Hypothesis testing
Two-population significance test

Two-variance significance test to infer that two variances are equal in two normal populations

We study two general populations with normal distributions N ($\mu 1$, $\sigma 1$) and N ($\mu 2$, $\sigma 2$).

The distribution normality condition is met - we use the F test for the equality of two variances.

Two-variance significance test to infer that two variances are equal in two normal populations

We check whether $\sigma^2_1 = \sigma^2_2$:

H_0: $\sigma^2_1 = \sigma^2_2$

H_1: $\sigma^2_1 \neq \sigma^2_2$

σ^2_1 variance of the first population

σ^2_2 variance of the second population

Significance test for two variances

Samples of n_1 and n_2 are drawn from both populations. The variance for each of trials s^2_1 and s^2_2 is then computed.

For the two-variance test, the F-Snedecor statistic is used as the test statistic.

Significance test for two variances

The numerator of the F statistic must include the greater of the variance due to the assumed null hypothesis, and the degrees of freedom n_L and n_M relate to the variance from the numerator and the denominator, respectively.

if $s^2_1 > s^2_2$

$$F = \frac{s_1^2}{s_2^2}$$

if $s^2_2 > s^2_1$

$$F = \frac{s_2^2}{s_1^2}$$

Significance test for two variances

Then we read the values of the $\mathbf{F_{a,\,nL,\,nM}}$ statistics from the tables (F statistics with the assumed significance level and n_L and n_M degrees of freedom $n_L = n_1 - 1$, $n_M = n_2 - 1$ or $n_L = n_2 - 1$, $n_M = n_1 - 1$).

If $\mathbf{F > F_{a,\,nL,\,nM}}$, there are grounds for rejecting the null hypothesis, otherwise there are no grounds for rejecting the null hypothesis.

Significance test for two variances - example

Assuming that the daily expenditure of students on entertainment in Krakow (x_1) is the same as in Rzeszów (x_2), 45 students were randomly selected from Krakow and 36 students from Rzeszów. Based on the collected information, the arithmetic mean and the variance of students' entertainment expenses in both cities were calculated.

$$\bar{x} = 22 \ zł$$

$$\bar{x} = 18 \ zł$$

$$s_1^2 = 2.5 \ (zł)^2$$

$$s_2^2 = 1.4 \ (zł)^2$$

Verify the hypothesis that the variance of entertainment expenses of both studied groups of students is the same, assuming that the significance level is 0.05, against the alternative hypothesis that the variance of expenses in the Krakow group is greater.

Significance test for two variances - example

$H_0: \sigma^2_1 = \sigma^2_2$

$H_1: \sigma^2_1 > \sigma^2_2$
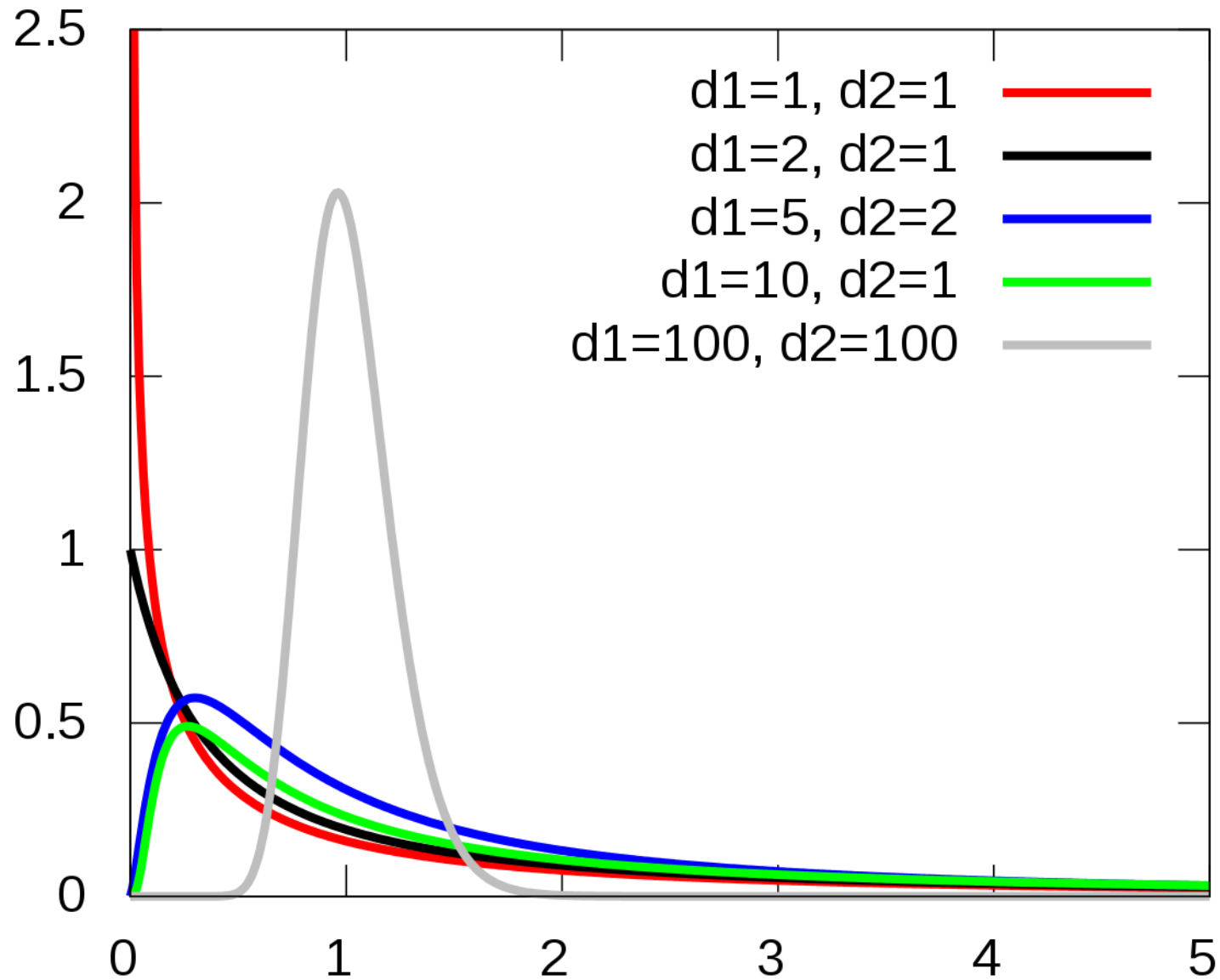
$$F = \frac{s_1^2}{s_2^2} = \frac{2.5}{1.4} = 1{,}786$$

Critical value $F_{0,05, \, 44, \, 35}$

the value of the F statistic at significance level of 0.05 and 44 and 35 degrees of freedom

LibreCalc:

=F.INV(0.05,44,35)

F Snedecor distribution

d1=1, d2=1
d1=2, d2=1
d1=5, d2=2
d1=10, d2=1
d1=100, d2=100

F Snedecor critical values

Wartości krytyczne rozkładu F-Snedecora

$X \sim F_{v1, v2}$ - X zmienna losowa o rozkładzie F- Snedecora z liczbami stopni swobody (v1, v2)

poziom istotności α =0,05,

$F_{\alpha, v1, v2}$ - wartość krytyczna - liczba taka, że $P(X > F_{\alpha, v1, v2}) = \alpha$

v2 \ v1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161,446	199,499	215,707	224,583	230,160	233,988	236,767	238,884	240,543	241,882	242,981	243,905	244,690	245,363	245,949
2	18,513	19,000	19,164	19,247	19,296	19,329	19,353	19,371	19,385	19,396	19,405	19,412	19,419	19,424	19,429
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,785	8,763	8,745	8,729	8,715	8,703
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,936	5,912	5,891	5,873	5,858
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,704	4,678	4,655	4,636	4,619
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,027	4,000	3,976	3,956	3,938
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,603	3,575	3,550	3,529	3,511
8	5,318	4,459	4,066	3,838	3,688	3,581	3,500	3,438	3,388	3,347	3,313	3,284	3,259	3,237	3,218
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,102	3,073	3,048	3,025	3,006
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,943	2,913	2,887	2,865	2,845
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,818	2,788	2,761	2,739	2,719
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,717	2,687	2,660	2,637	2,617

Significance test for two variances - example

Result:

$F > F_{0,05, 44, 35}$

H_0 must be rejected.

The variance of expenditure in the Krakow group is greater.

Online calculator

http://www.statskingdom.com/220VarF2.html

Significance test for two fractions

feature X_1 has a two-point distribution with unknown parameter p_1

feature X_2 has a two-point distribution with an unknown parameter p_2

the n_1 -element sample from the first population was taken and

n_2 - an element sample from the second population,

k_i - number of elements distinguished in the i -th sample;

Significance test for two fractions

We check if $p_1 = p_2$:

$H_0: p_1 = p_2$

$H_1: p_1 \neq p_2$

p1 fraction of the first population

p2 fraction of the second population

Significance test for two fractions

feature X_1 has a two-point distribution with unknown parameter p_1

feature X_2 has a two-point distribution with an unknown parameter p_2

the n1-element sample from the first population was taken and

n_2 - an element sample from the second population,

k_i - number of elements distinguished in the i -th sample

$$\overline{p_1} = \frac{k_1}{n_1}$$

$$\overline{p_2} = \frac{k_2}{n_2}$$

$$\bar{p} = \frac{k_1 + k_2}{n_1 + n_2}$$

Significance test for two fractions

$$u_{emp} = \frac{\overline{p_1} - \overline{p_2}}{\sqrt{\bar{p}(1 - \bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Significance test for two fractions of critical values u$_{crit}$

for $\alpha=0{,}05$

=NORM.INV(0.975,0,1)=1.98

for $\alpha=0{,}01$

= NORM.INV(0.995,0,1)=2.58

Significance test for two fractions

We check, if $p_1 = p_2$:

H_0: $p_1 = p_2$

H_1: $p_1 \neq p_2$

if $|u_{emp}| > u_{kryt}$ there is no reason to accept H_0

Significance test for two fractions - example

A survey on the sorting of waste in households was conducted in two city districts. The following results were obtained: in the first of the 210 surveyed farms, waste was sorted in 55, while in the second, out of 130 farms, in 51 household waste was sorted. At the significance level of 0.01, verify the hypothesis about the same fraction of households sorting waste in both cities.

Significance test for two fractions - example

$$\overline{p_1} = \frac{k_1}{n_1} = \frac{55}{210}$$

$$\overline{p_2} = \frac{k_2}{n_2} = \frac{51}{130}$$

$$\bar{p} = \frac{k_1 + k_2}{n_1 + n_2} = \frac{106}{340}$$

$$\alpha = 0.01$$

$$u_{emp} = 2.52$$
$$u_{kryt} = 2.58$$

$$|u_{emp}| < u_{kryt}$$

Significance test for two fractions - example

$p < \alpha$ we accept H_0

The fractions of households sorting waste in cities are equal.

Online calculator

https://www.socscistatistics.com/tests/ztest/default2.aspx

ANOVA

ANOVA

Analysis of variance, or ANOVA, is a statistical tool that is used to show the difference between more means or components through significance tests. It also allows to make multiple comparisons of several populations means. The ANOVA test is performed by comparing two types of variation, the variation between the sample means, as well as the variation within each of the samples. Analysis of variance is a parametric statistical test where the total observed variance is divided into variances from different sources of variation. The ANOVA checks if the means of several groups are equal or not.

ANOVA

The null hypothesis of H_0 assumes that there are no differences in the mean values among several datasets. The H_1 alternative hypothesis says that at least one mean is different from the others. If the null hypothesis is rejected, the alternative hypothesis H_1 is assumed, i.e., at least one set of results may come from a distribution with a different mean. ANOVA can be used to compare mean values for three or more data sets.

ANOVA

Depending on the number of analyzed factors, the analysis of variance can take various forms. One of them is one-way ANOVA in which only one factor is assessed. For example, comparing the yield of four wheat cultivars, where the variable is yield and the factor is variety, then it is one-way analysis of variance and multiple comparisons (completely random system).

ANOVA

An exemplary scheme of a one-factor experiment with four levels of the factor (4 subjects: A1, A2, A3, A4) and three repetitions. Objects are randomly distributed throughout the experiment area.

A2	A2	A1
A3	A3	A2
A1	A4	A3
A4	A4	A1

ANOVA

The purpose of the analysis of variance (ANOVA) is to compare the means across multiple normally distributed populations. ANOVA indicates whether the variance of at least one data set is significantly different from the others. However, to answer the question of which data set is different, post-hoc testing is performed after H_0 is rejected. These tests compare all combinations of mean pairs. Detailed comparisons provide information on sets of data significantly different from the others.

ANOVA

Assumptions: the variables have a normal distribution $X_i \sim$ N (μ, σ^2) variances (and thus standard deviations) for the studied populations are equal $\sigma_1 = \sigma_2 = \sigma_3 = \ldots = \sigma_i$

- Null hypothesis H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu i$ (means do not differ)
- Alternative hypothesis H_1: $\mu_i \neq \mu_j$ (at least two means are different)

ANOVA

Application examples: comparison of the yield of several wheat cultivars, comparison of starch content in potato tubers with different fertilization.

The results of the analysis of variance are most often presented in the form of the following table (ANOVA table):

ANOVA

Source of variation	Sum of Squares	Degrees of Freedom	Mean Squares (MS)	F
Within	$$SS_W = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$$	$k \cdot (n-1)$	$$MS_w = \frac{SS_W}{k \cdot (n-1)}$$	
Between	$$SS_B = \sum_{i=1}^{k} n \cdot (\bar{x}_i - \bar{x})^2$$	$k-1$	$$MS_b = \frac{SS_B}{k-1}$$	$$F = \frac{MS_B}{MS_W}$$
Total	$$SS_T = \sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij} - \bar{x})^2 = SS_w + SS_b$$	$n \cdot k - 1$		

ANOVA

Where, F = ANOVA Coefficient, a test statistic given by the F distribution, also known as Snedecor's F distribution or the Fisher–Snedecor distribution,

- SS_W = Sum of squares within the groups,
- SS_B = Sum of squares between the groups,
- SS_T = total Sum of squares,
- MS_W = Mean sum of squares within the groups,
- MS_B = Mean sum of squares between the groups,
- k = the total number of populations,
- n_i = the total number of samples in an i population.

ANOVA

The empirical value F_{emp} (F statistic) has an F distribution (it is a ration of two Chi-squared distributions with the same or different degrees of freedom). If the numerator in the equation for F statistic is much larger than the dominator, then the variation in this data is mostly due to the differences between the means and less due to the variation within the samples. So the probability of the null hypothesis to be correct is very small. The F statistic is then compared with the critical value $F_{k-1, k \cdot (n-1), \alpha}$. Then, a decision is made whether the samples come from the same population (the means do not differ) or different populations (with significantly different means).

ANOVA

If $F_{emp} > F_{k-1, k \cdot (n-1), \alpha}$ the null hypothesis should be rejected, there is a difference in populations means. If p $< \alpha$, we reject the null hypothesis and accept the alternative hypothesis, i.e. we find that at least two means differ significantly and we move on to multiple comparisons, i.e. comparisons of all possible pairs of means. Post-hoc tests are used to test statistically significant differences between the populations that were identified in ANOVA as having different means. Choosing the best test to compare means values depend on the researcher. Multiple (detailed) comparisons - it is a method that allows to determine which means differ significantly and which do not differ. Then homogeneous groups can be distinguished, it is subsets of means that can be considered the same (not significantly different).

ANOVA

Multiple comparison procedures are as follows: Tukey, Scheff, Bonfferroni, Duncan, Newman-Kuels and others. The choice of procedure is often quite arbitrary. If the difference of means is greater than NIR, then it can be recognized that the means differ (there is a statistically significant difference). Sample tests for multiple comparisons are presented below. Tukey HSD is a widely used test that allows you to compare pairs of means multiple times when the ANOVA test shows differences between them. The test checks if one mean value differs significantly from another. Compares all possible pairs of means. It is the most useful test for multiple comparisons. Another is the Fisher's LSD test is a statistical significance test used when sample sizes are small.

ANOVA example

The protein content of grain in four wheat cultivars was assessed. The grain was collected from 9 randomly selected fields for each cultivar. On the basis of the obtained results, determine whether there is a difference in protein content between these wheat cultivars (assume the significance level $\alpha = 0.05$)

ANOVA example

cultivar	protein content (%)	cultivar	protein content (%)
A	12.4	C	11.2
A	14	C	12.4
A	13.6	C	11.6
A	14.4	C	13.6
A	15.2	C	14.4
A	14	C	13.6
A	15.2	C	14
A	15.6	C	14.4
A	13.6	C	14
B	17.4	D	14.8
B	18.2	D	14.4
B	19.6	D	13.2
B	18.0	D	16
B	16.8	D	15.6
B	17.2	D	14
B	18.6	D	18
B	20.8	D	17.2
B	19.4	D	16.8

ANOVA calculations

Null hypothesis H_0: $\mu_A = \mu_B = \mu_C = \mu_D$ (means do not differ)

Alternative hypothesis H_1: $\mu_i \neq \mu_j$ (at least two means are different)

The total mean $\bar{x} = 15.4$

The mean for cultivar A $\bar{x}_A = 14.2$

The mean for cultivar B $\bar{x}_B = 13.2$

The mean for cultivar C $\bar{x}_C = 18.3$

The mean for cultivar D $\bar{x}_D = 15.5$

$$SS_W = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 =$$

$= (12.4 - 14.2)^2 + (14 - 14.2)^2 + (13.6 - 14.2)^2 + (14.4 - 14.2)^2 + (15.2 - 14.2)^2 +$

$(14 - 14.2)^2 + (15.2 - 14.2)^2 + (15.6 - 14.2)^2 + (13.6 - 14.2)^2 + (11.2 - 13.2)^2 +$

$(12.4 - 13.2)^2 + (11.6 - 13.2)^2 + (13.6 - 13.2)^2 + (14.4 - 13.2)^2 + (13.6 - 13.2)^2 +$

$(14 - 13.2)^2 + (14.4 - 13.2)^2 + (14 - 13.2)^2 + (17.4 - 18.3)^2 + (18.2 - 18.3)^2 +$

$(19.6 - 18.3)^2 + (18 - 18.3)^2 + (16.8 - 18.3)^2 + (17.2 - 18.3)^2 + (18.6 - 18.3)^2 +$

$(20.8 - 18.3)^2 + (19.4 - 18.3)^2 + (14.8 - 15.5)^2 + (14.4 - 15.5)^2 + (13.2 - 15.5)^2 +$

$(16 - 15.5)^2 + (15.6 - 15.5)^2 + (14 - 15.5)^2 + (18 - 15.5)^2 + (17.2 - 15.5)^2 +$

$(16.8 - 15.5)^2 = 53.4$

Degrees of Freedom – k(n-1) = 32

$$SS_B = \sum_{i=1}^{k} n \cdot (\overline{x_i} - \bar{x})^2 =$$

$$= 9 \cdot [(14.2 - 15.3)^2 + (13.2 - 15.3)^2 + (18.3 - 15.3)^2 + (15.5 - 15.3)^2] = 137.9$$

Degrees of Freedom – k-1 = 3

$$SS_T = \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x})^2 = SS_w + SS_b =$$

$= (12.4 - 15.4)^2 + (14 - 15.4)^2 + (13.6 - 15.4)^2 + (14.4 - 15.4)^2 + (15.2 - 15.4)^2 +$

$(14 - 15.4)^2 + (15.2 - 15.4)^2 + (15.6 - 15.4)^2 + (13.6 - 15.4)^2 + (11.2 - 15.4)^2 +$

$(12.4 - 15.4)^2 + (11.6 - 15.4)^2 + (13.6 - 15.4)^2 + (14.4 - 15.4)^2 + (13.6 - 15.4)^2 +$

$(14 - 15.4)^2 + (14.4 - 15.4)^2 + (14 - 15.4)^2 + (17.4 - 15.4)^2 + (18.2 - 15.4)^2 +$

$(19.6 - 15.4)^2 + (18 - 15.4)^2 + (16.8 - 15.4)^2 + (17.2 - 15.4)^2 + (18.6 - 15.4)^2 +$

$(20.8 - 15.4)^2 + (19.4 - 15.4)^2 + (14.8 - 15.4)^2 + (14.4 - 15.4)^2 + (13.2 - 15.4)^2 +$

$(16 - 15.4)^2 + (15.6 - 15.4)^2 + (14 - 15.4)^2 + (18 - 15.4)^2 + (17.2 - 15.4)^2 +$

$(16.8 - 15.4)^2 = 191.3$

Degrees of Freedom – k·n-1 = 35

Source of variation	Sum of Squares	Degrees of Freedom	Mean Squares (MS)	F
Within	$SS_W = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 = 53.4$	$k \cdot (n-1)$ $= 32$	$MS_w = \dfrac{SS_W}{k \cdot (n-1)}$ $= 1.7$	
Between	$SS_B = \sum_{i=1}^{k} n \cdot (\bar{x}_i - \bar{x})^2 = 137.9$	$k - 1 = 3$	$MS_b = \dfrac{SS_B}{k-1}\ 46.0$	$F = \dfrac{MS_B}{MS_W}$ $= 27.0$
Total	$SS_T = \sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij} - \bar{x})^2 = SS_w + SS_b$ $= 191.3$	$k \cdot n - 1$ $= 35$		

ANOVA – conclusions

The test statistic $F_{emp} = 27.0$. The F statistic is then compared with the critical value $F_{k-1, k \cdot (n-1), \alpha}$, which here is $F_{3,32,0.5} = 0.11$. $F_{emp} > F_{3,32,0.5}$ the null hypothesis should be rejected, there is a difference in populations means.

Online calculator:

https://www.socscistatistics.com/tests/anova/default2.aspx

ANOVA – nonparametric

When the assumptions of the parametric test are not met, non-parametric tests are used. These tests use ranking data, not actual data values. These tests rank data from the lowest to the highest. Each value is assigned a sequential integer from 1 to n, where n is the total sample size. Nonparametric the tests are applied when the assumption of normality of distributions is not met or there is no homogeneity of variance. Typically, these tests compare the sample medians, not the mean values. Nonparametric tests do not address clearly defined parameters and it is more difficult to infer from their results about differences between populations. These tests, using ranking, often lose information about real data. Therefore, nonparametric tests give less reliable results than their parametric counterparts. Nonparametric tests work well for nominal, qualitative, and ordinal data.

ANOVA – nonparametric

For ANOVA, its non-parametric counterpart is Kruskal-Wallis test – can be called as one-way nonparametric ANOVA. Kruskal-Wallis is a nonparametric comparison of the tendencies of the position (in this case the median) of the distributions for many populations. It is used as an alternative to one-way analysis of variance for three or more independent samples when the assumptions about the normality of the distribution are not met. The test works on the basis of ranking data. The test can be used for data with heterogeneous variances and for data from non-normal distributions. The null hypothesis in this test is that the distributions of data for each compared group are similar. It is assumed, however, that the distributions of the variables are of a similar shape, it is this method should not be used when the distribution of the dependent variable in one population is for example right-asymmetric and left-asymmetric in the other population. The decision to choose a test is made after the initial data analysis. If the data do not come from normal distributions or the assumption of equal variance is not met, the non-parametric test is selected.

ANOVA – nonparametric

Thus, Kruskal-Wallis test allows the comparison of many populations with any distribution. The test statistic is the K value (also denoted as H). The K value is the greater the greater the difference between the ranks for the studied groups. As in the case of the Mann-Whitney U test, the ranks of individual observation values are assigned after ordering in ascending order the values from all tested samples. If the null hypothesis in the Kruskall-Wallis test is rejected (i.e. if the p for the test is less than the value of α), it means that at least two populations differ in terms of the distributions of the trait under study.
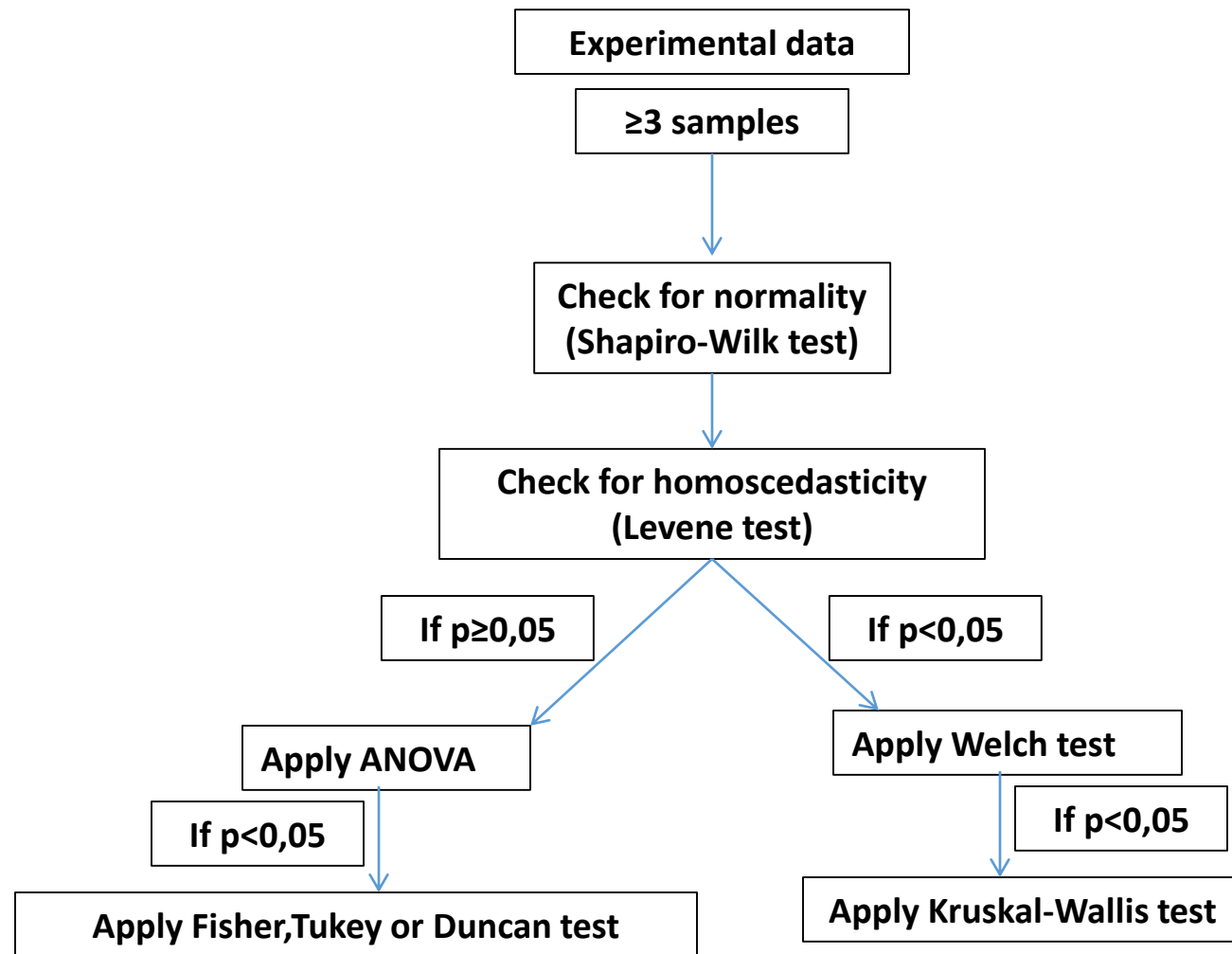
Based on:

Daniel Granato, Verônica Maria de Araújo Calado, Basil Jarvis „Observations on the use of statistical methods in Food Science and Technology" Food Research International 55 (2014) 137–149