod = aov(y~A+B)
> eta.sq(mod)
[1] 0.6710947
> eta.sq(mod,k=1)
[1] 0.4723673
> eta.sq(mod,k=2)
[1] 0.1987273
```

Example 3: Simple Regression

```
> set.seed(1)
> x = rnorm(100)
> e = rnorm(100)
> bs = c(0.05,0.5,5)
> R = Ra = rep(NA,3)
> for(k in 1:3){
+       y = 2 + bs[k]*x + e
+       smod = summary(lm(y~x))
+       R[k] = smod$r.squared
+       Ra[k] = smod$adj.r.squared
+ }
> R
[1] 0.002101513 0.179579589 0.956469774
> Ra
[1] -0.008081125  0.171207952  0.956025588
```

Example 4: Multiple Regression (GPA Data)

```
> gpa = read.csv(paste(myfilepath,"sat.csv",sep=""),header=TRUE)
>
> g1mod = lm(univ_GPA~high_GPA,data=gpa)
> Rsq1 = summary(g1mod)$r.squared
> g2mod = lm(univ_GPA~high_GPA+math_SAT,data=gpa)
> Rsq2 = summary(g2mod)$r.squared
> (Rsq2-Rsq1)/(1-Rsq2)    # f^2 (math_SAT given high_GPA)
[1] 0.02610875
>
> g1mod = lm(univ_GPA~math_SAT,data=gpa)
> Rsq1 = summary(g1mod)$r.squared
> g2mod = lm(univ_GPA~math_SAT+high_GPA,data=gpa)
> Rsq2 = summary(g2mod)$r.squared
> (Rsq2-Rsq1)/(1-Rsq2)    # f^2 (high_GPA given math_SAT)
[1] 0.4666959
```

Overview of Difference Effect Size Family

Measures of ES having to do with how different various quantities are. For two population means

$$\theta = \frac{\mu_1 - \mu_2}{\sigma}$$

measures standardized difference, where σ is standard deviation.

Some examples of difference ES include:

- Glass's Δ
- Cohen's d
- Hedges's g and g^*
- Root mean square standardized effect (RMSSE)

Gene Glass's Δ (1976)

If group 1 is the "control" group and group 2 is the "test" group use:

$$\Delta = \frac{\bar{x}_1 - \bar{x}_2}{s_1}$$

where s_1 is sample standard deviation of control group.

Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3–8.

Jacob Cohen's *d* (1969)

When $x_{i1} \sim (\mu_1, \sigma^2)$ and $x_{i2} \sim (\mu_2, \sigma^2)$ we can use

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_p}$$

where $s_p = \left\{ \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2} \right\}^{1/2}$ is MLE of the standard deviation σ and $s_j^2 = \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 / (n_j - 1)$ is the sample standard deviation.

Cohen, J. (1969). *Statistical power analysis of the behavioral sciences*. San Diego, CA: Academic Press.

Larry Hedges's g

Hedges's g modifies the denominator of Cohen's d

$$g = \frac{\bar{x}_1 - \bar{x}_2}{s_p^*}$$

where $s_p^* = \left\{ \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} \right\}^{1/2}$ is an unbiased estimate of σ.

Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107–128.

Larry Hedges's g^*

It can be shown that $E(g) = \delta / c(n_1 + n_2 - 2)$ where

$$\delta = \frac{\mu_1 - \mu_2}{\sigma} \qquad \text{and} \qquad c(m) = \frac{\Gamma(m/2)}{\sqrt{\frac{m}{2}}\,\Gamma\left(\frac{m-1}{2}\right)}$$

is a term that depends on the sample size.

Hedges's g^* corrects the bias of g by multiplying it by $c(n_1 + n_2 - 2)$:

$$g^* = c(n_1 + n_2 - 2)g$$

Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107–128.

Mean Difference Effect Sizes in R

```
diffES <-
  function(x,y,type=c("gs","g","d","D")){
    md = mean(x) - mean(y)
    nx = length(x)
    ny = length(y)
    if(type[1]=="gs"){
      m = nx + ny - 2
      cm = gamma(m/2)/(sqrt(m/2)*gamma((m-1)/2))
      spsq = ((nx-1)*sd(x)+(ny-1)*sd(y))/m
      theta = cm*md/sqrt(spsq)
    } else if(type[1]=="g"){
      spsq = ((nx-1)*sd(x)+(ny-1)*sd(y))/(nx+ny-2)
      theta = md/sqrt(spsq)
    } else if(type[1]=="d"){
      spsq = ((nx-1)*sd(x)+(ny-1)*sd(y))/(nx+ny)
      theta = md/sqrt(spsq)
    } else { theta = md/sd(x) }
}
```

Root Mean Square Standardized Effect (RMSSE)

In one-way ANOVA we can use the RMSSE:

$$\Psi = \sqrt{\frac{\frac{1}{k-1}\sum_{j=1}^{k}(\mu_j - \mu)^2}{\sigma^2}} = \sqrt{\frac{1}{k-1}\sum_{j=1}^{k}\delta_j^2}$$

where

- μ_j is j-th group's population mean
- μ is overall population mean
- σ^2 is common variance
- $\delta_j = \frac{\mu_j - \mu}{\sigma}$ is j-th group's standardized population difference

Mean Difference Effect Sizes in R (continued)

RMSSE function for one-way ANOVA model:

```
rmsse <- function(x,g){
  mx = tapply(x,g,mean)
  ng = nlevels(g)
  nx = length(x)
  msd = sum((mx-mean(x))^2)/(ng-1)
  mse = sum((mx[g]-x)^2)/(nx-ng)
  sqrt(msd/mse)
}
```

Example 5: Student's *t* test

```
> set.seed(1)
> e = rnorm(100)
> mu = rbind(c(0,0.05),c(0,0.5),c(0,1))
> gs = g = d = D = rep(NA,3)
> for(k in 1:3){
+     x = rnorm(100,mean=mu[k,1])
+     y = rnorm(100,mean=mu[k,2])
+     gs[k] = diffES(x,y)
+     g[k] = diffES(x,y,type="g")
+     d[k] = diffES(x,y,type="d")
+     D[k] = diffES(x,y,type="D")
+ }
> rtab = rbind(gs,g,d,D)
> rownames(rtab) = c("gs","g","d","D")
> colnames(rtab) = c("small","medium","large")
> rtab
        small     medium      large
gs -0.1172665 -0.3922036 -0.8320117
g  -0.1177130 -0.3936971 -0.8351799
d  -0.1183060 -0.3956804 -0.8393874
D  -0.1226476 -0.4126996 -0.8745153
```

Example 6: One-Way ANOVA (revisited)

```
> set.seed(1)
> g = factor(sample(c(1,2,3),100,replace=TRUE))
> e = rnorm(100)
> mu = rbind(c(0,0.05,0.1),c(0,0.5,1),c(0,5,10))
> rvec = rep(NA,3)
> for(k in 1:3){
+     y = 2 + mu[k,g] + e
+     rvec[k] = rmsse(y,g)
+ }
> rvec
[1] 0.2348609 0.6844976 5.4882692
```

Power Analyses

Some Classification Lingo

Classification Outcomes Table:

	Is Negative	Is Positive
Test Negative	True Negative	False Negative
Test Positive	False Positive	True Positive

Some vocabulary:

- Sensitivity = TP / (TP + FN) $\Longrightarrow$ True Positive Rate
- Specificity = TN / (TN + FP) $\Longrightarrow$ True Negative Rate
- Fall-Out = FP / (FP + TN) $\Longrightarrow$ False Positive Rate
- Miss Rate = FN / (FN + TP) $\Longrightarrow$ False Negative Rate

NHST and Statistical Power

NHST Outcomes Table:

	H_0 True	H_1 True
Accept H_0	True Negative	Type II Error
Reject H_0	Type I Error	True Positive

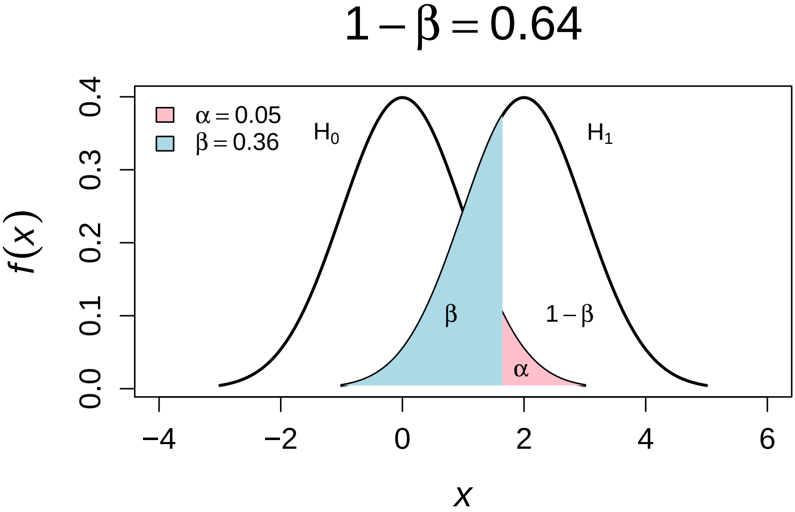
Note: Type I = False Positive, Type II = False Negative

$\alpha = P(\text{Reject } H_0 \mid H_0 \text{ true}) = P(\text{Type I Error})$

$\beta = P(\text{Accept } H_0 \mid H_1 \text{ true}) = P(\text{Type II Error})$

$\text{power} = P(\text{Reject } H_0 \mid H_1 \text{ true}) = 1 - \beta$

Visualizing Alpha, Beta, and Power



Things that Affect Power

Power is influenced by three factors:

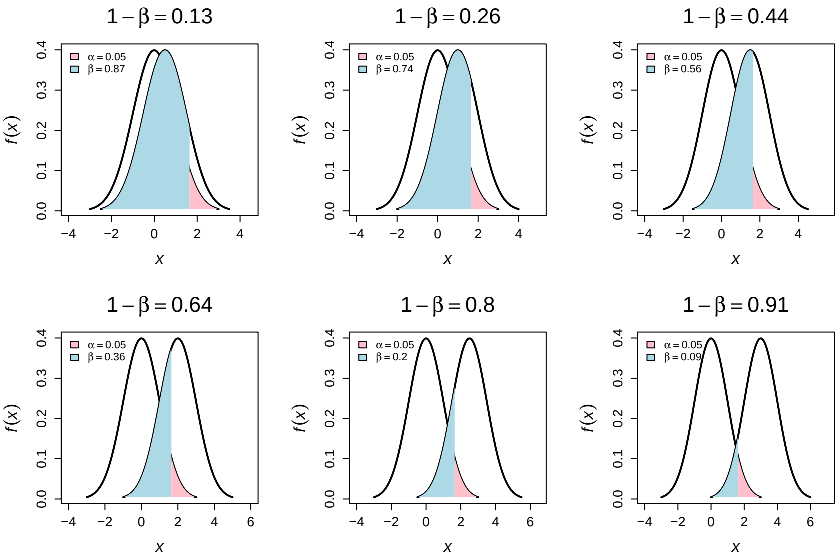
- Population effect size (larger ES = more power)
- Precision of estimate (more precision = more power)
- Significance level α (larger α = more power)

Note: you can control two of these three (precision and α)

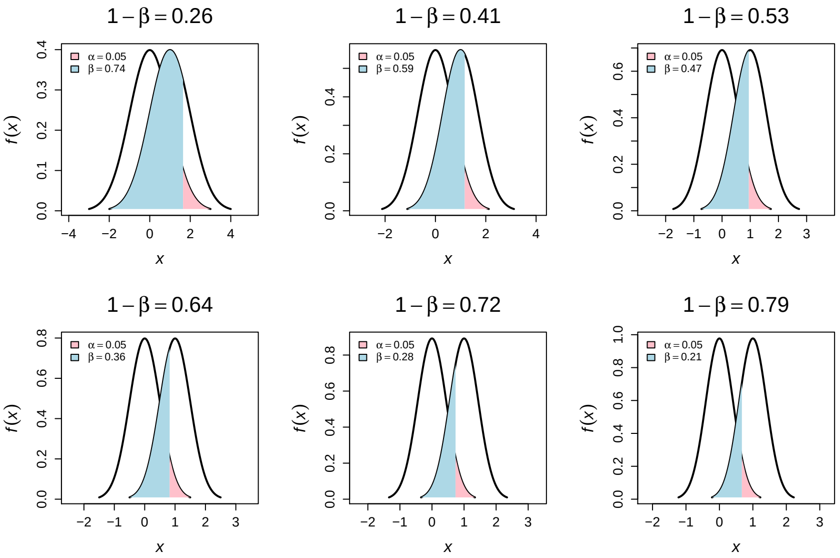
Precision of the estimate is controlled by the sample size *n*.

- Larger *n* gives you more precision (more power)
- Power analysis can give you needed *n* to find effect

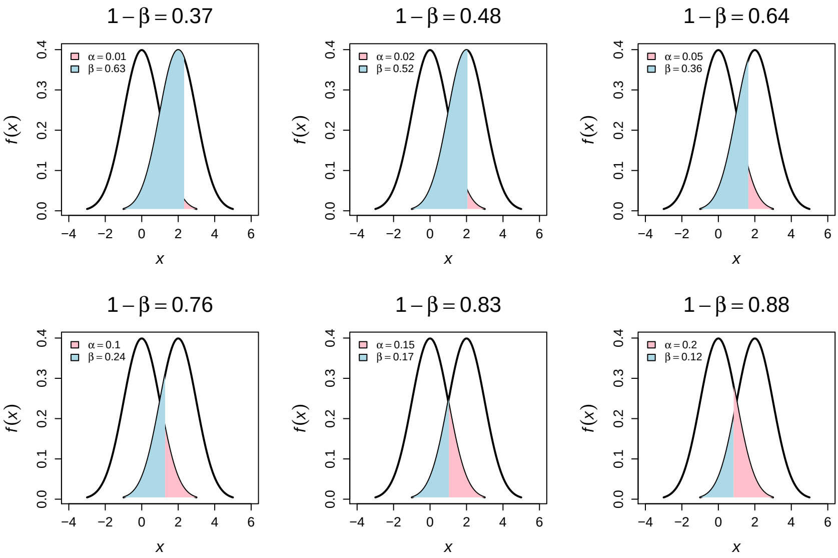
Alpha, Beta, and Power: Mean Differences



Alpha, Beta, and Power: SD Differences



Alpha, Beta, and Power: Alpha Differences



Alpha, Beta, and Power: R code

```
plotpower <- function(m0,m1,sd=1,alpha=0.05){
  x0seq = seq(m0-3*sd,m0+3*sd,length=500)
  x1seq = seq(m1-3*sd,m1+3*sd,length=500)
  cval = qnorm(1-alpha,m0,sd)
  power = round(pnorm(cval,m1,sd,lower=F),2)
  plot(x0seq,dnorm(x0seq,m0,sd),xlim=c(m0-3*sd-1,m1+3*sd+1),type="l",
       lwd=2,xlab=expression(italic(x)),ylab=expression(italic(f(x))),
       main=bquote(1-beta==.(power)),cex.main=2,cex.axis=1.25,cex.lab=1.5)
  lines(x1seq,dnorm(x1seq,m1,sd),lwd=2)
  px=c(rep(cval,2),seq(cval-0.01,m1-3*sd,length=50),m1-3*sd,cval)
  py=c(0,dnorm(cval,m1,sd),dnorm(seq(cval-0.01,m1-3*sd,length=50),m1,sd),
       rep(dnorm(m1-3*sd,m1,sd),2))
  polygon(px,py,col="lightblue",border=NA)
  px=c(rep(cval,2),seq(cval+0.1,m0+3*sd,length=50),m0+3*sd,cval)
  py=c(0,dnorm(cval,m0,sd),dnorm(seq(cval+0.1,m0+3*sd,length=50),m0,sd),
       rep(dnorm(m0+3*sd,m0,sd),2))
  polygon(px,py,col="pink",border=NA)
  legend("topleft",legend=c(as.expression(bquote(alpha==.(alpha))),
                            as.expression(bquote(beta==.(1-power)))),
         fill=c("pink","lightblue"),bty="n")
}
```

One Sample Location Problem

Suppose we have sample of data $x_i \stackrel{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$ and we want to make inferences about the mean μ.

Letting $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ denote the sample mean, we have

$$E(\bar{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{n\mu}{n} = \mu$$

$$V(\bar{x}) = V\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

where $E(\bar{x})$ and $V(\bar{x})$ are the expectation and variance of $\bar{x}$.

Implies that $\bar{x} \sim \mathrm{N}(\mu, \sigma^2/n)$

One Sample *t* Test Statistic

Suppose we want to test one the of the following sets of hypotheses

- $H_0: \mu = \mu_0$ versus $H_1: \mu > \mu_0$
- $H_0: \mu = \mu_0$ versus $H_1: \mu < \mu_0$
- $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$

and define $\delta = \mu_1 - \mu_0$ where μ_1 is the true mean under H_1.

Our *t* test statistic has the form: $T = \frac{\hat{\delta}}{\hat{\sigma}/\sqrt{n}}$

- $\hat{\delta} = \bar{x} - \mu_0$
- $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

Power Calculations for One Sample *t* Test

Given the test statistic $T = \frac{\hat{\delta}}{\hat{\sigma}/\sqrt{n}}$, power is influenced by

- Effect size $\hat{\delta}$
- Precision $\hat{\sigma}/\sqrt{n}$
- Significance level α

Note that there are four parameters that affect power: $\{\delta, \sigma, n, \alpha\}$

- If you know 3 of 4 (and desired power), you can solve for fourth
- The `power.t.test` function in R does this for you

One Sample *t* Test Power in R (one-sided)

```
> power.t.test(n=NULL,delta=-1,sd=1,sig.level=0.05,power=0.80,
+              type="one.sample",alternative="one.sided")
Error in uniroot(function(n) eval(p.body) - power, c(2, 1e+07),
  tol = tol,  :  no sign change found in 1000 iterations
> power.t.test(n=NULL,delta=1,sd=1,sig.level=0.05,power=0.80,
+              type="one.sample",alternative="one.sided")

     One-sample t test power calculation

              n = 7.727622
          delta = 1
             sd = 1
      sig.level = 0.05
          power = 0.8
    alternative = one.sided
```

One Sample *t* Test Power in R (two-sided)

```
> power.t.test(n=NULL,delta=1,sd=1,sig.level=0.05,power=0.80,
+              type="one.sample",alternative="two.sided")

     One-sample t test power calculation

              n = 9.937864
          delta = 1
             sd = 1
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

One Sample *t* Test Power in R (small effect size)

```
> power.t.test(n=NULL,delta=0.2,sd=1,sig.level=0.05,power=0.80,
+              type="one.sample",alternative="two.sided")

     One-sample t test power calculation

              n = 198.1513
          delta = 0.2
             sd = 1
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

One Sample *t* Test Power in R (really small effect size)

```
> power.t.test(n=NULL,delta=0.02,sd=1,sig.level=0.05,power=0.80,
+              type="one.sample",alternative="two.sided")

     One-sample t test power calculation

              n = 19624.12
          delta = 0.02
             sd = 1
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

Two Sample Location Problem

Suppose we have two independent samples of data $x_{i1} \stackrel{\text{iid}}{\sim} \text{N}(\mu_1, \sigma^2)$ and $x_{i2} \stackrel{\text{iid}}{\sim} \text{N}(\mu_2, \sigma^2)$ and we want to make inferences about $\delta = \mu_1 - \mu_2$.

Letting $\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{i1}$ and $\bar{x}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} x_{i2}$, we have

$$E(\bar{x}_1) = \mu_1 \quad \text{and} \quad E(\bar{x}_2) = \mu_2$$
$$V(\bar{x}_1) = \frac{\sigma^2}{n_1} \quad \text{and} \quad V(\bar{x}_2) = \frac{\sigma^2}{n_2}$$

where $E(\cdot)$ and $V(\cdot)$ are the expectation and variance operators.

Implies that $\bar{x}_1 \sim \text{N}(\mu_1, \sigma^2/n_1)$ and $\bar{x}_2 \sim \text{N}(\mu_2, \sigma^2/n_2)$.

- Note that $\bar{x}_1 - \bar{x}_2 \sim \text{N}(\delta, \sigma_*^2)$ where $\sigma_*^2 = \sigma^2(\frac{1}{n_1} + \frac{1}{n_2})$

Two Sample *t* Test Statistic

Suppose we want to test one the of the following sets of hypotheses

- $H_0: \ \delta = 0$ versus $H_1: \ \delta > 0$
- $H_0: \ \delta = 0$ versus $H_1: \ \delta < 0$
- $H_0: \ \delta = 0$ versus $H_1: \ \delta \neq 0$

where $\delta = \mu_1 - \mu_2$ is the population mean difference.

Our *t* test statistic has the form: $T = \frac{\hat{\delta}}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

- $\hat{\delta} = \bar{x}_1 - \bar{x}_2$
- $\hat{\sigma}^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ is the pooled variance estimate
- $s_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_{i1} - \bar{x}_1)^2$ and $s_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(x_{i2} - \bar{x}_2)^2$

Power Calculations for Two Sample *t* Test

Given the test statistic $T = \frac{\hat{\delta}}{\hat{\sigma}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$, power is influenced by

- Effect size $\hat{\delta}$
- Precision $\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \hat{\sigma}\sqrt{2/n}$ if $n_1 = n_2 = n$
- Significance level α

Note that the same four parameters affect power: $\{\delta, \sigma, n, \alpha\}$

- If you know 3 of 4 (and desired power), you can solve for fourth
- The `power.t.test` function in R does this for you

Two Sample *t* Test Power in R (one-sided)

```
> power.t.test(n=NULL,delta=-1,sd=1,sig.level=0.05,power=0.80,
+              alternative="one.sided")
Error in uniroot(function(n) eval(p.body) - power, c(2, 1e+07),
  tol = tol,  :  no sign change found in 1000 iterations
> power.t.test(n=NULL,delta=1,sd=1,sig.level=0.05,power=0.80,
+              alternative="one.sided")

     Two-sample t test power calculation

              n = 13.09777
          delta = 1
             sd = 1
      sig.level = 0.05
          power = 0.8
    alternative = one.sided

NOTE: n is number in *each* group
```

Two Sample *t* Test Power in R (two-sided)

```
> power.t.test(n=NULL,delta=1,sd=1,sig.level=0.05,power=0.80,
+              alternative="two.sided")

     Two-sample t test power calculation

              n = 16.71477
          delta = 1
             sd = 1
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Two Sample *t* Test Power in R (small effect size)

```
> power.t.test(n=NULL,delta=0.2,sd=1,sig.level=0.05,power=0.80,
+              alternative="two.sided")

     Two-sample t test power calculation

              n = 393.4067
          delta = 0.2
             sd = 1
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Two Sample t Test Power in R (really small effect size)

```
> power.t.test(n=NULL,delta=0.02,sd=1,sig.level=0.05,power=0.80,
+              alternative="two.sided")

     Two-sample t test power calculation

              n = 39245.36
          delta = 0.02
             sd = 1
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Multiple Sample Location Problem

Suppose we have $k > 2$ independent samples of data $x_{ij} \stackrel{\text{iid}}{\sim} \mathrm{N}(\mu_j, \sigma^2)$.

Letting $\bar{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}$ denote the j-th group's mean, we have

$$E(\bar{x}_j) = \mu_j \qquad \text{and} \qquad V(\bar{x}_j) = \frac{\sigma^2}{n_j}$$

where $E(\cdot)$ and $V(\cdot)$ are the expectation and variance operators.

Implies that $\bar{x}_j \sim \mathrm{N}(\mu_j, \sigma^2/n_j)$ for $j \in \{1, \ldots, k\}$

One-Way ANOVA F Test Statistic

Suppose we want to test the overall (omnibus) F test

- $H_0: \ \mu_j = \mu \ \forall j$ versus $H_1: \ \text{not all } \mu_j \text{ are equal}$

where μ is some common population mean.

Our F test statistic has the form: $F = \frac{MSB}{MSW} = \frac{\frac{1}{k-1}\sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2}{\frac{1}{n-k}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2}$

- If $n_j = n \ \forall j$, then $F = n\hat{\Psi}^2$ where $\hat{\Psi}$ is estimated RMSSE
- $\hat{\Psi}^2 = \frac{V(\text{between})}{V(\text{within})}$ is ratio of variances
- $MSW = \hat{\sigma}^2$ is the pooled variance estimate

Power Calculations for One-Way ANOVA

Given the test statistic $F = n\hat{\Psi}^2$, power is influenced by

- Effect sizes $\tilde{\delta}_j = \bar{y}_j - \bar{y}$
- Precisions $\hat{\sigma}$ and n
- Significance level α

Note that there are four parameters that affect power: $\{V(\tilde{\delta}), \sigma, n, \alpha\}$

- If you know 3 of 4 (and desired power), you can solve for fourth
- The `power.anova.test` function in R does this for you

One-Way ANOVA Power in R (large effect size)

```
> power.anova.test(groups=3,n=NULL,between.var=1,within.var=2,
+                  sig.level=0.05,power=0.80)

     Balanced one-way analysis of variance power calculation

         groups = 3
              n = 10.69938
    between.var = 1
     within.var = 2
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

One-Way ANOVA Power in R (medium effect size)

```
> power.anova.test(groups=3,n=NULL,between.var=1,within.var=4,
+                  sig.level=0.05,power=0.80)

     Balanced one-way analysis of variance power calculation

         groups = 3
              n = 20.30205
    between.var = 1
     within.var = 4
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

One-Way ANOVA Power in R (small effect size)

```
> power.anova.test(groups=3,n=NULL,between.var=1,within.var=100,
+                  sig.level=0.05,power=0.80)

     Balanced one-way analysis of variance power calculation

         groups = 3
              n = 482.7344
    between.var = 1
     within.var = 100
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

One-Way ANOVA Power in R (really small effect size)

```
> power.anova.test(groups=3,n=NULL,between.var=1,within.var=1000,
+                  sig.level=0.05,power=0.80)

     Balanced one-way analysis of variance power calculation

         groups = 3
              n = 4818.343
    between.var = 1
     within.var = 1000
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

Multiple Regression Problem

Suppose we have a multiple linear regression

$$y_i = b_0 + \sum_{j=1}^{p} b_j x_{ij} + e_i$$

for $i \in \{1, \ldots, n\}$ where

- $y_i \in \mathbb{R}$ is the real-valued response for the i-th observation
- $b_0 \in \mathbb{R}$ is the regression intercept
- $b_j \in \mathbb{R}$ is the j-th predictor's regression slope
- $x_{ij} \in \mathbb{R}$ is the j-th predictor for the i-th observation
- $e_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$ is a Gaussian error term

Multiple Regression F Test Statistic

Suppose we want to test the overall (omnibus) F test

- $H_0: \; b_1 = \cdots = b_p = 0$ versus $H_1: \;$ not all b_j equal 0

where the b_j terms are the unknown population slopes.

Our F test statistic has the form: $F = \frac{MSR}{MSE} = \frac{\frac{1}{p}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\frac{1}{n-(p+1)}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

- Reminder: $f^2 = \frac{R^2}{1-R^2} = \frac{SSR/SST}{1-SSR/SST} = \frac{SSR}{SSE}$
- Note that $f^2 = \left(\frac{p}{n-(p+1)}\right) F$

Power Calculations for Multiple Regression

Given the test statistic $F = \left(\frac{n-(p+1)}{p}\right) f^2$, power is influenced by

- Effect sizes f^2
- Degrees of freedom p and $n - (p+1)$
- Significance level α

Note that there are four parameters that affect power: $\{f^2, p, n, \alpha\}$

- If you know 3 of 4 (and desired power), you can solve for fourth
- See `pwr.f2.test` function in `pwr` R package

Multiple Regression Power in R (large effect size)

Assume that $p = 2$ and $R^2 = 0.8$, so that our ES is $f^2 = 4$.

```
> library(pwr)
> pwr.f2.test(u=2,v=NULL,f2=4,sig.level=0.05,power=0.80)

     Multiple regression power calculation

              u = 2
              v = 3.478466
             f2 = 4
      sig.level = 0.05
          power = 0.8
```

Need $n = \lceil v \rceil + (p + 1) = 4 + 3 = 7$ subjects.

Multiple Regression Power in R (medium effect size)

Assume that $p = 2$ and $R^2 = 0.5$, so that our ES is $f^2 = 1$.

```
> pwr.f2.test(u=2,v=NULL,f2=1,sig.level=0.05,power=0.80)

     Multiple regression power calculation

              u = 2
              v = 10.14429
             f2 = 1
      sig.level = 0.05
          power = 0.8
```

Need $n = \lceil v \rceil + (p + 1) = 11 + 3 = 14$ subjects.

Multiple Regression Power in R (small effect size)

Assume that $p = 2$ and $R^2 = 0.1$, so that our ES is $f^2 \approx 0.11$.

```
> pwr.f2.test(u=2,v=NULL,f2=0.11,sig.level=0.05,power=0.80)

     Multiple regression power calculation

              u = 2
              v = 87.65198
             f2 = 0.11
      sig.level = 0.05
          power = 0.8
```

Need $n = \lceil v \rceil + (p + 1) = 88 + 3 = 91$ subjects.

Multiple Regression Power in R (really small ES)

Assume that $p = 2$ and $R^2 = 0.01$, so that our ES is $f^2 \approx 0.01$.

```
> pwr.f2.test(u=2,v=NULL,f2=0.01,sig.level=0.05,power=0.80)

     Multiple regression power calculation

              u = 2
              v = 963.4709
             f2 = 0.01
      sig.level = 0.05
          power = 0.8
```

Need $n = \lceil v \rceil + (p + 1) = 964 + 3 = 967$ subjects.

Multiple Regression Power in R (more predictors)

Assume that $p = 4$ and $R^2 = 0.01$, so that our ES is $f^2 \approx 0.01$.

```
> pwr.f2.test(u=4,v=NULL,f2=0.01,sig.level=0.05,power=0.80)

     Multiple regression power calculation

              u = 4
              v = 1193.282
             f2 = 0.01
      sig.level = 0.05
          power = 0.8
```

Note: $F = \left(\frac{n-(p+1)}{p}\right) f^2$ so need larger n as p increases (for fixed f^2)