Lecture 14

Paired T-Tests; Analysis of Variance (ANOVA)

Text: Chapters 6, 8

STAT 8010 Statistical Methods I
March 3, 2020

Whitney Huang
Clemson University

Another Example

A simple random sample with sample size 37 is taken and are subjected to a treatment ($\bar{X}_1 = 19.45, s_1 = 4.3$). A Simple random sample with sample size 31 is taken and given a placebo ($\bar{X}_2 = 18.2, s_2 = 2.2$). At the 10% level can we say that the means are different between the two groups?

Paired T-Tests

Paired T-Test: Motivating Example

Insurance handlers are concerned about the high estimates they are receiving for auto repairs from garage I compared to garage II. To verify their suspicions, each of 15 cars recently involved in an accident was taken to both garages for separate estimates of repair costs. The estimates from the two garages are given in the following table

Garage I	Garage II	Garage I	Garage II	Garage I	Garage II
17.6	17.3	20.2	19.1	19.5	18.4
11.3	11.5	13.0	12.7	16.3	15.8
15.3	14.9	16.2	15.3	12.2	12.0
14.8	14.2	21.3	21.0	22.1	21.0
16.9	16.1	17.6	16.7	18.4	17.5

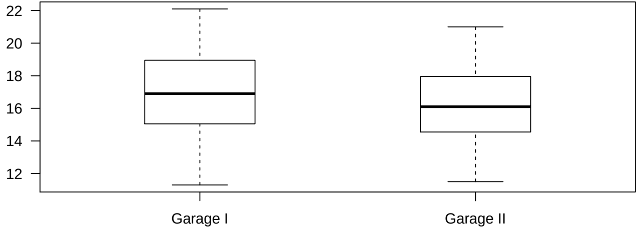
Example Cont'd

Suppose we perform a two-sample test

Sample statistics: $\bar{X}_1 = 16.85, \bar{X}_2 = 16.23, s_1 = 3.20, s_2 = 2.94$

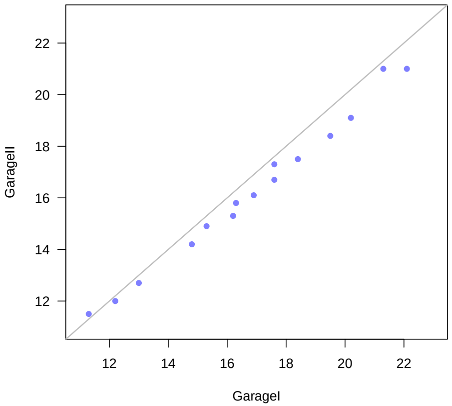
- $H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 > 0$
- $t_{obs} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{16.85 - 16.23}{\sqrt{\frac{3.2^2}{15} + \frac{2.94^2}{15}}} = \frac{0.62}{1.12} = 0.55$
- Critical value for rejection region: $t_{0.05, df=27} = 1.70$
- Since t_{obs} is not in the rejection region. We fail to reject H_0 at 0.05 level.

Boxplots and R Output




```
        Welch Two Sample t-test

data:  GarageI and GarageII
t = 0.54616, df = 27.797, p-value =
0.2947
alternative hypothesis: true difference in means
is greater than 0
95 percent confidence interval:
 -1.29749      Inf
sample estimates:
mean of x mean of y
 16.84667  16.23333
```



For all but one of the 15 cars, the estimates from garage I were higher than that from garage II.

Analyzing Matched Pairs

- Matched pairs are dependent samples where each unit in the first sample is directly linked with a unit in the second sample
- This could occur in several situations, for example, before/after study, study on twins, pairing subjects based on similar characteristics
- We need different strategy for testing two dependent samples $\Rightarrow$ Paired T-Tests

Paired T-Tests

- $H_0 : \mu_{diff} = 0$ vs. $H_a : \mu_{diff} > 0$ (Upper-tailed); $\mu_{diff} < 0$ (Lower-tailed); $\mu_{diff} \neq 0$ (Two-tailed)
- Test statistic: $t^* = \frac{\bar{X}_{diff} - 0}{\frac{s_{diff}}{\sqrt{n}}}$. If $\mu_{diff} = 0$, then $t^* \sim t_{df=n-1}$
- Use rejection region method or P-value method to make a decision

Car Repair Example Revisited

Garage I - Garage II	Garage I - Garage II	Garage I - Garage II
17.6 - 17.3 = 0.3	20.2 - 19.1 = 1.1	19.5 - 18.4 = 1.1
11.3 - 11.5 = -0.2	13.0 - 12.7 = 0.3	16.3 - 15.8 = 0.5
15.3 - 14.9 = 0.4	16.2 - 15.3 = 0.9	12.2 - 12.0 = 0.2
14.8 - 14.2 = 0.6	21.3 - 21.0 = 0.3	22.1 - 21.0 = 1.1
16.9 - 16.1 = 0.8	17.6 - 16.7 = 0.9	18.4 - 17.5 = 0.9

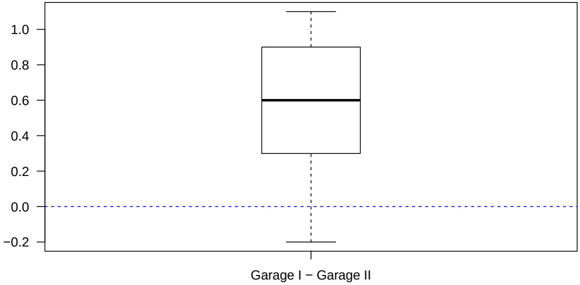
1. First, compute the difference in paired samples
2. Compute the sample mean and standard deviation for the differences
3. Then perform a one sample t-test

Car Repair Example Cont'd

$\bar{X}_{diff} = 0.61, s_{diff} = 0.39$

1. $H_0 : \mu_{diff} = 0$ vs. $H_a : \mu_{diff} > 0$
2. $t_{obs} = \frac{0.61}{\frac{0.39}{\sqrt{15}}} = 6.03$
3. Critical value for rejection region: $t_{0.05,df=14} = 1.76 \Rightarrow$ reject H_0
4. We do have enough evidence that the true mean repair cost difference for the garage I and II is greater than 0

Boxplot and R Output




```
        Paired t-test

data:  GarageI and GarageII
t = 6.0234, df = 14, p-value = 1.563e-05
alternative hypothesis: true difference in means
is greater than 0
95 percent confidence interval:
 0.4339886       Inf
sample estimates:
mean of the differences
              0.6133333
```

Analysis of Variance (ANOVA)

Testing for a Difference in More Than Two Means

- In the last few lectures we have seen how to test a difference in two means, using two sample t-test

Testing for a Difference in More Than Two Means

- In the last few lectures we have seen how to test a difference in two means, using two sample t-test

Testing for a Difference in More Than Two Means

- In the last few lectures we have seen how to test a difference in two means, using two sample t-test
- **Question**: what if we want to test if there are differences in a set of **more than two means**?

Testing for a Difference in More Than Two Means

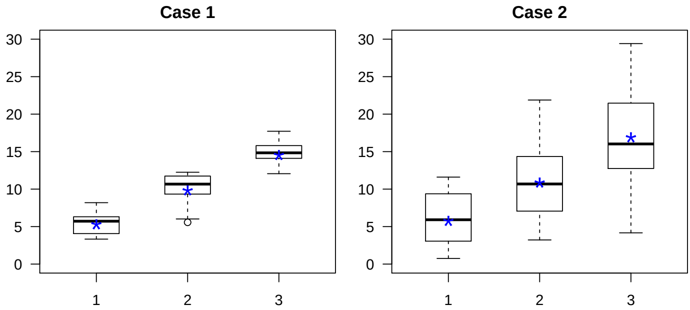
- In the last few lectures we have seen how to test a difference in two means, using two sample t-test
- **Question**: what if we want to test if there are differences in a set of **more than two means**?

Testing for a Difference in More Than Two Means

- In the last few lectures we have seen how to test a difference in two means, using two sample t-test
- **Question**: what if we want to test if there are differences in a set of **more than two means**?
- The statistical tool for doing this is called analysis of variance (ANOVA)

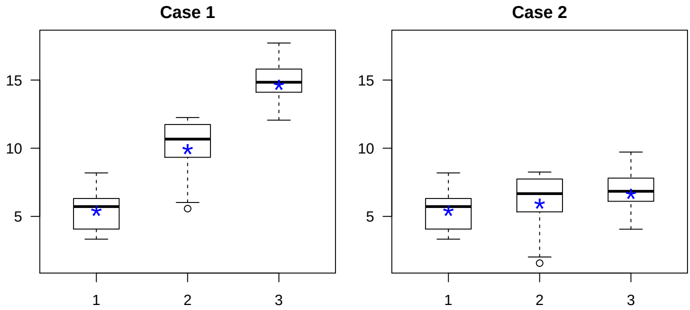
A Quick Quiz: To Detect Differences in Means

Question: Are group 1, 2, 3 for each case come from the same population?



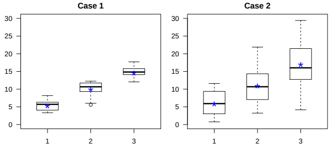
Another Quiz: To Detect Differences in Means

Question: Are group 1, 2, 3 for each case come from the same population?

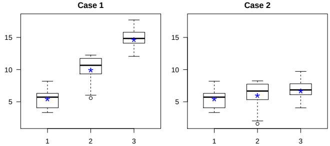


Decomposing Variance to Test for a Difference in Means

- In the first quiz, the data within each group is not very spread out for Case 1, while in Case 2 it is



- In the second quiz, the group means are quite different for Case 1, while they are not in Case 2



- In ANOVA, we compare between group variance (“signal”) to within group variance (“noise”) to detect a difference in means

Notation

$$X_{ij} = \mu_j + \varepsilon_{ij},\ \varepsilon_{ij} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2),\ i = 1, \cdots, n_j, 1 \leq j \leq J$$

- J: number of groups
- $\mu_j, j = 1, \cdots, J$: population mean for j_{th} group
- $\bar{X}_j, j = 1, \cdots, J$: sample mean for j_{th} group
- $s_j^2, j = 1, \cdots, J$: sample variance for j_{th} group
- $N = \sum_{j=1}^{J} n_j$: overall sample size
- $\bar{X} = \frac{\sum_{j=1}^{J} \sum_{i=1}^{n_j} X_{ij}}{N}$: overall sample mean

Partition of Sums of Squares

"Sums of squares" refers to sums of squared deviations from some mean. ANOVA decomposes the total sum of squares into treatment sum of squares and error sum of squares:

- **Total sum of square**: SSTo $= \sum_{j=1}^{J} \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2$
- **Treatment sum of square**: SSTr $= \sum_{j=1}^{J} n_j (\bar{X}_j - \bar{X})^2$
- **Error sum of square**: SSE $= \sum_{j=1}^{J} (n_j - 1) s_j^2$

We can show that $\boxed{\text{SSTo} = \text{SSTr} + \text{SSE}}$

Mean squares

A mean square is a sum of squares divided by its associated degrees of freedom

- **Mean square of treatments**: $\text{MSTr} = \frac{\text{SSTr}}{J-1}$
- **Mean square of error**: $\text{MSE} = \frac{\text{SSE}}{N-J}$

Think of MSTr as the "signal", and MSE as the "noise" when detecting a difference in means $(\mu_1, \cdots, \mu_J)$. A nature test statistic is the signal-to-noise ratio i.e.,

$$F^* = \frac{\text{MSTr}}{\text{MSE}}$$

ANOVA Table and F Test

Source	df	SS	MS	F statistic
Treatment	$J-1$	SSTr	$\text{MSTr} = \frac{\text{SSTr}}{J-1}$	$F = \frac{\text{MSTr}}{\text{MSE}}$
Error	$N-J$	SSE	$\text{MSE} = \frac{\text{SSE}}{N-J}$	
Total	$N-1$	SSTo		

F-Test

- $H_0 : \mu_1 = \mu_2 = \cdots = \mu_J$
  H_a : At least one mean is different

- Test Statistic: $F^* = \frac{\text{MSTr}}{\text{MSE}}$. Under H_0, $F^* \sim F_{df_1=J-1, df_2=N-J}$

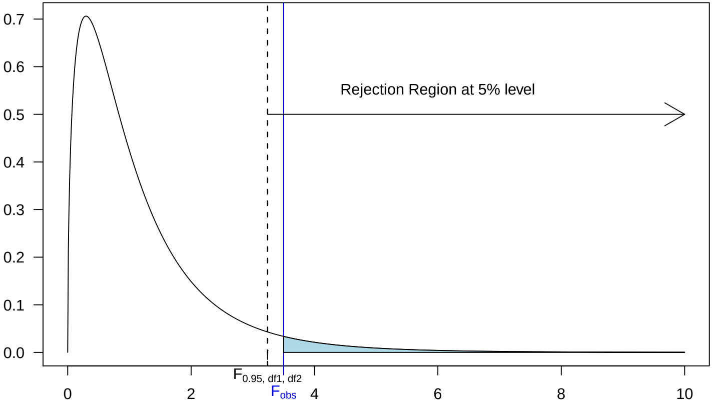
- **Assumptions:**
  - The distribution of each group is normal with equal variance (i.e. $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_J^2$)
  - Responses for a given group are independent to each other

F Distribution and the Overall F-Test

Consider the observed F test statistic: $F_{obs} = \frac{\text{MSTr}}{\text{MSE}}$

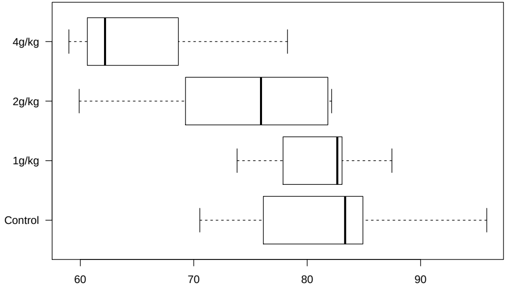
- Should be "near" 1 if the means are equal
- Should be "larger than" 1 if means are not equal

$\Rightarrow$ We use the null distribution of $F^* \sim F_{df_1=J-1, df_2=N-J}$ to quantify if F_{obs} is large enough to reject H_0



Example

A researcher who studies sleep is interested in the effects of ethanol on sleep time. She gets a sample of 20 rats and gives each an injection having a particular concentration of ethanol per body weight. There are 4 treatment groups, with 5 rats per treatment. She records Rapid eye movement (REM) sleep time for each rat over a 24-period. The results are plotted below:



Set Up Hypotheses and Compute Sums of Squares

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs.
  H_a : At least one mean is different

- Sample statistics:

Treatment	Control	1g/kg	2g/kg	4g/kg
Mean	82.2	81.0	73.8	65.7
Std	9.6	5.3	9.4	7.9

- Overall Mean $\bar{X} = \frac{\sum_{j=1}^{4} \sum_{i=1}^{5} X_{ij}}{20} = 75.67$

- SSTo $= \sum_{j=1}^{4} \sum_{i=1}^{5} (X_{ij} - \bar{X})^2 = 1940.69$

- SSTr $= \sum_{j=1}^{4} 5 \times (\bar{X}_j - \bar{X})^2 = 861.13$

- SSE $= \sum_{j=1}^{4} (5 - 1) \times s_j^2 = 1079.56$

ANOVA Table and F-Test

Source	df	SS	MS	F statistic
Treatment	$4-1=3$	861.13	$\frac{861.13}{3}=287.04$	$\frac{287.04}{67.47}=4.25$
Error	$20-4=16$	1079.56	$\frac{1079.56}{16}=67.47$	
Total	19	1940.69		

Suppose we use $\alpha = 0.05$

- Rejection Region Method:
  $F_{obs} = 4.25 > F_{0.95,df_1=3,df_2=16} = 3.24$
- P-value Method: $\mathbb{P}(F^* > F_{obs}) = \mathbb{P}(F^* > 4.25) = 0.022 < 0.05$

> Reject $H_0 \Rightarrow$ We do have enough evidence that not all of population means are equal at 5% level.

R Output

```
Analysis of Variance Table

Response: Response
          Df  Sum Sq Mean Sq
Treatment  3  861.13 287.044
Residuals 16 1079.56  67.472
          F value  Pr(>F)
Treatment  4.2542 0.02173 *
Residuals
---
Signif. codes:
  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’
  0.05 ‘.’ 0.1 ‘ ’ 1
```