Chapter 22

One-Way Analysis of Variance: Comparing Several Means

Comparing Means

◆ Chapter 17: compared the means of two populations or the mean responses to two treatments in an experiment
  – two-sample t tests
◆ This chapter: compare any number of means
  – Analysis of Variance
  ♦ Remember: we are comparing means even though the procedure is Analysis of Variance

Case Study

Gas Mileage for Classes of Vehicles


Do SUVs and trucks have lower gas mileage than midsize cars?
Case Study
Gas Mileage for Classes of Vehicles

Data collection
- Response variable: gas mileage (mpg)
- Groups: vehicle classification
  - 31 midsize cars
  - 31 SUVs
  - 14 standard-size pickup trucks
    - only two-wheel drive vehicles were used
    - four-wheel drive SUVs and trucks get poorer mileage

Means (\(\bar{x}\)):
Midsize: 27.903
SUV: 22.677
Pickup: 21.286
Case Study
Gas Mileage for Classes of Vehicles

Data analysis

Means (\(\bar{x}\)):
- Midsize: 27.903
- SUV: 22.677
- Pickup: 21.286

- Mean gas mileage for SUVs and pickups appears less than for midsize cars
- Are these differences statistically significant?

---

Null hypothesis:
The true means (for gas mileage)
- for all groups (the three vehicle classifications)

For example, could look at separate t tests to compare each pair of means to see if they are different:
- 27.903 vs. 22.677
- 27.903 vs. 21.286
- 22.677 vs. 21.286

H_0: \(\mu_1 = \mu_2\)
H_0: \(\mu_1 = \mu_3\)
H_0: \(\mu_2 = \mu_3\)

Problem of multiple comparisons!

---

Multiple Comparisons

- Problem of how to do many comparisons at the same time with some overall measure of confidence in all the conclusions
- Two steps:
  - overall test to test for any differences
  - follow-up analysis to decide which groups differ and how large the differences are
- Follow-up analyses can be quite complex; we will look only at the overall test for a difference in several means, and examine the data to make follow-up conclusions
Analysis of Variance $F$ Test

- $H_0$: $\mu_1 = \mu_2 = \mu_3$
- $H_a$: not all of the means are the same
- To test $H_0$, compare how much variation exists among the sample means (how much the $X$'s differ) with how much variation exists within the samples from each group
  - is called the analysis of variance $F$ test
  - test statistic is an $F$ statistic
  - use $F$ distribution ($F$ table) to find $P$-value
  - analysis of variance is abbreviated ANOVA

---

Case Study
Gas Mileage for Classes of Vehicles
Using Technology

One-way ANOVA: Midsize, SUV, Pickup

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td>666.48</td>
<td>2</td>
<td>333.24</td>
</tr>
<tr>
<td>Error</td>
<td>700.16</td>
<td>90</td>
<td>7.78</td>
</tr>
<tr>
<td>Total</td>
<td>1366.64</td>
<td>92</td>
<td></td>
</tr>
</tbody>
</table>

$F = \frac{333.24}{7.78} = 42.83$

$P$-value < 0.001

Significant differences

Follow-up analysis

Pooled SD = 1.27

Individual 95% CIs for Mean

Based on Pooled SD

---

Case Study
Gas Mileage for Classes of Vehicles
Data analysis

- $F = 31.61$
- $P$-value = 0.000 (rounded) (is <0.001)
  - there is significant evidence that the three types of vehicle do not all have the same gas mileage
  - from the confidence intervals (and looking at the original data), we see that SUVs and pickups have similar fuel economy and both are distinctly poorer than midsize cars

---
ANOVA Idea

- ANOVA tests whether several populations have the same mean by comparing how much variation exists among the sample means (how much the $\bar{X}$s differ) with how much variation exists within the samples from each group
  - the decision is not based only on how far apart the sample means are, but instead on how far apart they are relative to the variability of the individual observations within each group

ANOVA Idea

- Sample means for the three samples are the same for each set (a) and (b) of boxplots (shown by the center of the boxplots)
  - variation among sample means for (a) is identical to (b)
- Less spread in the boxplots for (b)
  - variation among the individuals within the three samples is much less for (b)

CONCLUSION: the samples in (b) contain a larger amount of variation among the sample means relative to the amount of variation within the samples, so ANOVA will find more significant differences among the means in (b)
  - assuming equal sample sizes here for (a) and (b)
  - larger samples will find more significant differences
Gas Mileage for Classes of Vehicles

Case Study

Variation among sample means
(how much the *x*'s differ from each other)

<table>
<thead>
<tr>
<th>Midsize</th>
<th>SUV</th>
<th>Pickup</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>22</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>26</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>27</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>28</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>29</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>32</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
</tbody>
</table>

ANOVA *F* Statistic

To determine statistical significance, we need a test statistic that we can calculate

- ANOVA *F* Statistic:
  
  \[ F = \frac{\text{variation among the sample means}}{\text{variation among individuals in the same sample}} \]

- must be zero or positive
  - only zero when all sample means are identical
  - gets larger as means move further apart
- large values of *F* are evidence against *H₀: equal means*
- the *F* test is upper one-sided
ANOVA $F$ Test

- Calculate value of $F$ statistic
  - by hand (cumbersome)
  - using technology (computer software, etc.)
- Find $P$-value in order to reject or fail to reject $H_0$
  - use $F$ table (Table D on pages 656-659 in text) for $F$ distribution (described in Chapter 17)
  - from computer output
- If significant relationship exists (small $P$-value):
  - follow-up analysis
    - observe differences in sample means in original data
    - formal multiple comparison procedures (not covered here)

$F$ test for comparing $I$ populations, with an SRS of size $n_i$ from the $i^{th}$ population (thus giving $N = n_1 + n_2 + \cdots + n_I$, total observations) uses critical values from an $F$ distribution with the following numerator and denominator degrees of freedom:

- numerator $df = I - 1$
- denominator $df = N - I$

$P$-value is the area to the right of $F$ under the density curve of the $F$ distribution

$P$-value:

- for particular numerator $df$ in the top margin of Table D and denominator $df$ in the left margin, locate the $F$ critical value ($F^*$) in the body of the table
- the corresponding probability ($p$) of lying to the right of this value is found in the left margin of the table (this is the $P$-value for an $F$ test)
Case Study
Gas Mileage for Classes of Vehicles

**Using Technology**

Gas Mileage for Classes of Vehicles

\[ F = 31.61 \]

\[ I = 3 \] classes of vehicle

\[ n_1 = 31 \text{ midsize}, n_2 = 31 \text{ SUVs}, n_3 = 14 \text{ trucks} \]

\[ N = 31 + 31 + 14 = 76 \]

\[ df_{num} = (I-1) = (3-1) = 2 \]

\[ df_{den} = (N-I) = (76-3) = 73 \]

Look up \( df_{num} = 2 \) and \( df_{den} = 73 \) (use 50) in Table D; the value \( F = 31.61 \) falls above the 0.001 critical value.

Thus, the **\( P \)-value** for this ANOVA \( F \) test is **less than 0.001**.

**\( P \)-value < .05, so we conclude significant differences**

---

**ANOVA Model, Assumptions**

- Conditions required for using ANOVA \( F \) test to compare population means
  1. have \( I \) independent SRSs, one from each population.
  2. the \( i \)th population has a Normal distribution with unknown mean \( \mu_i \) (means may be different).
  3. all of the populations have the same standard deviation \( \sigma \), whose value is unknown.
Robustness

- ANOVA F test is not very sensitive to lack of Normality (is robust)
  - what matters is Normality of the sample means
  - ANOVA becomes safer as the sample sizes get larger, due to the Central Limit Theorem
  - if there are no outliers and the distributions are roughly symmetric, can safely use ANOVA for sample sizes as small as 4 or 5

Robustness

- ANOVA F test is not too sensitive to violations of the assumption of equal standard deviations
  - especially when all samples have the same or similar sizes and no sample is very small
  - statistical tests for equal standard deviations are very sensitive to lack of Normality (not practical)
  - check that sample standard deviations are similar to each other (next slide)

Checking Standard Deviations

- The results of ANOVA F tests are approximately correct when the largest sample standard deviation (s) is no more than twice as large as the smallest sample standard deviation
Case Study
Gas Mileage for Classes of Vehicles

<table>
<thead>
<tr>
<th>Level</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midsize</td>
<td>31</td>
<td>27.503</td>
<td>3.563</td>
</tr>
<tr>
<td>SUV</td>
<td>31</td>
<td>22.677</td>
<td>3.479</td>
</tr>
<tr>
<td>Pickup</td>
<td>31</td>
<td>31.296</td>
<td>3.759</td>
</tr>
</tbody>
</table>

Pooled StDev = 3.097

ANOVA Details

◆ ANOVA F statistic:

$$F = \frac{\text{variation among the sample means}}{\text{variation among individuals in the same sample}}$$

– the measures of variation in the numerator and denominator are mean squares

◆ Numerator: Mean Square for Groups (MSG)

$$MSG = \frac{n_1(x_1 - \bar{X})^2 + n_2(x_2 - \bar{X})^2 + \cdots + n_I(x_I - \bar{X})^2}{I - 1}$$

✔ $n_i$ is the number of observations in the $i^{th}$ group

✔ $\bar{X} = \frac{n_1x_1 + n_2x_2 + \cdots + n_Ix_I}{N}$

$s_1 = 2.561$
$s_2 = 3.673$
$s_3 = 2.758$

largest $s = 3.673$
smallest $s = 2.561$

⇒ safe to use ANOVA $F$ test
ANOVA Details

- Denominator: Mean Square for Error (MSE)
  - an average of the individual sample variances ($s_i^2$) within each of the $I$ groups
  \[
  \text{MSE} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2}{N - I}
  \]
  \* MSE is also called the pooled sample variance, written as $s_p^2$ ($s_p$ is the pooled standard deviation)
  \* $s_p^2$ estimates the common variance $\sigma^2$

ANOVA Details

- the numerators of the mean squares are called the sums of squares (SSG and SSE)
- the denominators of the mean squares are the two degrees of freedom for the $F$ test, $(I-1)$ and $(N-I)$
- usually results of ANOVA are presented in an ANOVA table, which gives the source of variation, df, SS, MS, and $F$ statistic
  \* ANOVA $F$ statistic: \( F = \frac{\text{MSG}}{\text{MSE}} \frac{\text{SSG}}{\text{SSE}} \)

Case Study
Gas Mileage for Classes of Vehicles
Using Technology

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td>2</td>
<td>606.46</td>
<td>303.22</td>
<td>31.61</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>73</td>
<td>700.34</td>
<td>9.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>75</td>
<td>1306.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For detailed calculations, see Examples 22.7 and 22.8 on pages 618-619 of the textbook.
Summary

The ANOVA $F$-test is used to compare the means of more than two groups. The null hypothesis is that all group means are equal, and the alternative hypothesis is that at least one group mean is different.

Confidence Interval for the Mean $\mu_i$ of any group:

$$\bar{x}_i \pm t^* \frac{s_p}{\sqrt{n_i}}$$

- $t^*$ is the critical value from the $t$ distribution with $N-I$ degrees of freedom (because $s_p$ has $N-I$ degrees of freedom)
- $s_p$ (pooled standard deviation) is used to estimate $\sigma$ because it is better than any individual $s_i$

Case Study: Gas Mileage for Classes of Vehicles

Using Technology

<table>
<thead>
<tr>
<th>Level</th>
<th>N</th>
<th>Mean</th>
<th>$s_i$</th>
<th>$F$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUV</td>
<td>31</td>
<td>22.677</td>
<td>1.437</td>
<td>(---)</td>
</tr>
<tr>
<td>Pickup</td>
<td>14</td>
<td>21.286</td>
<td>1.724</td>
<td>(---)</td>
</tr>
</tbody>
</table>

Pooled $s_p = 3.037$