ANOVA: Comparing More Than Two Treatments

Sergey Tarima, PhD

Sponsored by the Clinical and Translational Science Institute (CTSI) and the Department of Population Health / Division of Biostatistics
Speaker Disclosure

In accordance with the ACCME policy on speaker disclosure, the speaker and planners who are in a position to control the educational activity of this program were asked to disclose all relevant financial relationships with any commercial interest to the audience. The speaker and program planners have no relationships to disclose.
Outline

• Refreshment: t-test
• Multiple Comparisons
• One-way ANOVA
• Two-way ANOVA
T-test (1)

• Suppose we have 2 groups (for example, the class of Mr. First and the class of Ms. Second)
• We want to compare mean response between groups (what is our response? suppose - the mean number of tardy slips)
• We assume that our means are approximately normally distributed
Recall what t-test does (2)

• What does it mean normally distributed? … bell-shaped curve 😊.
• Suppose Mr. First provided the following data: 1,10,12,0,0,5,6,2,3,7,8,9,12,12,10,0.
• Ms. Second reported: 10,20,3,5,6,2,13,19,8,2,5,2,5,7,9,22,1,10
• We can see that there were 16 students in Mr. First class and 18 in Ms. Second.
Normal Densities of the two classes

Mr. First Class

Ms. Second Class
Histograms (1)

Histogram: Mr. First Class

- Frequency
- Tardy slips

0 1 2 3 4 5

0 2 4 6 8 10 12
Histograms (2)

Histogram: Ms. Second Class

- Frequency
- Tardy slips

6/11/2009
Normal data?

- Our data do NOT look normal – it is obvious from the histograms.
- This means that our data cannot be fully described only through means and standard deviations.
- However, for large sample sizes averages approximately follow a normal distribution.
- Thus, t-test allows comparing means between two groups for large sample sizes (some claim that 30 per group is enough).
T-test (3)

- Our data are likely not normally distributed... but we will proceed with our illustrative example, as if they were
- We estimate the mean for Mr. First group as \((1+10+\ldots+12+10)/16=6.0625\). Standard deviation as 4.567549.
- The estimated mean and standard deviation are 8.3 and 6.5
Two Sample t-test

data: mrf and mss

\[ t = -1.1417, \text{df} = 32, p\text{-value} = 0.2621 \]

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-6.167687 1.737131

sample estimates:

mean of x  mean of y

6.062500  8.277778

0.2621 > 0.05

So we fail to show that group means are different

In this case 5% is our type I error (the probability that we wrongly conclude that there is a difference between group means)
ANOVA (R output)

the output is

Analysis of Variance Table
Response: data[, 1]

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>data[, 2]</td>
<td>1</td>
<td>41.57</td>
<td>41.57</td>
<td>1.3034</td>
<td>0.2621</td>
</tr>
<tr>
<td>Residuals</td>
<td>32</td>
<td>1020.55</td>
<td>31.89</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note, P-value is the same

So, ANOVA applied to two groups is the same as a two sample t-test
Population

- The population we are making inference about is restricted to two groups only:
  - Mr. First Class
  - Ms. Second Class
The third group!

- Mrs. Third comes! She claims that the number of “tardy slips” in her class is much smaller than in Mr. First and Ms. Second classes!

- Her data are:
  1,2,3,4,11,2,3,4,11,2,3,4,1,2,3,4,1,10,0,7,3,4,5,2,3,4,5,6,17,14,3,7

- She has a big class and seems like students are severely punished for not being in time for class.

- The estimated mean is 4.7 and the estimated standard deviation is 3.9.
Three classes

Mrs. Third class
Multiple comparisons

• Now we have 3 groups.
• We can use t-test to compare whether the number of “tardy slips” is different between Mr. First and Ms. Second classes.
• Similarly we can compare Mr. First and Mrs. Third classes, and Ms. Second and Mrs. Third classes.
• What would be our type I error?
Three t-tests

Comparing 1\textsuperscript{st} vs 3\textsuperscript{rd} classes P-value = 29.6%
Comparing 2\textsuperscript{nd} vs 3\textsuperscript{rd} classes P-value = 1.9%
Comparing 1\textsuperscript{st} vs 2\textsuperscript{nd} classes P-value = 26.2%

What can we say if we see these results?
…the mean number of “tardy slips” is significantly higher in Ms. Second than in Mrs. Third class!
but this is true only if we are looking at 5\% significance level
and testing only ONE hypothesis (2\textsuperscript{nd} vs 3\textsuperscript{rd})
Beware: Type I Error (1)

- One test: 5% significance level means if P-value is below 5% we reject the null of no difference, otherwise “fail to reject”
- In other words: 5 times out of 100 we falsely reject the null when it is true !!!
- Two (independent) tests: reject the null if P-value < 5%. No error in 95 out of 100 cases in each of the tests. So, the probability that we fail to reject the null is 0.95*0.95 = 0.91 !!! (91 out of 100)
Beware: Type I Error (2)

- Fifty tests: each is performed at 5% significance. The probability that the null is rejected at least once when the null is true is $0.95^{50} = 7.7\%$ (!!!) So, we erroneously reject the null in 92.3% cases.
- Did not we go too far from the initially stated 5% type I error rate?
- This why statisticians mention “multiple comparisons” problem. If you massage your data long enough eventually you will be able to find a significant result.
ANOVA (1)

• **The null hypothesis:** The mean number of “tardy slips” is the same across all three groups!!!

• This hypothesis can be tested using ANOVA test and the type I error will stay at a priori defined significance level (say 5%).

• ANOVA has a mathematical justification via sum of squares partitioning… not in this lecture
ANOVA (2)
(“tardy slips” example, now 3 groups)

R OUTPUT:
Analysis of Variance Table
Response: data[, 1]

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>as.factor(data[, 2])</td>
<td>2</td>
<td>145.97</td>
<td>72.98</td>
<td>3.0632</td>
<td>0.05376 .</td>
</tr>
<tr>
<td>Residuals</td>
<td>63</td>
<td>1501.02</td>
<td>23.83</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

P-value
ANOVA (3)

- ANOVA keeps 5% type I error.
- If ANOVA’s P-value is higher than 5%, we fail to conclude that at least one group is significantly different from others.
- If ANOVA’s P-value is smaller than 5%, we conclude that there exists at least one group and its mean is significantly different (in terms of means) from others.
ANOVA model (4)

- ANOVA can be represented as the following model, where

\[ y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \]

1. \( i=1,\ldots,I \) denotes groups (in our case there are three groups),
2. \( j=1,\ldots,J_i \) denotes the number of observations per group (the number of students per class),
3. \( y_{ij} \) are our observations (number of “tardy slips” per student),
4. \( \mu \) – overall mean (mean number of “tardy slips” among all three classes),
5. \( \alpha_i \) – \( i^{th} \) group effect (the difference between \( i^{th} \) group and overall means),
6. \( \varepsilon_{ij} \) is a measurement error (assumed to follow \( N(0,\sigma^2) \))
ANOVA model (5)

\[ y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \]

\( y_{ij} \) is what we observe (actual numbers)

\( \varepsilon_{ij} \) is a random noise or error

\( \mu \) and \( \alpha_i \) are not observed (this is what we are estimating)
ANOVA Model (8)  
"tardy slips" example

• We will use reference parameterization. So, we set $\alpha_1=0$ and we will need to estimate $\alpha_2$, $\alpha_3$, and $\mu$.

• In this case
  (1) $\mu$ represents the mean number of “tardy slips” in Mr. First class,
  (2) $\mu + \alpha_2$ is the mean for Ms. Second, and
  (3) $\mu + \alpha_3$ is the mean for Mrs. Third.
ANOVA Model (9) ("tardy slips" example)

• With reference coding the null hypothesis of no difference would require $\alpha_2=\alpha_3=0$, which corresponds to
  $H_0: y_{ij}=\mu+\epsilon_{ij}$. (here $\alpha_2=\alpha_3=0$)

• The alternative hypothesis will be
  $H_A: y_{ij}=\mu+\alpha_i+\epsilon_{ij}$. (here $\alpha_2$ or $\alpha_3$ not equal 0)
ANOVA Model (10) ("tardy slips" example)

It is good to explore our data first:

SCATTERPLOT

Note, some multiple observations are plotted as a single point.
ANOVA Model (11)
("tardy slips" example - R)

Another exploration tool:

BOX PLOTS

Some observations may be outliers and if excluded we may see a smaller variance in the 3rd group
ANOVA diagnostics (12)

In order to check validity of a model it is recommended to look at normality of model residuals !!!

QQ PLOT

On a Q-Q plot points should be close to a straight line
ANOVA model \((14)\)

If the data are not well described by ANOVA model the following can be considered:

- outcome transformation (for example, the natural logarithm of the number of “tardy slips”)

- outlier detection; outliers can be excluded if justified (suppose somebody lives in a different school attendance area)
### ANOVA Model (15) (‘‘tardy slips’’ example)

Analysis of Variance Table

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>as.factor(data[, 2])</td>
<td>2</td>
<td>145.97</td>
<td>72.98</td>
<td>3.0632</td>
</tr>
<tr>
<td>Residuals</td>
<td>63</td>
<td>1501.02</td>
<td>23.83</td>
<td></td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Significant at 10%
Population

- The population we are making inference about is restricted to three groups only

  Mr. First Class  Ms. Second Class  Mrs. Third Class
ANOVA Model (16)  
(“tardy slips” example)

If ANOVA tests allow us to say that there exists a difference, the question remains: “Where is the difference?”

Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.062 1.220 4.968 5.47e-06 ***
as.factor(data[, 2])2  2.215 1.677 1.321 0.191
as.factor(data[, 2])3 -1.344 1.495 -0.899 0.372
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 4.881 on 63 degrees of freedom
Multiple R-squared: 0.08863,  Adjusted R-squared: 0.05969
F-statistic: 3.063 on 2 and 63 DF,  p-value: 0.05376

α₂ is not different from 0

α₃ is not different from 0

Estimate of μ

Estimate of α₂

Estimate of α₃
ANOVA Model (17) ("tardy slips" example)

• So, ANOVA test says that the groups are different (at 10%), but when we compared classes 1 and 2 (P-val = 19.1%) and classes 1 and 3 (P-val = 37.2%)… not different (note, P-values are the same as in t-tests)

• The hypothesis of no difference between classes 2 and 3 corresponds to $H_0: \alpha_2 = \alpha_3$. 
ANOVA Model (18) (multiple comparisons)

• There many procedures developed for multiple comparisons – the simplest and the most conservative is the Bonferroni method.

• According to Bonferroni, we can divide the significance level (say 5%) on the number of tests (say 50). Then each test is tested at 5% / 50 = 0.1% significance level.
Summary of ANOVA (19)

- We talked about the simplest ANOVA model and test: one-way ANOVA
- ANOVA represents a generalization of a t-test for more than two groups.
- ANOVA assumes that noise follows $N(0, \sigma^2)$.
- ANOVA assumes equal variance within each group.
• Suppose Ms. Second claims that many of her students actually live in a different school attendance area, and she wants this to be taken into account !!!

• Two-way ANOVA allows to solve this problem.
Two-way ANOVA (1)

- One-way ANOVA considers only one way to classify subjects. In our example, all students were divided into three classes.
- Two-way ANOVA allows us to consider another factor. Suppose “attendance area” is our second factor. Each student is classified as living in the school attendance area or not.
Two-way ANOVA model (2)

$y_{ijk}$ is what we observe (actual numbers)

$\varepsilon_{ijk}$ is a random noise or error

$\mu, \alpha_i, \text{ and } \beta_j$ are not observed (this is what we are estimating)

$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$
Two-way ANOVA model (3) ("tardy slips" example)

Here i denotes classes, j denotes attendance area, k enumerates students

\[ y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \]

class effect

attendance area effect
# Two-way ANOVA model (4)

Analysis of Variance Table

<table>
<thead>
<tr>
<th>Response: data[,1]</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Df</td>
<td>Sum Sq</td>
<td>Mean Sq</td>
<td>F value</td>
<td>Pr(&gt;F)</td>
<td></td>
</tr>
<tr>
<td>data[,2]</td>
<td>2</td>
<td>145.97</td>
<td>72.98</td>
<td>3.0370</td>
<td>0.05517 .</td>
</tr>
<tr>
<td>data[,3]</td>
<td>1</td>
<td>11.05</td>
<td>11.05</td>
<td>0.4596</td>
<td>0.50032</td>
</tr>
<tr>
<td>Residuals</td>
<td>62</td>
<td>1489.97</td>
<td>24.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Effect of class continues being NOT significant in the adjusted for attendance area analysis.

The effect of attendance area is not significant.
Population

The population we are making inference about is restricted to three groups only, but we also control for “attendance area” factor.
Resources

- The Clinical and Translation Science Institute (CTSI) supports education, collaboration, and research in clinical and translational science: www.ctsi.mcw.edu
- The Biostatistics Consulting Service provides comprehensive statistical support: http://www.mcw.edu/biostatsconsult.htm
Free drop-in consulting

- **MCW/Froedtert/CHW:**
  - Monday, Wednesday, Friday 1 – 3 PM @ CTSI Administrative offices (LL772A)
  - Tuesday, Thursday 1 – 3 PM @ Health Research Center, H2400

- **VA:** 1\(^{st}\) and 3\(^{rd}\) Monday, 8:30-11:30 am
  - VA Medical Center, Building 70, Room D-21

- **Marquette:** 2\(^{nd}\) and 4\(^{th}\) Monday, 8:30-11:30 am
  - Olin Engineering Building, Room 338D
This video/audio file is owned by the Medical College of Wisconsin. Any copying or distribution is prohibited without our express written consent.

This multimedia program was developed by the Medical College of Wisconsin for the private use of individuals who are authorized to view its contents.

The Medical College of Wisconsin presents the information on the video/audio presentation as an educational service. While the information on this presentation is about health care issues, it is not medical advice. People seeking specific medical advice or assistance should contact their personal physician.

Although we believe the information in this presentation to be accurate and timely, because of the rapid advances in health care and our reliance on information provided by outside sources, we make no warranty or guarantee concerning the accuracy or reliability of the content or other material which we may reference. When clinical matters are discussed, the opinions presented are those of the discussants only.

The material discussed on the presentation is not intended to present the only or necessarily the best method or procedure, but rather presents the approach or opinion of the discussant. This presentation is provided in an “as is” format without warranties of any kind, expressed or implied, including but not limited to warranties of title, noninfringement or implied warranties of merchantability or fitness for a particular purpose.

Any questions concerning the content of this program may be directed to the Medical College of Wisconsin Department of Biostatistics at biostatistics@hpi.mcw.edu or 414-456-8280.