

Alternative analysis of postural sway data

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

This program was written by Steve Simon and Leroy Wheeler on 2024-10-16. It is placed in the public domain.

Libraries

```
library(tidyverse)
```

Read data

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

Rows: 17

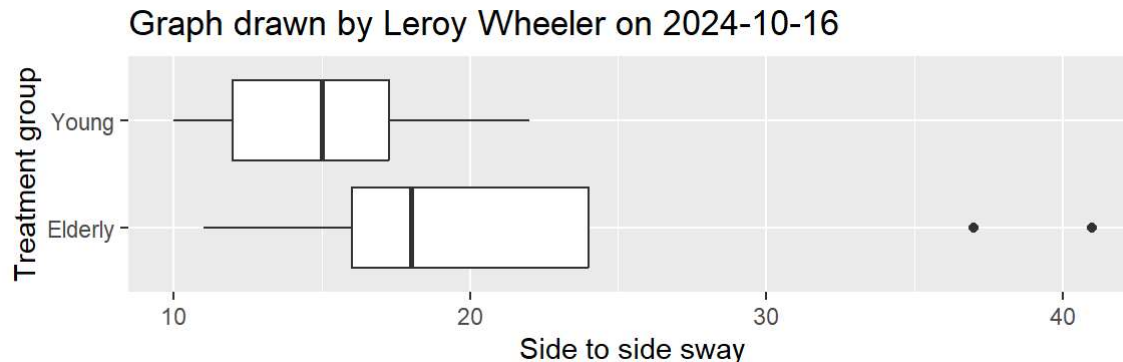
Columns: 3

```
$ age      <chr> "Elderly", "Elderly", "Elderly", "Elderly", "Elderly", "Elderly", "Elderly", "Elderly", "Elderly", "Elderly", "Elderly", "Elderly", "Elderly", "Elderly", "Elderly", "Elderly", "Elderly"  
$ fbssway  <dbl> 19, 30, 20, 19, 29, 25, 21, 24, 50, 25, 21, 17, 15, 14, 14, 21, 19  
$ sidesway <dbl> 14, 41, 18, 11, 16, 24, 18, 21, 37, 17, 10, 16, 22, 12, 14, 11, 10
```

Boxplot of side to side sway by age

```
sway |>  
  ggplot(aes(age, sidesway)) +  
    geom_boxplot() +  
    ggtitle("Graph drawn by Leroy Wheeler on 2024-10-16") +  
    xlab("Treatment group") +
```

```
ylab("Side to side sway") +
coord_flip()
```



There are two outliers in the side to side sway data as well as some heterogeneity in the data.

Descriptive statistics for side to side sway by age

```
sway |>
  group_by(age) |>
  summarize(
    side_mn=mean(sidesway),
    side_sd=sd(sidesway),
    n=n())
```

```
# A tibble: 2 × 4
  age      side_mn side_sd     n
<chr>    <dbl>   <dbl> <int>
1 Elderly    22.2    10.3     9
2 Young     15.1     3.91    8
```

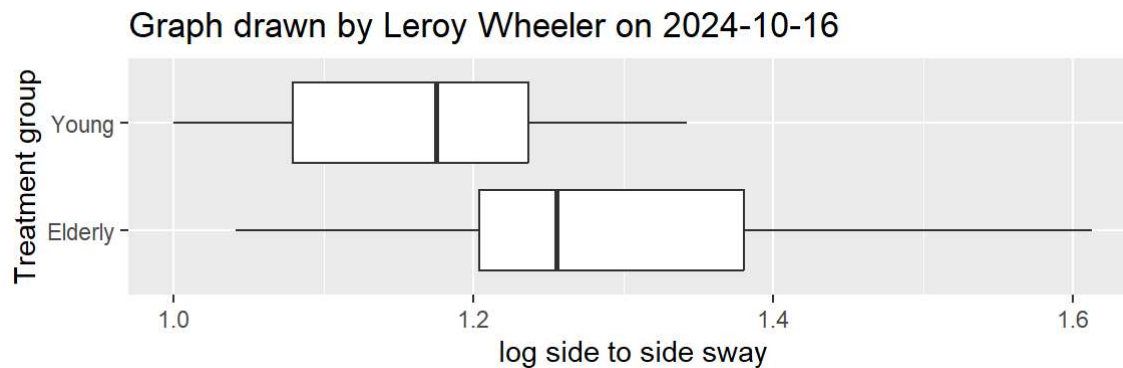
In addition to the outliers, we see that the larger mean in the elderly group brings along with it a larger standard deviation that is close to three times that of the young group. We will perform a log transformation to obtain more accurate results.

Log transformation, 1

```
sway |>
  mutate(log_sidesway=log10(sidesway)) -> log_sway
```

Log transformation, 2

```
log_sway |>
  ggplot(aes(age, log_sidesway)) +
    geom_boxplot() +
    ggtitle("Graph drawn by Leroy Wheeler on 2024-10-16") +
    xlab("Treatment group") +
    ylab("log side to side sway") +
    coord_flip()
```



There are no outliers after a log transformation and the distribution of the data between the two groups looks more similar.

Log transformation, 3

```
log_sway |>
  group_by(age) |>
```

```

summarize(
  log_mn=mean(log_sidesway),
  log_sd=sd(log_sidesway),
  n=n())

```

A tibble: 2 × 4

	age	log_mn	log_sd	n
	<chr>	<dbl>	<dbl>	<int>
1	Elderly	1.31	0.187	9
2	Young	1.17	0.112	8

The standard deviations on the log scale are more similar than they were on the original scale.

Question 1: The side-to-side sway data appears to have a pair of outliers and some evidence of heterogeneity. 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.

```

m2 <- t.test(
  log_sidesway ~ age,
  data=log_sway,
  alternative="two.sided",
  var.equal=TRUE)
m2

```

Two Sample t-test

data: log_sidesway by age

t = 1.8729, df = 15, p-value = 0.08071

alternative hypothesis: true difference in means between group Elderly and group Young is not equal to 0

95 percent confidence interval:

-0.01966132 0.30449529

sample estimates:

mean in group Elderly	mean in group Young
1.309511	1.167094

There is a no statistically significant difference between the log side to side sway between elderly patients and young patients because the 95% confidence interval includes the possibility that the difference in the log means includes zero. This conclusion is further supported by the p value which is greater than our alpha of 0.05.

Back-transform confidence interval to the original scale.

```
10^(m2$conf.int)
```

```
[1] 0.9557376 2.0160221  
attr(,"conf.level")  
[1] 0.95
```

We are 95% confident that the geometric mean side to side sway in elderly patients is somewhere between 0.96 times higher (4% decrease) and 2.02 times higher (102% increase) than the geometric mean for young patients. Because our 95% CI includes the possibility that the elderly patient values are either smaller or larger than the younger patient values, we do not reject the null hypothesis that the side sway values in both data sets are equivalent.

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?

```
power.t.test(  
  n=NULL,  
  delta=0.176,
```

```
sd=0.14945,  
sig.level=0.05,  
power=0.9,  
type="two.sample",  
alternative="two.sided")
```

Two-sample t test power calculation

```
      n = 16.17645  
delta = 0.176  
      sd = 0.14945  
sig.level = 0.05  
      power = 0.9  
alternative = two.sided
```

NOTE: n is number in *each* group

In order to detect a 1.5 fold (0.176 units on a log scale) change in side to side sway between the two groups with a 90% power and a two sided hypothesis with an $\alpha = 0.05$, we will need a sample size of at least 17.