MEDB 5501, Module09

2024-10-15

Topics to be covered

- What you will learn
    - Restructuring your data
    - One-tailed tests
    - Checking assumptions
    - Alternative tests
    - R code for alternative tests
    - Paired data
    - R code for paired data
    - Your homework

Two different data structures

- Wide format
  - Group 1 in first column
  - Group 2 in second column
- Long format
  - Group 1 in first n_1 rows
  - Group 2 in remaining rows
  - Could be different order or mixed
  - Additional column for group

There are two different data structures that you will encounter in research settings. The wide format places the data from each group in a different column. The long format places both groups in the same column, usually with the first group on the top and the second group on the bottom. It could be reversed, however, or it could be all jumbled together.

You will find that certain analyses are only available in the long format. Other analyses are only available in the wide format.

Being able to move from the wide format to the long format or vice versa is an important programming skill that you should learn early. I don't want to emphasize programming too much in this class, but I am making an exception here.

Hypothetical data in the wide format

```
# A tibble: 4 × 2
    trt    pbo
  <dbl> <dbl>
1    15    34
2    13    31
3    18    36
4    19    NA
```

Pivoting from wide form to long form

```
1  wide_form |>
2    pivot_longer(
3      cols=c("trt", "pbo"),
4      names_to="intervention",
5      values_to="outcome") |>
6    filter(!is.na(outcome))
```

```
# A tibble: 7 × 2
  intervention outcome
  <chr>          <dbl>
1 trt               15
2 pbo               34
3 trt               13
4 pbo               31
5 trt               18
6 pbo               36
7 trt               19
```

Hypothetical data in the long format

```
# A tibble: 8 × 3
      id result exposed
   <int>  <dbl> <chr>
1      1     45 n
2      2     43 n
3      3     48 n
4      4     49 n
5      1     64 y
6      2     61 y
7      3     66 y
8      4     62 y
```

Pivoting from long form to wide form

```
1  long_form |>
2    pivot_wider(
3      names_from=exposed,
4      values_from=result)
```

```
# A tibble: 4 × 3
      id     n     y
   <int> <dbl> <dbl>
1      1    45    64
2      2    43    61
3      3    48    66
4      4    49    62
```

Break #1

- What you have learned
    - Restructuring your data
- What's coming next
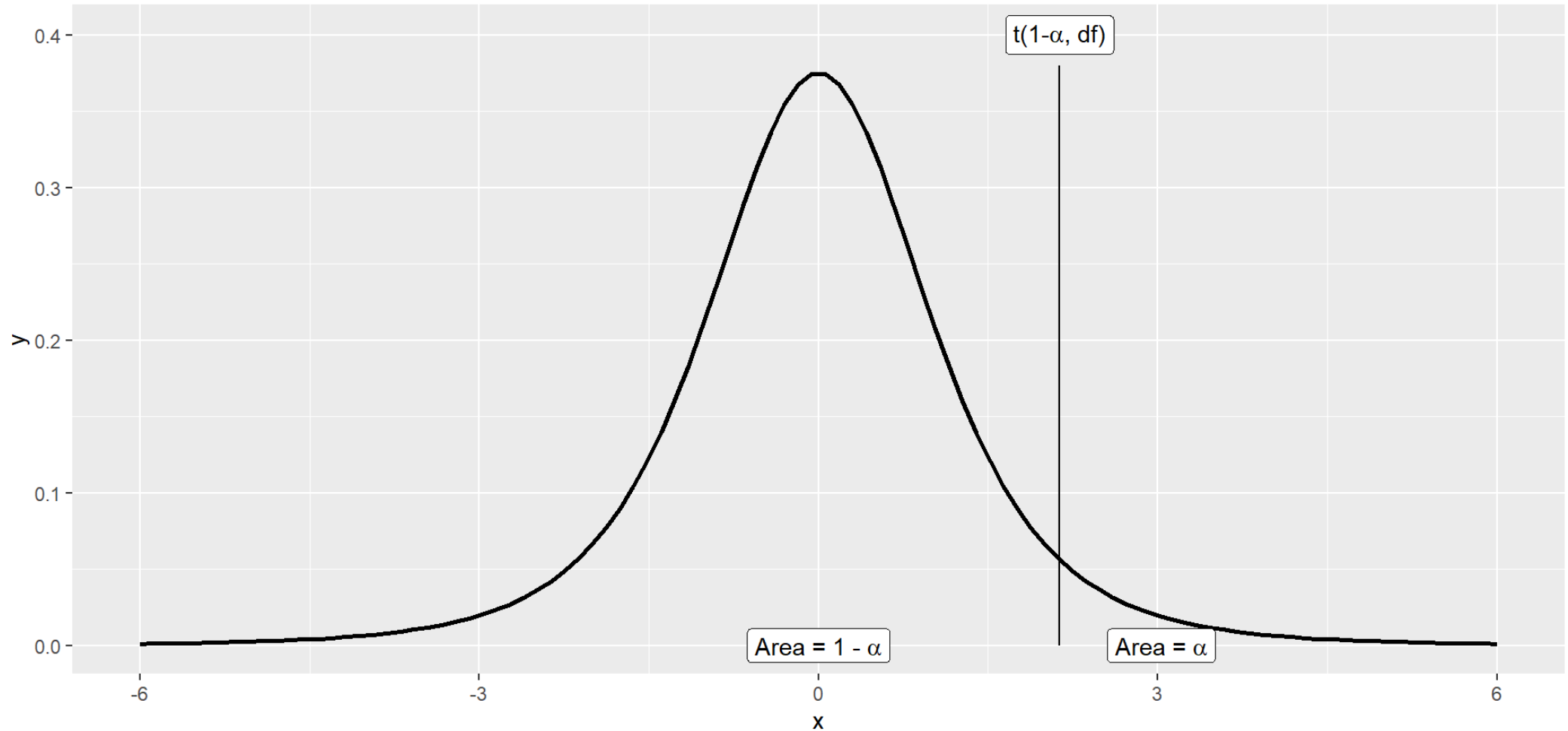   - One-tailed tests

One-tailed tests

- $H_0: \mu_1 - \mu_2 = 0$
- $H_1: \mu_1 - \mu_2 > 0$
  - Accept H_0 if $\bar{X}_1 - \bar{X}_2$ is large negative
  - Accept H_0 if $\bar{X}_1 - \bar{X}_2$ is close to zero
  - Reject H_0 if $\bar{X}_1 - \bar{X}_2$ is large positive
- More precisely,
  - Accept H_0 if $T \leq t(1 - \alpha, df)$
  - Reject H_0 if $T > t(1 - \alpha, df)$

Sometimes you are only interested in changes in a positive direction. This means the first population mean is larger than the second population mean.

Critical value for a one tailed test

Graph drawn by Steve Simon on 2024-10-06

$t(1-\alpha, df)$

Area = 1 - α

Area = α

The area to the right of the critical value is α and the area to the left is $1-\alpha$.

One-tailed tests

- $H_0 : \mu_1 - \mu_2 = 0$
- $H_1 : \mu_1 - \mu_2 < 0$
  - Reject H_0 if $\bar{X}_1 - \bar{X}_2$ is large negative
  - Accept H_0 if $\bar{X}_1 - \bar{X}_2$ is close to zero
  - Accept H_0 if $\bar{X}_1 - \bar{X}_2$ is large positive
- More precisely,
  - Accept H_0 if $T \geq t(\alpha, df)$
  - Reject H_0 if $T < t(\alpha, df)$

There is a second one-sided hypothesis. This tests the hypothesis that the difference is negative, or equivalently that the first population mean is smaller than the second population mean.

Critical value

Graph drawn by Steve Simon on 2024-10-06
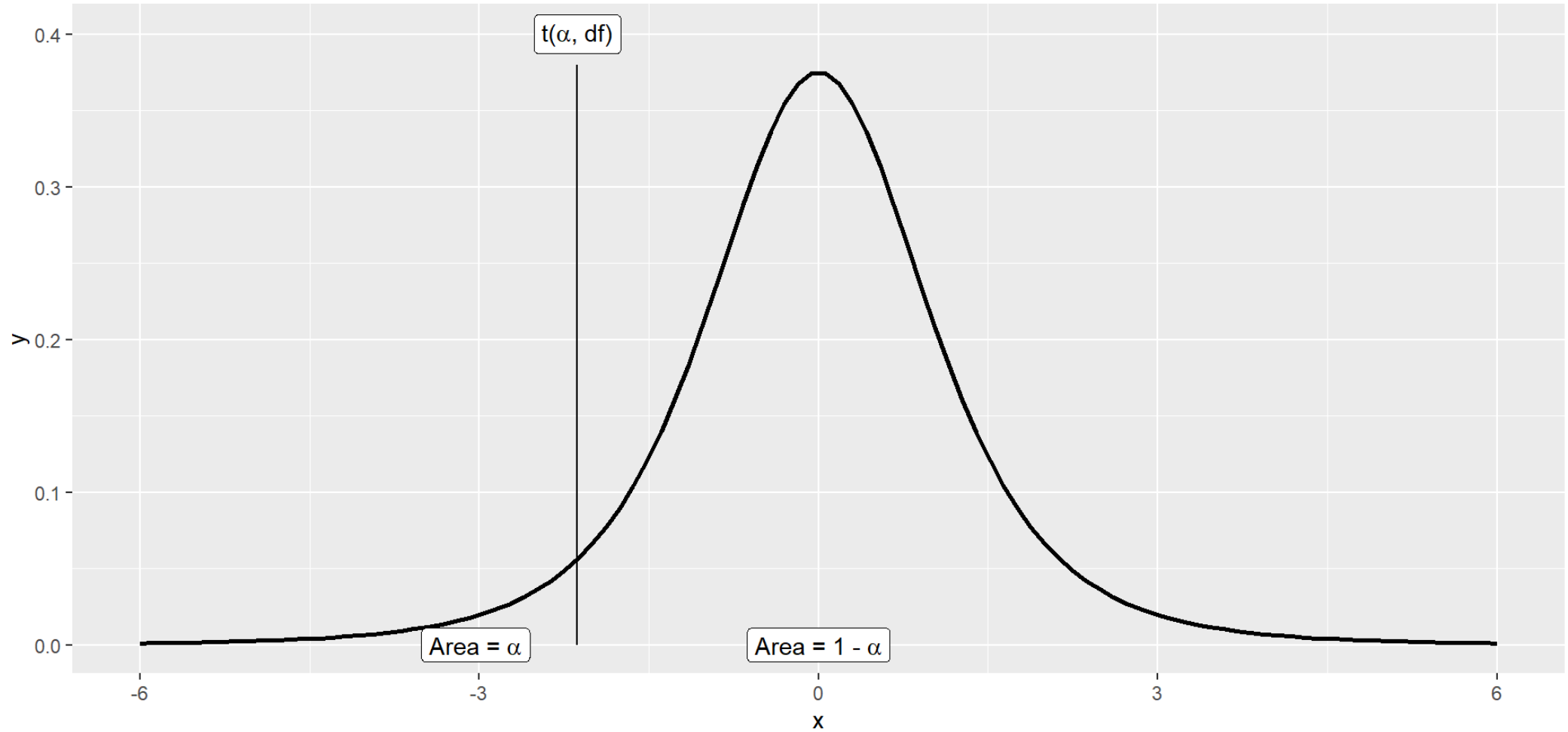
The area to the left is α and the area to the right is 1-α.

Calculation of the p-value

- For testing $H_1 : \mu_1 - \mu_2 \neq 0$

  - p-value $= 2P\big[t(n_1 + n_2 - 2) > |T|\big]$

- For testing $H_1 : \mu_1 - \mu_2 > 0$

  - p-value $= P\big[t(n_1 + n_2 - 2) > T\big]$

- For testing $H_1 : \mu_1 - \mu_2 < 0$

  - p-value $= P\big[t(n_1 + n_2 - 2) < T\big]$

- For any of these hypothesis,

  - Accept H_0 if p-value $> \alpha$

The calculation of the p-value changes slightly if you have a one-sided hypothesis. But the decision rule is always the same, accept the null hypothesis if the p-value is large, larger than alpha.

R code for one-sided hypotheses

- For testing $H_1 : \mu_1 - \mu_2 \neq 0$

  - `alternative="two-sided"`

- For testing $H_1 : \mu_1 - \mu_2 > 0$

  - `alternative="greater"`

- For testing $H_1 : \mu_1 - \mu_2 < 0$

  - `alternative="less"`

The t.test function in R has an argument, alternative. The default is "two-sided". Change this to "greater" or "less" to produce the appropriate one-sided hypothesis.

Break #2

- What you have learned

  - One-tailed tests

- What's coming next

  - Checking assumptions

Population model for the two-sample t-test

- Population 1
  - $X_{11}, X_{12}, \ldots, X_{1N_1}$
  - X_{1i} are independent $N(\mu_1, \sigma_1)$
- Population 2
  - $X_{21}, X_{22}, \ldots, X_{2N_2}$
  - X_{2i} are independent $N(\mu_2, \sigma_2)$
- Both populations independent from each other
- Both populations have the same standard deviation

Here is the population model for the two-sample t-test.

Violations of these population model

- Non-normality

- Heterogeneity

- Lack of independence

  - Within each group

  - Between groups

There are several possible violations of the population model

Checking for non-normality

- Boxplot

  - Might be sufficient by itself

- Non-normality is less concerning with large sample sizes

- Residual analysis

  - Normal probability plot

  - Histogram

  - Always residuals, never the original data

The boxplot is a great way to check for non-normality. Look for skewness and/or outliers. Other types of non-normality, such as a bimodal distribution or a light-tailed distribution are not a serious problem. Also, non-normality is less of a concern for large sample sizes.

You really don't need to look at residuals, but if you do, compute them the same way as in linear regression. Always examine the residuals and not the original data. If the two groups have different means (which is something that you actually hypothesize might happen), then any analysis of the original data will show a bimodal pattern.

Checking for heterogeneity

- Boxplot

- Compare group standard deviations

  - 3 to 1 or higher ratio

- Is the data unbalanced $(n_1 \neq n_2)$

The boxplot is also a simple way to check for heterogeneity. Is one box a lot bigger than the other?

You should also compute means and standard deviations for each group. This is a routine that I always recommend. Wade in from the shallow end of the pool.

If one standard deviation is a lot larger than another, you may have problems with heterogeneity. Don't worry about small disparities. Look for a three fold change or greater between the larger standard deviation and the smaller standard deviation.

Unbalanced data, where you have more observations in one group than the other is not a violation of assumptions, but it does tend to exacerbate problems with heterogeneity if it exists.

Checking for independence

- Qualitative assessments only
  - Clustering
  - Geographic proximity

The independence needs to be evaluated qualitatively. Look for features of the data such as clustering or geographic proximity, especially for infectious diseases.

Consequences of violations

- Loss of control of Type I error rate

  - Two possibilities

    - Liberal, actual alpha much larger than 0.05

    - Conservative, actual alpha much smaller than 0.05

- Possible increase of Type II error rate = loss of power

  - Especially if there are outliers

- Confidence intervals too wide or too narrow

What does a violation of assumptions mean from a practical perspective? First, it means that you lose control of the Type I error rate. It can change liberally, meaning that a test you intended to have a Type I error rate of 0.05 actually has a much larger Type I error rate. It can change conservatively, meaning that a test that you intended to have a Type I error rate of 0.05 actually might have a much smaller Type I error rate.

Now you might think that being conservative sounds pretty good. Why wouldn't you mind having a smaller Type I error rate? It turns out that you are robbing Peter to pay Paul. What you gain in a reduction of the Type I error rate gets taken away by an increase in the Type II error rate. This is equivalent to a loss in power.

The problem with increased Type II error rate and a loss of power is especially likely if your data has serious outliers.

Break #3

- What you have learned
  - Checking assumptions
- What's coming next
  - Alternative tests

Alternative tests

- If you have heterogeneity

    - Welch's test

- If you have non-normality

    - Mann-Whitney-Wilcoxon test

- If you have both:

    - Log transformation

- If you have lack of independence

    - Random effects models

        - Beyond the scope of this class

Speaker notes

There are several approaches that you should consider if you have trouble with the assumptions of the two-sample t-test.

Welch's test is a simple modification to the test if you have unequal variances. The Wilcoxon-Mann-Whitney test is a non-parametric test that does not require the assumption of normality.

If you have both heterogeneity and non-normality, and it has to be specific type of heterogeneity and non-normality, then a log transformation is good. It stretches the small values and squeezes the large values.

Lack of independence typically requires a fairly complex approach like a random effects model that is well beyond the scope of this class. This model is covered in quite a bit of detail by Anlin Cheng in MEDB 5503, Applied Biostatistics III.

Welch's test, 1

- $T = \frac{\bar{X}_1 - \bar{X}_2}{se}$

- se = standard error changes slightly

  - $se = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

- df = degrees of freedom changes slightly

  - $df = \dfrac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{(n_1-1)n_1^2} + \frac{s_2^4}{(n_2-1)n_2^2}}$

    - Also known as Satterthwaite's approximation

Welch's test makes minor modification to the standard error and the degrees of freedom. The formula for the degrees of freedom looks difficult, but it is just involves addition, division and raising to power. It is one of those formulas that has very little intuition and it is also something that makes us glad that the computer does all the work.

Welch's test, 2

- R code
  - `var.equal=FALSE`

The R code is fairly easy. Change the argument var.equal from TRUE to FALSE.

Controversy over the use of Welch's test, 1

- Some advocate using it all the time.
  - Welch's test is valid for homogeneity OR heterogeneity
- Some advocate using a preliminary hypothesis
  - $H_0 : \sigma_1 = \sigma_2$
    - Levene's test
    - Bartlett's test
    - Brown-Forsythe test

Controversy over the use of Welch's test, 2

- My opinion:
  - Avoid ANY test of heterogeneity
  - Make decision based on prior experience

Tests of heterogeneity are worthless. All are excessively sensitive to the normality assumption (some more than others). Worse, for very large sample sizes, these tests tend to conclude that you have heterogeneity even for trivial changes in variation from one group to another.

I do not use Welch's test, unless there are strong a priori reasons to believe that there is heterogeneity.
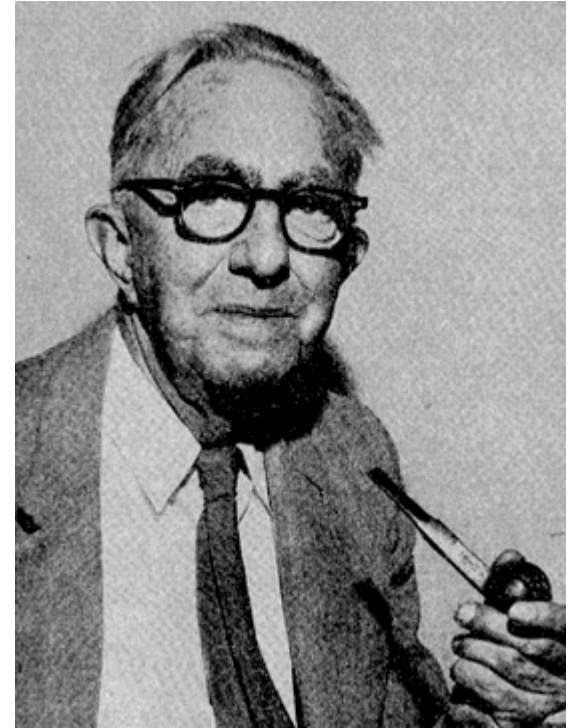
Mann-Whitney-Wilcoxon test

Henry Mann, PhD in Mathematics in 1935

Donald Ransom Whitney, PhD in Mathematics in 1946

Frank Wilcoxon, PhD in Chemistry, 1924

Here are images behind the three researchers who developed the Mann-Whitney-Wilcoxon test. The websites listed below also provide nice biographies of these three people.

Henry Mann: https://math.osu.edu/about-us/history/henry-berthold-mann Donald Ransom Whitney: http://sections.maa.org/ohio/ohio_masters/whitney.pdf Frank Wilcoxon: https://en.wikipedia.org/wiki/Frank_Wilcoxon

Other names

- Wilcoxon-Mann-Whitney

- Mann-Whitney U test

- Wilcoxon rank sum test

  - Not to be confused with Wilcoxon signed rank test

There are alternate names for this test. Most people actually give Frank Wilcoxon first billing. It turns out that Frank Wilcoxon published a paper in 1945 that outlined both the rank sum test and the signed rank test. The rank sum test was derived under a limited assumption of equal sample sizes in the two groups. Mann and Whitney extended this test in 1947 to make it work with unequal sample sizes.

Theory behind the Mann-Whitney-Wilcoxon test

- Combine the two groups

- Assign ranks $R(X_{ij})$

  - 1 to smallest value, 2 to second smallest, etc.

- Compute average rank, $\bar{R}$

- Compute the sum of the ranks in first group, $\Sigma R(X_{1j})$

- T = $\dfrac{\Sigma R(X_{1j}) - n_1 \bar{R}}{se}$

  - se = $\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$

Alternate theory

- Count the times that X_{1j} "wins" compared to all the X_2's

  - X_{1j} wins if it is larger than X_{2k}

  - Count ties as 0.5 (half of a "win")

  - There are $n_1 n_2$ contests

  - U = number of wins

  - T = $\dfrac{U - \frac{n_1 n_2}{2}}{se}$

  - Same se as earlier slide

Hypothetical data

- X1: 34, 1695, 1193
- X2: 652, 11, 24, 16, 1543, 39

Rank sum for hypothetical data

```
Group   Outcome   Rank
    1        34       4
    1      1695       9
    1      1193       7
    2       652       6
    2        11       1
    2        24       3
    2        16       2
    2      1543       8
    2        39       5
```
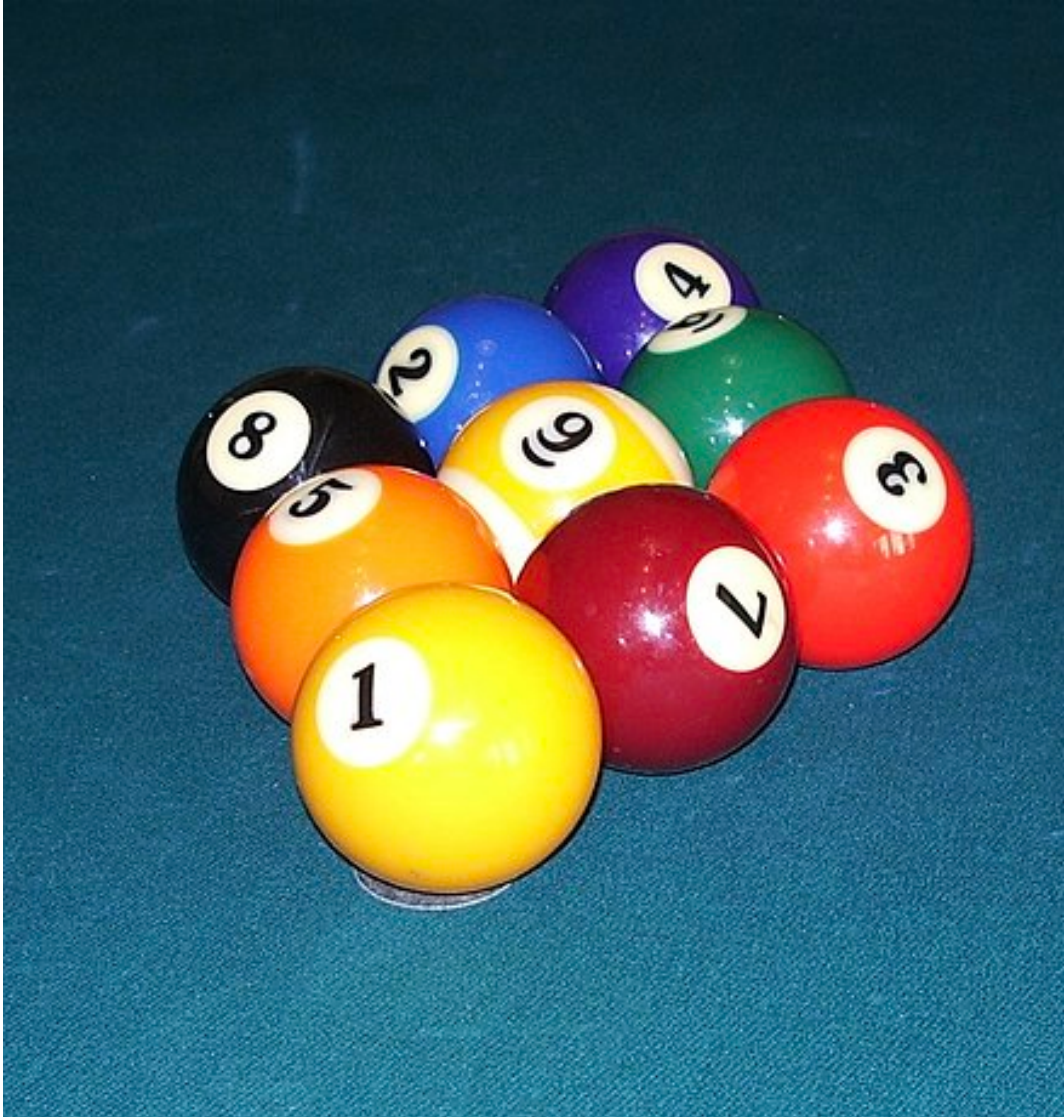
- $\Sigma R(X_{1j})$ = 4 + 9 + 7 = 20

- Expected sum = 3*5 = 15

Assign the rank of 1 to the smallest value (11), a rank of 2 to the second smallest value (16),…, and a rank of 9 to the largest value (1695)

Counting "wins" for hypothetical data

```
        652    11    24    16 1543    39
  34   Loss   Win   Win   Win Loss  Loss
1695    Win   Win   Win   Win  Win   Win
1193    Win   Win   Win   Win Loss   Win
```

- U = 14 (out of 18 possible)

- Expected wins = 9

The probability basis for ranking

You randomly draw three balls from an urn containing balls numbered 1 through 9. What are the chances that you draw the a particular combination of these three balls?

List all the possible outcomes, 1

```
1,2,3   1,2,4   1,2,5   1,2,6   1,2,7   1,2,8   1,2,9
1,3,4   1,3,5   1,3,6   1,3,7   1,3,8   1,3,9   1,4,5
1,4,6   1,4,7   1,4,8   1,4,9   1,5,6   1,5,7   1,5,8
1,5,9   1,6,7   1,6,8   1,6,9   1,7,8   1,7,9   1,8,9
2,3,4   2,3,5   2,3,6   2,3,7   2,3,8   2,3,9   2,4,5
2,4,6   2,4,7   2,4,8   2,4,9   2,5,6   2,5,7   2,5,8
2,5,9   2,6,7   2,6,8   2,6,9   2,7,8   2,7,9   2,8,9
3,4,5   3,4,6   3,4,7   3,4,8   3,4,9   3,5,6   3,5,7
3,5,8   3,5,9   3,6,7   3,6,8   3,6,9   3,7,8   3,7,9
3,8,9   4,5,6   4,5,7   4,5,8   4,5,9   4,6,7   4,6,8
4,6,9   4,7,8   4,7,9   4,8,9   5,6,7   5,6,8   5,6,9
5,7,8   5,7,9   5,8,9   6,7,8   6,7,9   6,8,9   7,8,9
```

Here are all the possible outcomes. There are 84 outcomes and each has probability 1/84.

List all the possible outcomes, 2

```
 6: 1,2,3
 7: 1,2,4
 8: 1,2,5  1,3,4
 9: 1,2,6  1,3,5  2,3,4
10: 1,2,7  1,3,6  1,4,5  2,3,5
11: 1,2,8  1,3,7  1,4,6  2,3,6  2,4,5
12: 1,2,9  1,3,8  1,4,7  1,5,6  2,3,7  2,4,6  3,4,5
13: 1,3,9  1,4,8  1,5,7  2,3,8  2,4,7  2,5,6  3,4,6
14: 1,4,9  1,5,8  1,6,7  2,3,9  2,4,8  2,5,7  3,4,7  3,5,6
```

List all the possible outcomes, 3

```
15: 1,5,9  1,6,8  2,4,9  2,5,8  2,6,7  3,4,8  3,5,7  4,5,6
16: 1,6,9  1,7,8  2,5,9  2,6,8  3,4,9  3,5,8  3,6,7  4,5,7
17: 1,7,9  2,6,9  2,7,8  3,5,9  3,6,8  4,5,8  4,6,7
18: 1,8,9  2,7,9  3,6,9  3,7,8  4,5,9  4,6,8  5,6,7
19: 2,8,9  3,7,9  4,6,9  4,7,8  5,6,8
20: 3,8,9  4,7,9  5,6,9  5,7,8
21: 4,8,9  5,7,9  6,7,8
22: 5,8,9  6,7,9
23: 6,8,9
24: 7,8,9
```

Controversy over the use of nonparametric tests

- Unclear what your hypothesis is
  - Inequality in population means
  - Inequality in population medians
  - $P[X > Y] \neq \frac{1}{2}$
- Loss of power
  - Multiply t-test sample size by pi/3.

Log transformation

- Arithmetic mean of original data
  - $\frac{1}{n}\Sigma X_i$
- Arithmetic mean of log transformed data
  - $\frac{1}{n}\Sigma log(X_i)$
  - $\frac{1}{n}log(X_1 \times X_2 \times \ldots \times X_n)$
- Arithmetic mean of log transformed data back transformed
  - $(X_1 \times X_2 \times \ldots \times X_n)^{\frac{1}{n}}$
  - This is known as the geometric mean

Break #4

- What you have learned

  - Alternative tests

- What's coming next

  - R code for alternative tests

postural-sway data dictionary, 1

```
data_dictionary: postural-sway.txt

copyright: |
  Permission is granted to freely use material on the OzDASL site, although
acknowledgement of the OzDASL site as the data source would be appreciated.
Please be aware though that some of the OzData data sets were originally
published elsewhere; original sources are cited on the description page. If
you intend to re-publish one of these data sets it may be necessary to obtain
permission from the original copyright holder.

description: |
  Postural sway is a measure of how well patients can balance. The postural
sway was measured using a force plate in two groups of subjects, elderly or
young. Sway was measured in the forward/back direction and in the side-to-side
direction.

additional_description: https://gksmyth.github.io/ozdasl/general/balaconc.html
```

postural-sway data dictionary, 2

```
download_url: https://gksmyth.github.io/ozdasl/general/balaconc.txt

format: tab delimited

varnames: Included in the first line

missing_value_code: not needed

size:
  rows: 17
  columns: 3
```

postural-sway data dictionary, 3

```
Age:
  scale: nominal
  values:
  - Elderly
  - Young

FBSway:
  label: Sway in the forward-backward direction
  scale: ratio
  range: positive real

SideSway:
  label: Sway in the side-to-side direction
  scale: ratio
  range: positive real
```

simon-5501-09-sway.qmd, 1

```
---
title: "Alternative analysis of postural sway data"
format:
  html:
    embed-resources: true
---

This program runs some alternatives to the two-sample t-test. Consult the
[data dictionary][dic] for information about the data itself.

[dic]: https://github.com/pmean/data/blob/main/files/postural-sway.yaml

This program was written by Steve Simon on 2024-10-13. It is placed in the
public domain.
```

simon-5501-09-sway.qmd, 2

```
## Libraries

```{r setup}
#| message: false
#| warning: false
library(tidyverse)
```
```

simon-5501-09-sway.qmd, 3

```
## Read data

```{r read-sway}
sway <- read_tsv(
  file="../data/postural-sway.txt",
  col_types="cnn")
names(sway) <- tolower(names(sway))
glimpse(sway)
```
```

simon-5501-09-sway.qmd, 4

```
## Boxplot of front-to-back sway by age

```{r boxplot-1}
#| fig.height: 2
#| fig.width: 6
sway |>
  ggplot(aes(age, fbsway)) +
    geom_boxplot() +
    ggtitle("Graph drawn by Steve Simon on 2024-10-13") +
    xlab("Treatment group") +
    ylab("Front to back sway") +
    coord_flip()
```

The outlier causes some concern about the validity of the two-sample t-test.
```

```
## Descriptive statistics for front-to-back sway by age

```{r group-means}
sway |>
  group_by(age) |>
  summarize(
    fb_mn=mean(fbsway),
    fb_sd=sd(fbsway),
    n=n())
```

In addition to the outlier, notice that the group with the larger mean
(elderly) has the larger standard deviation. This indicates that a log
transformation may produce better results.
```

simon-5501-09-sway.qmd, 6

```
## Log transformation, 1

```{r log-transform-1}
sway |>
  mutate(log_fbsway=log10(fbsway)) -> log_sway
```
```

simon-5501-09-sway.qmd, 7

```
## Log transformation, 2

```{r log-transform-2}
#| fig.height: 2
#| fig.width: 6
log_sway |>
  ggplot(aes(age, log_fbsway)) +
    geom_boxplot() +
    ggtitle("Graph drawn by Steve Simon on 2024-10-13") +
    xlab("Treatment group") +
    ylab("Front to back sway") +
    coord_flip()
```

There are no outliers after a log transformation.
```

simon-5501-09-sway.qmd, 8

```
## Log transformation, 3

```{r compare-means-on-log-scale}
log_sway |>
  group_by(age) |>
  summarize(
    log_mn=mean(log_fbsway),
    log_sd=sd(log_fbsway),
    n=n())
```

The standard deviations on the log scale are quite a bit more similar than
they were on the original scale.
```

simon-5501-09-sway.qmd, 9

```
## Two-sample t-test using the log transformation

```{r t-test}
m2 <- t.test(
  log_fbsway ~ age,
  data=log_sway,
  alternative="two.sided",
  var.equal=TRUE)
m2
```

There is a statistically significant difference between the log front-to-back
sway between elderly patients and young patients. The confidence interval will
be interpreted after transforming back to the original scale of measurement.
```

simon-5501-09-sway.qmd, 10

```
## Back-transform confidence interval to the original scale.

```{r back-transform}
10^(m2$conf.int)
```
```

We are 95% confident that the geometric mean front-to-back sway in elderly patients is somewhere between 1.24 times higher and 8 times higher than the geometric mean for young patients. This indicates a statistically significantly higher mean for elderly patients. The confidence interval is still very wide, indicating a lot of sampling error in these two small samples.

simon-5501-09-sway.qmd, 11

```
## Mann-Whitney-Wilcoxon

```{r nonparametric-test}
wilcox.test(fbsway ~ age, data=sway)
```
```

Since the p-value is small, you would reject the null hypothesis and conclude
that there is a statistically significant difference in front-to-back sway
values between elderly and young patients.

Break #5

- What you have learned

    - R code for alternative tests

- What's coming next

    - Paired data

Paired data

- Sometimes you collect data in pairs
  - Drug for left arm, placebo for right arm
  - Measure patient before intervention and after
  - Each exposed patient matched on age/sex with control

There are lots of situations where the two groups in a research study are paired together. In this case, the test changes

Calculate differences

- $D_i = X_{1i} - X_{2i}$
  - $\bar{D} = \frac{1}{n}\Sigma D_i$
  - $S_D = \sqrt{\frac{1}{n-1}\Sigma(D_i - \bar{D})^2}$

Test of hypothesis

- $H_0: \ \mu_D = 0$
- $H_1: \ \mu_D \neq 0$
  - T = $\dfrac{\bar{D}}{se}$
  - se = $\dfrac{S_D}{\sqrt{n}}$
- Accept H_0 if T is close to zero

Confidence interval

- $\bar{D} \pm t(1 - \alpha/2, n - 1)se$
    - Range of plausible values for μ_D

Break #6

- What you have learned

  - Paired data

- What's coming next

  - R code for paired data

simon-5501-09-paired.qmd, 1

```
---
title: "Alternative analysis of postural sway data"
format:
  html:
    embed-resources: true
---
```

This program runs a paired t-test. Consult the [data dictionary][dic] for information about the data itself.

[dic]: https://github.com/pmean/data/blob/main/files/postural-sway.yaml

This program was written by Steve Simon on 2024-10-15. It is placed in the public domain.

simon-5501-09-paired.qmd, 2

```
## Libraries

```{r setup}
#| message: false
#| warning: false
library(tidyverse)
```
```

simon-5501-09-paired.qmd, 3

```
## Read data

```{r read-sway}
sway <- read_tsv(
  file="../data/postural-sway.txt",
  col_types="cnn")
names(sway) <- tolower(names(sway))
glimpse(sway)
```
```

simon-5501-09-paired.qmd, 4

```
## Paired t-test

The research team is also interested in whether the front-to-back sway is
larger than the side-to-side sway, combining both the elderly and young
patients into one group.
```

simon-5501-09-paired.qmd, 5

```
## Compute and graph differences

```{r compute-difference-1}
sway |>
  mutate(diff_sway=fbsway-sidesway) -> paired_differences
paired_differences |>
    summarize(
      diff_mn=mean(diff_sway),
      diff_sd=sd(diff_sway))
```

The average difference is positive, indicating that front-to-back-sways are
larger on average. The standard deviation of the differences is large.
```

simon-5501-09-paired.qmd, 6

```
## Boxplot of differences

```{r compute-difference-2}
#| fig.height: 1.5
#| fig.width: 6
paired_differences |>
  ggplot(aes(diff_sway, "Combined")) +
    geom_boxplot() +
    ggtitle("Graph drawn by Steve Simon on 2024-10-13") +
    xlab("Front-to-back sway minus side-to-side sway") +
    ylab(" ")
```
```

More than 75% of the differences are positive, also indicating that front-to-back sways tend to be larger.

simon-5501-09-paired.qmd, 7

```
## Paired t-test

```{r paired-t-test}
t.test(
  sway$fbsway,
  sway$sidesway,
  paired=TRUE,
  alternative="two.sided")
```

The p-value is small, indicating that front-to-back sways are significantly
larger than side-to-side sways. The 95% confidence intervals shows that the
mean difference is at least 0.21 units larger for front-to-back sway and
possibly as large as 7.0 units.
```

Break #7

- What you have learned

  - R code for paired data

- What's coming next

  - Your homework

simon-5501-09-directions.md, 1

```
---
title: "Directions for 5501-08 programming assignment"
format:
  html:
    embed-resources: true
---

This programming assignment was written by Steve Simon on 2024-10-08 and is
placed in the public domain.
```

simon-5501-09-directions.md, 2

```
## Program

-    Download the [program][tem]
    -    Store it in your src folder
-    Modify the file name
    -    Use your last name instead of "simon"
-    Modify the documentation header
    -    Add your name to the author field
    -    Optional: change the copyright statement

[tem]: https://github.com/pmean/classes/blob/master/biostats-1/09/src/simon-
5501-09-sway.qmd
```

simon-5501-09-directions.md, 3

```
## Data

-    Download the [data][dat] file
     -    Store it in your data folder
-    Refer to the [data dictionary][dic], if needed.

[dat]: https://github.com/pmean/data/blob/main/files/postural-sway.txt
[dic]: https://github.com/pmean/data/blob/main/files/postural-sway.yaml
```

simon-5501-09-directions.md, 4

Question 1

The side-to-side sway data appears to have a pair of outliers and some evidence of heterogenity. Use a log transformation to see if this makes things better. Analyze the data on a log scale using a two-sample t-test and report the confidence interval after transforming back to the original scale of measurement.

simon-5501-09-directions.md, 5

```
## Question 2
```

You want to replicate these findings in a new population, using the log transformation on the new data. Select a standard deviation from log transformed side-to-side sway values in the current study (go halfway between the two standard deviations). You want to have 90% power for detecting a 1.5 fold change (which is 0.176 units on a log base 10 scale). If you use a two-sided hypothesis with alpha = 0.05, what sample size would you need?

simon-5501-09-directions.md, 6

```
## Your submission

-    Save the output in html format
-    Convert it to pdf format.
-    Make sure that the pdf file includes
     -    Your last name
     -    The number of this course
     -    The number of this module
-    Upload the file
```

simon-5501-09-directions.md, 7

```
## If it doesn't work

Please review the [suggestions if you encounter an error page][sim3].

[sim3]: https://github.com/pmean/classes/blob/master/general/suggestions-if-
you-encounter-an-error.md
```

Summary

- What you have learned
  - Restructuring your data
  - One-tailed tests
  - Checking assumptions
  - Alternative tests
  - R code for alternative tests
  - Paired data
  - R code for paired data
  - Your homework