

STA305/1004 - Class 5

January 20-21, 2020

Class Outline

- ▶ Randomization test with Monte Carlo Sampling
- ▶ Two-sample t-tests
- ▶ Paired Comparisons

Randomization Test via Monte Carlo Sampling

Randomization Test

- ▶ We could calculate the difference in means for every possible way to split the data into two samples of size 6.
- ▶ This would result in $\binom{12}{6} = 924$ differences.
- ▶ If there were 30 observations split evenly into two groups then there are $\binom{30}{15} = 155,117,520$ differences.
- ▶ So unless the sample sizes are small these exhaustive calculations are not practical.

Randomization Test

Instead we can create a permutation resample (Monte Carlo Sampling).

1. Draw 6 observations from the pooled data without replacement. (fert A)
2. The remaining 6 observations will be the second sample (fert B)
3. Calculate the difference in means of the two samples
4. Repeat 1-3 at least 250000 times.
5. P-value is the fraction of times the random statistics exceeds the original statistic.

Estimate P-value via Monte Carlo Sampling

If M test statistics, t_i , $i = 1, \dots, M$ are randomly sampled from the permutation distribution, a one-sided Monte Carlo p value for a test of $H_0 : \mu_T = 0$ versus $H_1 : \mu_T > 0$ is

$$\hat{p} = \frac{1 + \sum_{i=1}^M I(t_i \geq t^*)}{M + 1}.$$

Including the observed value t^* there are $M + 1$ test statistics.

Estimate P-value via Monte Carlo Sampling

```
N <- 250000 # number of times to repeat this process
result <- numeric(N) # space to save random diffs.
for (i in 1:N)
{ #sample of size 6, from 1 to 12, without replacement
  index <- sample(12,size=6,replace=F)
  result[i] <- mean(fert[index])-mean(fert[-index])
}

#store observed mean difference
observed <- mean(yA)-mean(yB)

#P-value - mean - results will vary
pval <- (sum(result <= observed)+1)/(N+1)
round(pval,4)
```

```
## [1] 0.0279
```

The Randomization P-value

- ▶ An achievable P-value of the randomization test must be a multiple of $\frac{k}{\binom{12}{6}} = \frac{k}{924}$, where $k = 1, 2, \dots, 924$.
- ▶ If we choose a significance level of $\alpha = \frac{k}{924}$ that is one of the achievable P-values then $P(\text{type I error}) = \alpha$.
- ▶ The randomization test is an exact test.
- ▶ If α is not chosen to be one of the achievable P-values but $\alpha = \frac{k}{924}$ is the largest achievable P-value less than α then $P(\text{type I error}) < \alpha$.

Choosing a Test Statistic

A test statistic should be able to differentiate between H_0 and H_a in ways that are scientifically relevant.

Other Test Statistics

Other Test Statistics

- ▶ Other test statistics could be used instead of $T = \bar{Y}_A - \bar{Y}_B$ to measure the effectiveness of fertilizer A.
- ▶ The difference in group medians

$$\text{median}(Y_A) - \text{median}(Y_B)$$

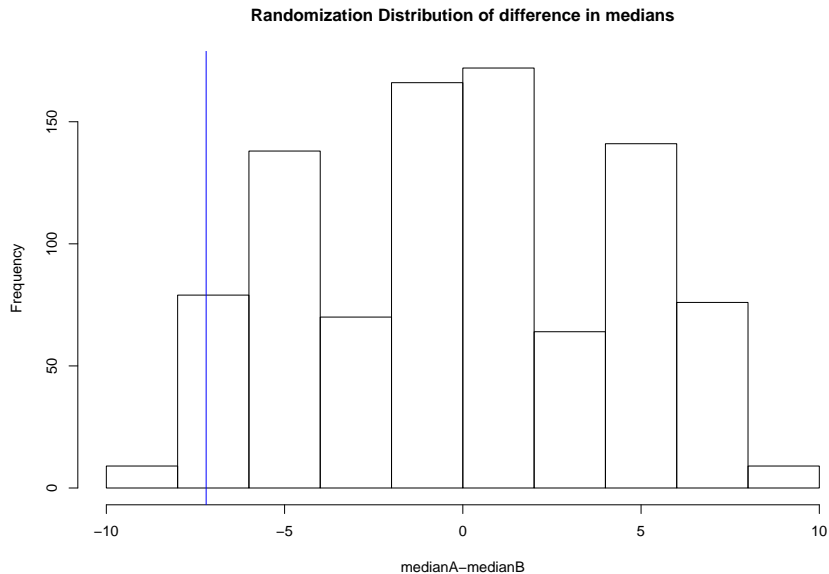
or trimmed means are examples of other test statistics.

Other Test Statistics

The randomization distribution of the difference in group medians can be obtained by modifying the R code used for the difference in group means.

```
fert <- c(yA,yB) #pool data
N <- choose(12,6)
res <- numeric(N) # store the results
index <-combn(1:12,6) # Generate N treatment assignments
for (i in 1:N)
{
  res[i] <- median(fert[index[,i]])-median(fert[-index[,i]])
}
```

Other Test Statistics



Other Test Statistics

The p-value of the randomization test can be calculated

```
# of times values from the median randomization  
# distribution less than observed value  
sum(res<=observed)
```

```
## [1] 36
```

```
N # Number of randomizations
```

```
## [1] 924
```

```
pval <- sum(res<=observed)/N # Randomization p value  
round(pval,2)
```

```
## [1] 0.04
```

Two-Sample T test

The two-sample t-test

If the two wheat yield samples are independent random samples from a normal distribution with means μ_A and μ_B but the same variance then the statistic

$$\bar{y}_A - \bar{y}_b \sim N(\mu_A - \mu_B, \sigma^2(1/n_A + 1/n_B)).$$

So,

$$\frac{\bar{y}_A - \bar{y}_b - \delta}{\sigma \sqrt{(1/n_A + 1/n_B)}} \sim N(0, 1),$$

where $\delta = \mu_A - \mu_B$.

If we substitute

$$S^2 = \frac{\sum_{i=1}^{n_A} (y_{iA} - \bar{y}_A) + \sum_{i=1}^{n_B} (y_{iB} - \bar{y}_B)}{n_A + n_B - 2}$$

for σ^2 then

$$\frac{\bar{y}_A - \bar{y}_b - \delta}{s \sqrt{(1/n_A + 1/n_B)}} \sim t_{n_A+n_B-2},$$

is called the two sample t-statistic.

The two-sample t-test

In the wheat yield example $H_0 : \mu_A = \mu_B$ and suppose that $H_1 : \mu_A < \mu_B$. The p-value of the test is obtained by calculating the observed value of the two sample t-statistic under H_0 .

$$t^* = \frac{\bar{y}_A - \bar{y}_B}{s \sqrt{(1/n_A + 1/n_B)}} = \frac{18.37 - 24.3}{4.72 \sqrt{(1/6 + 1/6)}} = -2.18$$

The p-value is $P(t_{18} < -2.18) = 0.03$.

The calculation was done in R.

```
s <- sqrt((5*var(yA)+5*var(yB))/10)
tstar <- (mean(yA)-mean(yB))/(s*sqrt(1/6+1/6)); round(tstar,2)
```

```
## [1] -2.18
```

```
pval <- pt(tstar,10); round(pval,5)
```

```
## [1] 0.02715
```

The two-sample t-test

In R the command to run a two-sample t-test is `t.test()`.

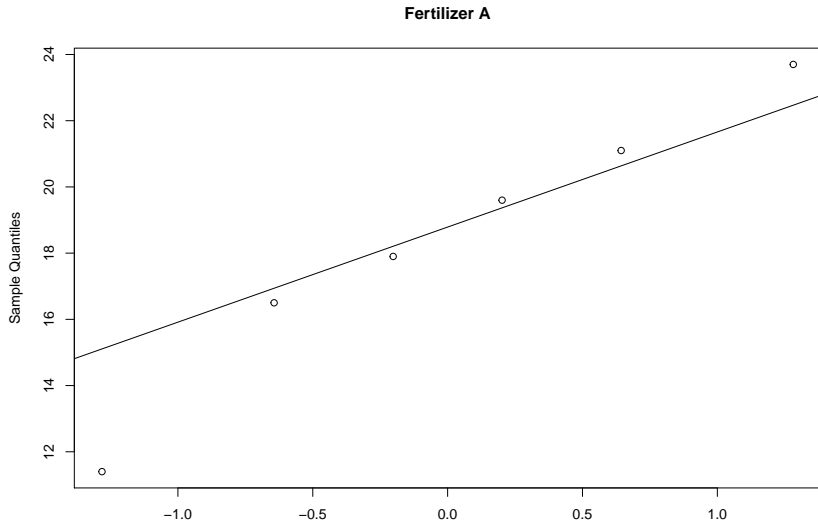
```
t.test(yA,yB,var.equal = TRUE,alternative = "less")
```

```
##  
## Two Sample t-test  
##  
## data: yA and yB  
## t = -2.1793, df = 10, p-value = 0.02715  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
##      -Inf -0.9987621  
## sample estimates:  
## mean of x mean of y  
## 18.36667 24.30000
```

The two-sample t-test

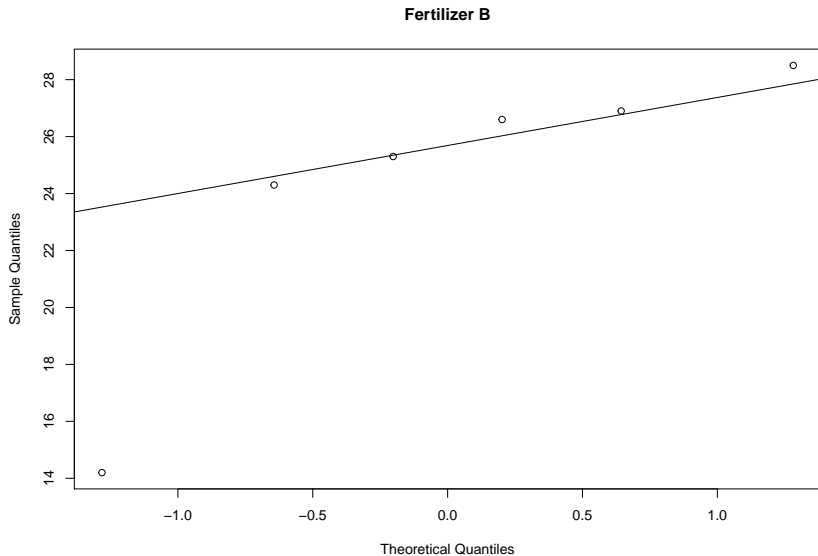
The assumption of normality can be checked using normal quantile plots, although the t-test is robust against non-normality.

```
qqnorm(yA,main = "Fertilizer A");qqline(yA)
```



The two-sample t-test

```
qqnorm(yB,main = "Fertilizer B");qqline(yB)
```



Two-Sample t-test versus Randomization Test

- ▶ The p-value from the randomization test and the p-value from two-sample t-test are almost identical.
- ▶ The randomization test does not depend on normality or independence.

Two-Sample t-test versus Randomization Test

- ▶ The randomization test does depend on Fisher's concept that after randomization, if the null hypothesis is true, the two results obtained from each particular plot will be exchangeable.
- ▶ The randomization test tells you what you could say if exchangeability were true.

Paired Comparisons

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.
- ▶ Data: Material used by foot for each boy
- ▶ 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.

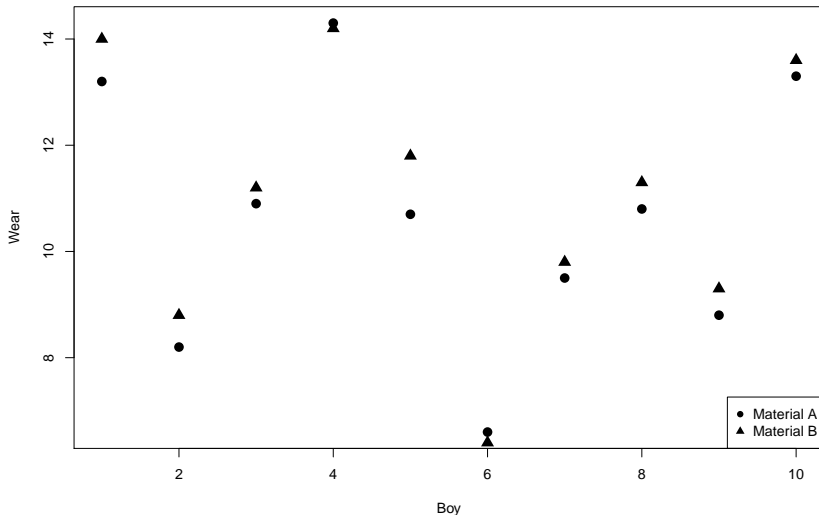
Randomized paired comparison

```
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
```

```
##      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  
## 4      4 14.3      L 14.2      R  0.1  
## 5      5 10.7      R 11.8      L -1.1  
## 6      6  6.6      L  6.4      R  0.2  
## 7      7  9.5      L  9.8      R -0.3  
## 8      8 10.8      L 11.3      R -0.5  
## 9      9  8.8      R  9.3      L -0.5  
## 10     10 13.3      L 13.6      R -0.3
```

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,]
```

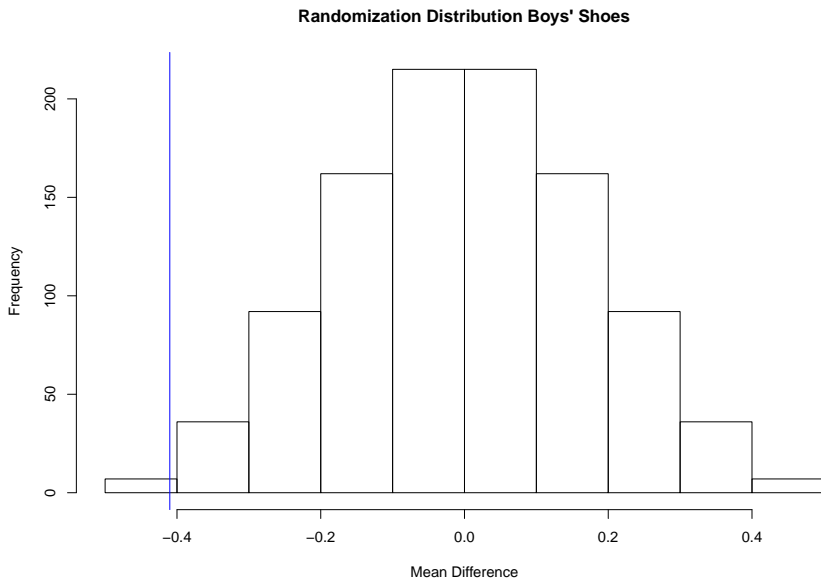
```
##      Var1 Var2 Var3 Var4 Var5 Var6 Var7 Var8 Var9 Var10
## 1     -1   -1   -1   -1   -1   -1   -1   -1   -1   -1
## 2      1   -1   -1   -1   -1   -1   -1   -1   -1   -1
```

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

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

Randomized paired comparison

```
hist(res, xlab="Mean Difference",main="Randomization Distribution Boys'  
abline(v = meandiff,col="blue")
```



Randomized paired comparison

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

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

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

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

Left-tailed 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

$$P(t_9 < t).$$

Left-tailed 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 the option `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)
```

