

11.3: Analysis of Variance (ANOVA)

There are times where you want to compare three or more population means. One idea is to just test different combinations of two means. The problem with that is that your chance for a type I error increases. Instead you need a process for analyzing all of them at the same time. This process is known as **analysis of variance (ANOVA)**. The test statistic for the ANOVA is fairly complicated, you will want to use technology to find the test statistic and p-value. The test statistic is distributed as an F-distribution, which is skewed right and depends on degrees of freedom. Since you will use technology to find these, the distribution and the test statistic will not be presented. Remember, all hypothesis tests are the same process. Note that to obtain a statistically significant result there need only be a difference between any two of the k means.

Before conducting the hypothesis test, it is helpful to look at the means and standard deviations for each data set. If the sample means with consideration of the sample standard deviations are different, it may mean that some of the population means are different. However, do realize that if they are different, it doesn't provide enough evidence to show the population means are different. Calculating the sample statistics just gives you an idea that conducting the hypothesis test is a good idea.

Hypothesis test using ANOVA to compare k means

1. State the random variables and the parameters in words

x_1 = random variable 1

x_2 = random variable 2

\vdots

x_k = random variable k

μ_1 = mean of random variable 1

μ_2 = mean of random variable 2

\vdots

μ_k = mean of random variable k

2. State the null and alternative hypotheses and the level of significance

$H_o : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$

H_A : at least two of the means are not equal

Also, state your α level here.

3. State and check the assumptions for the hypothesis test

- a. A random sample of size n_i is taken from each population.
- b. All the samples are independent of each other.
- c. Each population is normally distributed. The ANOVA test is fairly robust to the assumption especially if the sample sizes are fairly close to each other. Unless the populations are really not normally distributed and the sample sizes are close to each other, then this is a loose assumption.
- d. The population variances are all equal. If the sample sizes are close to each other, then this is a loose assumption.

4. Find the test statistic and p-value

The test statistic is $F = \frac{MS_B}{MS_W}$, where $MS_B = \frac{SS_B}{df_B}$ is the mean square between the groups (or factors), and $MS_W = \frac{SS_W}{df_W}$

is the mean square within the groups. The degrees of freedom between the groups is $df_B = k - 1$ and the degrees of freedom within the groups is $df_W = n_1 + n_2 + \cdots + n_k - k$. To find all of the values, use technology such as the TI-83/84 calculator or R.

The test statistic, F , is distributed as an F-distribution, where both degrees of freedom are needed in this distribution. The p-value is also calculated by the calculator or R.

5. Conclusion

This is where you write reject H_o or fail to reject H_o . The rule is: if the p-value $< \alpha$, then reject H_o . If the p-value $\geq \alpha$, then fail to reject H_o .

6. Interpretation

This is where you interpret in real world terms the conclusion to the test. The conclusion for a hypothesis test is that you either have enough evidence to show H_A is true, or you do not have enough evidence to show H_A is true.

If you do in fact reject H_0 , then you know that at least two of the means are different. The next question you might ask is which are different? You can look at the sample means, but realize that these only give a preliminary result. To actually determine which means are different, you need to conduct other tests. Some of these tests are the range test, multiple comparison tests, Duncan test, Student-Newman-Keuls test, Tukey test, Scheffé test, Dunnett test, least significant different test, and the Bonferroni test. There is no consensus on which test to use. These tests are available in statistical computer packages such as Minitab and SPSS.

Example 11.3.1 hypothesis test involving several means

Cancer is a terrible disease. Surviving may depend on the type of cancer the person has. To see if the mean survival time for several types of cancer are different, data was collected on the survival time in days of patients with one of these cancer in advanced stage. The data is in Example 11.3.1 ("Cancer survival story," 2013). (Please realize that this data is from 1978. There have been many advances in cancer treatment, so do not use this data as an indication of survival rates from these cancers.) Do the data indicate that at least two of the mean survival time for these types of cancer are not all equal? Test at the 1% level.

Table 11.3.1: Survival Times in Days of Five Cancer Types

| Stomach | Bronchus | Colon | Ovary | Breast |
|---------|----------|-------|-------|--------|
| 124 | 81 | 248 | 1234 | 1235 |
| 42 | 461 | 377 | 89 | 24 |
| 25 | 20 | 189 | 201 | 1581 |
| 45 | 450 | 1843 | 356 | 1166 |
| 412 | 246 | 180 | 2970 | 40 |
| 51 | 166 | 537 | 456 | 727 |
| 1112 | 63 | 519 | | 3808 |
| 46 | 64 | 455 | | 791 |
| 103 | 155 | 406 | | 1804 |
| 876 | 859 | 365 | | 3460 |
| 146 | 151 | 942 | | 719 |
| 340 | 166 | 776 | | |
| 396 | 37 | 372 | | |
| | 223 | 163 | | |
| | 138 | 101 | | |
| | 72 | 20 | | |
| | 245 | 283 | | |

Solution

1. State the random variables and the parameters in words

x_1 = survival time from stomach cancer
 x_2 = survival time from bronchus cancer
 x_3 = survival time from colon cancer
 x_4 = survival time from ovarian cancer
 x_5 = survival time from breast cancer
 μ_1 = mean survival time from breast cancer
 μ_1 = mean survival time from bronchus cancer
 μ_3 = mean survival time from colon cancer
 μ_4 = mean survival time from ovarian cancer
 μ_5 = mean survival time from breast cancer

Now before conducting the hypothesis test, look at the means and standard deviations.

$\bar{x}_1 = 286$ $s_1 \approx 346.31$
 $\bar{x}_2 \approx 211.59$ $s_2 \approx 209.86$
 $\bar{x}_3 \approx 457.41$ $s_3 \approx 427.17$
 $\bar{x}_4 \approx 884.33$ $s_4 \approx 1098.58$
 $\bar{x}_5 \approx 1395.91$ $s_5 \approx 1238.97$

There appears to be a difference between at least two of the means, but realize that the standard deviations are very different. The difference you see may not be significant.

Notice the sample sizes are not the same. The sample sizes are

$n_1 = 13, n_2 = 17, n_3 = 17, n_4 = 6, n_5 = 11$

2. State the null and alternative hypotheses and the level of significance

$H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

H_A : at least two of the means are not equal

$\alpha = 0.01$

3. State and check the assumptions for the hypothesis test

- A random sample of 13 survival times from stomach cancer was taken. A random sample of 17 survival times from bronchus cancer was taken. A random sample of 17 survival times from colon cancer was taken. A random sample of 6 survival times from ovarian cancer was taken. A random sample of 11 survival times from breast cancer was taken. These statements may not be true. This information was not shared as to whether the samples were random or not but it may be safe to assume that.
- Since the individuals have different cancers, then the samples are independent.
- Population of all survival times from stomach cancer is normally distributed.
Population of all survival times from bronchus cancer is normally distributed.
Population of all survival times from colon cancer is normally distributed.
Population of all survival times from ovarian cancer is normally distributed.
Population of all survival times from breast cancer is normally distributed.
Looking at the histograms, box plots and normal quantile plots for each sample, it appears that none of the populations are normally distributed. The sample sizes are somewhat different for the problem. This assumption may not be true.
- The population variances are all equal. The sample standard deviations are approximately 346.3, 209.9, 427.2, 1098.6, and 1239.0 respectively. This assumption does not appear to be met, since the sample standard deviations are very different. The sample sizes are somewhat different for the problem. This assumption may not be true.

4. Find the test statistic and p-value

To find the test statistic and p-value using the TI-83/84, type each data set into L1 through L5. Then go into STAT and over to TESTS and choose ANOVA(. Then type in L1,L2,L3,L4,L5 and press enter. You will get the results of the ANOVA test.

| L1 | L2 | L3 | 1 |
|-----------------------|-----|------|---|
| 42 | 81 | 248 | |
| 25 | 461 | 377 | |
| 45 | 20 | 189 | |
| 412 | 450 | 1843 | |
| 51 | 246 | 180 | |
| 1112 | 166 | 537 | |
| | 63 | 519 | |
| L1()=124 | | | |
| ANOVA(L1,L2,L3,L4,L5) | | | |

Figure 11.3.1: Setup for ANOVA on TI-83/84

| |
|-----------------|
| One-way ANOVA |
| F=6.433436865 |
| P=2.2945316E-4 |
| Factor |
| df=4 |
| SS=11535760.5 |
| ↓ MS=2883940.13 |
| ■ |
| One-way ANOVA |
| ↑ MS=2883940.13 |
| Error |
| df=59 |
| SS=26448144.5 |
| MS=448273.635 |
| SxP=669.5324 |
| ■ |

Figure 11.3.2: Results of ANOVA on TI-83/84

The test statistic is $F \approx 6.433$ and p - value $\approx 2.29 \times 10^{-4}$

Just so you know, the Factor information is between the groups and the Error is within the groups. So

$MS_B \approx 2883940.13$, $SS_B \approx 11535760.5$, and $df_B = 4$ and

$MS_W \approx 448273.635$, $SS_W \approx 26448144.5$, and $df_W = 59$

To find the test statistic and p-value on R:

The commands would be:

variable=c(type in all data values with commas in between) – this is the response variable

factor=c(rep("factor 1", number of data values for factor 1), rep("factor 2", number of data values for factor 2), etc) – this separates the data into the different factors that the measurements were based on.

`data_name = data.frame(variable, factor)` – this puts the data into one variable. `data_name` is the name you give this variable
`aov(variable ~ factor, data = data_name)` – runs the ANOVA analysis

For this example, the commands would be:

```
time=c(124, 42, 25, 45, 412, 51, 1112, 46, 103, 876, 146, 340, 396, 81, 461, 20, 450, 246, 166, 63, 64, 155, 859, 151, 166, 37,
223, 138, 72, 245, 248, 377, 189, 1843, 180, 537, 519, 455, 406, 365, 942, 776, 372, 163, 101, 20, 283, 1234, 89, 201, 356,
2970, 456, 1235, 24, 1581, 1166, 40, 727, 3808, 791, 1804, 3460, 719)
```

```
factor=c(rep("Stomach", 13), rep("Bronchus", 17), rep("Colon", 17), rep("Ovary", 6), rep("Breast", 11))
```

```
survival=data.frame(time, factor)
```

```
results=aov(time~factor, data=survival)
```

```
summary(results)
```

```

              Df  Sum Sq  Mean Sq  F value    Pr(>F)
factor         4  11535761  2883940   6.4333 0.000229 ***
Residuals     59  26448144   448274
---

```

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

The test statistic is $F = 6.433$ and the p-value = 0.000229.

5. Conclusion

Reject H_0 since the p-value is less than 0.01.

6. Interpretation

There is evidence to show that at least two of the mean survival times from different cancers are not equal.

By examination of the means, it appears that the mean survival time for breast cancer is different from the mean survival times for both stomach and bronchus cancers. It may also be different for the mean survival time for colon cancer. The others may not be different enough to actually say for sure.

Homework

Exercise 11.3.1

In each problem show all steps of the hypothesis test. If some of the assumptions are not met, note that the results of the test may not be correct and then continue the process of the hypothesis test.

1. Cuckoo birds are in the habit of laying their eggs in other birds' nest. The other birds adopt and hatch the eggs. The lengths (in cm) of cuckoo birds' eggs in the other species nests were measured and are in Example 11.3.2("Cuckoo eggs in," 2013). Do the data show that the mean length of cuckoo bird's eggs is not all the same when put into different nests? Test at the 5% level.

Table 11.3.2: Lengths of Cuckoo Bird Eggs in Different Species Nests

| Meadow Pipit | | Tree Pipit | Hedge Sparrow | Robin | Pied Wagtail | Wren |
|--------------|-------|------------|---------------|-------|--------------|-------|
| 19.65 | 22.25 | 21.05 | 20.85 | 21.05 | 21.05 | 19.85 |
| 20.05 | 22.25 | 21.85 | 21.65 | 21.85 | 21.85 | 20.05 |
| 20.65 | 22.25 | 22.05 | 22.05 | 22.05 | 21.85 | 20.25 |
| 20.85 | 22.25 | 22.45 | 22.85 | 22.05 | 21.85 | 20.85 |
| 21.65 | 22.65 | 22.65 | 23.05 | 22.05 | 22.05 | 20.85 |
| 21.65 | 22.65 | 23.25 | 23.05 | 22.25 | 22.45 | 20.85 |
| 21.65 | 22.85 | 23.25 | 23.05 | 22.45 | 22.65 | 21.05 |
| 21.85 | 22.85 | 23.25 | 23.05 | 22.45 | 23.05 | 21.05 |

| Meadow Pipit | | Tree Pipit | Hedge Sparrow | Robin | Pied Wagtail | Wren |
|--------------|-------|------------|---------------|-------|--------------|-------|
| 21.85 | 22.85 | 23.45 | 23.45 | 22.65 | 23.05 | 21.05 |
| 21.85 | 22.85 | 23.45 | 23.85 | 23.05 | 23.25 | 21.25 |
| 22.05 | 23.05 | 23.65 | 23.85 | 23.05 | 23.45 | 21.45 |
| 22.05 | 23.25 | 23.85 | 23.85 | 23.05 | 24.05 | 22.05 |
| 22.05 | 23.25 | 24.05 | 24.05 | 23.05 | 24.05 | 22.05 |
| 22.05 | 23.45 | 24.05 | 25.05 | 23.05 | 24.05 | 22.05 |
| 22.05 | 23.65 | 24.05 | | 23.25 | 24.85 | 22.25 |
| 22.05 | 23.85 | | | 23.85 | | |
| 22.05 | 24.25 | | | | | |
| 22.05 | 24.45 | | | | | |
| 22.05 | 22.25 | | | | | |
| 22.05 | 22.25 | | | | | |
| 22.25 | 22.25 | | | | | |
| 22.25 | 22.25 | | | | | |
| 22.25 | | | | | | |

2. Levi-Strauss Co manufactures clothing. The quality control department measures weekly values of different suppliers for the percentage difference of waste between the layout on the computer and the actual waste when the clothing is made (called run-up). The data is in Example 11.3.3 and there are some negative values because sometimes the supplier is able to layout the pattern better than the computer ("Waste run up," 2013). Do the data show that there is a difference between some of the suppliers? Test at the 1% level.

Table 11.3.3: Run-ups for Different Plants Making Levi Strauss Clothing

| Plant 1 | Plant 2 | Plant 3 | Plant 4 | Plant 5 |
|---------|---------|---------|---------|---------|
| 1.2 | 16.4 | 12.1 | 11.5 | 24 |
| 10.1 | -6 | 9.7 | 10.2 | -3.7 |
| -2 | -11.6 | 7.4 | 3.8 | 8.2 |
| 1.5 | -1.3 | -2.1 | 8.3 | 9.2 |
| -3 | 4 | 10.1 | 6.6 | -9.3 |
| -0.7 | 17 | 4.7 | 10.2 | 8 |
| 3.2 | 3.8 | 4.6 | 8.8 | 15.8 |
| 2.7 | 4.3 | 3.9 | 2.7 | 22.3 |
| -3.2 | 10.4 | 3.6 | 5.1 | 3.1 |
| -1.7 | 4.2 | 9.6 | 11.2 | 16.8 |
| 2.4 | 8.5 | 9.8 | 5.9 | 11.3 |
| 0.3 | 6.3 | 6.5 | 13 | 12.3 |

| Plant 1 | Plant 2 | Plant 3 | Plant 4 | Plant 5 |
|---------|---------|---------|---------|---------|
| 3.5 | 9 | 5.7 | 6.8 | 16.9 |
| -0.8 | 7.1 | 5.1 | 14.5 | |
| 19.4 | 4.3 | 3.4 | 5.2 | |
| 2.8 | 19.7 | -0.8 | 7.3 | |
| 13 | 3 | -3.9 | 7.1 | |
| 42.7 | 7.6 | 0.9 | 3.4 | |
| 1.4 | 70.2 | 1.5 | 0.7 | |
| 3 | 8.5 | | | |
| 2.4 | 6 | | | |
| 1.3 | 2.9 | | | |

3. Several magazines were grouped into three categories based on what level of education of their readers the magazines are geared towards: high, medium, or low level. Then random samples of the magazines were selected to determine the number of three-plus-syllable words were in the advertising copy, and the data is in Example 11.3.4("Magazine ads readability," 2013). Is there enough evidence to show that the mean number of three-plus-syllable words in advertising copy is different for at least two of the education levels? Test at the 5% level.

Table 11.3.4: Number of Three Plus Syllable Words in Advertising Copy

| High Education | Medium Education | Low Education |
|----------------|------------------|---------------|
| 34 | 13 | 7 |
| 21 | 22 | 7 |
| 37 | 25 | 7 |
| 31 | 3 | 7 |
| 10 | 5 | 7 |
| 24 | 2 | 7 |
| 39 | 9 | 8 |
| 10 | 3 | 8 |
| 17 | 0 | 8 |
| 18 | 4 | 8 |
| 32 | 29 | 8 |
| 17 | 26 | 8 |
| 3 | 5 | 9 |
| 10 | 5 | 9 |
| 6 | 24 | 9 |
| 5 | 15 | 9 |
| 6 | 3 | 9 |

| High Education | Medium Education | Low Education |
|----------------|------------------|---------------|
| 6 | 8 | 9 |

4. A study was undertaken to see how accurate food labeling for calories on food that is considered reduced calorie. The group measured the amount of calories for each item of food and then found the percent difference between measured and labeled food, $\frac{(\text{measured} - \text{labeled})}{\text{labeled}} * 100\%$. The group also looked at food that was nationally advertised, regionally distributed, or locally prepared. The data is in Example 11.3.5("Calories datafile," 2013). Do the data indicate that at least two of the mean percent differences between the three groups are different? Test at the 10% level.

Table 11.3.5: Percent Differences Between Measured and Labeled Food

| National Advertised | Regionally Advertised | Locally Prepared |
|---------------------|-----------------------|------------------|
| 2 | 41 | 15 |
| -28 | 46 | 60 |
| -6 | 2 | 250 |
| 8 | 25 | 145 |
| 6 | 39 | 6 |
| -1 | 16.5 | 8- |
| 1- | 17 | 95 |
| 13 | 28 | 3 |
| 15 | -3 | |
| -4 | 14 | |
| -4 | 34 | |
| -18 | 42 | |
| 10 | | |
| 5 | | |
| 3 | | |
| -7 | | |
| 3 | | |
| -0.5 | | |
| -10 | | |
| 6 | | |

5. The amount of sodium (in mg) in different types of hotdogs is in Example 11.3.6("Hot dogs story," 2013). Is there sufficient evidence to show that the mean amount of sodium in the types of hotdogs are not all equal? Test at the 5% level.

Table 11.3.6: Amount of Sodium (in mg) in Beef, Meat, and Poultry Hotdogs

| Beef | Meat | Poultry |
|------|------|---------|
| 495 | 458 | 430 |

| Beef | Meat | Poultry |
|------|------|---------|
| 477 | 506 | 375 |
| 425 | 473 | 396 |
| 322 | 545 | 383 |
| 482 | 496 | 387 |
| 587 | 360 | 542 |
| 370 | 387 | 359 |
| 322 | 386 | 357 |
| 479 | 507 | 528 |
| 375 | 393 | 513 |
| 330 | 405 | 426 |
| 300 | 372 | 513 |
| 386 | 144 | 358 |
| 401 | 511 | 581 |
| 645 | 405 | 588 |
| 440 | 428 | 522 |
| 317 | 339 | 545 |
| 319 | | |
| 298 | | |
| 253 | | |

Answer

For all hypothesis tests, just the conclusion is given. See solutions for the entire answer.

1. Reject H_0
3. Reject H_0
5. Fail to reject H_0

Data Source:

Aboriginal deaths in custody. (2013, September 26). Retrieved from <http://www.statsci.org/data/oz/custody.html>

Activities of dolphin groups. (2013, September 26). Retrieved from <http://www.statsci.org/data/general/dolpacti.html>

Boyle, P., Flowerdew, R., & Williams, A. (1997). Evaluating the goodness of fit in models of sparse medical data: A simulation approach. *International Journal of Epidemiology*, 26(3), 651-656. Retrieved from <http://ije.oxfordjournals.org/conten...3/651.full.pdf> html

Calories datafile. (2013, December 07). Retrieved from lib.stat.cmu.edu/DASL/Datafiles/Calories.html

Cancer survival story. (2013, December 04). Retrieved from lib.stat.cmu.edu/DASL/Stories...rSurvival.html

Car preferences. (2013, September 26). Retrieved from <http://www.statsci.org/data/oz/carprefs.html>

Cuckoo eggs in nest of other birds. (2013, December 04). Retrieved from lib.stat.cmu.edu/DASL/Stories/cuckoo.html

Education by age datafile. (2013, December 05). Retrieved from lib.stat.cmu.edu/DASL/Datafil...tionbyage.html

Encyclopedia Titanica. (2013, November 09). Retrieved from www.encyclopediatitanica.org/

Global health observatory data respository. (2013, October 09). Retrieved from http://apps.who.int/gho/athena/data/...t=GHO/MORT_400&profile=excel&filter=AGEGROUP:YEARS05-14;AGEGROUP:YEARS15-29;AGEGROUP:YEARS30-49;AGEGROUP:YEARS50-69;AGEGROUP:YEARS70;MGHEREG:REG6_AFR;GHECAUSES:*;SEX:*

Hot dogs story. (2013, November 16). Retrieved from lib.stat.cmu.edu/DASL/Stories/Hotdogs.html

Leprosy: Number of reported cases by country. (2013, September 04). Retrieved from <http://apps.who.int/gho/data/node.main.A1639>

Magazine ads readability. (2013, December 04). Retrieved from lib.stat.cmu.edu/DASL/Datafiles/magadsdat.html

Popular kids datafile. (2013, December 05). Retrieved from lib.stat.cmu.edu/DASL/Datafil...pularKids.html

Schultz, S. T., Klonoff-Cohen, H. S., Wingard, D. L., Askhoomoff, N. A., Macera, C. A., Ji, M., & Bacher, C. (2006). Breastfeeding, infant formula supplementation, and autistic disorder: the results of a parent survey. *International Breastfeeding Journal*, 1(16), doi: 10.1186/1746-4358-1-16

Waste run up. (2013, December 04). Retrieved from lib.stat.cmu.edu/DASL/Stories/wasterunup.html

This page titled [11.3: Analysis of Variance \(ANOVA\)](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Kathryn Kozak](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- **11.3: Analysis of Variance (ANOVA)** by [Kathryn Kozak](#) is licensed [CC BY-SA 4.0](#). Original source: <https://s3-us-west-2.amazonaws.com/oerfiles/statsusingtech2.pdf>.