Contents

1 ANOVA

1.1 What is it?

ANOVA stands for ANalysis Of VAriance.
It is a collection of tools for comparing the means of more than two populations. We have already seen how to compare two means based on the information from two independent samples using a 2-sample t-test. ANOVA will allow us to simultaneously compare more than two means.

Example: A school board is interested in comparing test scores on a standardized reading test for fourth–grade students in their district. They select random samples of children from each of the elementary schools, and use the information to test if the mean score is the same across all elementary schools. They want to include confounding variable gender, and are interested in the effect of the Social Economic Status (SES) on the mean test scores. A rejection of the hypothesis that all means are equal implies that at least two means differ.
We call reading score the *response variable* and school (identifying which school a student attends), gender, and SES the *factors*. The different schools included with the analysis are the levels of the variable school, the *factor levels*.
What are the levels of gender and SES?

Definition:

- A *factor variable* is an independent variable whose values indicate the sub-populations of interest.
- A *factor level* is a specific value chosen for the factor.
- The *response variable* is the variable being measured.

Example (Holding Times – adjusted from Agresti & Franklin(2007) *Statistics*)
The university conducted an experiment to find out if callers to the technology help desk remain longer on hold, on average, if they hear (1) some random information about the university and technology, (2) Muzak, or (3) a recording by the university Jazz Orchestra.
They randomly selected one out of every 20 calls in a particular week and measured the number of minutes a caller stayed on the line before hanging up (the selected calls were never answered). In order not annoy too many clients from the university community the sample sizes was kept low.
response variable = time callers wait in line
factor = message played
factor levels = random information, Muzak, Jazz

The data

Information	Muzak	Jazz
5	0	13
1	1	9
11	4	8
2	6	15
8	3	7

Do the data provide sufficient evidence that the mean waiting time is different for the three messages?
ANOVA will help answering this question.

1.2 One-way ANOVA

One-way ANOVA allows only one factor in the analysis. In principle, it is possible to include more than one factor in an analysis of variance, allowing to assess the effect of several treatment variables and to account for confounding variables. This will be studied in upper level courses.

But why is it called Analysis of Variance, if it is about means?
We will see how different variances can be used to indicate if the means are different.

1.2.1 Types of Variances

Assume we want to test if k population means are the same. For each (sub)population i we take a random sample of size n_i, and calculate the sample mean, $\bar{x}_i$, and the sample standard deviation, s_i.

The statistic used in ANOVA compares the variance **between** the k sample means with the variance **within** the k samples.
The statistics is based on the fact that the total variance in the sample can be decomposed (analyzed) into the variance between and the variance within.

$$\underbrace{\sum_{i,j}(x_{ij}-\bar{x})^2}_{SST} = \underbrace{\sum_{i} n_i(\bar{x}_i-\bar{x})^2}_{SSTR} + \underbrace{\sum_{i,j}(x_{ij}-\bar{x}_i)^2}_{SSE}$$

where
SST is the total Sum of Squares
$SSTR$ is the Sum of Squares of Treatment (between), and
SSE is the Sum of Squares of Error (within)
and the one-way ANOVA identity is always true:

$$SST = SSTR + SSE$$

Let's visualize:

Assume we have data from two different populations a and b:

```
Set A
       a a a a a                                      b b b b b
----------------------------------------------------------------------


Set B
        a b           a b         a b           a b         a b
----------------------------------------------------------------------
```

The **total variance** in the two data sets is almost the same, but for set A the **variability within** is much smaller than in set B, and the **variability between** is much smaller in Set B than in Set A.

Set A
Since the **variance within** in relationship to the **total variance** is relatively small, the total variance can only be explained by the **variance between**, so that one would conclude that the means of the groups must be different.

SET B
The **variance within** is close to the total variance, so that the total variance is explained by this variance. We conclude that the **variance between** the groups is relatively small, the group means are not different.

Computing Formulas

Applying the formulas above for computation is very tedious. The following table provides shortcut formulas

Sum of Squares	Defining Formula	Shortcut Formula
Total, SST	$\sum_{i,j}(x_{ij} - \bar{x})^2$	$\sum_{i,j} x_{ij}^2 - (\sum_{i,j} x_{ij})^2/n$
Treatment, SSTR	$\sum_i n_i(\bar{x}_i - \bar{x})^2$	$\sum_i n_i\bar{x}_i^2 - (\sum_{i,j} x_{ij})^2/n$
Error, SSE	$\sum_{i,j}(x_{ij} - \bar{x}_i)^2 = \sum_i(n_i - 1)s_i^2$	$SST - SSTR$

with the following notations:

- n is the total sample size
- x_{ij} is the j-th measurement from sample i
- $\bar{x}$ is the overall sample mean, mean of all n observations
- n_i is the sample size of sample i
- $\bar{x}_i$ is the sample mean of sample i
- s_i is the sample standard deviation of sample i

The sum of squares are standardized by division by their degrees of freedom.
They are:
the total degree of freedom are $df_T = n - 1$
the degrees of freedom for treatments are $df_{TR} = k - 1$
the degree of freedom for error are $df_E = n - k$

It is easy to confirm that for the degrees of freedon we have a similar identity as for the Sum of Squares

$$df_T = df_{TR} + df_E$$

Continue Example (Holding Times):

Information	Muzak	Jazz
5	0	13
1	1	9
11	4	8
2	6	15
8	3	7
$n_I = 5$	$n_M = 5$	$n_J = 5$
$\bar{x}_I = 5.4$	$\bar{x}_M = 2.8$	$\bar{x}_J = 10.4$
$s_I = 4.2$	$s_M = 2.4$	$s_J = 3.4$

First calculate:
$k = 3$
$n = 5 + 5 + 5 = 15$
$\sum_{i,j} x_{ij} = 5 + 1 + 11 + 2 + 8 + 0 + 1 + 4 + 6 + 3 + 13 + 9 + 8 + 15 + 7 = 93$
$\sum_{i,j} x_{ij}^2 = 25 + 1 + 121 + 4 + 64 + 0 + 1 + 16 + 36 + 9 + 169 + 81 + 64 + 225 + 49 = 865$

Using these values:

$$\begin{aligned}
SST &= \sum_{i,j} x_{ij}^2 - \frac{(\sum_{i,j} x_{ij})^2}{n} = 865 - \frac{93^2}{15} = 288.4 \\
SSTR &= \sum_{i} n_i \bar{x}_i^2 - \frac{(\sum_{i,j} x_{ij})^2}{n} = 5(5.4)^2 + 5(2.8)^2 + 5(10.4)^- \frac{93^2}{15} \\
&= 725.8 - 576.6 = 149.2 \\
SSE &= SST - SSTR = 139.2 \\
df_T &= n - 1 = 14 \\
df_{TR} &= k - 1 = 2 \\
df_E &= n - k = 15 - 3 = 12
\end{aligned}$$

1.2.2 One-way ANOVA Table

All these quantities are organized in the one-way ANOVA table and results on mean squares and F-statistic are added.
Generally, means squares are calculated from the sum of squares and the degrees of freedom, $MS = SS/df$, and the F-statistic is calculated as $F = MS/MSE$

one-way ANOVA Table

Source	df	SS	$MS = SS/df$	F-statistic
Treatments	$k-1$	$SSTR$	$MSTR = SSTR/(k-1)$	$F = MSTR/MSE$
Error	$n-k$	SSE	$MSE = SSE/(n-k)$	
Total	$n-1$	SST		

Continue Example (Holding Times):

one-way ANOVA Table

Source	df	SS	$MS = SS/df$	F-statistic
Treatments	2	149.2	149.2/2=74.6	74.6/11.6=11.90
Error	12	139.2	139.2/12=11.6	
Total	14	288.4		

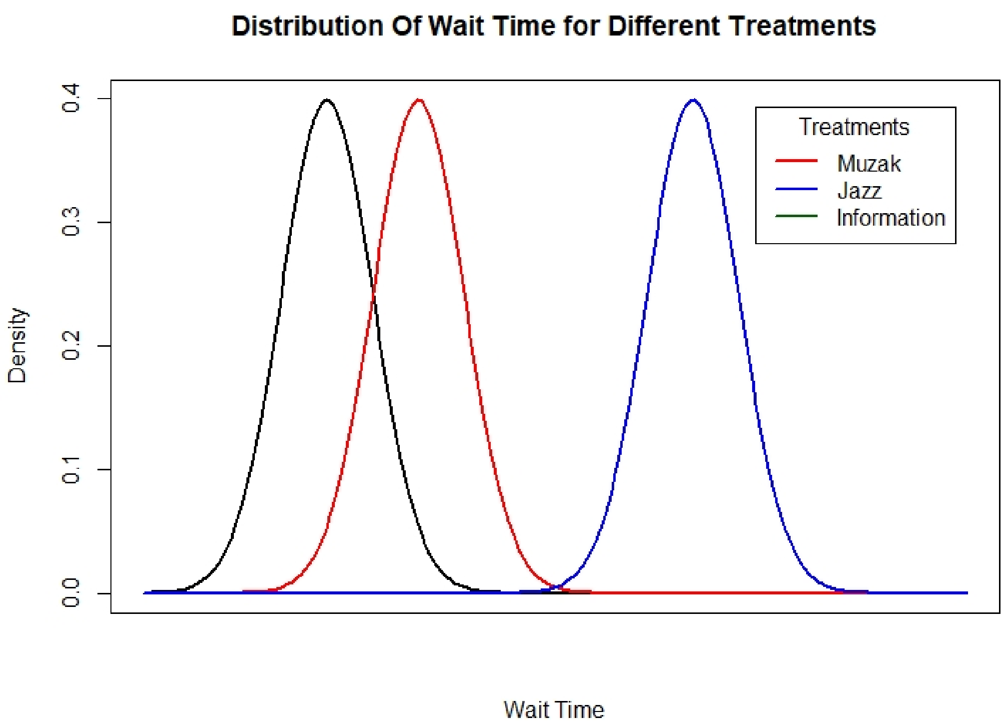
How can this information be used to decide if there is evidence from the sample that the mean holding times are not the same for the three conditions (information, muzak, jazz)?

1.2.3 Testing the Equality of Treatment Means

Assumptions for an ANOVA:
For the argument of the F-test to work we have to assume that

- the response variable is normally distributed within every (sub)population (treatment, group)
- the standard deviation of the response variable is the same across the (sub)populations (treatments, groups), σ^2

Continue Example:
We have to assume that the waiting times for treatments "information", "muzak", and "jazz" are normally distributed with the same standard deviation σ.



The graph implies that the means are different, but this we still have to test.

The hypotheses of interest are

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k \quad \text{versus} \quad H_a : \text{at least one of the means differs from the others}$$

Use the following argument for developing the test:

- The variances in the k populations are all the same σ^2. The statistic

$$MSE = \frac{SSE}{n-k}$$

  is an estimate of σ^2, whether or not H_0 is true.

- **If H_0 is true**, i.e. $\mu_1 = \mu_2 = \ldots = \mu_k$, then the variation between the sample means, measured by
$$MSTR = \frac{SSTR}{k-1}$$
also provides an unbiased estimate of σ^2.

  However, if H_0 is false and the population means are not the same, then $MSTR$ is larger than σ^2.

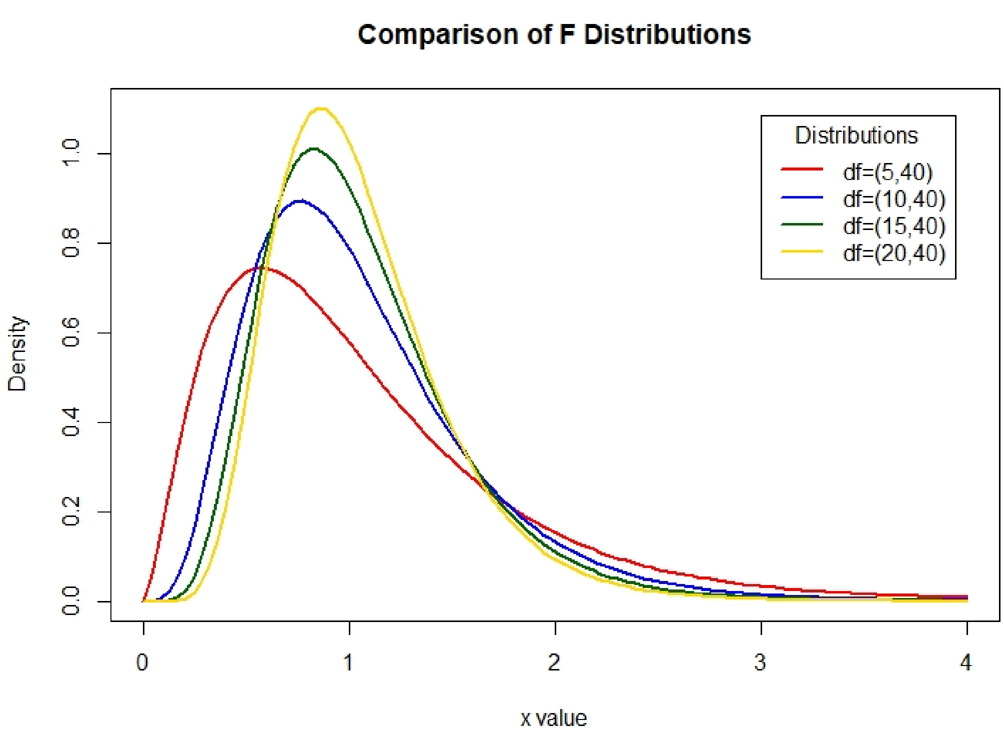
- The test statistic
$$F = \frac{MSTR}{MSE}$$
tends to be much larger than 1, if H_0 is false. Hence, H_0 should be rejected for large values of F.

  What values of F have to be considered large, we learn from the distribution of F.

- When H_0 is true, the statistic
$$F = \frac{MSTR}{MSE}$$
has an F distribution with $df_1 = (k-1)$ and $df_2 = (n-k)$ degrees of freedom.

  Upper tailed critical values of the F distribution can be found in Table E.

The F-distribution
The F-distribution has two parameters, degrees of freedom for the numerator, df_1, and degrees of freedom for the denominator, df_2



The ANOVA F-Test

1. **Hypotheses**:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k \quad \text{versus} \quad H_a : \text{at least one of the means differs from the others}$$

   Choose α

2. **Assumption**: Random samples from each population/treatment/group. The response variable follows a normal distribution with means $\mu_1, \mu_2, \ldots, \mu_k$ in the k populations and equal variances, σ^2.

3. **Test statistic**:

$$F_0 = \frac{MSTR}{MSE}$$

   based on $df_1 = df_{TR} = (k-1)$ and $df_2 = df_E = (n-k)$.

4. **P-value**: $P(F > F_0)$, where F follows an F-distribution with $df_1 = (k-1)$ and $df_2 = (n-k)$.

5. **Conclusion**: Is the $P - value < \alpha$, reject H_0

6. **Context**:

Continue Example (Holding Times):

1. **Hypotheses**:

$$H_0 : \mu_I = \mu_M = \mu_J \quad \text{versus} \quad H_a : \text{at least one of the means differs from the others}$$

   Choose $\alpha = 0.05$

2. **Assumption**: Participants were randomly chosen and assigned to a holding message. We will assume that the assumptions of normality and equal standard deviations are met.

3. **Test statistic**:

$$F_0 = \frac{MSTR}{MSE} = 11.9$$

   based on $df_1 = 2$ and $df_2 = 12$.

4. **P-value**: $P - value = P(F > 11.9)$. Use the F-distribution table, and find that for $df_1 = 2$, and $df_2 = 12$ $P - value = P(F > 11.9) < 0.005$

5. **Conclusion**: Since $P - value < 0.05 = \alpha$, reject H_0

6. **Context**: At significance level of 5% the data provide sufficient evidence that the mean holding times for callers either listening to some information, Muzak, or Jazz are not the same.

Example:
In a effort to improve the quality of recording tapes, the effects of four kinds of coatings A, B, C, D on the reproducing quality of sound are compared.
The following values on distortion are obtained:

Coating	Observations	Sample Sizes
A	10, 15, 8, 12, 15	5
B	14, 18, 21, 15	4
C	17, 16, 14, 15, 17, 15, 18	7
D	12, 15, 17, 15, 16, 15	6

With the help of such a sample we want to decide if the four different coatings result in different mean distortions. Let
μ_A= mean distortion of tapes with coating A,
μ_B= mean distortion of tapes with coating B,
μ_C= mean distortion of tapes with coating C,
μ_D= mean distortion of tapes with coating D,

Check (meaning test) if $\mu_A = \mu_B = \mu_C = \mu_D$.
To create the ANOVA table compute:
We get $k = 4$, $n_1 = 5, n_2 = 4, n_3 = 7, n_4 = 6$, $n = n_1 + n_2 + n_3 + n_4 = 22$.
It is $\bar{x}_1 = 12$, $\bar{x}_2 = 17$, $\bar{x}_3 = 16$, $\bar{x}_4 = 15$.
$\sum_{i,j} x_{ij} = 182$
$\sum_{i,j} x_{ij}^2 = 2338$
From this we find:

$$\begin{aligned} SST &= 162 \\ SSTR &= 68 \\ SSE &= Total\ SS - SST = 162 - 68 = 94 \end{aligned}$$

The ANOVA table for the tape coating data:

Source	*df*	SS	MS	F
Coating	3	68	$MSTR = 22.67$	4.343
Error	18	94	$MSE = 5.22$	
Total	21	162		

Continue tape coating example:
The data resulted in the following ANOVA table:

Source	*df*	SS	MS	F
Coating	3	68	$MST = 22.67$	4.343
Error	18	94	$MSE = 5.22$	
Total	21	162		

Conduct an ANOVA F-test:

1. Hypotheses:

   $$H_0 : \mu_A = \mu_B = \mu_C = \mu_D \quad \text{versus} \quad H_a : \text{at least one of the means differs from the others}$$

   choose $\alpha = 0.05$

2. Assumptions: assume to be met

3. Test statistic:
   $$F = 4.343 \quad \text{with} \quad df_1 = 3, df_2 = 18$$

4. P-value: Use $df_1 = 3$ and $df_2 = 18$, in the F-distribution table and find that the uppertail area for 3.95 equals 0.025 and for 5.09 equals 0.010, so that $0.010 <$ P-value< 0.025.

5. Decision: The P-value$< 0.025 < 0.05 = \alpha$, we reject H_0.
   At significance level 0.05 the data provide sufficient evidence that the mean distortion for the four coatings are not the same.

One more Example: This study was designed to investigate the impact of nutrition on the attention span (in minutes) of elementary school students. A group of 15 students was randomly assigned to one of three groups. The first group will get no breakfast, the second a light breakfast and the third a full breakfast.
This experiment contains one factor which is the meal with factor levels no breakfast, light breakfast, and full breakfast. The response variable is the attention span.

μ_1 is the mean attention span of an elementary school student after no breakfast.
μ_2 is the mean attention span of an elementary school student after a light breakfast.
μ_3 is the mean attention span of an elementary school student after a full breakfast.

The following numbers were reported:

no breakfast	light breakfast	full breakfast
8	14	10
7	16	12
9	12	16
13	17	15
10	11	12

Give the ANOVA table for these data:

We get $k = 3$, $n_1 = n_2 = n_3 = 5$, $n = n_1 + n_2 + n_3 = 15$.
It is $\bar{x}_1 = 9.4$, $\bar{x}_2 = 14$, $\bar{x}_3 = 13$, so that $\sum_{i,j} x_{ij} = 182$, $\sum_{i,j} x_{ij}^2 = 2338$.
From this we find:

$$\begin{aligned} SST &= 129.7333 \\ SSTR &= 58.5333 \\ SSE &= Total\ SS - SST = 129.7333 - 58.5333 = 71.2 \end{aligned}$$

This results are summarized in the following ANOVA table:

Source	df	SS	MS	F
Meal	2	58.53	$MSTR = 29.27$	4.93
Error	12	71.2	5.93	
Total	14	129.7333		

The data resulted in the following ANOVA table:

Source	df	SS	MS	F
Meal	2	58.53	$MST = 29.27$	4.93
Error	12	71.2	5.93	
Total	14	129.7333		

1. Hypotheses: The hypotheses of interest to test at a level of $\alpha = 0.05$ are

$$H_0 : \mu_1 = \mu_2 = \mu_3 \text{ versus } H_a : \text{at least one of the means differs from the others}$$

2. Assumptions: assume to be met

3. Test statistic:

$$F_0 = \frac{MST}{MSE} = 4.93$$

   based on $df_1 = 2$ and $df_2 = 12$.

4. Rejection Region: $F_0 > F_\alpha$, where $F_{0.05} = 3.89$. Since the value of the test statistic exceeds 3.89, H_0 is rejected

5. At significance level of 5% we find enough evidence to assume that the mean attention span is not the same for the three types of breakfast.