

STA305/1004 - Class 5

September 19, 2019

Today's Class

- ▶ Paired Comparisons
- ▶ Introduction to Phase III Clinical Trials
- ▶ Introduction to power

Paired Comparisons

- ▶ Increase precision by making comparisons within matched pairs of experimental material.
- ▶ Randomize within a pair.

Boy's Shoe Experiment

- ▶ Two materials to make boy's shoes, A and B, are tested to evaluate if B is more sturdy compared to A.
- ▶ During the experimental test some boys scuffed their shoes more than others.
- ▶ Each boy's two shoes were subjected to the same treatment by having each boy wear both materials.
- ▶ Working with 10 differences $B-A$ most of the boy-to-boy variation could be eliminated.
- ▶ Called a randomized paired comparison design.

Boy's Shoe Experiment

- ▶ Toss a coin to randomize material to L/R foot of a boy.
- ▶ Head: Material A used on right foot.
- ▶ Null hypothesis: amount of wear associated with material A and B are the same.
- ▶ So labelling given to a pair of results only affects the sign of the difference.

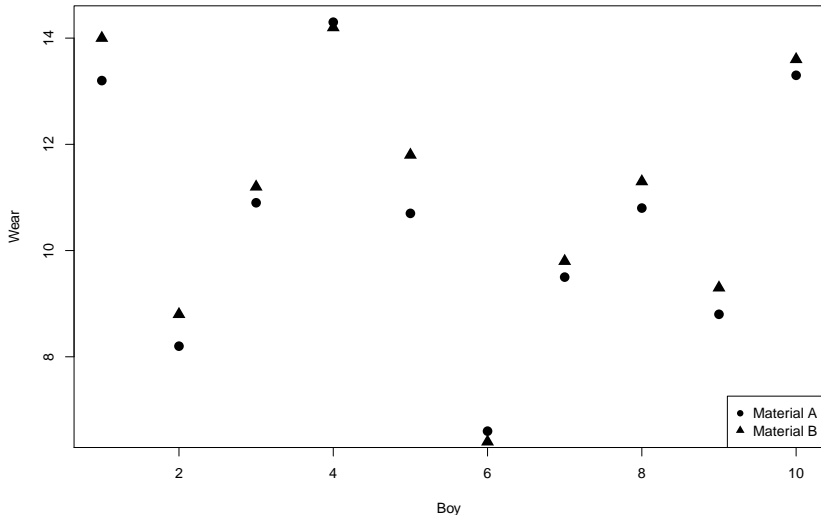
Randomized paired comparison

```
library(BHH2)
data(shoes.data)
shoes.data
```

##	boy	matA	sideA	matB	sideB
## 1	1	13.2	L	14.0	R
## 2	2	8.2	L	8.8	R
## 3	3	10.9	R	11.2	L
## 4	4	14.3	L	14.2	R
## 5	5	10.7	R	11.8	L
## 6	6	6.6	L	6.4	R
## 7	7	9.5	L	9.8	R
## 8	8	10.8	L	11.3	R
## 9	9	8.8	R	9.3	L
## 10	10	13.3	L	13.6	R

Randomized paired comparison

```
plot(shoes.data$boy,shoes.data$matA,pch=16,cex=1.5,  
     xlab="Boy",ylab="Wear")  
points(shoes.data$boy,shoes.data$matB,pch=17,cex=1.5)  
legend("bottomright",legend=c("Material A","Material B"),pch=c(16,17))
```



Randomized paired comparison

```
diff <- shoes.data$matA-shoes.data$matB  
meandiff <- mean(diff); meandiff
```

```
## [1] -0.41
```

```
shoe.dat2 <- data.frame(shoes.data,diff)  
shoe.dat2[1:3,]
```

```
##   boy matA sideA matB sideB diff  
## 1   1 13.2     L 14.0     R -0.8  
## 2   2   8.2     L   8.8     R -0.6  
## 3   3 10.9     R 11.2     L -0.3
```

```
mean(shoe.dat2$matA) - mean(shoe.dat2$matB)
```

```
## [1] -0.41
```

```
sd(diff)
```

```
## [1] 0.3871549
```

```
sqrt(sd(shoe.dat2$matA)^2/length(shoe.dat2$matA) +  
      sd(shoe.dat2$matB)^2/length(shoe.dat2$matB))
```

```
## [1] 1.111381
```


Boy's Shoe Experiment

- ▶ The sequence of coin tosses is one of $2^{10} = 1024$ equiprobable outcomes.
- ▶ To test H_0 the average difference of -0.41 observed observed can be compared with the other 1023 averages by calculating the average difference for each of 1024 arrangements of signs in:

$$\bar{d} = \frac{\pm 0.8 \pm 0.6 \cdots \pm 0.3}{10}$$

Randomized paired comparison

```
N <- 2^(10) # number of treatment assignments
res <- numeric(N) #vector to store results
LR <- list(c(-1,1)) # difference is multiplied by -1 or 1
# generate all possible treatment assign
trtassign <- expand.grid(rep(LR, 10))
```

```
for (i in 1:N) {
res[i] <- mean(as.numeric(trtassign[i,])*diff)
}
trtassign[1:2,]
```

```
##   Var1 Var2 Var3 Var4 Var5 Var6 Var7 Var8 Var9 Var10
## 1   -1  -1  -1  -1  -1  -1  -1  -1  -1  -1
## 2    1  -1  -1  -1  -1  -1  -1  -1  -1  -1
```

```
res[1:2]
```

```
## [1] 0.41 0.25
```

R Notation

res is a vector of length 2^{10} .

```
length(res)
```

```
## [1] 1024
```

- ▶ The i^{th} element of res can be accessed using the notation `res[i]`, and `res[-i]` will return all the elements of res except the i^{th} element.
- ▶ `trtassign` is a 1024×10 matrix.

```
dim(trtassign)
```

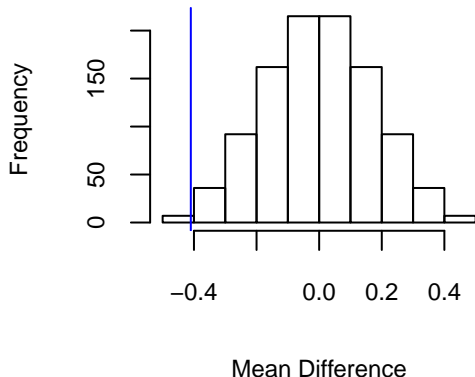
```
## [1] 1024 10
```

- ▶ `trtassign[i,j]` will return the value of `trtassign` that is in the i^{th} and the j^{th} column.

Randomized paired comparison

```
hist(res, xlab = "Mean Difference",  
     main = "Randomization Distribution \n Boys' Shoes",  
     cex.main = 0.75, cex.lab = 0.75, cex.axis = 0.75)  
abline(v = meandiff, col = "blue")
```

**Randomization Distribution
Boys' Shoes**



Randomized paired comparison

```
# number of differences less than observed diff  
sum(res <= meandiff)
```

```
## [1] 7
```

```
# p-value  
sum(res <= meandiff)/N
```

```
## [1] 0.006835938
```

R Notation

- ▶ `res <= meandiff` compares each element of `res` to `meandiff` and returns TRUE or FALSE which is *coerced* to a numeric via TRUE → 1 and FALSE → 0.
- ▶ `sum` counts the number of times elements of the `res` vector are less than or equal to `meandiff`.

```
res[1:4] <= meandiff
```

```
## [1] FALSE FALSE FALSE FALSE
```

```
# explicit coercion to type numeric
```

```
as.numeric(res[1:4] <= meandiff)
```

```
## [1] 0 0 0 0
```

Paired t-test

If we assume that the differences -0.8, -0.6, -0.3, 0.1, -1.1, 0.2, -0.3, -0.5, -0.5, -0.3 are a random sample from a normal distribution then the statistic

$$t = \frac{\bar{d}}{s_{\bar{d}}/\sqrt{10}} \sim t_{10-1},$$

where, $s_{\bar{d}}$ is the sample standard deviation of the paired differences. The p-value for testing if $\bar{D} < 0$ is

$$P(t_9 < t).$$

Paired t-test

In general if there are n differences then

$$t = \frac{\bar{d}}{s_{\bar{d}}/\sqrt{n}} \sim t_{n-1},$$

where, $s_{\bar{d}}$ is the sample standard deviation of the paired differences. The p-value for testing if $\bar{D} < 0$ is

$$P(t_{n-1} < t).$$

NB: This is the same as a one-sample t-test of the differences.

Paired t-test

In R a paired t-test can be obtained by using the command `t.test()` with `paired=T`.

```
t.test(shoes.data$matA,shoes.data$matB,paired = TRUE,  
       alternative = "less")
```

```
##  
## Paired t-test  
##  
## data: shoes.data$matA and shoes.data$matB  
## t = -3.3489, df = 9, p-value = 0.004269  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
##      -Inf -0.1855736  
## sample estimates:  
## mean of the differences  
##                -0.41
```

Paired t-test

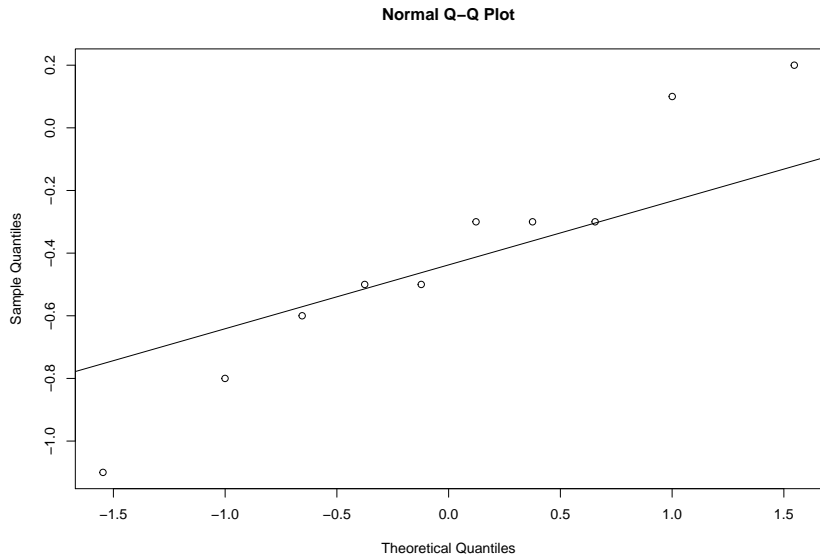
This is the same as a one-sample t-test on the difference.

```
# same as a one-sample t-test on the diff  
t.test(diff, alternative = "less")
```

```
##  
## One Sample t-test  
##  
## data: diff  
## t = -3.3489, df = 9, p-value = 0.004269  
## alternative hypothesis: true mean is less than 0  
## 95 percent confidence interval:  
##      -Inf -0.1855736  
## sample estimates:  
## mean of x  
##      -0.41
```

Paired t-test

```
qqnorm(diff); qqline(diff)
```



What are clinical trials?

Clinical trials are prospective intervention studies with human subjects to investigate experimental drugs, new treatments, medical devices, or clinical procedures (Yin, 2012).

Phases of clinical trials

Developing a new drug for cancer.

- ▶ **Preclinical studies:** In vitro (e.g. slides, test tubes) and in vivo (living organism such as rodents) studies on wide range of doses of experimental agents. This stage of study provides preliminary toxicity and efficacy data including pharmacokinetics (PK) and pharmacodynamics (PD) information.
- ▶ **Phase I:** Usually first study in humans to investigate the toxicity and side effects of the new agent. Identify MTD.
- ▶ **Phase II:** Assess if drug has sufficient efficacy. The drug is usually administered around the MTD. If drug does not show efficacy or is too toxic then further testing is discontinued.

Phases of clinical trials

- ▶ **Phase III:** If drug passes phase II testing then it is compared to the current standard of care or placebo. These are long-term, large scale randomized studies that may involve hundreds or thousands of patients.
- ▶ If the drug is proven to be effective (e.g. two positive phase III trials required for FDA approval) the company will file an application with regulatory agencies to sell the drug. If approved then the drug will be available to the general population in the country where it was approved.
- ▶ **Phase IV:** After approval a study might follow a large number of patients over a longer period of time to monitor side effects and drug interactions. For example, findings from these studies might add a warning label to the drug.

Phases of clinical trials

- ▶ The four phases are usually conducted sequentially and separately.
- ▶ Each trial requires an independent study design and a study protocol.
- ▶ Every aspect of trial design, monitoring, and data analysis call upon statistical methods.
- ▶ In randomized clinical trials a treatment group is often referred to as an **arm**.

Phases of clinical trials

- ▶ Experimental design plays a very important role in the design of clinical trials.
- ▶ Two arm clinical trials use all of theory of randomization that we learned about last week. Randomization is used to design phase III clinical trials since causation can usually be assessed using a randomized design.

How can causation be assessed using a randomized design?

- ▶ Suppose that patients are randomized in a two arm clinical trial where one of the arms is the standard treatment and the other arm is an experimental treatment
- ▶ A statistically significant difference in the outcome between the two arms is observed showing the experimental treatment is more efficacious.
- ▶ The interpretation is that the experimental treatment *caused* patients to have a better outcome since the only difference between the two arms is the treatment. Randomization is supposed to ensure that the groups will be similar with respect to all the factors measured in the study and all the factors that are not measured.

How many patients should be enrolled in a Phase III clinical trial?

- ▶ In a phase III trial sample size is the most critical component of the study design. The sample size has implications for how many subjects will be exposed to a drug that has no proven efficacy.
- ▶ The investigator needs to specify type I, II error rates, and the effect sizes.
- ▶ Standard practice is to compute the smallest sample size required to detect a clinically important/significant treatment difference with sufficient.

How many patients should be enrolled in a Phase III clinical trial?

- ▶ If the sample size is too small then the trial might fail to discover a truly effective drug because the statistical test cannot reach the significance level (5%) due to a lack of power.
- ▶ If the sample size is overestimated then resources wasted and drug development delayed since patient enrollment is often the main factor in time to complete a trial.

Statistical hypotheses

Suppose that subjects are randomized to treatments A or B with equal probability. Let μ_A be the mean response in the group receiving drug A and μ_B be the mean response in the group receiving drug B. The null hypothesis is that there is no difference between A and B, the alternative claims there is a clinically meaningful difference between them.

$$H_0 : \mu_A = \mu_B \text{ versus } H_0 : \mu_A \neq \mu_B$$

Statistical hypotheses

The type I error rate is defined as:

$$\begin{aligned}\alpha &= P(\text{type I error}) \\ &= P(\text{Reject } H_0 | H_0 \text{ is true}).\end{aligned}$$

Statistical hypotheses

The type II error rate is defined as:

$$\begin{aligned}\beta &= P(\text{type II error}) \\ &= P(\text{Accept } H_0 | H_1 \text{ is true}).\end{aligned}$$

Statistical hypotheses

Power is define as:

$$\begin{aligned}\text{power} &= 1 - \beta \\ &= 1 - P(\text{Accept } H_0 | H_1 \text{ is true}) \\ &= P(\text{Reject } H_0 | H_1 \text{ is true}).\end{aligned}$$

Power

The probability that a fixed level α test will reject H_0 when a particular alternative value of the parameter is true is called power of the test to detect that alternative.

Power

Can a 6-month exercise program increase the total body bone mineral content (TBBMC) of young women? Based on results of a previous study $\sigma = 2$ for the percent change in TBBMC over the 6-month period. A change in TBBMC of 1% would be considered important. Is 25 subjects a large enough sample size for this project?

Power of the one sample z-test

Let X_1, \dots, X_n be a random sample from a $N(\mu, \sigma^2)$ distribution. A test of the hypothesis

$$H_0 : \mu = \mu_0 \text{ versus } H_0 : \mu \neq \mu_0$$

will reject at level α if and only if

$$\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq z_{\alpha/2},$$

or

$$|\bar{X} - \mu_0| \geq \frac{\sigma}{\sqrt{n}} z_{\alpha/2},$$

where $z_{\alpha/2}$ is the $100(1 - \alpha/2)^{th}$ percentile of the $N(0, 1)$.

Power of the one sample z-test

The power of the test at $\mu = \mu_1$ is

$$\begin{aligned}1 - \beta &= 1 - P(\text{type II error}) \\&= P(\text{Reject } H_0 | H_1 \text{ is true}) \\&= P(\text{Reject } H_0 | \mu = \mu_1) \\&= P\left(|\bar{X} - \mu_0| \geq \frac{\sigma}{\sqrt{n}} z_{\alpha/2} | \mu = \mu_1\right)\end{aligned}$$

Subtract the mean μ_1 and divide by σ/\sqrt{n} to obtain (why?):

$$1 - \beta = 1 - \Phi\left(z_{\alpha/2} - \left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)\right) + \Phi\left(-z_{\alpha/2} - \left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)\right),$$

where $\Phi(\cdot)$ is the $N(0, 1)$ CDF.

Power of the one sample z-test

The power function of the one-sample z-test is:

$$1 - \beta = 1 - \Phi \left(z_{\alpha/2} - \left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} \right) \right) + \Phi \left(-z_{\alpha/2} - \left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} \right) \right).$$

What is the limit of the power function as:

- ▶ $n \rightarrow \infty$
- ▶ $\mu_1 \rightarrow \mu_0$