Lecture 18: Paired T-test

STAT 324

Ralph Trane
University of Wisconsin–Madison

Spring 2020

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Paired T-test

Musculoskeletal disorders of the neck and shoulders are common in office workers because of repetitive tasks. Long periods of upper-arm elevation above 30 degrees have been shown to be related to disorders. It was thought that varying working conditions over the course of the day could alleviate some of these problems. Eight office workers were randomly selected. They were observed for one work day under the standard conditions, and the percentage of time that their dominant upper-arm was below 30 degrees was recorded. The next day, these same individuals had their work diversified, and again were observed.

Question: does more diversified work increase the percent of time a worker's upper-arm is below 30 degrees?

```
library(tidyverse); library(distributior

work_data <- tibble(participant = rep(1:
                    work_condition = fac

                    time = c(78, 91, 79,
                             81, 87, 86,

DT::datatable(work_data,
              options = list(dom = "t",
              rownames = FALSE)
```

participant	work_condition	time
1	Diverse	78
1	Standard	81
2	Diverse	91
2	Standard	87
3	Diverse	79
3	Standard	86

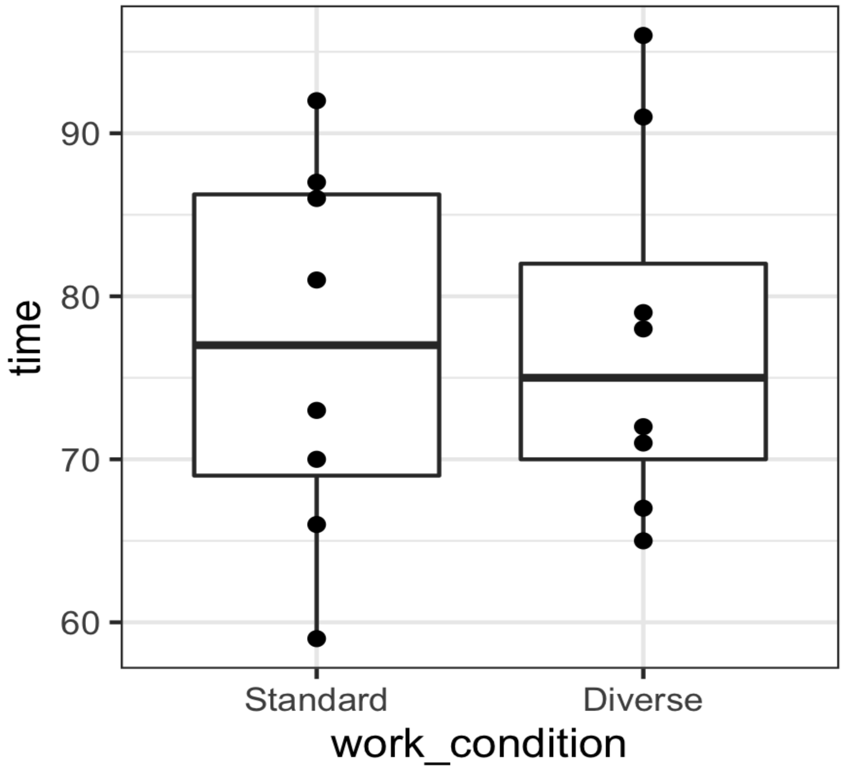
Paired T-test

In our familiar statistical jargon: we want to test the hypothesis $H_0 : \mu_{\text{diverse}} - \mu_{\text{standard}} = 0$ against the alternative $H_A : \mu_{\text{diverse}} - \mu_{\text{standard}} < 0$. (Null: no difference. Alternative: diversified *decreases* time upper-arm below 30 degrees.)

First step in any analysis: pretty plots!

Paired T-test

```
ggplot(work_data,
       aes(x = work_condition, y = time)
  geom_boxplot() +
  geom_point()
```



From this plot: seems to be little to no difference. Comparing medians suggests that more diverse conditions lowers the time measured. But overall, not convincing.

Paired T-test

Problem: this is not the entire story. We have another variable, `participant`. Here are the observations connected by participant.

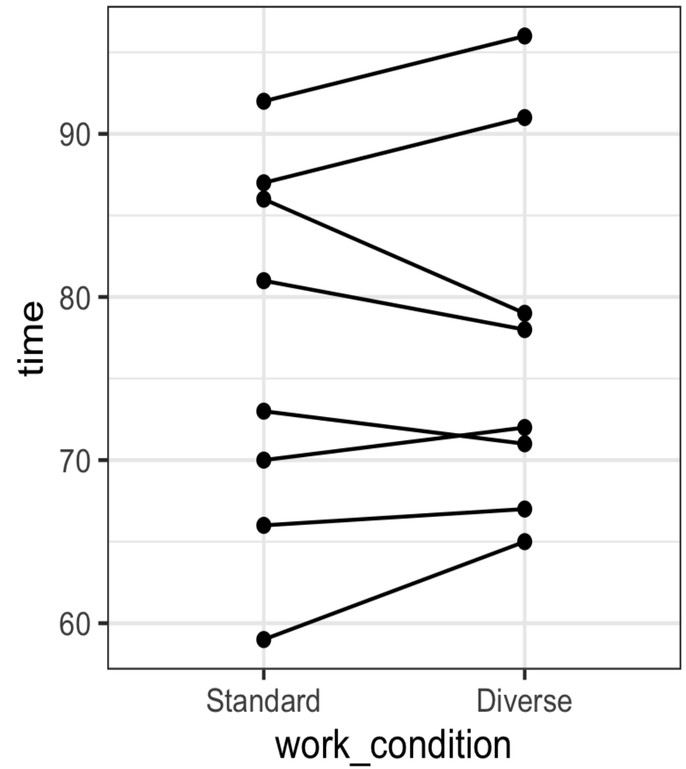
```
ggplot(data = work_data,
       aes(x = work_condition, y = time)) +
  geom_point() +
  geom_line(aes(group = participant))
```

Paired T-test

How do we test the hypothesis? One "obvious" (= the one we've spend the most time on) choice would be a two sample t-test.

Assumptions to check:

- independent groups
- independent observations
- normal averages



Groups not independent!

Notice how, in general, lines that start low end low, lines that start high end high.

Paired T-test

So how do we deal with this scenario? Many ways to do so, but by far the simplest is the paired t-test.

The "trick" is to realize that we are really interested in the *difference* from one day to the next. So, why not look at differences?

participant	Diverse	Standard	Difference
1	78	81	-3
2	91	87	4
3	79	86	-7
4	65	59	6
5	67	66	1
6	72	70	2
7	71	73	-2

Now we have one observations per individual. Let's call the corresponding random variable D_i, and let's say that the true expected difference (i.e. true expected value of D_i) is μ_D.

Paired T-test

So, $D_i = X_{\text{Diverse},i} - X_{\text{Standard},i}$.

Notice that

$\mu_D = E(D_i) = E(X_{\text{Diverse},i} - X_{\text{Standard},i}) = E(X_{\text{Diverse},i}) - E(X_{\text{Standard},i}) = \mu_{\text{Diverse}} - \mu_{\text{Standard}}$.

So, testing $H_0 : \mu_{\text{Diverse}} - \mu_{\text{Standard}} = 0$ against $H_A : \mu_{\text{Diverse}} - \mu_{\text{Standard}} < 0$ is the same as testing $H_0 : \mu_D = 0$ against $H_A : \mu_D < 0$.

So, we are actually back in a one-sample setting, and can therefore use a one-sample t-test!

Paired T-test

First, we'll setup the data. That is, we have to calculate the differences for each individual.

```
work_data_differences <- work_data %>%
  group_by(participant) %>%
  summarize(d = diff(time))

work_data_differences
```

```
## # A tibble: 8 x 2
##   participant     d
## *       <int> <dbl>
## 1           1     3
## 2           2    -4
## 3           3     7
## 4           4    -6
## 5           5    -1
## 6           6    -2
## 7           7     2
## 8           8    -4
```

Next, we'll see if we can do a one-sample t-test on these new observations!

Paired T-test

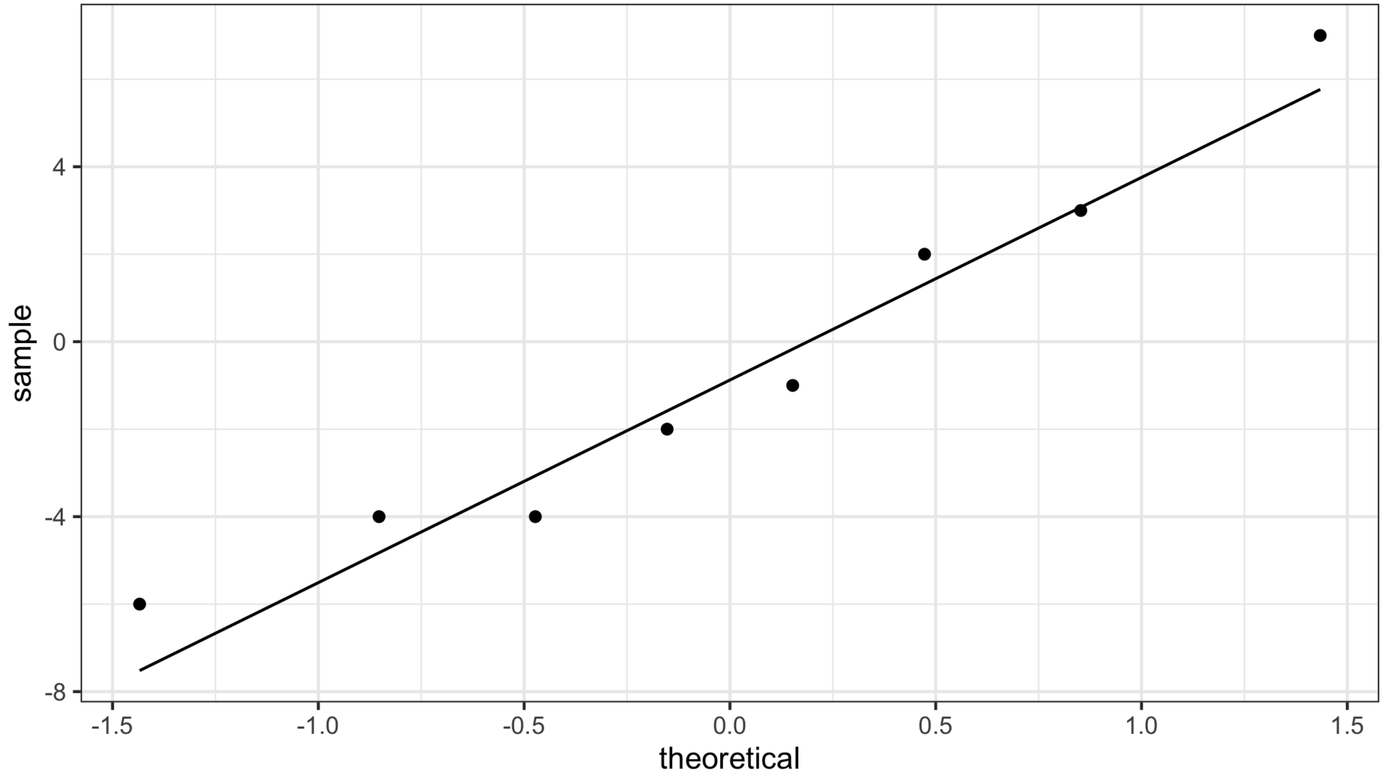
Check assumptions:

- independence?
- normal average?
  - small sample size, so no CLT. Have to check if data is reasonably normal

Paired T-test

QQ-plot to check if the *differences* are reasonably normal:

```
ggplot(work_data_differences,
       aes(sample = d)) +
  geom_qq() +
  geom_qq_line()
```



Paired T-test

By hand:

```
work_data_differences %>%
  summarize(average = mean(d),
            s = sd(d),
            n = n(),
            T_obs = average/(s/sqrt(n)),
            p_value = cdf(StudentsT(df =
```

```
## # A tibble: 1 x 5
##   average     s     n  T_obs p_value
##     <dbl> <dbl> <int>  <dbl>   <dbl>
## 1  -0.625  4.34     8 -0.407   0.348
```

Using built-in function:

```
t.test(work_data_differences$d,
       alternative = "less")
```

```
##
##      One Sample t-test
##
## data:  work_data_differences$d
## t = -0.40728, df = 7, p-value = 0.348
## alternative hypothesis: true mean is l
## 95 percent confidence interval:
##       -Inf 2.282367
## sample estimates:
## mean of x
##    -0.625
```

Paired T-test

Note: It matters if you do $X_{\text{Diverse},i} - X_{\text{Standard},i}$ or $X_{\text{Standard},i} - X_{\text{Diverse},i}$ …

… BUT the difference is only a sign. If you want to specify direction in `R`:

```
tmp <- work_data %>%
  mutate(work_condition =
           factor(work_condition,
                  levels = c("Standard",
  group_by(participant) %>%
  arrange(work_condition) %>%
  summarize(d = diff(time))

t.test(tmp$d, alternative = "less")
```

```
##
##     One Sample t-test
##
## data:  tmp$d
## t = 0.40728, df = 7, p-value = 0.652
## alternative hypothesis: true mean is l
## 95 percent confidence interval:
##      -Inf 3.532367
## sample estimates:
## mean of x
##     0.625
```

Notice that you now are testing against the reverse hypothesis: $H_A : \mu_{\text{Standard}} - \mu_{\text{Diverse}} < 0$.

Paired T-test

If you want same result as before, you need to flip to greater, instead of less:

```r
t.test(tmp$d, alternative = "greater")
```

```
## 
##      One Sample t-test
## 
## data:  tmp$d
## t = 0.40728, df = 7, p-value = 0.348
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  -2.282367       Inf
## sample estimates:
## mean of x 
##     0.625
```

Paired T-test

Sometimes, you might get data in a slightly different format:

```
work_data1 %>% DT::datatable(options = list(dom = "t", paging = FALSE, scrollY = "20v
```

	participant	Diverse	Standard
1	1	78	81
2	2	91	87
3	3	79	86
4	4	65	59
5	5	67	66
6	6	72	70
7	7	71	73
8	8	96	92

Paired T-test

We can then use the built-in function right away:

```
t.test(x = work_data1$Standard, y = work_data1$Diverse,
       paired = TRUE, alternative = "less")
```

```
##
##      Paired t-test
##
## data:  work_data1$Standard and work_data1$Diverse
## t = -0.40728, df = 7, p-value = 0.348
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf 2.282367
## sample estimates:
## mean of the differences
##                  -0.625
```

Paired T-test

Summary: Paired t-test

When two samples are dependent, but there exists a *natural* pairing of observations.

Paired t-test is simply a one-sample t-test using the *differences* as the observations, i.e. for each "subject", $D_i = X_{1,\mathrm{i}} - X_{2,\mathrm{i}}$.

Assuming that $D_1, D_2, \ldots, D_n$ are independent observations, and that $\bar{D} \sim N$, we can test $H_0 : \mu_D = \mu_0$ against any of the three alternatives using $T = \frac{\bar{D}-\mu_0}{\widehat{\mathrm{SD}}(\bar{D})} = \frac{\bar{D}-\mu_0}{S_D/\sqrt{n}}$.

IF the null hypothesis is true, $T \sim t_{n-1}$, and p-values are obtained as

- $P(T_{n-1} > T_{\mathrm{obs}})$ if $H_A : \mu_D > \mu_0$,
- $P(T_{n-1} < T_{\mathrm{obs}})$ if $H_A : \mu_D < \mu_0$,
- $2 \cdot P(T_{n-1} > |T_{\mathrm{obs}}|)$ if $H_A : \mu_D \neq \mu_0$.