

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.

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", "Elder..."  
$ fbsway   <dbl> 19, 30, 20, 19, 29, 25, 21, 24, 50, 25, 21, 17, 15, 14, 14, 2...  
$ sidesway <dbl> 14, 41, 18, 11, 16, 24, 18, 21, 37, 17, 10, 16, 22, 12, 14, 1...
```

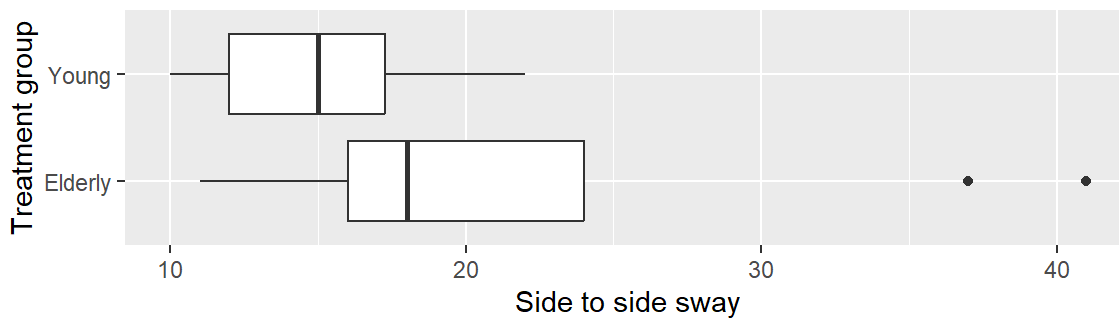
Question 1

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.

Boxplot of side-to-side sway by age

```
sway |>  
  ggplot(aes(age, sidesway)) +  
    geom_boxplot() +  
    ggtitle("Graph drawn by Michael Dang on 2024-10-22") +  
    xlab("Treatment group") +  
    ylab("Side to side sway") +  
    coord_flip()
```

Graph drawn by Michael Dang on 2024-10-22



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

Descriptive statistics for side-to-side sway by age

```
sway |>
  group_by(age) |>
  summarize(
    sts_mn=mean(sidesway),
    sts_sd=sd(sidesway),
    n=n())
```

A tibble: 2 × 4

	age	sts_mn	sts_sd	n
	<chr>	<dbl>	<dbl>	<int>
1	Elderly	22.2	10.3	9
2	Young	15.1	3.91	8

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.

Log transformation, 1

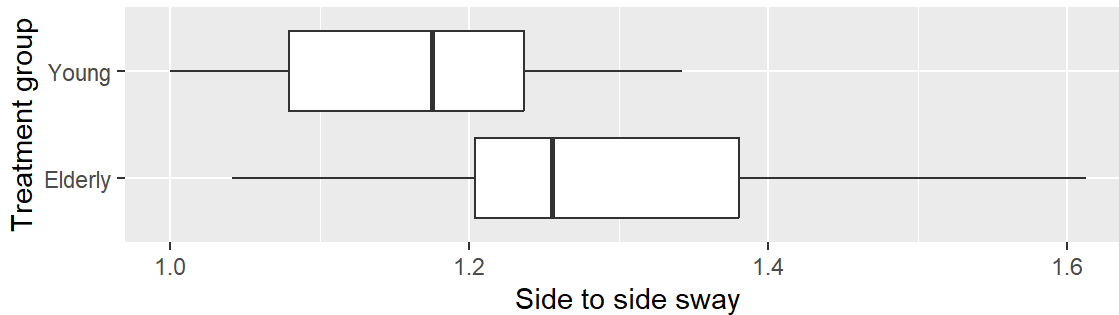
```
sway |>
  mutate(log_sidesway=log10(sidesway)) -> log_sway
```

Log transformation, 2

```
log_sway |>
  ggplot(aes(age, log_sidesway)) +
  geom_boxplot() +
  ggtitle("Graph drawn by Steve Simon on 2024-10-13") +
  xlab("Treatment group") +
```

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

Graph drawn by Steve Simon on 2024-10-13



There are no outliers after a log transformation.

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 quite a bit more similar than they were on the original scale.

Two-sample t-test using the log transformation

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

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 and 2.02 times that of the geometric mean for young patients. This suggests that there is no statistically significant difference between the groups, as the interval includes values below 1, which implies that the sway in elderly patients could potentially be lower than that in younger patients.

Mann-Whitney-Wilcoxon

```
wilcox.test(sidesway ~ age, data=sway)
```

```
Warning in wilcox.test.default(x = DATA[[1L]], y = DATA[[2L]], ...): cannot  
compute exact p-value with ties
```

```
Wilcoxon rank sum test with continuity correction
```

```
data: sidesway by age
```

```
W = 53, p-value = 0.1108
```

```
alternative hypothesis: true location shift is not equal to 0
```

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

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=0.149, # (0.1119424 + 0.1869584) / 2  
  sig.level=0.05,  
  power=0.9,  
  type="two.sample",  
  alternative="two.sided")
```

Two-sample t test power calculation

```
n = 16.08571  
delta = 0.176  
sd = 0.149  
sig.level = 0.05  
power = 0.9  
alternative = two.sided
```

NOTE: n is number in *each* group

The sample size is 16.