Lecture 15

Analysis of Variance (ANOVA)

Readings: IntroStat Chapter 8; OpenIntro Chapter 7.5

STAT 8010 Statistical Methods I

June 6, 2023

Whitney Huang
Clemson University

Testing for a Difference in More Than Two Means

- In the last lecture 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

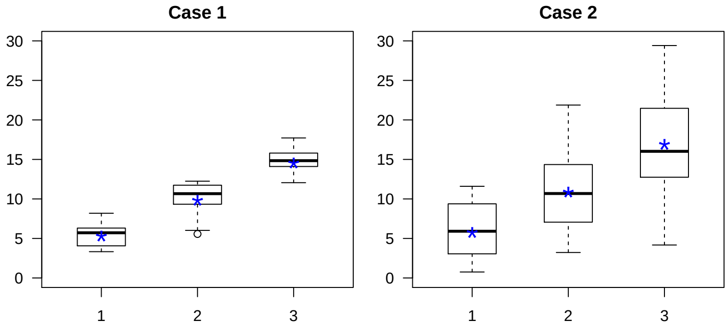
- In the last lecture we have seen how to test a difference in two means, using two sample t-test
- **Question**: what if we want to test if there are differences in a set of **more than two means**?

Testing for a Difference in More Than Two Means

- In the last lecture 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)

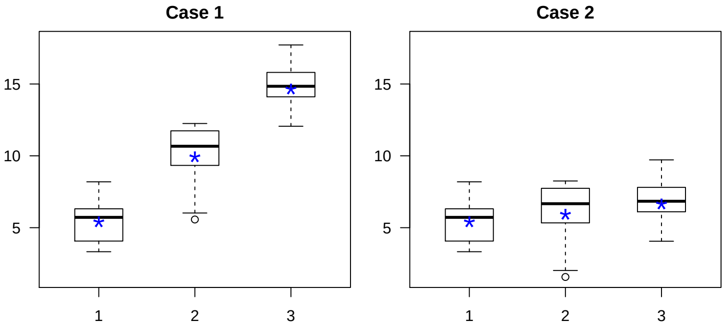
A Quick Quiz: To Detect Differences in Means

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



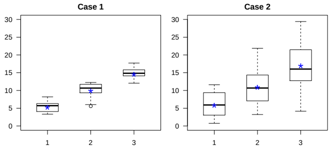
Another Quiz: To Detect Differences in Means

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

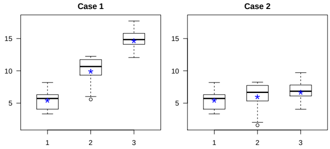


Decomposing Variance to Test for a Difference in Means

- 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

Model and 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	MSTr $= \frac{\text{SSTr}}{J-1}$	$F = \frac{\text{MSTr}}{\text{MSE}}$
Error	$N-J$	SSE	MSE $= \frac{\text{SSE}}{N-J}$	
Total	$N-1$	SSTo		

F-Test

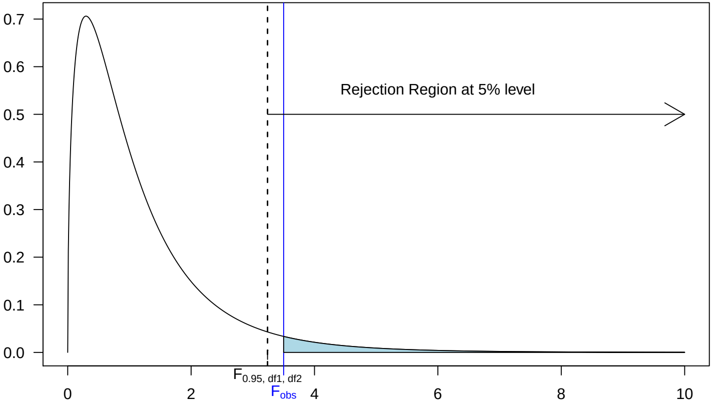
- $H_0 : \mu_1 = \mu_2 = \cdots = \mu_J$
  H_a : At least one mean is different
- Test Statistic: $F^* = \frac{\text{MSTr}}{\text{MSE}}$. Under H_0, $F^* \sim F_{df_1=J-1, df_2=N-J}$
- **Assumptions:**
  - The distribution of each group is normal with equal variance (i.e. $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_J^2$)
  - Responses for a given group are independent to each other

F Distribution and the Overall F-Test

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

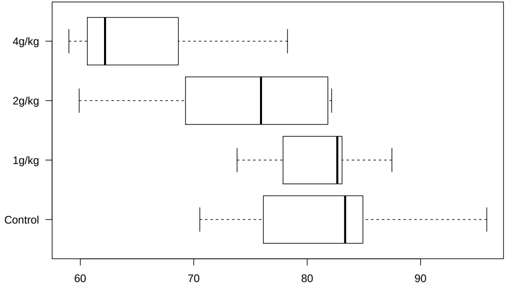
- Should be "near" 1 if the means are equal
- Should be "larger than" 1 if means are not equal

$\Rightarrow$ 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
```

One-Way ANOVA & Overall F-Test

- We use one-way ANOVA to compare means of **J (≥ 3) groups/conditions**

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_J$$
$$H_a : \text{at least a pair } \mu\text{'s differ}$$

- If H_0 is rejected, ANOVA just states that there is a significant difference between the groups **but not where those differences occur**

- We need to perform additional post hoc tests, multiple comparisons, to determine where the group differences are

Pairwise t-Tests

- Suppose we have 4 groups, i.e. $J = 4$, then we need to perform $\binom{4}{2} = 6$ two-sample tests to locate where the group differences are

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_a : \mu_1 \neq \mu_2$$
$$H_0 : \mu_1 = \mu_3 \text{ vs. } H_a : \mu_1 \neq \mu_3$$
$$H_0 : \mu_1 = \mu_4 \text{ vs. } H_a : \mu_1 \neq \mu_4$$
$$H_0 : \mu_2 = \mu_3 \text{ vs. } H_a : \mu_2 \neq \mu_3$$
$$H_0 : \mu_2 = \mu_4 \text{ vs. } H_a : \mu_2 \neq \mu_4$$
$$H_0 : \mu_3 = \mu_4 \text{ vs. } H_a : \mu_3 \neq \mu_4$$

- What if we simply perform these tests using, say, $\alpha = 0.05$ for each test?

$$\mathrm{P}(\text{making a least one type I error}) = 1 - (1 - 0.05)^6 = 0.265$$

if each test was independent

Family-Wise Error Rate (FWER)

Family-Wise Error Rate (FWER) $\bar{\alpha}$: the probability of making 1 or more type I errors in a set of hypothesis tests

For m independent tests, each with individual type I error rate α, then we have

$$\bar{\alpha} = 1 - (1 - \alpha)^m$$

	α		
m	0.1	0.05	0.01
1	0.100	0.050	0.010
3	0.271	0.143	0.030
6	0.469	0.265	0.059
10	0.651	0.401	0.096
15	0.794	0.537	0.140
21	0.891	0.659	0.190

The Bonferroni Correction

If we would like to control the FWER to be α, then we adjust the significant level for each of the m tests to be $\frac{\alpha}{m}$

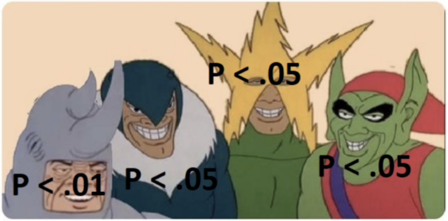
$$FWER = \mathrm{P}(\cup_{i=1}^{m} p_i \leq \frac{\alpha}{m}) \leq \sum_{i=1}^{m} \mathrm{P}(p_i \leq \frac{\alpha}{m}) = m\frac{\alpha}{m} = \alpha$$

where p_i is the p-value for the i_{th} test

If we have 4 treatment groups, then we need to perform 6 tests $(m = 6) \Rightarrow$ will need to set the significant level for each individual pairwise t-test to be $0.05/6 = 0.0083$ to ensure that FWER is less than 0.05

Remark: Bonferroni procedure can be very conservative but gives guaranteed control over FWER at the risk of reducing statistical power. Does not assume independence of the comparisons.

Me and the significant boys



Me and the significant boys after Bonferroni correction

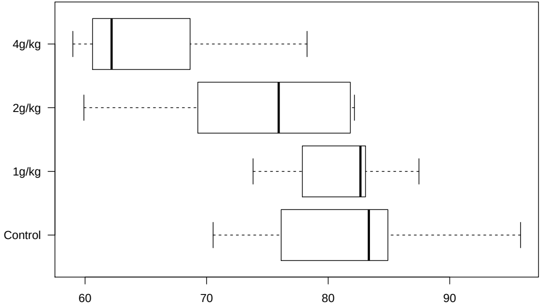
Example

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.

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

Recall in last lecture we reject $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ at 0.05 level. But where these differences are?

Example: Multiple Testing with Bonferroni Correction



P-value

Test	μ_1, μ_2	μ_1, μ_3	μ_1, μ_4	μ_2, μ_3	μ_2, μ_4	μ_3, μ_4
Pooled	0.816	0.202	0.018	0.175	0.007	0.179
Non-pooled	0.818	0.202	0.019	0.185	0.009	0.180

Fisher's Protected Least Significant Difference (LSD) Procedure

- We conclude that μ_i and μ_j differ at α significance level if $|\bar{X}_i - \bar{X}_j| > LSD$, where

$$LSD = t_{\alpha/2, df=N-J} \sqrt{\text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

- This procedure builds on the equal variances t-test of the difference between two means
- The test statistic is improved by using MSE rather than s_p^2

Tukey's Honest Significance Difference (HSD) Test

- The test procedure:
  - Requires equal sample size n per populations
  - Find a critical value ω as follows:

$$\omega = q_\alpha(J, N-J)\sqrt{\frac{\text{MSE}}{n}}$$

  where $q_\alpha(J, N-J)$ can be obtained from the studentized range table
- If $\bar{X}_{max} - \bar{X}_{min} > \omega \Rightarrow$ there is sufficient evidence to conclude that $\mu_{max} > \mu_{min}$
- Repeat this procedure for each pair of samples. Rank the means if possible

Summary

In this lecture, we learned

- Sums of Squares Decomposition of ANOVA
- ANOVA Table and F-Test
- Multiple Testing