

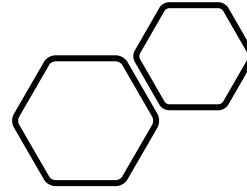
Sample Size Calculation with R

Dr. Mark Williamson, Statistician

Biostatistics, Epidemiology, and Research Design Core

DaCCoTA

Purpose



- This Module was created to provide instruction and examples on sample size calculations for a variety of statistical tests on behalf of BERDC
- The software used is R a free, open-source package



Background

- The Biostatistics, Epidemiology, and Research Design Core (BERDC) is a component of the DaCCoTA program
- Dakota Cancer Collaborative on Translational Activity has as its goal to bring together researchers and clinicians with diverse experience from across the region to develop unique and innovative means of combating cancer in North and South Dakota
- If you use this Module for research, please reference the DaCCoTA project

DaCCoTA
DAKOTA CANCER COLLABORATIVE
ON TRANSLATIONAL ACTIVITY

The Why of Sample Size Calculations



- In designing an experiment, a key question is:
How many animals/subjects do I need for my experiment?
- Too small of a sample size can under detect the effect of interest in your experiment
- Too large of a sample size may lead to unnecessary wasting of resources and animals
- Like Goldilocks, we want our sample size to be 'just right'
- **The answer: *Sample Size Calculation***
- **Goal:** We strive to have enough samples to reasonably detect an effect if it really is there without wasting limited resources on too many samples.

Key Bits of Sample Size Calculation

Effect size: magnitude of the effect under the alternative hypothesis

- The larger the effect size, the easier it is to detect an effect and require fewer samples

Power: probability of correctly rejecting the null hypothesis if it is false

- AKA, probability of detecting a true difference when it exists
- Power = $1 - \beta$, where β is the probability of a Type II error (false negative)
- The higher the power, the more likely it is to detect an effect if it is present and the more samples needed
- Standard setting for power is 0.80

Significance level (α): probability of falsely rejecting the null hypothesis even though it is true

- AKA, probability of a Type I error (false positive)
 - The lower the significance level, the more likely it is to avoid a false positive and the more samples needed
 - Standard setting for α is 0.05
- Given those three bits, and other information based on the specific design, you can calculate sample size for most statistical tests



Effect Size in detail

- While **Power** and **Significance level** are usually set irrespective of the data, the effect size is a property of the sample data
- It is essentially a function of the difference between the means of the null and alternative hypotheses over the variation (standard deviation) in the data

How to estimate Effect Size:

- A. Use background information in the form of preliminary/trial data to get means and variation, then calculate effect size directly
- B. Use background information in the form of similar studies to get means and variation, then calculate effect size directly
- C. With no prior information, make an estimated guess on the effect size expected, then use an effect size that corresponds to the size of the effect
 - Broad effect sizes categories are small, medium, and large
 - Different statistical tests will have different values of effect size for each category

$$\text{Effect Size} \approx \frac{\text{Mean}_A - \text{Mean}_0}{\text{Std. deviation}}$$

Effect Size Calculation within R

- As opposed to GPower, which allows you to enter details such as means and standard deviations into the program and it will calculate effect size for you, that is not the case for R
- Most R functions for sample size only allow you to enter effect size
- If you want to estimate effect size from background information, you'll need to calculate it yourself first
- Throughout this Module, I will provide an equation to calculate effect size for each of the statistical tests

❖ **Disclaimer:** Most of the examples and practice problems are the same as an earlier GPower Module. However, it was not always clear how effect size was calculated in GPower or in R, so sometimes the sample size calculated was different between the two. When in doubt, I would go with the result that gives the higher sample size to avoid undersampling.

Statistical Rules of the Game

Here are a few pieces of terminology to refresh yourself with before embarking on calculating sample size:

- **Null Hypothesis (H0):** default or 'boring' state; your statistical test is run to either Reject or Fail to Reject the Null
 - **Alternative Hypothesis (H1):** alternative state; usually what your experiment is interested in retaining over the Null
-
- **One Tailed Test:** looking for a deviation from the H0 in only one direction (ex: *Is variable X larger than 0?*)
 - **Two-tailed Test:** looking for a deviation from the H0 in either direction (ex: *Is variable Y different from 0?*)
-
- **Parametric data:** approximately fits a normal distribution; needed for many statistical tests
 - **Non-parametric data:** does not fit a normal distribution; alternative and less powerful tests available
-
- **Paired (dependent) data:** categories are related to one another (often result of before/after situations)
 - **Un-paired (independent) data:** categories are not related to one another
-
- **Dependent Variable:** Depends on other variables; the variable the experimenter cares about; also known as the Y or response variable
 - **Independent Variable:** Does not depend on other variables; usually set by the experimenter; also known as the X or predictor variable

Using R: Basics

- This module assumes the user is familiar with R
- For an introduction or refresher, please check out the following material
 - <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>
 - <http://www.r-tutor.com/r-introduction>
 - <https://www.statmethods.net/>
- R can be downloaded here: <https://cran.r-project.org/>
- I strongly suggest also getting RStudio, an integrated development environment: <https://rstudio.com/>

Organization of tests

- As opposed to the earlier GPower Module, which organized tests taxonomically based on types of variables, this module will follow a different order
- The order will be based on the packages available in R
- We will start with basic statistical tests that are easily calculated
- For each test:
 - Introduction slide: description, example, R code, and effect size calculation
 - Result slide: shows R code and results for the example question
 - Practice: 2-3 questions to practice on your own
 - Answers: parameters, R-code, and resulting sample size for practice questions

#	Name of Test	in R?	Package	Function
1	One Mean T-test	Yes	pwr	pwr.t.test
2	Two Means T-test	Yes	pwr	pwr.t.test
3	Paired T-test	Yes	pwr	pwr.t.test
4	One-way ANOVA	Yes	pwr	pwr.anova.test
5	Single Proportion Test	Yes	pwr	pwr.p.test
6	Two Proportions Test	Yes	pwr	pwr.2p.test
7	Chi-Squared Test	Yes	pwr	pwr.chisq.test
8	Simple Linear Regression	Yes	pwr	pwr.f2.test
9	Multiple Linear Regression	Yes	pwr	pwr.f2.test
10	Correlation	Yes	pwr	pwr.r.test
11	One Mean Wilcoxon Test	Yes*	pwr	pwer.t.test + 15%
12	Mann-Whitney Test	Yes*	pwr	pwer.t.test + 15%
13	Paired Wilcoxon Test	Yes*	pwr	pwer.t.test + 15%
14	Kruskal Wallace Test	Yes*	pwr	pwr.anova.test + 15%
15	Repeated Measures ANOVA	Yes	WebPower	wp.rmanova
16	Multi-way ANOVA (1 Category of interest)	Yes	WebPower	wp.kanova
17	Multi-way ANOVA (>1 Category of interest)	Yes	WebPower	wp.kanova
18	Non-Parametric Regression (Logistic)	Yes	WebPower	wp.logistic
19	Non-Parametric Regression (Poisson)	Yes	WebPower	wp.poisson
20	Multilevel modeling: CRT	Yes	WebPower	wp.crt2arm/wp.crt3arm
21	Multilevel modeling: MRT	Yes	WebPower	wp.mrt2arm/wp.mrt3arm
22	GLMM	Yes^	Simr & lme4	n/a

*-parametric test with non-parametric correction

^-detailed in future Module

One Mean T-Test

Description: This tests if a sample mean is any different from a set value for a normally distributed variable.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
1	0	0	0	Yes	N/A

Example: Is the average body temperature of college students any different from 98.6°F?

- $H_0=98.6^\circ\text{F}$, $H_1\neq 98.6^\circ\text{F}$
- We will guess that the effect sizes will be **medium**
- For t-tests:
 - 0.2=small, 0.5=medium, and 0.8 large effect sizes
- Selected Two-tailed, because we were asking if temp differed, not whether it was simply lower or higher

R Code: `pwer -> pwr.t.test`

```
pwr.t.test(d = , sig.level = , power = , type = c("two.sample",  
"one.sample", "paired"))
```

- **d**=effect size
- **sig.level**=significant level
- **power**=power of test
- **type**=type of test

Effect size calculation

- Cohen's $D = (M_2 - M_1) / SD$
 - M_2 =Mean 2
 - M_1 =Mean 1
 - SD =Standard deviation

One Mean T-Test

Results:

> #sample number

> `pwr.t.test(d=0.50, sig.level=0.05, power=0.80, type="one.sample", alternative="two.sided")`

One-sample t test power calculation

n = 33.36713 → Round up to 34

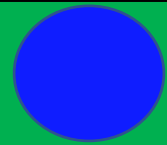
d = 0.5

sig.level = 0.05

power = 0.8

alternative = two.sided

One Mean T-Test: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if the average income of college freshman is less than \$20,000. You collect trial data and find that the mean income was \$14,500 (SD=6000).
2. You are interested in determining if the average sleep time change in a year for college freshman is different from zero. You collect the following data of sleep change (in hours).

Sleep Change	-0.55	0.16	2.6	0.65	-0.23	0.21	-4.3	2	-1.7	1.9
--------------	-------	------	-----	------	-------	------	------	---	------	-----

1. You are interested in determining if the average weight change in a year for college freshman is greater than zero.

One Mean T-Test: Answers

1. You are interested in determining if the average income of college freshman is less than \$20,000. You collect trial data and find that the mean income was \$14,500 (SD=6000).

- Effect size = $(\text{Mean}_{H1} - \text{Mean}_{H0}) / \text{SD} = (14,500 - 20,000) / 6000 = -0.917$
- One-tailed test
- **`pwr.t.test(d=-0.917, sig.level=0.05, power=0.80, type="one.sample", alternative="less")`**
- $n = 8.871645 \rightarrow$ **9 samples**

2. You are interested in determining if the average sleep time change in a year for college freshman is different from zero. You collect the following data of sleep change (in hours).

Sleep Change	-0.55	0.16	2.6	0.65	-0.23	0.21	-4.3	2	-1.7	1.9
--------------	-------	------	-----	------	-------	------	------	---	------	-----

- Effect size = $(\text{Mean}_{H1} - \text{Mean}_{H0}) / \text{SD} = (-0.446 - 0) / 1.96 = -0.228$
- Two-tailed test
- **`pwr.t.test(d=-0.228, sig.level=0.05, power=0.80, type="one.sample", alternative="two.sided")`**
- $n = 152.91 \rightarrow$ **153 samples**

3. You are interested in determining if the average weight change in a year for college freshman is greater than zero.

- Guessed a large effect size (0.8), and used one-tailed test
- **`pwr.t.test(d=0.80, sig.level=0.05, power=0.80, type="one.sample", alternative="greater")`**
- $n = 11.14 \rightarrow$ **12 samples**

Two Means T-test

Description: this tests if a mean from one group is different from the mean of another group for a normally distributed variable. AKA, testing to see if the difference in means is different from zero.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
1	1	2	1	Yes	No

Example: Is the average body temperature higher in women than in men?

- $H_0=0^\circ\text{F}$, $H_1>0^\circ\text{F}$
- We will guess that the effect sizes will be **medium**
- For t-tests:
 - 0.2=small, 0.5=medium, and 0.8 large effect sizes
- Selected greater, because we only cared to test if women's temp was higher, not lower (group 1 is women, group 2 is men)

R Code: `pwr -> pwr.t.test`

```
pwr.t.test(d = , sig.level = , power = , type = c("two.sample", "one.sample", "paired"))
```

- **d**=effect size
- **sig.level**=significant level
- **power**=power of test
- **type**=type of test

Effect size calculation

- Cohen's $D = (M_2 - M_1) / SD_{\text{pooled}}$
 - M_2 =Mean 2
 - M_1 =Mean 1
 - SD_{pooled} =Pooled standard deviation
- $SD_{\text{pooled}} = \sqrt{((SD_1^2 + SD_2^2) / 2)}$

Two Means T-test

Results:

> #sample number

> pwr.t.test(d=0.5, sig.level=0.05, power=0.80, type="two.sample", alternative="greater")

Two-sample t test power calculation

n = 50.1508 → Round up to 51, per group

d = 0.5

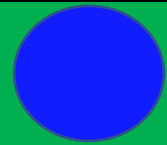
sig.level = 0.05

power = 0.8

alternative = greater

NOTE: n is number in *each* group

Two Means T-Test: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if the average daily caloric intake different between men and women. You collected trial data and found the average caloric intake for males to be 2350.2 (SD=258), while females had intake of 1872.4 (SD=420).
2. You are interested in determining if the average protein level in blood different between men and women. You collected the following trial data on protein level (grams/deciliter).

Male Protein	1.8	5.8	7.1	4.6	5.5	2.4	8.3	1.2
Female Protein	9.5	2.6	3.7	4.7	6.4	8.4	3.1	1.4

3. You are interested in determining if the average glucose level in blood is lower in men than women

Two Means T-Test: Answers

1. You are interested in determining if the average daily caloric intake different between men and women. You collected trial data and found the average caloric intake for males to be 2350.2 (SD=258), while females had intake of 1872.4 (SD=420).

- Effect size = $(\text{Mean}_{H1} - \text{Mean}_{H0}) / \text{SD}_{\text{pooled}} = (2350.2 - 1872.4) / \sqrt{((258^2 + 420^2) / 2)} = 477.8 / 348.54 = 1.37$
- two-tailed test
- **pwr.t.test(d=1.37, sig.level=0.05, power=0.80, type="two.sample", alternative="two-sided")**
- $n = 9.43 \rightarrow$ **10 samples per group**

2. You are interested in determining if the average protein level in blood different between men and women. You collected the following trial data on protein level (grams/deciliter).

Male Protein	1.8	5.8	7.1	4.6	5.5	2.4	8.3	1.2
Female Protein	9.5	2.6	3.7	4.7	6.4	8.4	3.1	1.4

- Effect size = $(\text{Mean}_{H1} - \text{Mean}_{H0}) / \text{SD}_{\text{pooled}} = (4.59 - 4.98) / \sqrt{((2.58^2 + 2.88^2) / 2)} = -0.14$
- two-tailed test
- **pwr.t.test(d=-0.14, sig.level=0.05, power=0.80, type="two.sample", alternative="two-sided")**
- $n = 801.87 \rightarrow$ **802 samples per group**

3. You are interested in determining if the average glucose level in blood is lower in men than women

- Guessed a small effect (0.20), then used a one-tailed test
- **pwr.t.test(d=-0.20, sig.level=0.05, power=0.80, type="two.sample", alternative="less")**
- $n = 309.8 \rightarrow$ **310 samples per group**

Paired T-test

Description: this tests if a mean from one group is different from the mean of another group, where the groups are dependent (not independent) for a normally distributed variable. Pairing can be leaves on same branch, siblings, the same individual before and after a trial, etc.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
1	1	2	1	Yes	Yes

Example: Is heart rate higher in patients after a run compared to before a run?

- H_0 ; bpm (after) – bpm (before) ≤ 0
- H_1 ; bpm (after) – bpm (before) > 0
- We will guess that the effect sizes will be **large**
- For t-tests:
 - 0.2=small, 0.5=medium, and 0.8 large effect sizes
- Selected One-tailed, because we only cared if bpm was higher after a run
- Group 1 is after the run, while group 2 is before the run

R Code: pwr -> pwr.t.test

```
pwr.t.test(d = , sig.level = , power = , type = c("two.sample", "one.sample", "paired"))
```

- **d**=effect size
- **sig.level**=significant level
- **power**=power of test
- **type**=type of test

Effect size calculation

- Cohen's $D = (M_2 - M_1) / SD_{pooled}$
 - M_2 =Mean 2
 - M_1 =Mean 1
 - SD_{pooled} =Pooled standard deviation
- $SD_{pooled} = \sqrt{((SD_1^2 + SD_2^2) / 2)}$

Paired T-test

Results:

> #sample number

> pwr.t.test(d=0.8, sig.level=0.05, power=0.80, type="paired", alternative="greater")

Paired t test power calculation

n = 11.14424 → Round up to 12 pairs

d = 0.8

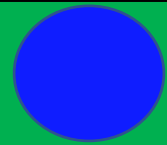
sig.level = 0.05

power = 0.8

alternative = greater

NOTE: n is number of *pairs*

Paired T-Test: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if heart rate is higher in patients after a doctor's visit compared to before a visit. You collected the following trial data and found mean heart rate before and after a visit.

BPM before	126	88	53.1	98.5	88.3	82.5	105	41.9
BPM after	138.6	110.1	58.44	110.2	89.61	98.6	115.3	64.3

2. You are interested in determining if metabolic rate in patients after surgery is different from before surgery. You collected trial data and found a mean difference of 0.73 (SD=2.9).
3. You are interested in determining if glucose levels in patients after surgery are lower compared to before surgery.

Paired T-Test: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if heart rate is higher in patients after a doctor's visit compared to before a visit. You collected the following trial data and found mean heart rate before and after a visit.

BPM before	126	88	53.1	98.5	88.3	82.5	105	41.9
BPM after	138.6	110.1	58.44	110.2	89.61	98.6	115.3	64.3

- Effect size = $(\text{Mean}_{H_1} - \text{Mean}_{H_0}) / \text{SD}_{\text{pooled}} = (98.1 - 85.4) / \sqrt{((26.8^2 + 27.2^2) / 2)} = 12.7 / 27 = 0.47$
 - one-tailed test
 - `pwr.t.test(d=0.47, sig.level=0.05, power=0.80, type="paired", alternative="greater")`
 - $n = 29.39 \rightarrow$ **30 pairs**
2. You are interested in determining if metabolic rate in patients after surgery is different from before surgery. You collected trial data and found a mean difference of 0.73 (SD=2.9).
- Effect size = $(\text{Mean}_{H_1} - \text{Mean}_{H_0}) / \text{SD} = (0.73) / 2.9 = 0.25$
 - two-tailed test
 - `pwr.t.test(d=0.25, sig.level=0.05, power=0.80, type="paired", alternative="two.sided")`
 - $n = 127.52 \rightarrow$ **128 pairs**
3. You are interested in determining if glucose levels in patients after surgery are lower compared to before surgery.
- Gussed a small effect (-0.20), then used a one tail-test *{used a negative effect to match the 'less' alternative}*
 - `pwr.t.test(d=-0.20, sig.level=0.05, power=0.80, type="paired", alternative="less")`
 - $n = 155.92 \rightarrow$ **156 pairs**

One-Way ANOVA

Description: this tests if at least one mean is different among groups, where the groups are larger than two, for a normally distributed variable. ANOVA is the extension of the Two Means T-test for more than two groups.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
1	1	>2	1	Yes	No

Example: Is there a difference in new car interest rates across 6 different cities?

- $H_0=0\%$, $H_1 \neq 0\%$
- There are a total of 6 groups (cities)
- We will guess that the effect sizes will be **small**
- For f-tests:
 - 0.1=small, 0.25=medium, and 0.4 large effect sizes
- No Tails in ANOVA
- Groups assumed to be the same size

R Code: pwr -> pwr.anova.test

pwr.anova.test(k = , f = , sig.level = , power =)

- **k**=number of groups
- **f**=effect size
- **sig.level**=significant level
- **power**=power of test

Effect size calculation

- $\eta^2 = SS_{\text{treat}} / SS_{\text{total}}$
 - SS_{treat} =treatment sum of squares
 - SS_{total} =total sum of squares
- $f = \sqrt{(\eta^2 / (1 - \eta^2))}$

One-Way ANOVA

Results:

```
> pwr.anova.test(k = 6 , f = 0.1 , sig.level = 0.05 , power = 0.80 )
```

Balanced one-way analysis of variance power calculation

k = 6

n = 214.7178 → Round up to 215 samples per group

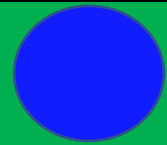
f = 0.1

sig.level = 0.05

power = 0.8

NOTE: n is number in each group

One-way ANOVA: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining there is a difference in weight lost between 4 different surgery options. You collect the following trial data of weight lost in pounds (shown on right)
2. You are interested in determining if there is a difference in white blood cell counts between 5 different medication regimes.

Option 1	Option 2	Option 3	Option 4
6.3	9.9	5.1	1.0
2.8	4.1	2.9	2.8
7.8	3.9	3.6	4.8
7.9	6.3	5.7	3.9
4.9	6.9	4.5	1.6

One-way ANOVA: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining there is a difference in weight lost between 4 different surgery options. You collect the following trial data of weight lost in pounds (shown on right)

- $\eta^2 = SS_{\text{treat}} / SS_{\text{total}} = 31.47 / (31.47 + 62.87) = 0.33$
- $f = \sqrt{(0.33 / (1 - 0.33))} = 0.7$
- 4 groups
- **pwr.anova.test(k = 4 , f = 0.7 , sig.level = 0.05 , power = 0.80)**
- $n = 6.63 \rightarrow$ **7 samples per group (28 total)**

2. You are interested in determining if there is a difference in white blood cell counts between 5 different medication regimes.

- Gussed a medium effect size (0.25)
- 5 groups
- **pwr.anova.test(k = 5 , f = 0.25 , sig.level = 0.05 , power = 0.80)**
- $n = 39.15 \rightarrow$ **40 samples per group (200 total)**

Option 1	Option 2	Option 3	Option 4
6.3	9.9	5.1	1.0
2.8	4.1	2.9	2.8
7.8	3.9	3.6	4.8
7.9	6.3	5.7	3.9
4.9	6.9	4.5	1.6

Single Proportion Test

Description: this tests when you only have a single proportion and you want to know if the proportions of certain values differ from some constant proportion.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
0	1	2	1	N/A	N/A

Example: Is there a significance difference in cancer prevalence of middle-aged women who have a sister with breast cancer (5%) compared to the general population prevalence (2%)?

- $H_0=0$, $H_1 \neq 0$
- You don't have background info, so you guess that there is a **small** effect size
- For h-tests:
0.2=small, 0.5=medium, and 0.8 large effect sizes
- Selected Two-sided, because we don't care about directionality

R Code: `pwr -> pwr.p.test`

`pwr.p.test(h = , sig.level = , power = , alternative="two.sided", "less", or "greater")`

- **h**=effect size
- **sig.level**=significant level
- **power**=power of test
- **alternative**=type of tail

Effect size calculation

- $h = 2 * \text{asin}(\sqrt{p_1}) - 2 * \text{asin}(\sqrt{p_2})$
 - p_1 =proportion 1
 - p_2 =proportion 2

Single Proportion Test

Results:

```
> #sample number
```

```
> pwr.p.test(h=0.2, sig.level=0.05, power=0.80, alternative="two.sided")
```

proportion power calculation for binomial distribution (arcsine transformation)

h = 0.2

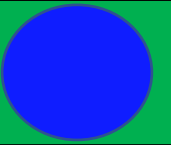
n = 196.2215 → Round up to 197

sig.level = 0.05

power = 0.8

alternative = two.sided

Single Proportion: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if the male incidence rate proportion of cancer in North Dakota is higher than the US average (prop=0.00490). You find trial data cancer prevalence of 0.00495.
2. You are interested in determining if the female incidence rate proportion of cancer in North Dakota is lower than the US average (prop=0.00420).

Single Proportion: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if the male incidence rate proportion of cancer in North Dakota is higher than the US average (prop=0.00490). You find trial data cancer prevalence of 0.00495.
 - $h = 2 * \text{asin}(\sqrt{0.00495}) - 2 * \text{asin}(\sqrt{0.00490}) = 0.0007$
 - `pwr.p.test(h=0.0007, sig.level=0.05, power=0.80, alternative="greater")`
 - $n = 12617464 \rightarrow$ **12,617,464 samples**
2. You are interested in determining if the female incidence rate proportion of cancer in North Dakota is lower than the US average (prop=0.00420).
 - Guess a very low effect size (0.001)
 - `pwr.p.test(h=-0.001, sig.level=0.05, power=0.80, alternative="less")`
 - $n = 6182557 \rightarrow$ **6,182,557 samples**

Two Proportions Test

Description: this tests when you only have two groups and you want to know if the proportions of each group are different from one another.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
0	2	2	2	N/A	No

Example: Is the expected proportion of students passing a stats course taught by psychology teachers different from the observed proportion of students passing the same stats class taught by mathematics teachers?

- $H_0=0, H_1 \neq 0$
- You don't have background info, so you guess that there is a **small** effect size
- For h-tests:
0.2=small, 0.5=medium, and 0.8 large effect sizes
- Selected Two-sided, because we don't care about directionality

R Code: `pwer -> pwr.2p.test`

`pwr.2p.test(h = , sig.level =, power =, alternative="two.sided", "less", or "greater")`

- **h**=effect size
- **sig.level**=significant level
- **power**=power of test
- **alternative**=type of tail

Effect size calculation

- $h = 2 * \text{asin}(\sqrt{p_1}) - 2 * \text{asin}(\sqrt{p_2})$
 - p_1 =proportion 1
 - p_2 =proportion 2

Two Proportions Test

Results:

```
> #sample number  
> pwr.2p.test(h=0.2, sig.level=0.05, power=.80, alternative="two.sided")
```

Difference of proportion power calculation for binomial distribution (arcsine transformation)

h = 0.2

n = 392.443 → Round up to 393

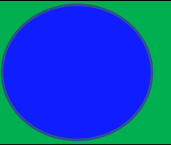
sig.level = 0.05

power = 0.8

alternative = two.sided

NOTE: same sample sizes

Two Proportions: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if the expected proportion (P1) of students passing a stats course taught by psychology teachers is different than the observed proportion (P2) of students passing the same stats class taught by biology teachers. You collected the following data of passed tests.

Psychology	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No
Biology	No	No	Yes	Yes	Yes	No	Yes	No	Yes	Yes

2. You are interested in determining of the expected proportion (P1) of female students who selected YES on a question was higher than the observed proportion (P2) of male students who selected YES. The observed proportion of males who selected yes was 0.75.

Two Proportions: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if the expected proportion (P1) of students passing a stats course taught by psychology teachers is different than the observed proportion (P2) of students passing the same stats class taught by biology teachers. You collected the following data of passed tests.

- $P1=7/10=0.70$, $P2=6/10=0.60$
- $h= 2*\text{asin}(\text{sqrt}(0.60))-2*\text{asin}(\text{sqrt}(0.70))=-0.21$
- **pwr.2p.test(h=-0.21, sig.level=0.05, power=0.80, alternative="two.sided")**
- $n = 355.96 \rightarrow$ **356 samples**

Psychology	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No
Biology	No	No	Yes	Yes	Yes	No	Yes	No	Yes	Yes

2. You are interested in determining of the expected proportion (P1) of female students who selected YES on a question was higher than the observed proportion (P2) of male students who selected YES. The observed proportion of males who selected yes was 0.75.

- Guess that the expected proportion (P1) =0.85
- $h= 2*\text{asin}(\text{sqrt}(0.85))-2*\text{asin}(\text{sqrt}(0.75))=0.25$
- **pwr.2p.test(h=0.25, sig.level=0.05, power=0.80, alternative="greater")**
- $n = 197.84 \rightarrow$ **198 samples**

Chi-Squared Test

Description: Extension of proportions test, which asks if table of observed values are any different from a table of expected ones. Also called Goodness-of-fit test.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
0	≥ 1	≥ 2	1	N/A	No

Example: Does the observed proportions of phenotypes from a genetics experiment different from the expected 9:3:3:1?

- $H_0=0, H_1 \neq 0$
- You don't have background info, so you guess that there is a **medium** effect size
- For w-tests:
0.1=small, 0.3=medium, and 0.5 large effect sizes
- Degrees of freedoms is the number of proportions minus 1
4 (phenotypes) – 1 = 3

R Code: pwer -> pwr.chisq.test

pwr.chisq.test(w = , df = , sig.level = , power =)

- **w**=effect size
- **df**=degrees of freedom
- **sig.level**=significant level
- **power**=power of test

Effect size calculation

- $w = \sqrt{X^2 / (n * df)}$
 - $X^2 = \text{Chi-squared} = \sum (O-E)^2 / E$
 - O=observed
 - E=expected
 - n=number of samples
 - df= degrees of freedom

Chi-Squared Test

Results:

> #sample number

> `pwr.chisq.test(w=0.3, df=3, sig.level=0.05, power=0.80)`

Chi squared power calculation

w = 0.3

N = 121.1396 → **Round up to 122**

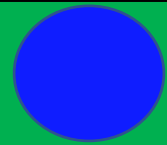
df = 3

sig.level = 0.05

power = 0.8

NOTE: N is the number of observations

Chi-Squared: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if the ethnic ratios in a company differ by gender. You collect the following trial data from 200 employees.

Gender	White	Black	Am. Indian	Asian
Male	0.60	0.25	0.01	0.14
Female	0.65	0.21	0.11	0.03

2. You are **interested** in determining if the proportions of student by year (Freshman, Sophomore, Junior, Senior) is any different from 1:1:1:1. You collect the following trial data.

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Grade	Frs	Frs	Frs	Frs	Frs	Frs	Frs	Soph	Soph	Soph	Soph	Soph	Jun	Jun	Jun	Jun	Jun	Sen	Sen	Sen

Chi-Squared: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if the ethnic ratios in a company differ by gender. You collect the following trial data from 200 employees.

Gender	White	Black	Am. Indian	Asian
Male	60	25	1	14
Female	65	21	11	3

- If they were equal the expected ratios should be the same as the overall ethnic ratios (62.5, 23.0, 6.0, 8.5)
- Will just focus on males
- X^2 (Chi-squared) = $\sum(O-E)^2/E = (60-62.5)^2/62.5 + (25-23)^2/23 + (1-6)^2/6 + (14-8.5)^2/8.5$
- $= 0.10 + 0.17 + 4.17 + 3.56 = 8$
- $w = \sqrt{X^2 / (n * df)} = \sqrt{8 / (200 * 3)} = 0.115$
- **`pwr.chisq.test(w=0.115, df=3, sig.level=0.05, power=0.80)`**
- $n = 824.39 \rightarrow$ **825 samples**

2. You are **interested** in determining if the proportions of student by year (Freshman, Sophomore, Junior, Senior) is any different from 1:1:1:1. You collect the following trial data.

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Grade	Frs	Frs	Frs	Frs	Frs	Frs	Frs	Soph	Soph	Soph	Soph	Soph	Jun	Jun	Jun	Jun	Jun	Sen	Sen	Sen

- X^2 (Chi-squared) = $\sum(O-E)^2/E = (7-5)^2/5 + (5-5)^2/5 + (5-5)^2/5 + (3-5)^2/5 = 0.8 + 0 + 0 + 0.8 = 1.6$
- $w = \sqrt{X^2 / (n * df)} = \sqrt{1.6 / (20 * 3)} = 0.163$
- **`pwr.chisq.test(w=0.163, df=3, sig.level=0.05, power=0.80)`**
- $n = 410.34 \rightarrow$ **411 samples**

Simple Linear Regression

Description: this test determines if there is a significant relationship between two normally distributed numerical variables. The predictor variable is used to try to predict the response variable.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
2	0	N/A	N/A	Yes	N/A

Example: Is there a relationship between height and weight in college males?

- $H_0=0$, $H_1 \neq 0$
- You don't have background info, so you guess that there is a **large** effect size
- For f2-tests:
 - 0.02=small, 0.15=medium, and 0.35 large effect sizes
- For simple regression (only one predictor variable) = numerator df=1
- Output will be denominator degrees of freedom rather than sample size; will need to round up and add 2 to get sample size

R Code: pwer -> pwr.f2.test

pwr.f2.test(u =, v= , f2=, sig.level =, power =)

- **u**=numerator degrees of freedom
- **v**=denominator degrees of freedom
- **f2**=effect size
- **sig.level**=significant level
- **power**=power of test

Effect size calculation

- $f2=R=\sqrt{R^2}$
 - R=correlation coefficient
 - R^2 =goodness-of-fit
 - Use adjusted R^2

Simple Linear Regression

Results:

```
> #sample number
```

```
> pwr.f2.test(u=1, f2=0.35, sig.level=0.05, power=0.80)
```

Multiple regression power calculation

```
u = 1
```

```
v = 22.50313
```

```
f2 = 0.35
```

```
sig.level = 0.05
```

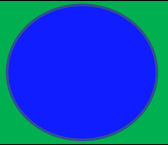
```
power = 0.8
```

```
> #denominator df to sample size
```

```
> round(22.5031,0)+2
```

```
[1] 25 → Sample size
```

Simple Linear Regression: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if height (meters) in plants can predict yield (grams of berries). You collect the following trial data.

Yield	46.8	48.7	48.4	53.7	56.7
Height	14.6	19.6	18.6	25.5	20.4

2. You are interested in determining if the size of a city (in square miles) can predict the population of the city (in # of individuals).

Simple Linear Regression: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if height (meters) in plants can predict yield (grams of berries). You collect the following trial data.

Yield	46.8	48.7	48.4	53.7	56.7
Height	14.6	19.6	18.6	25.5	20.4

- Created variables in R
 - `yield<-c(46.8, 48.7, 48.4, 53.7, 56.7)`
 - `height<-c(14.6, 19.6, 18.6, 25.5, 20.4)`
- Ran linear model to find R-squared
 - `linearMod <- lm(height~yield)`
 - `summary(linearMod) -> adj R2=0.2784`
 - $f^2=R^2=V(\text{adj } R^2)=V(0.4588)=0.53$
 - `pwr.f2.test(u=1, f2=0.53, sig.level=0.05, power=0.80)`
 - $v=14.96 \rightarrow 15+ 2(\text{variables}) \rightarrow$ **17 samples**

2. You are interested in determining if the size of a city (in square miles) can predict the population of the city (in # of individuals).

- Guessed a large effect size (0.35); for 1 predictor so 1 df
- `pwr.f2.test(u=1, f2=0.35, sig.level=0.05, power=0.80)`
- $v=22.5 \rightarrow 23+ 2(\text{variables}) \rightarrow$ **25 samples**

Multiple Linear Regression

Description: The extension of simple linear regression. The first major change is there are more predictor variables. The second change is that interaction effects can be used. Finally, the results typically can't be plotted.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
>2	0	N/A	N/A	Yes	N/A

Example: Can height, age, and time spent at the gym, predict weight in adult males?

- $H_0=0$, $H_1 \neq 0$
- You don't have background info, so you guess that there is a **medium** effect size
- For f2-tests:
 - 0.02=small, 0.15=medium, and 0.35 large effect sizes
- Numerator degrees of freedom is the number of predictor variables (3)
- Output will be denominator degrees of freedom rather than sample size; will need to round up and add the total number of variables (4)

R Code: pwr -> pwr.f2.test

pwr.f2.test(u =, v =, f2 =, sig.level =, power =)

- **u**=numerator degrees of freedom
- **v**=denominator degrees of freedom
- **f2**=effect size
- **sig.level**=significant level
- **power**=power of test

Effect size calculation

- $f2 = R = \sqrt{R^2}$
 - R=correlation coefficient
 - R^2 =goodness-of-fit
 - Use adjusted R^2

Multiple Linear Regression

Results:

```
> #sample number
```

```
> pwr.f2.test(u=3, f2=0.15, sig.level=0.05, power=0.80)
```

Multiple regression power calculation

```
u = 3
```

```
v = 72.70583
```

```
f2 = 0.15
```

```
sig.level = 0.05
```

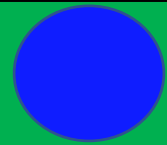
```
power = 0.8
```

```
> #denominator df to sample size
```

```
> round(72.70583,0)+4
```

```
[1] 77 → Sample Size
```

Multiple Linear Regression: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if height (meters), weight (grams), and fertilizer added (grams) in plants can predict yield (grams of berries). You collect the following trial data.

Yield	46.8	48.7	48.4	53.7	56.7
Height	14.6	19.6	18.6	25.5	20.4
Weight	95.3	99.5	94.1	110	103
Fertilizer	2.1	3.2	4.3	1.1	4.3

2. You are interested in determining if the size of a city (in square miles), number of houses, number of apartments, and number of jobs can predict the population of the city (in # of individuals).

Multiple Linear Regression: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if height (meters), weight (grams), and fertilizer added (grams) in plants can predict yield (grams of berries). You collect the following trial data.

- Created variables in R

- `yield<-c(46.8, 48.7, 48.4, 53.7, 56.7)`
- `height<-c(14.6, 19.6, 18.6, 25.5, 20.4)`
- `weight<-c(95.3, 99.5, 94.1, 110, 103)`
- `Fert<-c(2.1, 3.2, 4.3, 1.1, 4.3)`

Yield	46.8	48.7	48.4	53.7	56.7
Height	14.6	19.6	18.6	25.5	20.4
Weight	95.3	99.5	94.1	110	103
Fertilizer	2.1	3.2	4.3	1.1	4.3

- Ran linear model to find R-squared

- `linearMod2 <-lm(height~yield + weight + Fert)`
- `summary(linearMod2) -> Adj R2 = 0.6765`
- $f2=R=\sqrt{\text{adj } R^2} = \sqrt{0.6765}=0.822$
- `pwr.f2.test(u=3, f2=0.822, sig.level=0.05, power=0.80)`
- $v=13.7 \rightarrow 14+ 4(\text{variables}) \rightarrow$ **18 samples**

2. You are interested in determining if the size of a city (in square miles), number of houses, number of apartments, and number of jobs can predict the population of the city (in # of individuals).

- Guessed a large effect size (0.35); for 4 variables (df=3)
- `pwr.f2.test(u=3, f2=0.35, sig.level=0.05, power=0.80)`
- $v=31.31 \rightarrow 32+ 4(\text{variables}) \rightarrow$ **36 samples**

Correlation

Description: this test determines if there is a difference between two numerical values. It is like simple regression, but is not identical.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
2	0	N/A	N/A	Yes	No

Example: Is there a correlation between hours studied and test score?

- $H_0=0$, $H_1 \neq 0$
- You don't have background info, so you guess that there is a **large** correlation
- For correlation levels (r):
0.1=small, 0.3=medium, and 0.5 large correlations

R Code: `pwr -> pwr.r.test`

`pwr.r.test(r = , sig.level = , power =)`

- **r**=correlation
- **sig.level**=significant level
- **power**=power of test

Effect size calculation

- **r**=correlation coefficient

Correlation

Results:

> #sample number

> pwr.r.test(r=0.5, sig.level=0.05, power=0.80)

approximate correlation power calculation (arctangh transformation)

n = 28.24841 → Round up to 29

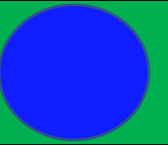
r = 0.5

sig.level = 0.05

power = 0.8

alternative = two.sided

Correlation: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if there is a correlation between height and weight in men

Males	Height	178	166	172	186	182
	Weight	165	139	257	225	196

2. You are interested in determining if, in lab mice, the correlation between longevity (in months) and average protein intake (grams).

Correlation: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if there is a correlation between height and weight in men

Males	Height	178	166	172	186	182
	Weight	165	139	257	225	196

- Created variables in R and ran correlation test
 - `MH <-c(178,166,172,186,182)`
 - `MW <-c(165,139,257,225,196)`
 - `cor(MH, MW) -> 0.37`
- `pwr.r.test(r=0.37, sig.level=0.05, power=0.80)`
- $n = 54.19 \rightarrow$ **55 samples**

2. You are interested in determining if, in lab mice, the correlation between longevity (in months) and average protein intake (grams).

- Guessed large (0.5) correlation
- `pwr.r.test(r=0.5, sig.level=0.05, power=0.80)`
- $n = 28.24 \rightarrow$ **29 samples**

Non-Parametric T-tests

Description: versions of the t-tests for non-parametric data.

- One Mean Wilcoxon: sample mean against set value
- Mann-Whitney: two sample means (unpaired)
- Paired Wilcoxon: two sample means (paired)

Name	Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
One Mean Wilcoxon	1	0	0	0	No	N/A
Mann-Whitney	1	1	2	1	No	No
Paired Wilcoxon	1	1	2	1	No	Yes

- There aren't any R packages that had useful non-parametric t-tests
- I suggest using the parametric + 15% approach

Effect size calculation

- Cohen's D: $(M_2 - M_1) / SD$; $(M_2 - M_1) / Sd_{pooled}$; $(Mean_{diff}) / SD_{diff}$

Examples: (for t-tests, 0.2=small, 0.5=medium, and 0.8 large effect sizes)

One Mean Wilcoxon:

Is the average number of children in Grand Forks families different than 1?

- $H_0 = 1$ child
- $H_1 > 1$ child
- You don't have background info, so you guess that there is a **medium** effect size
- Select one-tailed (greater)

Mann-Whitney:

Does the average number of snacks per day for individuals on a diet differ between young and old persons?

- $H_0 = 0$ difference in snack number,
- $H_1 \neq 0$ difference in snack number
- You don't have background info, so you guess that there is a **small** effect size
- Select two-sided

Paired Wilcoxon:

Is genome methylation patterns different between identical twins?

- $H_0 = 0\%$ methylation
- $H_1 \neq 0\%$ methylation
- You don't have background info, so you guess that there is a **large** effect size
- Select one-tailed (greater)

Non-parametric Tests

Results:

>#One Mean Wilcoxon

```
> pwr.t.test(d=0.5, sig.level=0.05, power=0.80, type="one.sample", alternative="greater")
```

One-sample t test power calculation

n = 26.13753

d = 0.5

sig.level = 0.05

power = 0.8

alternative = greater

> #Non-parametric correction

```
> round(26.13753*1.15,0)
```

```
[1] 30 → Total sample size
```

>#Mann-Whitney

```
> pwr.t.test(d=0.2, sig.level=0.05, power=0.80, type="two.sample", alternative="two.sided")
```

Two-sample t test power calculation

n = 198.1508

d = 0.2

sig.level = 0.05

power = 0.8

alternative = two.sided

> #Non-parametric correction

```
> round(198.1508*1.15,0)
```

```
[1] 228 → Total sample size
```

>#Paired Wilcoxon

```
> pwr.t.test(d=0.8, sig.level=0.05, power=0.80, type="paired", alternative="greater")
```

Paired t test power calculation

n = 11.14424

d = 0.8

sig.level = 0.05

power = 0.8

alternative = greater

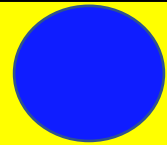
NOTE: n is number of *pairs*

> #Non-parametric correction

```
> round(11.14424*1.15,0)
```

```
[1] 13 → Total number of pairs
```

Non-Parametric T-tests: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if the average number of pets in Grand Forks families is greater than 1. You collect the following trial data for pet number.

Pets	1	1	1	3	2	1	0	0	0	4
------	---	---	---	---	---	---	---	---	---	---

2. You are interested in determining if the number of meals per day for individuals on a diet is higher in younger people than older. You collected trial data on meals per day.

Young meals	1	2	2	3	3	3	3	4
Older meals	1	1	1	2	2	2	3	3

3. You are interested in determining if genome methylation patterns are higher in the first fraternal twin born compared to the second. You collected the following trial data on methylation level difference (in percentage).

Methy. Diff (%)	5.96	5.63	1.25	1.17	3.59	1.64	1.6	1.4
-----------------	------	------	------	------	------	------	-----	-----

Non-Parametric T-tests: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if the average number of pets in Grand Forks families is greater than 1. You collect the following trial data for pet number.

Pets	1	1	1	3	2	1	0	0	0	4
------	---	---	---	---	---	---	---	---	---	---

- Effect size = $(\text{Mean}_{H_1} - \text{Mean}_{H_0}) / \text{SD} = (1.3 - 1.0) / 1.34 = 0.224$
- One-tailed test
- **pwr.t.test(d=0.224, sig.level=0.05, power=0.80, type="one.sample", alternative="greater")**
- $n = 124.58 * 1.15$ (then round up) -> **143 samples**

2. You are interested in determining if the number of meals per day for individuals on a diet is higher in younger people than older. You collected trial data on meals per day.

Young meals	1	2	2	3	3	3	3	4
Older meals	1	1	1	2	2	2	3	3

- Effect size = $(\text{Mean}_{H_1} - \text{Mean}_{H_0}) / \text{SD}_{\text{pooled}} = (2.625 - 1.875) / \sqrt{((0.92^2 + 0.83^2) / 2)} = 0.856$
- One-tailed test
- **pwr.t.test(d=0.856, sig.level=0.05, power=0.80, type="two.sample", alternative="greater")**
- $n = 17.59 * 1.15$ (then round up) -> **20 samples per group**

3. You are interested in determining if genome methylation patterns are different in the first fraternal twin born compared to the second. You collected the following trial data on methylation level difference (in percentage).

Methy. Diff (%)	5.96	5.63	1.25	1.17	3.59	1.64	1.6	1.4
-----------------	------	------	------	------	------	------	-----	-----

- Effect size = $(\text{Mean}_{\text{diff}}) / \text{SD}_{\text{diff}} = (2.78) / 2.01 = 1.38$
- Two-tailed test
- **pwr.t.test(d=1.38, sig.level=0.05, power=0.80, type="paired", alternative="two.sided")**
- $n = 6.29 * 1.15$ (then round up) -> **7 pairs**

Kruskal Wallance Test

Description: this tests if at least one mean is different among groups, where the groups are larger than two for a non-normally distributed variable. (AKA, non-parametric ANOVA). There really isn't a good way of calculating sample size in R, but you can use a rule of thumb:

1. Run Parametric Test
2. Add 15% to total sample size

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
1	1	>2	1	No	No

Example: Is there a difference in draft rank across 3 different months?

- $H_0=0$, $H_1 \neq 0$
- There will be a total of 3 groups (months)
- You don't have background info, so you guess that there is a **medium** effect size
- For f-tests:
0.1=small, 0.25=medium, and 0.4 large effect sizes
- No Tails in ANOVA
- Groups assumed to be the same size

R Code: pwr -> pwr.anova.test

pwr.anova.test(k = , f = , sig.level = , power =)

- **k**=number of groups
- **f**=effect size
- **sig.level**=significant level
- **power**=power of test

Effect size calculation

- $\eta^2 = SS_{\text{treat}} / SS_{\text{total}}$
 - SS_{treat} =treatment sum of squares
 - SS_{total} =total sum of squares
- $f = \sqrt{(\eta^2 / (1 - \eta^2))}$

Kruskal Wallace Test

Results:

> #sample number of ANOVA

> pwr.anova.test(k = 3 , f = 0.25 , sig.level = 0.05 , power = 0.80)

Balanced one-way analysis of variance power calculation

k = 3

n = 52.3966

f = 0.25

sig.level = 0.05

power = 0.8

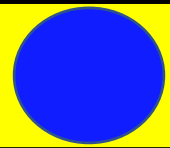
NOTE: n is number in each group

> #15% correction factor

> 52.3996 * 1.15

[1] 60.25954 → Round up to 61 samples per group

Kruskal Wallace Test: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining there is a difference in hours worked across 3 different groups (faculty, staff, and hourly workers). You collect the following trial data of weekly hours (shown on right).
2. You are interested in determining there is a difference in assistant professor salaries across 25 different departments.

Faculty	Staff	Hourly
42	46	29
45	45	42
46	37	33
55	42	50
42	40	23

Kruskal Wallance Test: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining there is a difference in hours worked across 3 different groups (faculty, staff, and hourly workers). You collect the following trial data of weekly hours (shown on right).

- $\eta^2 = SS_{\text{treat}} / SS_{\text{total}} = 286.5 / (286.5 + 625.2) = 0.314$
- $f = \sqrt{(0.314 / (1 - 0.314))} = 0.677$
- 3 groups
- **pwr.anova.test(k = 3, f = 0.677, sig.level = 0.05, power = 0.80)**
- $n = 8.09 * 1.15$ (then round up) -> **10 samples per group**

2. You are interested in determining there is a difference in assistant professor salaries across 25 different departments.

- Guess small effect size (0.10)
- 25 groups
- **pwr.anova.test(k = 25, f = 0.10, sig.level = 0.05, power = 0.80)**
- $n = 90.67 * 1.15$ (then round up) -> **105 samples per group**

Faculty	Staff	Hourly
42	46	29
45	45	42
46	37	33
55	42	50
42	40	23

Repeated Measures ANOVA

Description: this tests if at least one mean is different among groups, where the groups are repeated measures (more than two) for a normally distributed variable. Repeated Measures ANOVA is the extension of the Paired T-test for more than two groups.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
1	1	>2	1	Yes	Yes

Example: Is there a difference in blood pressure at 1, 2, 3, and 4 months post-treatment?

- $H_0=0$, $H_1 \neq 0$
- **1 group, 4 measurements**
- You don't have background info, so you guess that there is a **small** effect size
- For f-tests:
0.1=small, 0.25=medium, and 0.4 large effect sizes
- For the **nonsphericity correction coefficient**, 1 means sphericity is met. There are methods to estimate this but will go with 1 for this example.
- **Type** will be 1, as we want **within-effect**

R Code: WebPower -> wp.rmanova

wp.rmanova(ng = NULL, nm = NULL, f = NULL, nscor = 1, alpha = 0.05, power = NULL, type = 0)

- **ng**=number of groups
- **nm**=number of measurements
- **f**=effect size
- **nscor**=nonsphericity correction coefficient
- **alpha**=significant level of test
- **power**=statistical power
- **type**=(0,1,2) The value "0" is for between-effect; "1" is for within-effect; and "2" is for interaction effect

Effect size calculation

- $f = \frac{\sigma_m}{\sigma}$
 - σ_m = standard deviation of group means
 - $\sigma_m = \sqrt{\frac{\sum_{k=1}^K (m_k - m)^2}{k}}$
 - m_k = group mean
 - m = overall mean
 - k = number of groups,
 - σ = overall standard deviation

NOTE:

- **Within-effects:** variability of a particular value for individuals in a sample
- **Between-effects:** examines differences between individuals

Repeated Measures ANOVA

Results:

> #sample size

```
> wp.rmanova(n=NULL, ng=1, nm=4, f=0.1, nscor=1,  
+ alpha=0.05, power=0.80, type=1)
```

Repeated-measures ANOVA analysis

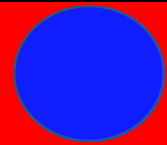
	n	f	ng	nm	nscor	alpha	power
	1091.559	0.1	1	4	1	0.05	0.8

→ Round up to 1092 samples total

NOTE: Power analysis for within-effect test

URL: <http://psychstat.org/rmanova>

Repeated Measures ANOVA: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if there is a difference in blood serum levels at 6, 12, 18, and 24 months post-treatment. You collect the following trial data of blood serum in mg/dL (shown on right).
2. You are interested in determining if there is a difference in antibody levels at 1, 2, and 3 months post-treatment.

6 months	12 months	18 months	24 months
38	38	46	52
13	44	15	29
32	35	53	60
35	48	51	44
21	27	29	36

Repeated Measures ANOVA: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if there is a difference in blood serum levels at 6, 12, 18, and 24 months post-treatment. You collect the following trial data of blood serum in mg/dL (shown on right).

$$f = \frac{\sigma_m}{\sigma} = \sqrt{\frac{\sum_{k=1}^K (m_k - m)^2}{k}} / \sigma$$

$$f = \sqrt{\frac{(27.8-37.3)^2 + (38.4-37.3)^2 + (38.8-37.3)^2 + (25.2-37.3)^2}{4}} / 12.74 = 0.608$$

- To get sphericity, ran ANOVA
 - `library(ez)`
 - `anova3 <- ezANOVA(ex3, dv=Serum, wid=Patient, within=(Month),detailed=TRUE)`
 - `print(anova3$ANOVA)`
- Sphericity was non-significant (0.43), so coefficient of 1
- One group, four measurements, within-effects so type 1
- `wp.rmanova(n=NULL, ng=1, nm=4, f=0.608, nscor=1, alpha=0.05, power=0.80, type=1)`
- $n = 30.81 \rightarrow$ **31 samples total**

2. You are interested in determining if there is a difference in antibody levels at 1, 2, and 3 months post-treatment.

- Guess a nonsphericity correction of of **1** and medium effect **0.25**
- One group, three measurements, type 1
- `wp.rmanova(n=NULL, ng=1, nm=3, f=0.25, nscor=1, alpha=0.05, power=0.80, type=1)`
- $n = 155.66 \rightarrow$ **156 samples total**

6 months	12 months	18 months	24 months
38	38	46	52
13	44	15	29
32	35	53	60
35	48	51	44
21	27	29	36

Multi-Way ANOVA (1 Category of Interest)

Description: this test is an extension of ANOVA, where there is more than one category, but only one category is of interest. The other category/categories are things that need to be controlled for (blocking/nesting/random effects/etc.).

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
1	≥2	≥2	1	Yes	No

Example: Is there difference in treatment (Drug A, B, and C) from a series of four different hospital sections (Block 1, 2, 3, and 4)?

- $H_0=0$, $H_1 \neq 0$
- Category of interest: Treatment
- Want to control for the Sections (Blocking)
- Numerator df (Treatment) = 3-1=2
- Number of groups (Treatment *Section)=3*4=12
- You don't have background info, so you guess that there is a **medium** effect size
- For f-tests:
 - 0.1=small, 0.25=medium, and 0.4 large effect sizes

R Code: WebPower -> wp.kanova

wp.kanova(ndf = NULL, f = NULL, ng = NULL, alpha = 0.05, power = NULL)

- **ndf**=numerator degrees of freedom
- **f**=effect size
- **ng**=number of groups
- **alpha**=significance level
- **power**=statistical power

Effect size calculation

- $f = \frac{\sigma_b}{\sigma_w}$
- σ_b =standard deviation of blocking sections
- $\sigma_b = \sqrt{\frac{\sum_{k=1}^K (\mu_k - \mu)^2}{K}}$, where μ_k =mean of section, μ =overall mean, and K =number of sections
- σ_w =standard deviation of all groups (treatment*section)
- $\sigma_w = \sqrt{\frac{\sum_{j=1}^J \sum_{k=1}^K (\mu_{jk} - \mu)^2}{JK}}$, where μ_{jk} =mean of group, μ =overall mean, J =number of treatments, and K =number of sections

Multi-Way ANOVA (1 Category of Interest)

Results:

> #sample size

> wp.kanova(ndf=2, f=0.25, ng=12, alpha=0.05, power=0.80)

Multiple way ANOVA analysis

n	ndf	ddf	f	ng	alpha	power
157.3764	2	145.3764	0.25	12	0.05	0.8

→ Round up to 158 total samples

NOTE: Sample size is the total sample size

URL: <http://psychstat.org/kanova>

Multi-Way ANOVA (>1 Category of Interest)

Description: this test is an extension of ANOVA, where there is more than one category, and each category is of interest. If there is two categories, it is 2-way ANOVA; three categories, 3-way ANOVA, etc.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
1	≥2	≥2	>1	Yes	No

Example: Is there difference in treatment (Drug A, B, and C) across age (child, adult, elder) and cancer stage (I, II, III, IV, V)?

- $H_0=0$, $H_1 \neq 0$
- Categories of interest: Treatment, Age, and Cancer Stage
- Numerator df = Treat DF * Age DF * Stage DF = $(3-1)*(3-1)*(5-1)=2*2*4=16$
- Number of groups = Treat*Age*Stage = $3*3*5=45$
- You don't have background info, so you guess that there is a **small** effect size
- For f-tests:
0.1=small, 0.25=medium, and 0.4 large effect sizes

R Code: WebPower -> wp.kanova

wp.kanova(ndf = NULL, f = NULL, ng = NULL, alpha = 0.05, power = NULL)

- **ndf**=numerator degrees of freedom
- **f**=effect size
- **ng**=number of groups
- **alpha**=significance level
- **power**=statistical power

Effect size calculation (sort of)

- $\eta^2 = \sigma_m^2 / \sigma_t^2$
 - σ_m^2 =between-group variance
 - σ_t^2 =total variance
- $f = \sqrt{\eta^2 / (1 - \eta^2)}$

Multi-Way ANOVA (>1 Category of Interest)

Results:

> #sample size

> wp.kanova(ndf=16, f=0.10, ng=45, alpha=0.05, power=0.80)

Multiple way ANOVA analysis

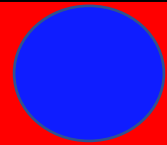
n	ndf	ddf	f	ng	alpha	power
1940.159	16	1895.159	0.145	0.05	0.8	

→ Round up to 1941 total samples

NOTE: Sample size is the total sample size

URL: <http://psychstat.org/kanova>

Multi-Way ANOVA: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if there is a difference in treatment (Drug A, B, and C), while controlling for age (child=c, adult=a, elder=e). You collect the following trial data for treatment (shown on right).
2. You are interested in determining if there is a difference in treatment (Drug A, B, and C) across age (child, adult, elder) and cancer stage (I, II, III, IV, V). You collect trial data and find that the between-group variance is 27.3, while the total variance is 85.2.

Drug A			Drug B			Drug C		
c	a	e	c	a	e	c	a	e
-6.4	8.7	-3.1	1.3	-6.0	6.8	-2.0	-4.3	-1.2
-8.2	-6.3	-6.5	3.6	1.3	2.4	1.5	1.3	1.1
7.9	-1	-1.5	3.9	-1.9	1.3	2.5	-8.2	-9.7

Multi-Way ANOVA: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if there is a difference in treatment (Drug A, B, and C), while controlling for age (child=c, adult=a, elder=e). You collect the following trial data for treatment (shown on right).

- Only care about Drug, so focus on treatment only

$$f = \frac{\sigma_b}{\sigma_w} = \sqrt{\frac{\sum_{k=1}^K (\mu_k - \mu)^2}{K}} / \sqrt{\frac{\sum_{j=1}^J \sum_{k=1}^K (\mu_{jk} - \mu)^2}{JK}}$$

- J= drug groups, K=age groups

$$f = \sqrt{((-1.82 - (-0.84))^2 + ((1.41 - (-0.84))^2 + ((-2.11 - (-0.84))^2) / 3) /$$

$$\sqrt{\frac{(-2.23 - (-0.84))^2 + (0.47 - (-0.84))^2 + (-3.7 - (-0.84))^2 + (2.93 - (-0.84))^2 + (-2.2 - (-0.84))^2 + (3.5 - (-0.84))^2 + (0.67 - (-0.84))^2 + (-3.73 - (-0.84))^2 + (-3.27 - (-0.84))^2}{3 * 3}}$$

$$f = 1.74/2.65 = 0.657$$

- Numerator df = 3 (Drug treatments) - 1 = 2
- Number of groups = 3*3 = 9
- **wp.kanova(ndf=2, f=0.657, ng=9, alpha=0.05, power=0.80)**
- n = 26.6 -> **27 samples total (3 per group)**

2. You are interested in determining if there is a difference in treatment (Drug A, B, and C) across age (child, adult, elder) and cancer stage (I, II, III, IV, V). You collect trial data and find that the between-group variance is 27.3, while the total variance is 85.2.

- Care about treatment, age, and cancer stage
- Numerator df = (3-1)*(3-1)*(5-1)=2*2*4=16
- Number of groups is 3*3*5=45
- $\eta^2 = \sigma_m^2 / \sigma_t^2 = 27.3/85.2 = 0.32$
- $f = \sqrt{\eta^2 / (1 - \eta^2)} = \sqrt{0.32 / (1 - 0.32)} = 0.686$
- **wp.kanova(ndf=16, f=0.686, ng=45, alpha=0.05, power=0.80)**
- n 67.03 -> **68 samples, need 90 samples to have even groups (2 per group)**

Drug A			Drug B			Drug C		
c	a	e	c	a	e	c	a	e
-6.4	8.7	-3.1	1.3	-6.0	6.8	-2.0	-4.3	-1.2
-8.2	-6.3	-6.5	3.6	1.3	2.4	1.5	1.3	1.1
7.9	-1	-1.5	3.9	-1.9	1.3	2.5	-8.2	-9.7

Logistic Regression

Description: Tests whether a predictor variable is a significant predictor of a binary outcome, with or without other covariates. It is a type of non-parametric regression: numerical variables are not normally distributed. In Logistic regression, the response variable (Y) is binary (0/1).

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
≥2	0	N/A	N/A	No	N/A

Example: Does body mass index (BMI) influences mortality (yes 1, no 0)?

- $H_0=0$, $H_1 \neq 0$
- You must have at least some background (or good guess) on the **p0** and **p1** probabilities; let's use **0.15** and **0.25**
- Will use 'two-sided' because we don't care about direction
- BMI seems normally distributed, so will go with **normal** for the family (but should confirm the distribution for whatever predictor variable you use)
- Can leave the parameter empty at the default of mean=0, SD=1

R Code: WebPower -> wp.logistic

```
wp.logistic(n = NULL, p0 = NULL, p1 = NULL, alpha = 0.05, power = NULL, alternative = c("two.sided", "less", "greater"), family = c("Bernoulli", "exponential", "lognormal", "normal", "Poisson", "uniform"), parameter = NULL)
```

- **p0**= $\text{Prob}(Y=1|X=0)$: the probability of observing 1 for the outcome variable Y when the predictor X equals 0
- **p1**= $\text{Prob}(Y=1|X=1)$: the probability of observing 1 for the outcome variable Y when the predictor X equals 1
- **alpha**= significance level
- **power**= statistical power
- **alternative**= direction of the alternative hypothesis ("two.sided" or "less" or "greater")
- **family**= distribution of the predictor ("Bernoulli", "exponential", "lognormal", "normal", "Poisson", "uniform"). The default is "Bernoulli"
- **parameter**= corresponding parameter for the predictor's distribution. The default is 0.5 for "Bernoulli", 1 for "exponential", (0,1) for "lognormal" or "normal", 1 for "Poisson", and (0,1) for "uniform"

Effect size calculation

- N/A, uses probability information instead

Logistic Regression

Results:

> #sample size

> wp.logistic(p0=0.15, p1=0.25, alpha=0.05, power=0.80, alternative="two.sided", family="normal")

Power for logistic regression

p0	p1	beta0	beta1	n	alpha	power
0.15	0.25	-1.734601	0.6359888	165.3687	0.05	0.8

→ Round up to 166 total samples

URL: <http://psychstat.org/logistic>

Poisson Regression

Description: Tests whether a predictor variable influences the rate of events over a set period, with or without other covariates. It is a type of non-parametric regression: numerical variables are not normally distributed. In Poisson regression, the events within the rate are assumed to be independent. Subjects can have multiple events, as long as they are independent.

Numeric. Var(s)	Cat. Var(s)	Cat. Var Group #	Cat Var. # of Interest	Parametric	Paired
≥2	0	N/A	N/A	No	N/A

Example: Does a change in drug dose decrease the rate of adverse affects?

- $H_0=0$, $H_1 \neq 0$
- You must have at least some background (or good guess) on the **exp0** and **exp1** rate; let's use **1.0** and **0.80**
- Will use '**less**' because we're asking if the alternative hypothesis has a lower rate than the null
- Because I have no idea about the distribution of drug dosage, I will go with **uniform** (but should confirm the distribution for whatever predictor variable you use)
- Can leave the parameter empty at the default of mean=0, SD=1

R Code: WebPower -> wp.poisson

```
wp.poisson(n = NULL, exp0 = NULL, exp1 = NULL, alpha = 0.05, power = NULL, alternative = c("two.sided", "less", "greater"), family = c("Bernoulli", "exponential", "lognormal", "normal", "Poisson", "uniform"), parameter = NULL)
```

- **exp0**= the base rate under the null hyp.(must be positive value)
- **exp1**= the relative increase of the event rate. It is used for calculation of the effect size
- **alpha**= significance level
- **power**= statistical power
- **alternative**= direction of the alternative hypothesis ("two.sided" or "less" or "greater")
- **family**= distribution of the predictor ("Bernoulli", "exponential", "lognormal", "normal", "Poisson", "uniform"). The default is "Bernoulli"
- **parameter**= corresponding parameter for the predictor's distribution. The default is 0.5 for "Bernoulli", 1 for "exponential", (0,1) for "lognormal" or "normal", 1 for "Poisson", and (0,1) for "uniform"

Effect size calculation

- N/A, uses rate information instead

Poisson Regression

Results:

> #sample size

> wp.poisson(exp0=1.0, exp1=0.80, alpha=0.05, power=0.80, alternative ="less", family="uniform")

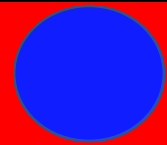
Power for Poisson regression

n	power	alpha	exp0	exp1	beta0	beta1
1666.539	0.8	0.05	1	0.8	0	-0.2231436

→ Round up to 1667 total samples

URL: <http://psychstat.org/poisson>

Logistic/Poisson Regression: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if body temperature influences sleep disorder prevalence (yes 1, no 0). You collect the following trial data.

Temperature	98.6	98.5	99.0	97.5	98.8	98.2	98.5	98.4	98.1
Sleep Disorder?	No	No	Yes	No	Yes	No	No	Yes	No

2. You are interested in determining if the rate of lung cancer incidence changes with a drug treatment.

Logistic/Poisson Regression: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if body temperature influences sleep disorder prevalence (yes 1, no 0). You collect the following trial data.

Temperature	98.6	98.5	99.0	97.5	98.8	98.2	98.5	98.4	98.1
Sleep Disorder?	No	No	Yes	No	Yes	No	No	Yes	No

- Logistic Regression (two.sided)
- Mean temp is 98.4 (SD=0.436) -> range of one SD=(97.964 --98.836)
- $P_0=0.33$ (as only one had sleep disorder at ranges outside one SD); $P_1=0.67$
- Temperature is normally distributed
- `wp.logistic(p0=0.33, p1=0.67, alpha=0.05, power=0.80, alternative="two.sided", family="normal")`
- $n = 40.80$ -> **41 samples total**

2. You are interested in determining if the rate of lung cancer incidence changes with a drug treatment.

- Poisson Regression (two.sided)
- Expect the base rate (intercept) for male lung cancer is 57.8 (per 100,000), so $\exp_0 = \exp(57.8/100000) = 1.0005$
- Expect the relative increase of the event rate (slope) to be -1.02, so $\exp_1 = \exp(-1.02) = 0.36$
- Go with default distribution of Bernoulli
- `wp.poisson(exp0=1.0005, exp1=0.36, alpha=0.05, power=0.80, alternative="two.sided", family="Bernoulli")`
- $n = 56.8$ -> **59 samples total**

Multilevel Modeling: Cluster Randomized Trials

Description: Multilevel models are used when data are clustered within a hierarchical structure that will make them non-independent. Also known as linear mixed models.

Cluster randomized trials (CRT) are a type of multilevel design where the entire cluster is randomly assigned to a control arm or one or more treatment arms.

Example: Is there a difference in blood glucose levels between a treatment and control?

- $H_0=0$, $H_1 \neq 0$
- You don't have background info, so you guess that there is a **medium** effect size
- For f-tests:
 - 0.1=small, 0.25=medium, and 0.4 large effect sizes
- Don't know the **icc**, so will guess at 0.1 (0.5 is the default for repeated measures, but we expect this to be lower, since the observations are from different people)
- **Alternative** is "two.sided" as we only care about difference
- We can test for two sizes: number per cluster or cluster number
 1. Try for 100 clusters
 2. Try for 15 individuals per cluster to get cluster number

R Code: WebPower -> wp.crt2arm

```
wp.crt2arm(n=NULL, f = NULL, J = NULL, icc = NULL, power = NULL, alpha = 0.05, alternative = c("two.sided", "one.sided"))
```

- **n**= sample size (number of individuals per cluster)
- **f**= effect size (either main effect of treatment, or mean difference between treatment clusters and control clusters)
- **J**= number of clusters/sides. It tells how many clusters are considered in the study design. At least two clusters are required
- **icc**= intra-class correlation (degree to which two randomly drawn observations within a cluster are correlated)
- **alpha**= significance level
- **power**= statistical power
- **alternative**= direction of the alternative hypothesis ("two.sided" or "less" or "greater")

Effect size calculation

- $f = \mu_D / \sqrt{(\sigma_B^2 + \sigma_W^2)}$
 - μ_D =mean difference between treatment and control clusters
 - σ_B^2 =between-cluster variance
 - σ_W^2 =within-cluster variance

NOTE: here we show a 2 arm example (treatment, control); to use a 3 arm design (treatment1, treatment2, control), use **wp.crt3arm**

Multilevel Modeling: Cluster Randomized Trials

Results:

> #Multilevel Modeling

>

> #CRT sample size (number per cluster)

> wp.crt2arm(f=0.25, J=100, icc=0.1, alpha=0.05, power=0.80, alternative="two.sided")

Cluster randomized trials with 2 arms

J	n	f	icc	power	alpha
100	9.456102	0.25	0.1	0.8	0.05

→ Round up to 10 individuals per cluster

NOTE: n is the number of subjects per cluster.

URL: <http://psychstat.org/crt2arm>

>

> #CRT sample size (cluster number)

> wp.crt2arm(f=0.25, n=15, icc=0.1, alpha=0.05, power=0.80, alternative="two.sided")

Cluster randomized trials with 2 arms

J	n	f	icc	power	alpha
82.33782	15	0.25	0.1	0.8	0.05

→ Round up to 84 clusters (42 per arm)

NOTE: n is the number of subjects per cluster.

URL: <http://psychstat.org/crt2arm>

Multilevel Modeling: Multisite Randomized Trials

Description: Multilevel models are used when data are clustered within a hierarchical structure that will make them non-independent. Also known as linear mixed models.

Multisite randomized trials (MRT) are a type of multilevel design where the entire cluster is randomly assigned to a control arm or one or more treatment arms, but then can be analyzed in a two-level hierarchical linear model. Can look at three types of tests: (1) The "**main**" type tests treatment main effect; (2) The "**site**" type tests the variance of cluster/site means; and (3) The "**variance**" type tests variance of treatment effects

Example: Is there a difference in blood glucose levels between a treatment and control?

- $H_0=0$, $H_1 \neq 0$
- You don't have background info, so you guess that there is a **medium** effect size
- For f-tests:
 - 0.1=small, 0.25=medium, and 0.4 large effect sizes
- Try a **main** effect, with a **tau11** of 0.5 and a **sg2** of 1.0
- **Alternative** is "two.sided" as we only care about difference
- We can test for two sizes: number per cluster or cluster number
 1. Try for 100 clusters
 2. Try for 15 individuals per cluster to get cluster number

R Code: WebPower -> wp.mrt2arm

```
wp.mrt2arm(n = NULL, f = NULL, J = NULL, tau00 = NULL, tau11 = NULL, sg2 = NULL, power = NULL, alpha = 0.05, alternative = c("two.sided", "one.sided"), type = c("main", "site", "variance"))
```

- **f**= effect size (either main effect of treatment, or mean difference between treatment clusters and control clusters)
- **J**= number of clusters/sides. It tells how many clusters are considered in the study design. At least two clusters are required
- **tau00**= variance of cluster/site means (must be positive); one of the residual variances in the second level
- **tau11**= variance of treatment effects across sites (must be positive); one of the residual variances in the second level
- **sg2**= level-one error variance; variance in the first level
- **alpha**= significance level
- **power**= statistical power
- **alternative**= direction of the alternative hypothesis ("two.sided" or "less" or "greater")
- **type**= type of effect ("main", "site", or "variance") with main as default. No tau00 needed for main effect; no tau11 needed for site effect; no tau or f needed for variance effect

Effect size calculation

- $f = \mu_D / \sqrt{\sigma^2}$
 - μ_D =mean difference between treatment and control clusters
 - σ^2 =sample-specific variance

NOTE: here we show a 2 arm example (treatment, control); to use a 3 arm design (treatment1, treatment2, control), use **wp.mrt3arm**

Multilevel Modeling: Multisite Randomized Trials

Results:

> #MRT sample size (number per cluster)

> wp.mrt2arm(f=0.25, J=100, tau11=0.5, sg2=1.0, alpha=0.05, power=0.80, alternative="two.sided")

Multisite randomized trials with 2 arms

J	n	f	tau11	sg2	power	alpha
100	14.24177	0.25	0.5	1	0.8	0.05

→ Round up to 15 individuals per cluster

NOTE: n is the number of subjects per cluster

URL: <http://psychstat.org/mrt2arm>

>

> #MRT sample size (cluster number)

> wp.mrt2arm(f=0.25, n=15, tau11=0.5, sg2=1.0, alpha=0.05, power=0.80, alternative="two.sided")

Multisite randomized trials with 2 arms

J	n	f	tau11	sg2	power	alpha
98.2174	15	0.25	0.5	1	0.8	0.05

→ Round up to 100 clusters (50 per arm)

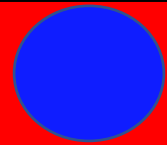
NOTE: n is the number of subjects per cluster

URL: <http://psychstat.org/mrt2arm>

Multilevel Modeling: Cluster/Site Size

- **While the webpower documentation says it can be used for clusters or sites for 2+, it cannot be used for small cluster number unless effect size (f) is large enough and the inter correlation coefficient (icc) is low enough**

Multilevel Modeling: Practice



Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if a drug A could lower blood pressure for patients with hypertension using 50 hospitals across the county, separated by cluster. From trial data, you found blood pressure to be lowered by 6.90, with a between-cluster variance of 58 and a within-cluster variance of 243.
2. You are interested in determining if a drug B changes blood pressure for patients with hypertension using 6 hospitals the state, randomizing at each site. From trial data, you found blood pressure to be different by 2.5, with a variance of treatment effect across site of 2 and a person-specific variance of 1.

Multilevel Modeling: Answers

Calculate the sample size for the following scenarios (with $\alpha=0.05$, and power=0.80):

1. You are interested in determining if a drug A could lower blood pressure for patients with hypertension using 50 hospitals across the county, separated by cluster. From trial data, you found blood pressure to be lowered by 6.90, with a between-cluster variance of 58 and a within-cluster variance of 243.
 - Number of clusters (J) is 50
 - One-tailed test -> "less"
 - Effect size= $f = \mu_D / \sqrt{(\sigma_B^2 + \sigma_W^2)} = 6.90 / \sqrt{(58 + 243)} = 0.40$
 - Inter-class correlation = $\rho = \sigma_B^2 / (\sigma_B^2 + \sigma_W^2) = 58 / (58 + 243) = 0.19$
 - **wp.crt2arm(f=0.40, J=50, icc=0.19, alpha=0.05, power=0.80, alternative="less")**
 - $n = 16.45$ -> **17 samples per cluster**
2. You are interested in determining if a drug B changes blood pressure for patients with hypertension using 6 hospitals the state, randomizing at each site. From trial data, you found blood pressure to be different by 2.5, with a variance of treatment effect across site of 2 and a person-specific variance of 1.
 - Number of sites (J) is 6
 - Two-tailed test -> "two.sided"
 - effect size= $f = \mu_D / \sqrt{\sigma^2} = 2.5 / \sqrt{1} = 2.5$
 - tau11= 2
 - sg2= 1
 - **wp.mrt2arm(f=2.5, J=6, tau11=2, sg2=1, alpha=0.05, power=0.80, alternative="two.sided")**
 - $n = 3.86$ -> **4 samples per site**

Generalized Linear Mixed Models

Description:

Combination of a Generalized Linear Model (GLM) and Mixed Model

- GLM: can be used with non-normal data
- Mixed Model: include both **fixed** and **random** effects

These models can be made very sophisticated and cover a very large range of models

- Need to understand how to create model and define variables
- Therefore, it requires a Module of their own
- Look for the second sample size module in R: **Sample Size Calculation with R: GLMMs**

Acknowledgements



- The DaCCoTA is supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729.
- For the labs that use the Biostatistics, Epidemiology, and Research Design Core in any way, including this Module, please acknowledge us for publications. **"Research reported in this publication was supported by DaCCoTA (the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729).**

References

- In this module, many of the functions that I show, I've refrained from including all the options for simplicity
- More detailed descriptions (and sometimes examples) can be found in the package manuals

General References:

<https://www.statmethods.net/stats/power.html>

https://www.graphpad.com/guides/prism/7/statistics/index.htm?stat_sample_size_for_nonparametric.htm

Packages:

<https://cran.r-project.org/web/packages/pwr/pwr.pdf>

<https://cran.r-project.org/web/packages/WebPower/WebPower.pdf>

https://webpower.psychstat.org/wiki/_media/grant/webpower_manual_book.pdf