American University of Beirut

STAT 210: Elementary Statistics for Sciences

2022–2023 Fall

Siamak Taati

Chapter 12

Analysis of variance

One-factor analysis of variance

One-factor analysis of variance

Analysis of variance (ANOVA for short) is an approach to testing the influence of factors (i.e., categorical variables) on a numerical variable.

For instance, we may want to test:

- Whether the BMI (body mass index) of adults is associated with their ethnicity.
- Whether there is a difference in income among three different professions.
- ...

One-factor analysis of variance

Analysis of variance (ANOVA for short) is an approach to testing the influence of factors (i.e., categorical variables) on a numerical variable.

For instance, we may want to test:

- Whether the BMI (body mass index) of adults is associated with their ethnicity.
- Whether there is a difference in income among three different professions.
- ...

Here, we only consider the influence of a single factor. The method can be extended to multiple factors.

One-factor analysis of variance

Analysis of variance (ANOVA for short) is an approach to testing the influence of factors (i.e., categorical variables) on a numerical variable.

We can view analysis of variance as

- an analogue of the "chi-squared test of independence" in which instead of testing the dependence of two categorical variables on each other, we test the dependence of a numerical variable on one (or more) categorical variables.

One-factor analysis of variance

Analysis of variance (ANOVA for short) is an approach to testing the influence of factors (i.e., categorical variables) on a numerical variable.

We can view analysis of variance as

- an analogue of the "chi-squared test of independence" in which instead of testing the dependence of two categorical variables on each other, we test the <u>dependence of a numerical variable on one (or more) categorical variables</u>.

Alternatively, we can view analysis of variance as

- an extension of the "t-test for comparing the means of two populations based on independent samples", where we wish to <u>compare the means of more than two populations</u>.

One-factor analysis of variance

The strategy will be as usual:

We use a statistic F and compare

1. The observed value of F,
2. The distribution of F as suggested by H_0.

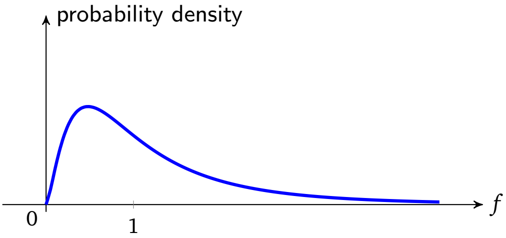
We ask ourselves:

Q Is the observed value of F too extreme for H_0 to be plausible?

The statistics we will use turns out to have the so-called F-distribution under the null hypothesis.

The F-distribution

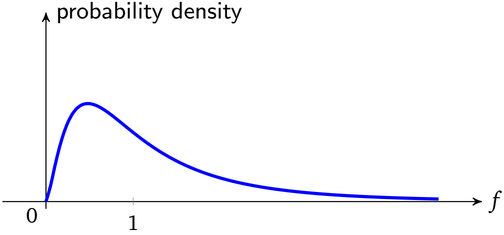
The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



A continuous RV with an F-distribution is called an F-RV.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):

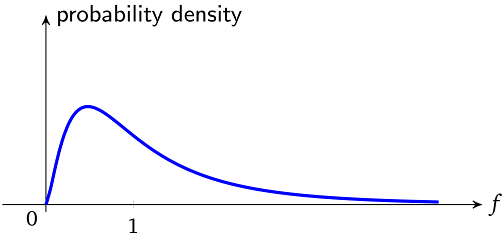


A continuous RV with an F-distribution is called an F-RV.

The F-distribution is unimodal and right-skewed. An F-RV can only take non-negative values.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



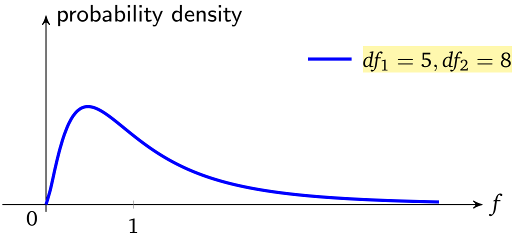
A continuous RV with an F-distribution is called an F-RV.

The F-distribution is unimodal and right-skewed. An F-RV can only take non-negative values.

The possible values of an F-RV are all non-negative numbers $0 \leq f < +\infty$.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



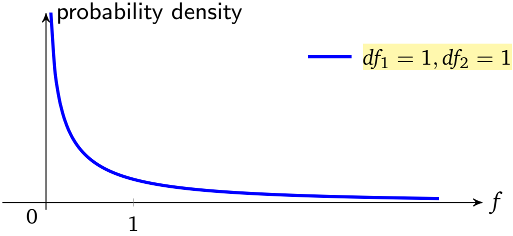
Remark 1

The mean of an F-RV is $\frac{df_2}{df_2+2}$. [less than 1]
The mode of an F-RV is $\frac{df_1-2}{df_1} \times \frac{df_2}{df_2+2}$ [less than the mean]

The larger df_1 and df_2, the less skewed the F-distribution, and the more shifted towards right.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



Remark 1

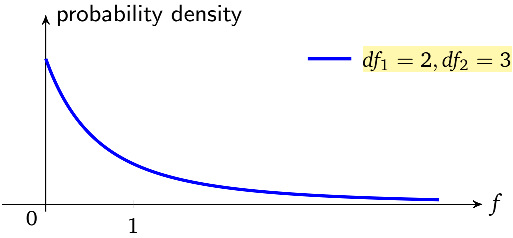
The mean of an F-RV is $\frac{df_2}{df_2+2}$. [less than 1]

The mode of an F-RV is $\frac{df_1-2}{df_1} \times \frac{df_2}{df_2+2}$ [less than the mean]

The larger df_1 and df_2, the less skewed the F-distribution, and the more shifted towards right.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



Remark 1

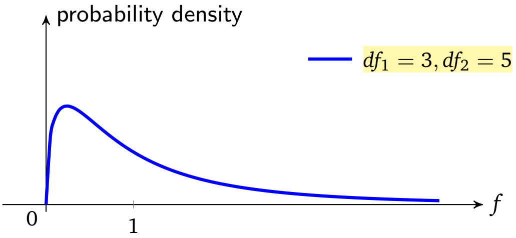
The mean of an F-RV is $\frac{df_2}{df_2+2}$. [less than 1]

The mode of an F-RV is $\frac{df_1-2}{df_1} \times \frac{df_2}{df_2+2}$ [less than the mean]

The larger df_1 and df_2, the less skewed the F-distribution, and the more shifted towards right.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



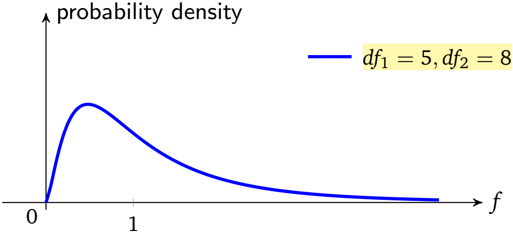
Remark 1

The mean of an F-RV is $\frac{df_2}{df_2+2}$. [less than 1]
The mode of an F-RV is $\frac{df_1-2}{df_1} \times \frac{df_2}{df_2+2}$ [less than the mean]

The larger df_1 and df_2, the less skewed the F-distribution, and the more shifted towards right.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



Remark 1

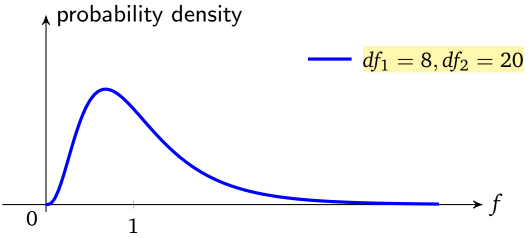
The mean of an F-RV is $\frac{df_2}{df_2+2}$. [less than 1]

The mode of an F-RV is $\frac{df_1-2}{df_1} \times \frac{df_2}{df_2+2}$ [less than the mean]

The larger df_1 and df_2, the less skewed the F-distribution, and the more shifted towards right.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



Remark 1

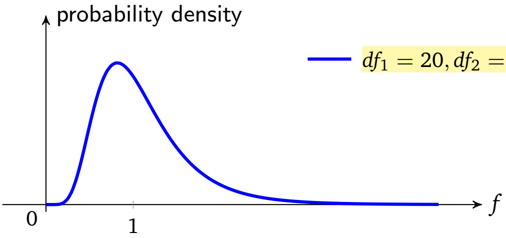
The mean of an F-RV is $\frac{df_2}{df_2+2}$. [less than 1]

The mode of an F-RV is $\frac{df_1-2}{df_1} \times \frac{df_2}{df_2+2}$ [less than the mean]

The larger df_1 and df_2, the less skewed the F-distribution, and the more shifted towards right.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



Remark 1

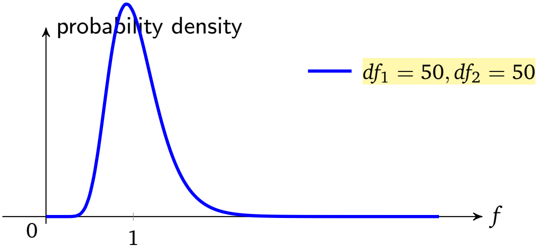
The mean of an F-RV is $\frac{df_2}{df_2+2}$. [less than 1]

The mode of an F-RV is $\frac{df_1-2}{df_1} \times \frac{df_2}{df_2+2}$ [less than the mean]

The larger df_1 and df_2, the less skewed the F-distribution, and the more shifted towards right.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



Remark 1

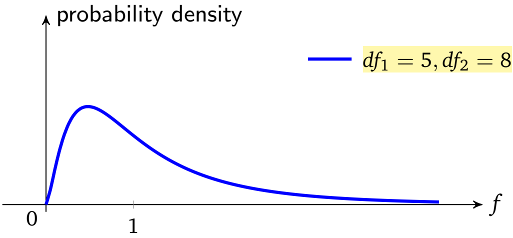
The mean of an F-RV is $\frac{df_2}{df_2+2}$. [less than 1]

The mode of an F-RV is $\frac{df_1-2}{df_1} \times \frac{df_2}{df_2+2}$ [less than the mean]

The larger df_1 and df_2, the less skewed the F-distribution, and the more shifted towards right.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



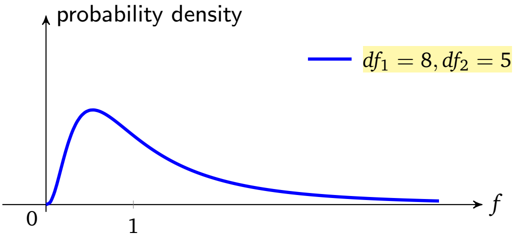
Remark 1

The mean of an F-RV is $\frac{df_2}{df_2+2}$. [less than 1]
The mode of an F-RV is $\frac{df_1-2}{df_1} \times \frac{df_2}{df_2+2}$ [less than the mean]

The larger df_1 and df_2, the less skewed the F-distribution, and the more shifted towards right.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



Remark 1

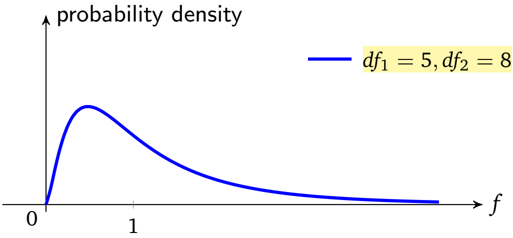
The mean of an F-RV is $\frac{df_2}{df_2+2}$. [less than 1]

The mode of an F-RV is $\frac{df_1-2}{df_1} \times \frac{df_2}{df_2+2}$ [less than the mean]

The larger df_1 and df_2, the less skewed the F-distribution, and the more shifted towards right.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



Remark 1

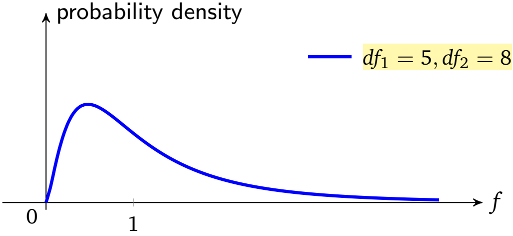
The mean of an F-RV is $\frac{df_2}{df_2+2}$. [less than 1]

The mode of an F-RV is $\frac{df_1-2}{df_1} \times \frac{df_2}{df_2+2}$ [less than the mean]

The larger df_1 and df_2, the less skewed the F-distribution, and the more shifted towards right.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



Remark 2 (for those curious)

The F-distribution is inter-connected with the chi-squared distribution.

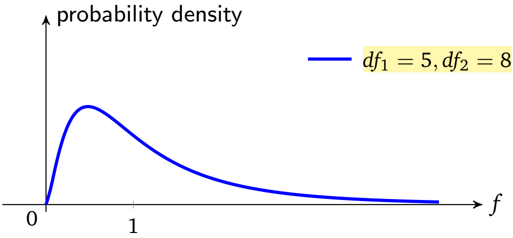
If X_1 and X_2 are independent chi-squared RVs with 5 and 8 degrees of freedom respectively, then

$$\frac{X_1/5}{X_2/8}$$

has the F-distribution with parameters $df_1 = 5$ and $df_2 = 8$.

The F-distribution

The F-distribution with parameters df_1 and df_2 (# of degrees of freedom for the numerator and denominator):



Remark 2 (for those curious)

The F-distribution is inter-connected with the chi-squared distribution.

If X_1 and X_2 are independent chi-squared RVs with df_1 and df_2 degrees of freedom respectively, then

$$\frac{X_1/df_1}{X_2/df_2}$$

has the F-distribution with parameters df_1 and df_2.

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

Nutrition researchers have conducted an experiment on a sample of 29 rats. Each of the 29 subjects is randomly assigned one of four possible diets A, B, C, D. The following table contains the liver weight expressed as percentage of body weight.

diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84	
diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44
diet C	3.34	3.72	3.81	3.66	3.55	3.51		
diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91

The researchers would like to know whether the diet influences the liver weight.

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

Nutrition researchers have conducted an experiment on a sample of 29 rats. Each of the 29 subjects is randomly assigned one of four possible diets A, B, C, D. The following table contains the liver weight expressed as percentage of body weight.

diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84	
diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44
diet C	3.34	3.72	3.81	3.66	3.55	3.51		
diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91

The researchers would like to know whether the diet influences the liver weight.

Remark

Since this is a randomized experiment, we can potentially conclude causality from it. In this context, the categories are often called treatments.

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

The researchers would like to know whether the diet influences the liver weight.

Q What are the competing hypotheses?

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

The researchers would like to know whether the diet influences the liver weight.

Q What are the competing hypotheses?

A H_0 Liver weight is not influenced by diet. [i.e., is independent of]

H_1 Liver weight is influenced by diet. [i.e., is not independent of]

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

The researchers would like to know whether the diet influences the liver weight.

Q What are the competing hypotheses?

A H_0 Liver weight is not influenced by diet. [i.e., is independent of]

H_1 Liver weight is influenced by diet. [i.e., is not independent of]

We will assume that the population in each category is normally distributed, each with the same (unknown) standard deviation σ.

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

The researchers would like to know whether the diet influences the liver weight.

Q What are the competing hypotheses?

A
- H_0 Liver weight is not influenced by diet. [i.e., is independent of]
- H_1 Liver weight is influenced by diet. [i.e., is not independent of]

We will assume that the population in each category is normally distributed, each with the same (unknown) standard deviation σ.

With these assumptions, the competing hypotheses can be rephrased as follows:

- H_0 $\mu_A = \mu_B = \mu_C = \mu_D$.
- H_1 Not all population means are the same.

(μ_A is the population mean liver weight within category A, etc.)

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

Q What is a suitable test statistic?

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

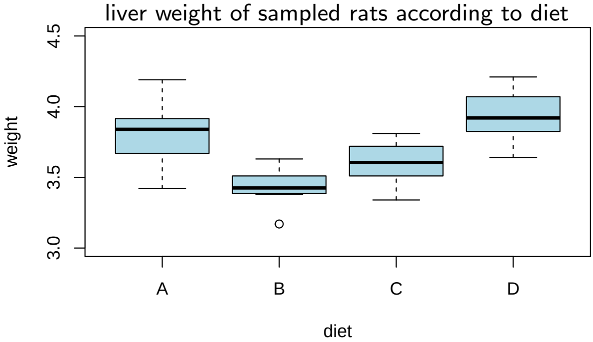
H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

- Q What is a suitable test statistic?
- Q Where should we look to find evidence against the null hypothesis?

One-factor analysis of variance

Example (Effect of diet on liver weight)

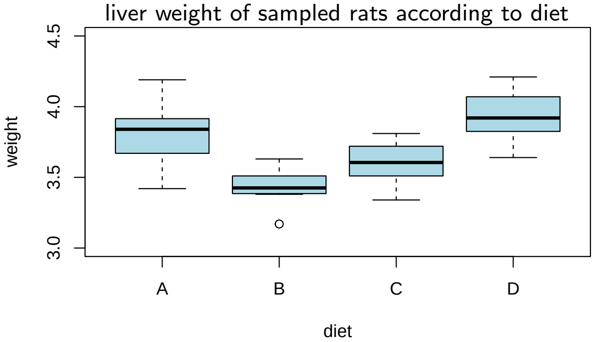
[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]



One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

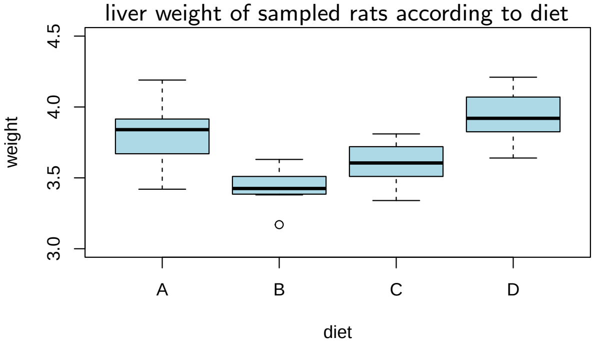


If the variations between categories are “large” compared to the variations within categories, then we have evidence in favor of the influence of diet on liver weight.

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]



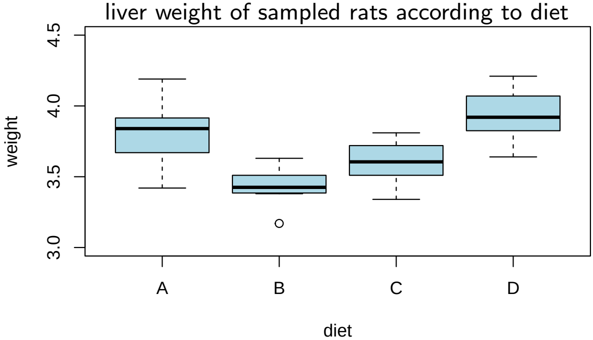
If the variations between categories are "large" compared to the variations within categories, then we have evidence in favor of the influence of diet on liver weight.

Q How can we quantify such a comparison?

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]



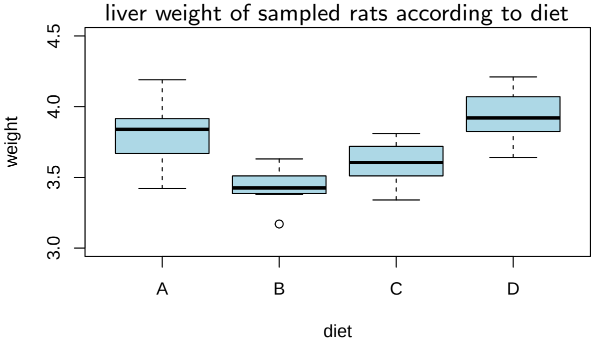
Two "variance-like" quantities to be introduced soon:

- MSB (mean square between samples)
  measures the variations between categories.
- MSW (mean square within samples)
  measures the variations within categories.

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]



If MSB is “large” compared to MSW, then that counts as evidence against the null hypothesis.

[i.e., in favor of the influence of the diet on liver weight]

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

Q What is a suitable test statistic?

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

Q What is a suitable test statistic?

A The statistic

$$F := \frac{\mathrm{MSB}}{\mathrm{MSW}}$$

measures the variations between categories relative to the variations within categories.

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

Q What is a suitable test statistic?

A The statistic

$$F := \frac{\mathsf{MSB}}{\mathsf{MSW}}$$

measures the variations between categories relative to the variations within categories.

Q What is the distribution of F according to H_0

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

Q What is a suitable test statistic?

A The statistic

$$F := \frac{\text{MSB}}{\text{MSW}}$$

measures the variations between categories relative to the variations within categories.

Q What is the distribution of F according to H_0

A It turns out that F has the F-distribution with
$df_1 = 4 - 1$ (number of categories minus 1) and
$df_2 = 29 - 4$ (total sample size minus number of categories).

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

Q What is the observed value of F?

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

Q What is the observed value of F?

We should first introduce MSB and MSW and learn how to calculate them.

Partition of sum of squares

Consider a sample of size n from a population with k distinct categories.

In our example, $n = 29$ and $k = 4$.

diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84	
diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44
diet C	3.34	3.72	3.81	3.66	3.55	3.51		
diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91

Partition of sum of squares

Consider a sample of size n from a population with k distinct categories.

- Let $X_{i,j}$ be the jth sampled entity within category i.

In our example, $n = 29$ and $k = 4$.

diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84	
diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44
diet C	3.34	3.72	3.81	3.66	3.55	3.51		
diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91

- $X_{2,3} = 3.38$ and $X_{4,1} = 3.64$.

Partition of sum of squares

Consider a sample of size n from a population with k distinct categories.

- Let $X_{i,j}$ be the jth sampled entity within category i.

In our example, $n = 29$ and $k = 4$.

diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84	
diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44
diet C	3.34	3.72	3.81	3.66	3.55	3.51		
diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91

- $X_{2,3} = 3.38$ and $X_{4,1} = 3.64$.

Partition of sum of squares

Consider a sample of size n from a population with k distinct categories.

- Let $X_{i,j}$ be the jth sampled entity within category i.

In our example, $n = 29$ and $k = 4$.

diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84	
diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44
diet C	3.34	3.72	3.81	3.66	3.55	3.51		
diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91

- $X_{2,3} = 3.38$ and $X_{4,1} = 3.64$.

Partition of sum of squares

Consider a sample of size n from a population with k distinct categories.

- Let $X_{i,j}$ be the jth sampled entity within category i.
- Let $\overline{X}_i$ be the mean of the sampled entities within category i.

In our example, $n = 29$ and $k = 4$.

diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84	
diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44
diet C	3.34	3.72	3.81	3.66	3.55	3.51		
diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91

- $X_{2,3} = 3.38$ and $X_{4,1} = 3.64$.
- $\overline{X}_1 = \frac{3.42+3.96+3.87+4.19+3.58+3.76+3.84}{7} = 3.802857$

Partition of sum of squares

Consider a sample of size n from a population with k distinct categories.

- Let $X_{i,j}$ be the jth sampled entity within category i.
- Let $\overline{X}_i$ be the mean of the sampled entities within category i.
- Let n_i be the number of sampled entities within category i.

In our example, $n = 29$ and $k = 4$.

diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84	
diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44
diet C	3.34	3.72	3.81	3.66	3.55	3.51		
diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91

- $X_{2,3} = 3.38$ and $X_{4,1} = 3.64$.
- $\overline{X}_1 = \frac{3.42+3.96+3.87+4.19+3.58+3.76+3.84}{7} = 3.802857$
- $n_1 = 7$ and $n_2 = 8$.

Partition of sum of squares

Consider a sample of size n from a population with k distinct categories.

- Let $X_{i,j}$ be the jth sampled entity within category i.
- Let $\overline{X}_i$ be the mean of the sampled entities within category i.
- Let n_i be the number of sampled entities within category i.

In our example, $n = 29$ and $k = 4$.

diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84	
diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44
diet C	3.34	3.72	3.81	3.66	3.55	3.51		
diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91

- $X_{2,3} = 3.38$ and $X_{4,1} = 3.64$.
- $\overline{X}_1 = \frac{3.42+3.96+3.87+4.19+3.58+3.76+3.84}{7} = 3.802857$
- $n_1 = 7$ and $n_2 = 8$.

Partition of sum of squares

Consider a sample of size n from a population with k distinct categories.

Partition of sum of squares

Consider a sample of size n from a population with k distinct categories.

In this setting, the sample variance is

$$S^2 = \frac{1}{n-1} \sum_i \sum_j (X_{i,j} - \overline{X})^2 \ .$$

Partition of sum of squares

Consider a sample of size n from a population with k distinct categories.

In this setting, the sample variance is

$$S^2 = \frac{1}{n-1} \sum_i \sum_j (X_{i,j} - \bar{X})^2 \ .$$

A straightforward algebraic manipulation shows that

$$\underbrace{\sum_i \sum_j (X_{i,j} - \bar{X})^2}_{\text{SST}} = \underbrace{\sum_i \sum_j (X_{i,j} - \bar{X}_i)^2}_{\text{SSW}} + \underbrace{\sum_i n_i (\bar{X}_i - \bar{X})^2}_{\text{SSB}} \ .$$

- SST: total sum of squares
- SSW:sum of squares within samples
- SSB: sum of squares between samples

Partition of sum of squares

Consider a sample of size n from a population with k distinct categories.

In this setting, the sample variance is

$$S^2 = \frac{1}{n-1} \sum_i \sum_j (X_{i,j} - \overline{X})^2 \; .$$

A straightforward algebraic manipulation shows that

$$\underbrace{\sum_i \sum_j (X_{i,j} - \overline{X})^2}_{\text{SST}} = \underbrace{\sum_i \sum_j (X_{i,j} - \overline{X}_i)^2}_{\text{SSW}} + \underbrace{\sum_i n_i (\overline{X}_i - \overline{X})^2}_{\text{SSB}} \; .$$

This is a variant of the so-called law of total variance.

Mean squares between and within samples

Consider a sample of size n from a population with k distinct categories.

$$\underbrace{\sum_i \sum_j (X_{i,j} - \overline{X})^2}_{\text{SST}} = \underbrace{\sum_i \sum_j (X_{i,j} - \overline{X}_i)^2}_{\text{SSW}} + \underbrace{\sum_i n_i(\overline{X}_i - \overline{X})^2}_{\text{SSB}} .$$

Mean squares between and within samples

Consider a sample of size n from a population with k distinct categories.

$$\underbrace{\sum_i \sum_j (X_{i,j} - \overline{X})^2}_{\text{SST}} = \underbrace{\sum_i \sum_j (X_{i,j} - \overline{X}_i)^2}_{\text{SSW}} + \underbrace{\sum_i n_i (\overline{X}_i - \overline{X})^2}_{\text{SSB}} \ .$$

The mean squares between and within samples are

$$\boxed{\text{MSB} := \frac{\text{SSB}}{k-1}} \qquad \boxed{\text{MSW} := \frac{\text{SSW}}{n-k}}$$

Computing the sums of squares

The sums of squares between and within sample can alternatively be calculated as follows:

$$\mathrm{SSB} = \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i \sum_j X_{i,j}\Big)^2$$

$$\mathrm{SSW} = \sum_i \sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2$$

The above formulas are somewhat more efficient than the original definitions. [i.e., they take less computation]

Computing the sums of squares

In our example,

i										total	$\sum x^2$
1	diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84		26.62	101.6106
2	diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44	27.44	94.2474
3	diet C	3.34	3.72	3.81	3.66	3.55	3.51			21.59	77.8283
4	diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91	31.48	124.1280
total										107.13	397.8143

$$\text{SSB} = \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i\sum_j X_{i,j}\Big)^2$$

$$\text{SSW} = \sum_i\sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2$$

Computing the sums of squares

In our example,

i										total	$\sum x^2$
1	diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84		26.62	101.6106
2	diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44	27.44	94.2474
3	diet C	3.34	3.72	3.81	3.66	3.55	3.51			21.59	77.8283
4	diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91	31.48	124.1280
total										107.13	397.8143

$$\begin{aligned}\text{SSB} &= \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i \sum_j X_{i,j}\Big)^2 \\ &= \left[\frac{1}{7} \times 26.62^2 + \frac{1}{8} \times 27.44^2 + \frac{1}{6} \times 21.59^2 + \frac{1}{8} \times 31.48^2\right] - \left[\frac{1}{29} \times 107.13^2\right]\end{aligned}$$

$$\text{SSW} = \sum_i \sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2$$

Computing the sums of squares

In our example,

i										total	$\sum x^2$
1	diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84		26.62	101.6106
2	diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44	27.44	94.2474
3	diet C	3.34	3.72	3.81	3.66	3.55	3.51			21.59	77.8283
4	diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91	31.48	124.1280
total										107.13	397.8143

$$\begin{aligned}\text{SSB} &= \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i\sum_j X_{i,j}\Big)^2 \\ &= \left[\frac{1}{7}\times 26.62^2 + \frac{1}{8}\times 27.44^2 + \frac{1}{6}\times 21.59^2 + \frac{1}{8}\times 31.48^2\right] - \left[\frac{1}{29}\times 107.13^2\right]\end{aligned}$$

$$\text{SSW} = \sum_i\sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2$$

Computing the sums of squares

In our example,

i										total	$\sum x^2$
1	diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84		26.62	101.6106
2	diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44	27.44	94.2474
3	diet C	3.34	3.72	3.81	3.66	3.55	3.51			21.59	77.8283
4	diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91	31.48	124.1280
total										107.13	397.8143

$$\text{SSB} = \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i\sum_j X_{i,j}\Big)^2$$

$$= \left[\frac{1}{7}\times 26.62^2 + \frac{1}{8}\times 27.44^2 + \frac{1}{6}\times 21.59^2 + \frac{1}{8}\times 31.48^2\right] - \left[\frac{1}{29}\times 107.13^2\right]$$

$$\text{SSW} = \sum_i\sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2$$

Computing the sums of squares

In our example,

i										total	$\sum x^2$
1	diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84		26.62	101.6106
2	diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44	27.44	94.2474
3	diet C	3.34	3.72	3.81	3.66	3.55	3.51			21.59	77.8283
4	diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91	31.48	124.1280
total										107.13	397.8143

$$\text{SSB} = \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i \sum_j X_{i,j}\Big)^2$$

$$= \left[\frac{1}{7} \times 26.62^2 + \frac{1}{8} \times 27.44^2 + \frac{1}{6} \times 21.59^2 + \frac{1}{8} \times 31.48^2\right] - \left[\frac{1}{29} \times 107.13^2\right]$$

$$\text{SSW} = \sum_i \sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2$$

Computing the sums of squares

In our example,

i										total	$\sum x^2$
1	diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84		26.62	101.6106
2	diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44	27.44	94.2474
3	diet C	3.34	3.72	3.81	3.66	3.55	3.51			21.59	77.8283
4	diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91	31.48	124.1280
total										107.13	397.8143

$$\begin{aligned}\text{SSB} &= \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i\sum_j X_{i,j}\Big)^2 \\ &= \left[\frac{1}{7}\times 26.62^2 + \frac{1}{8}\times 27.44^2 + \frac{1}{6}\times 21.59^2 + \frac{1}{8}\times 31.48^2\right] - \left[\frac{1}{29}\times 107.13^2\right]\end{aligned}$$

$$\text{SSW} = \sum_i\sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2$$

Computing the sums of squares

In our example,

i										total	$\sum x^2$
1	diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84		26.62	101.6106
2	diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44	27.44	94.2474
3	diet C	3.34	3.72	3.81	3.66	3.55	3.51			21.59	77.8283
4	diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91	31.48	124.1280
total										107.13	397.8143

$$\text{SSB} = \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i \sum_j X_{i,j}\Big)^2$$
$$= \left[\frac{1}{7} \times 26.62^2 + \frac{1}{8} \times 27.44^2 + \frac{1}{6} \times 21.59^2 + \frac{1}{8} \times 31.48^2\right] - \left[\frac{1}{29} \times 107.13^2\right]$$

$$\text{SSW} = \sum_i \sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2$$

Computing the sums of squares

In our example,

i										total	$\sum x^2$
1	diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84		26.62	101.6106
2	diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44	27.44	94.2474
3	diet C	3.34	3.72	3.81	3.66	3.55	3.51			21.59	77.8283
4	diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91	31.48	124.1280
total										107.13	397.8143

$$
\begin{aligned}
\text{SSB} &= \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i\sum_j X_{i,j}\Big)^2 \\
&= \left[\frac{1}{7}\times 26.62^2 + \frac{1}{8}\times 27.44^2 + \frac{1}{6}\times 21.59^2 + \frac{1}{8}\times 31.48^2\right] - \left[\frac{1}{29}\times 107.13^2\right] \\
&= \boxed{1.160077}\,.
\end{aligned}
$$

$$
\text{SSW} = \sum_i\sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2
$$

Computing the sums of squares

In our example,

i										total	$\sum x^2$
1	diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84		26.62	101.6106
2	diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44	27.44	94.2474
3	diet C	3.34	3.72	3.81	3.66	3.55	3.51			21.59	77.8283
4	diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91	31.48	124.1280
total										107.13	397.8143

$$\begin{aligned}
\text{SSB} &= \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i\sum_j X_{i,j}\Big)^2 \\
&= \left[\frac{1}{7}\times 26.62^2 + \frac{1}{8}\times 27.44^2 + \frac{1}{6}\times 21.59^2 + \frac{1}{8}\times 31.48^2\right] - \left[\frac{1}{29}\times 107.13^2\right] \\
&= \boxed{1.160077}\,. \\
\text{SSW} &= \sum_i\sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 \\
&= 397.8143 - \left[\frac{1}{7}\times 26.62^2 + \frac{1}{8}\times 27.44^2 + \frac{1}{6}\times 21.59^2 + \frac{1}{8}\times 31.48^2\right]
\end{aligned}$$

Computing the sums of squares

In our example,

i										total	$\sum x^2$
1	diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84		26.62	101.6106
2	diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44	27.44	94.2474
3	diet C	3.34	3.72	3.81	3.66	3.55	3.51			21.59	77.8283
4	diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91	31.48	124.1280
total										107.13	397.8143

$$
\begin{aligned}
\text{SSB} &= \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i\sum_j X_{i,j}\Big)^2 \\
&= \left[\frac{1}{7}\times 26.62^2 + \frac{1}{8}\times 27.44^2 + \frac{1}{6}\times 21.59^2 + \frac{1}{8}\times 31.48^2\right] - \left[\frac{1}{29}\times 107.13^2\right] \\
&= \boxed{1.160077}\,.
\end{aligned}
$$

$$
\begin{aligned}
\text{SSW} &= \sum_i\sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 \\
&= 397.8143 - \left[\frac{1}{7}\times 26.62^2 + \frac{1}{8}\times 27.44^2 + \frac{1}{6}\times 21.59^2 + \frac{1}{8}\times 31.48^2\right]
\end{aligned}
$$

Computing the sums of squares

In our example,

i										total	$\sum x^2$
1	diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84		26.62	101.6106
2	diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44	27.44	94.2474
3	diet C	3.34	3.72	3.81	3.66	3.55	3.51			21.59	77.8283
4	diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91	31.48	124.1280
total										107.13	397.8143

$$\begin{aligned}
\text{SSB} &= \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i \sum_j X_{i,j}\Big)^2 \\
&= \left[\frac{1}{7}\times 26.62^2 + \frac{1}{8}\times 27.44^2 + \frac{1}{6}\times 21.59^2 + \frac{1}{8}\times 31.48^2\right] - \left[\frac{1}{29}\times 107.13^2\right] \\
&= \boxed{1.160077}\,.
\end{aligned}$$

$$\begin{aligned}
\text{SSW} &= \sum_i \sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 \\
&= 397.8143 - \left[\frac{1}{7}\times 26.62^2 + \frac{1}{8}\times 27.44^2 + \frac{1}{6}\times 21.59^2 + \frac{1}{8}\times 31.48^2\right]
\end{aligned}$$

Computing the sums of squares

In our example,

i										total	$\sum x^2$
1	diet A	3.42	3.96	3.87	4.19	3.58	3.76	3.84		26.62	101.6106
2	diet B	3.17	3.63	3.38	3.47	3.39	3.41	3.55	3.44	27.44	94.2474
3	diet C	3.34	3.72	3.81	3.66	3.55	3.51			21.59	77.8283
4	diet D	3.64	3.93	3.77	4.18	4.21	3.88	3.96	3.91	31.48	124.1280
total										107.13	397.8143

$$
\begin{aligned}
\text{SSB} &= \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 - \frac{1}{n}\Big(\sum_i \sum_j X_{i,j}\Big)^2 \\
&= \left[\frac{1}{7}\times 26.62^2 + \frac{1}{8}\times 27.44^2 + \frac{1}{6}\times 21.59^2 + \frac{1}{8}\times 31.48^2\right] - \left[\frac{1}{29}\times 107.13^2\right] \\
&= \boxed{1.160077}\,.
\end{aligned}
$$

$$
\begin{aligned}
\text{SSW} &= \sum_i \sum_j X_{i,j}^2 - \sum_i \frac{1}{n_i}\Big(\sum_j X_{i,j}\Big)^2 \\
&= 397.8143 - \left[\frac{1}{7}\times 26.62^2 + \frac{1}{8}\times 27.44^2 + \frac{1}{6}\times 21.59^2 + \frac{1}{8}\times 31.48^2\right] \\
&= \boxed{0.9012262}\,.
\end{aligned}
$$

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet. H_1 Liver weight is influenced by diet.

Q What is the observed value of F?

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

Q What is the observed value of F?

A The mean squares between and within samples are

$$\text{MSB} = \frac{\text{SSB}}{k-1} = \frac{1.160077}{4-1} \approx 0.3866924$$

$$\text{MSW} = \frac{\text{SSW}}{n-k} = \frac{0.9012262}{29-4} \approx 0.03604905$$

Therefore, the observed value of F is

$$f = \frac{\text{MSB}}{\text{MSW}} \approx \frac{0.3866924}{0.03604905} \approx \boxed{10.72684}\,.$$

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

Q At significance level 1%, how should the researchers conclude?

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet. H_1 Liver weight is influenced by diet.

Q At significance level 1%, how should the researchers conclude?

A We have

$$\text{p-value} = \mathbb{P}(F \geq 10.72684 \mid H_0) \approx$$

using the app for the F-distribution with $df_1 = 4 - 1$ and $df_2 = 29 - 4$.

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

Q At significance level 1%, how should the researchers conclude?

A We have

$$\text{p-value} = \mathbb{P}(F \geq 10.72684 \mid H_0) \approx \boxed{0.0001}$$

using the app for the F-distribution with $df_1 = 4 - 1$ and $df_2 = 29 - 4$.

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet.	H_1 Liver weight is influenced by diet.

Q At significance level 1%, how should the researchers conclude?

A We have

$$\text{p-value} = \mathbb{P}(F \geq 10.72684 \mid H_0) \approx \boxed{0.0001}$$

using the app for the F-distribution with $df_1 = 4 - 1$ and $df_2 = 29 - 4$.
Since p-value $< \alpha$, the researchers must reject the null hypothesis.

One-factor analysis of variance

Example (Effect of diet on liver weight)

[This example is taken from S.C. Campbell, Statistics for Biologists, Cambridge, 1989]

H_0 Liver weight is not influenced by diet. H_1 Liver weight is influenced by diet.

Q At significance level 1%, how should the researchers conclude?

A We have

$$\text{p-value} = \mathbb{P}(F \geq 10.72684 \mid H_0) \approx \boxed{0.0001}$$

using the app for the F-distribution with $df_1 = 4 - 1$ and $df_2 = 29 - 4$.

Since p-value $< \alpha$, the researchers must reject the null hypothesis.

Conclusion: The evidence is statistically significant (p-value $\approx 0.01\%$) in favor of the claim that the liver weight of the rats is influenced by their diets.

One-factor analysis of variance (formulation 1)

Suppose C is a categorical variable with k possible values, and X is a numerical variable.

One-factor analysis of variance (formulation 1)

Suppose C is a categorical variable with k possible values, and X is a numerical variable.

Competing hypotheses (v1)

H_0 X is independent on C. [X is not influenced by C]

H_1 X is dependent on C. [X is influenced by C]

One-factor analysis of variance (formulation 1)

Suppose C is a categorical variable with k possible values, and X is a numerical variable.

Competing hypotheses (v1)

H_0 X is independent on C. [X is not influenced by C]

H_1 X is dependent on C. [X is influenced by C]

Evidence

The values of C and X for a random sample of size n.

One-factor analysis of variance (formulation 1)

Suppose C is a categorical variable with k possible values, and X is a numerical variable.

Competing hypotheses (v1)

H_0 X is independent on C. [X is not influenced by C]

H_1 X is dependent on C. [X is influenced by C]

Evidence

The values of C and X for a random sample of size n.

Test statistic

$$F := \frac{\text{MSB}}{\text{MSW}}$$

where

- $\text{MSB} := \text{SSB}/(k-1)$ is the mean square between samples,
- $\text{MSW} := \text{SSW}/(n-k)$ is the mean square within samples.

One-factor analysis of variance (formulation 1)

Suppose C is a categorical variable with k possible values, and C is a numerical variable.

Competing hypotheses (v1)

H_0 X is independent on C. [X is not influenced by C]

H_1 X is dependent on C. [X is influenced by C]

Covered scenario

The population in each category is normal with the same (unknown) variance σ^2.

One-factor analysis of variance (formulation 1)

Suppose C is a categorical variable with k possible values, and C is a numerical variable.

Competing hypotheses (v1)

H_0 X is independent on C. [X is not influenced by C]

H_1 X is dependent on C. [X is influenced by C]

Covered scenario

The population in each category is normal with the same (unknown) variance σ^2.

Strategy

We compare the observed value of F against the F-distribution with $df_1 = k - 1$ and $df_2 = n - k$. We can follow either the rejection region approach or the p-value approach.

One-factor analysis of variance (formulation 1)

Suppose C is a categorical variable with k possible values, and C is a numerical variable.

Competing hypotheses (v1)

H_0 X is independent on C. [X is not influenced by C]

H_1 X is dependent on C. [X is influenced by C]

Covered scenario

The population in each category is normal with the same (unknown) variance σ^2.

Strategy

We compare the observed value of F against the F-distribution with $df_1 = k - 1$ and $df_2 = n - k$. We can follow either the rejection region approach or the p-value approach.

Remark

We always use a right-tailed test.

One-factor analysis of variance (formulation 2)

Suppose X is a numerical variable, and the means of X in k distinct populations are $\mu_1, \mu_2, \ldots, \mu_k$.

One-factor analysis of variance (formulation 2)

Suppose X is a numerical variable, and the means of X in k distinct populations are $\mu_1, \mu_2, \ldots, \mu_k$.

Competing hypotheses (v2)

H_0 $\mu_1 = \mu_2 = \cdots = \mu_k$.

H_1 Not all k means are equal.

One-factor analysis of variance (formulation 2)

Suppose X is a numerical variable, and the means of X in k distinct populations are $\mu_1, \mu_2, \ldots, \mu_k$.

Competing hypotheses (v2)

H_0 $\mu_1 = \mu_2 = \cdots = \mu_k$.

H_1 Not all k means are equal.

Evidence

k random samples, one from each population, with total size n.

One-factor analysis of variance (formulation 2)

Suppose X is a numerical variable, and the means of X in k distinct populations are $\mu_1, \mu_2, \ldots, \mu_k$.

Competing hypotheses (v2)

H_0 $\mu_1 = \mu_2 = \cdots = \mu_k$.

H_1 Not all k means are equal.

Evidence

k random samples, one from each population, with total size n.

Test statistic

$$F := \frac{\mathsf{MSB}}{\mathsf{MSW}}$$

where

- $\mathsf{MSB} := \mathsf{SSB}/(k-1)$ is the mean square between samples,
- $\mathsf{MSW} := \mathsf{SSW}/(n-k)$ is the mean square within samples.

One-factor analysis of variance (formulation 2)

Suppose X is a numerical variable, and the means of X in k distinct populations are $\mu_1, \mu_2, \ldots, \mu_k$.

Competing hypotheses (v2)

H_0 $\mu_1 = \mu_2 = \cdots = \mu_k$.

H_1 Not all k means are equal.

Covered scenario

The samples are independent of one another. Each population is normal. All populations have the same (unknown) variance σ^2.

One-factor analysis of variance (formulation 2)

Suppose X is a numerical variable, and the means of X in k distinct populations are $\mu_1, \mu_2, \ldots, \mu_k$.

Competing hypotheses (v2)

H_0 $\mu_1 = \mu_2 = \cdots = \mu_k$.

H_1 Not all k means are equal.

Covered scenario

The samples are independent of one another. Each population is normal. All populations have the same (unknown) variance σ^2.

Strategy

We compare the observed value of F against the F-distribution with $df_1 = k - 1$ and $df_2 = n - k$. We can follow either the rejection region approach or the p-value approach.

One-factor analysis of variance (formulation 2)

Suppose X is a numerical variable, and the means of X in k distinct populations are $\mu_1, \mu_2, \ldots, \mu_k$.

Competing hypotheses (v2)

H$_0$ $\mu_1 = \mu_2 = \cdots = \mu_k$.

H$_1$ Not all k means are equal.

Covered scenario

The samples are independent of one another. Each population is normal. All populations have the same (unknown) variance σ^2.

Strategy

We compare the observed value of F against the F-distribution with $df_1 = k - 1$ and $df_2 = n - k$. We can follow either the rejection region approach or the p-value approach.

Remark

We always use a right-tailed test.