## Class 22: Analysis of Variance (Text: Sections 12.1)

**The F-test: Tests whether the means of two or more populations are equal.**
Now we do examples with more than two samples, which cannot be done by the *T*-test.

> **Null hypothesis:** The means of all the populations are equal.
> **Alternative hypothesis:** The means of all the populations are not all equal. Thus at least one population has a different mean.

**Assumptions Underlying the F-Test**
For the $F$-statistic to have the $F$- distribution, we have to assume

- Each sample comes from a normal population
- The population standard deviations are all equal

The means of the normal distributions do *not* have to all be equal—that's what we want to decide.

Fortunately, the $F$-Test is not very sensitive to unequal standard deviations,

- We can use the $F$-Test if the largest sample standard deviation is less than twice the smallest one.
- In other words, the largest variance should be no more than 4 times the smallest one

**Pooled Variance**
Since we assume all variances are equal, let $s_p^2$ represent the variance of all the samples pooled together. The estimate for this pooled variance is

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \cdots + (n_k-1)s_k^2}{(n_1-1)+(n_2-1)+\cdots+(n_k-1)}.$$

The common population standard deviation is $\sqrt{s_p^2}$.

**Relationship Between $s_p^2$ and SSE and MSE**
For each sample, we have

$$s_i^2 = \frac{1}{n_i-1}\sum_{Group\ i}(x_{ij}-\bar{x}_i)^2 \qquad \text{so} \qquad (n_i-1)s_i^2 = \sum_{Group\ i}(x_{ij}-\bar{x}_i)^2 .$$

Thus

$$SSE = \sum_{Groups}\sum_{j}(x_{ij}-\bar{x}_i)^2 = \sum_{i}(n_i-1)s_i^2.$$

Now $(n_1-1)+(n_2-1)+\cdots+(n_k-1) = N-k =$ DFE, so we have

$$\boxed{MSE = \frac{SSE}{N-k} = s_p^2}$$

**Example**: **The "fog index" measures the difficulty in a passage or writing.[1] A sample of passages from three magazines gave the following results.**

| | | | | | | |
|---|---|---|---|---|---|---|
| *Scientific American* | 15.75 | 11.55 | 11.16 | 9.92 | 9.23 | 8.20 |
| *Newsweek* | 10.21 | 9.66 | 7.67 | 5.12 | 4.88 | 3.12 |
| *Sports Illustrated* | 9.17 | 8.44 | 6.10 | 5.78 | 5.58 | 5.36 |

**Ex: Complete the table below:**

Anova: Single Factor

SUMMARY

| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
|---|---|---|---|---|
| Scientific American | 6 | 65.81 | 10.96833 | 7.004777 |
| Newsweek | 6 | 40.66 | 6.776667 | 8.122507 |
| Sports Illustrated | 6 | 40.43 | 6.738333 | 2.675217 |

ANOVA

| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
|---|---|---|---|---|---|---|
| Between Groups | 70.92888 | 2 | 35.46444 | 5.976313 | 0.012336 | 3.68232 |
| Within Groups | 89.0125 | 15 | 5.934167 | | | |
| Total | 159.9414 | 17 | | | | |

There are 3 groups, with 6 observations in each, so there are 18 observations in all.
DF groups $= 3 - 1 = 2$ and DF total $= 18 - 1 = 17$ and DF within $= 17 - 2 = 15$.
The mean squares are given by $MSG = {}^{70.9288}/_{2} = 35.46$ and $MSE = {}^{89.0125}/_{15} = 5.93$.
Then $F = {}^{35.46}/_{5.93} = 5.98$.
Since DF numerator $= 2$ and DF denominator $= 15$,
$P$-value $=$ Fcdf(5.98,1000,2,15) $= 0.0123$
The tables shows more exact values calculated by Excel.

**Ex: Does the data meet the conditions for the F-Test?**
Yes, since the largest SD $= \sqrt{8.1} = 2.8$ and the smallest SD $= \sqrt{2.7} = 1.6$ and $2 \cdot 1.6 = 3.8$.
**Ex: Is there a significant difference in fog indices between these magazines?**
**Step 1**: Null hypothesis: All mean fog indices are the same.
Alternate hypothesis: Not all mean fog indices are equal.
**Step 2**: The test statistic is $F = 5.98$ with DF numerator $= 2$ and DF denominator $= 15$.
**Step 3**: From the table: $P$-value is between 0.025 and 0.010; that is between 1% and 2.5%.
From the calculator: $P$-value $=$ Fcdf(5.9763,1000,2,15) $= 0.012336 = 1.2336\%$.
**Step 4**: We reject the null hypothesis. One or more of the magazines has a different mean fog index.

**Ex: How do we tell which magazine(s) have different fog indices?**
ANOVA cannot tell us this; we look back at the original data and at the summary statistics in the Excel output. *Newsweek* and *Sports Illustrated* have very similar means; *Scientific American* is higher. To find out if the differences are significant, we can do pair wise $T$-tests.

[1] From Shruptine and McVicar, reported by Wilde and Seber in *Chance Encounters* .The fog index is defined by
Fog Index $= 0.4 \cdot$ Average #words per sentence $+$ %words with $>$ 3syllables.

**Ex: Does the type of cooking pot affect the iron content of food?**[2]
Iron deficiency leads to anemia. In developing countries, iron has traditionally got into the food from iron cooking pots. But as heavy iron pots are replaced by lighter, cheaper aluminum pots, there is a concern that anemia and malnutrition may result. Use the data to decide if there is a significant relationship between type of pot and iron content in food.

| | Iron Content (mg per 100 gm of food) | | | |
|---|---|---|---|---|
| Aluminum | 1.77 | 2.36 | 1.96 | 2.14 |
| Clay | 2.27 | 1.28 | 2.48 | 2.68 |
| Iron | 5.27 | 5.17 | 4.06 | 4.22 |

Anova: Single Factor

SUMMARY

| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
|---|---|---|---|---|
| Aluminum | 4 | 8.23 | 2.0575 | 0.063492 |
| Clay | 4 | 8.71 | 2.1775 | 0.386025 |
| Iron | 4 | 18.72 | 4.68 | 0.394733 |

ANOVA

| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
|---|---|---|---|---|---|---|
| Between Groups | 17.53922 | 2 | 8.769608 | 31.16236 | 9.01E-05 | 4.256495 |
| Within Groups | 2.53275 | 9 | 0.281417 | | | |
| Total | 20.07197 | 11 | | | | |

**Ex: What do the summary statistics suggest?** That iron pots are better than either clay or aluminum

**Ex: Are the conditions for the F-Test met?**
Largest variance is 0.4 and smallest is 0.06, so conditions not well met.

**Ex: Perform the hypothesis test. What can you conclude?**

**Step 1**: Null hypothesis: All mean iron contents are the same.
Alternate hypothesis: Not all mean iron contents are equal.
**Step 2**: The test statistic is $F = 31.2$ with DF numerator $= 2$ and DF denominator $= 9$.
**Step 3**: From the table: $P$-value is less than $0.001 = 0.1\%$.
From the calculator: $P$-value $=$ Fcdf$(31.2,1000,2,9) = 0.00009 = 0.009\%$.
**Step 4**: We reject the null hypothesis. One or more of the pots has a different mean iron level.
The test does not tell us the iron content of food is higher in the iron pots, but the original data and the summary statistics support this conclusion.

[2] Text, Problem 12.49, reported from A. A. Adish, "Effect of consumption of food cooked in iron pots on iron status and growth of young children: a randomized trial", *The Lancet* (1999).

**<u>Ex</u>: If we combine clay and aluminum, does the food cooked in iron pots have significantly higher iron content?**

We can do a 2-sample T-test or ANOVA with two groups.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Iron | 5.27 | 5.17 | 4.06 | 4.22 | | | | |
| Aluminum and Clay | 1.77 | 2.36 | 1.96 | 2.14 | 2.27 | 1.28 | 2.48 | 2.68 |

Anova: Single Factor

SUMMARY

| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
|---|---|---|---|---|
| Iron | 4 | 18.72 | 4.68 | 0.394733 |
| Aluminum and Clay | 8 | 16.94 | 2.1175 | 0.196764 |

ANOVA

| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
|---|---|---|---|---|---|---|
| Between Groups | 17.51042 | 1 | 17.51042 | 68.35868 | 8.81E-06 | 4.964603 |
| Within Groups | 2.56155 | 10 | 0.256155 | | | |
| Total | 20.07197 | 11 | | | | |

From the $P$-value, we see that there *is* a significant difference between the iron content of the food cooked in iron pots.

**Ex: Does bread loose vitamins when it is stored?**[3]

| | Days 0 | Days 1 | Days 3 | Days 5 | Days 7 |
|---|---|---|---|---|---|
| Vitamin C | 47.62 | 40.45 | 21.25 | 13.18 | 8.51 |
| | 49.79 | 43.46 | 22.34 | 11.65 | 8.13 |

Anova: Single Factor

SUMMARY

| Groups | Count | Sum | Average | Variance |
|---|---|---|---|---|
| Days 0 | 2 | 97.41 | 48.705 | 2.35445 |
| Days 1 | 2 | 83.91 | 41.955 | 4.53005 |
| Days 3 | 2 | 43.59 | 21.795 | 0.59405 |
| Days 5 | 2 | 24.83 | 12.415 | 1.17045 |
| Days 7 | 2 | 16.64 | 8.32 | 0.0722 |

ANOVA

| Source of Variation | SS | df | MS | F | P-value | F crit |
|---|---|---|---|---|---|---|
| Between Groups | 2565.721 | 4 | 641.4302 | 367.742 | 2.33E-06 | 5.192168 |
| Within Groups | 8.7 | 5 | 1.74 | | | |
| Total | 2574.442 | 9 | | | | |

There are 5 groups, with two observations each, so $k = 5$ and $N = 10$. The average Vitamin C certainly seems to be going down. However, variances suggest that the population SDs are not equal.

To fill in the table, we will need to find

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_5 - 1)s_5^2}{(n_1 - 1) + (n_2 - 1) + \cdots + (n_5 - 1)}$$
$$= \frac{(2-1)2.35 + (2-1)4.53 + (2-1)0.59 + (2-1)1.17 + (2-1)0.07}{(2-1) + (2-1) + (2-1) + (2-1) + (2-1)} = \frac{8.7}{5}$$
$$= 1.74 = MSE$$

Thus

$$SSE = 5 \cdot MSE = 5 \cdot 1.74 = 8.7$$

Now we find $SST = 2565.72 + 8.7 = 2574.42$ and $MSG = 2565.72/4 = 641.43$.
Thus we have $F = 641.43/1.74 = 368$.
**Step 1**: Null hypothesis: Vitamin C content does not change.
Alternate hypothesis: Vitamin C content does change.
**Step 2**: The test statistic is $F = 368$ with DF numerator $= 4$ and DF denominator $= 5$.
**Step 3**: The table: The F-value is off the charts (too big), so the $P$-value is approximately 0.
From the calculator: $P$-value $=$ Fcdf(367.7,1000,4,5) $= 0.000002$, so very small.
**Step 4**: We reject the null hypothesis. Vitamin C content does change.

[3] Text, Problem 12.29, reported from H. Park et al "Fortifying bread with each of three antioxidants, *Cereal Chemistry* (1997).