Probabilities and Statistics - Lecture 12

Olariu E. Florentin

May 18, 2026

Table of contents

Tests of significance

The main steps while performing a test of significance:

1-2. Formulate the two hypotheses: H_0 and H_a: H_a is accepted if H_0 is rejected.

3. Choose a significance level α - how significant must be the evidence of rejecting H_0.
4. Compute the statistic or the score of the test.
5. Compute the critical value.
6. Compare the score and the critical value, and if it is the case reject H_0 and accept H_a, otherwise don't reject H_0 nor accept H_a.

Z-test

- A Z-test is a statistical hypothesis test for statistics (scores) that follow a normal distribution if the null hypothesis is true.
- Due to the Central Limit Theorem (CLT) we can use a Z-test even when the population is approximately normally distributed, but only for large samples ($n \geqslant 30$).
- We already used a Z-test for testing hypotheses on proportions (because of de Moivre-Laplace Theorem which is consequence of CLT).
- A Z-test is based on the normal distribution; for small samples, this significance test works best if your sample was drawn from a normal distribution or from one that is very close to normal.

Z-test - Inference for the mean of a population (σ known)

- We consider a statistical population whose variance (σ^2) is known.
- The population is (approximatively) normally distributed and we want to test a hypothesis about the mean of the population.
- The test can be performed even if the population hasn't a normal distribution, but if the used sample is large enough.
- If μ_0 is the mean of the population (known from the null hypothesis), then the following statistic $\dfrac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}}$ is standard normally distributed: $N(0, 1)$.
- We conduct the test like follows:

Z-test - Inference for the mean of a population (σ known)

1. We first formulate the *null hypothesis*, which says that the mean of the population has a certain value:

$$\boxed{H_0 : \mu = \mu_0}$$

2. We second formulate the *alternative hypothesis* according to the information gathered from the sample. We can have three different types of alternative hypothesis

$$\boxed{H_a : \mu < \mu_0}$$ (*left asymmetric*) or

$$\boxed{H_a : \mu > \mu_0}$$ (*right asymmetric*) or

$$\boxed{H_a : \mu \neq \mu_0}$$ (*symmetric hypothesis*).

The asymmetric hypotheses correspond to *one-tailed tests*, while the symmetric one corresponds to a *two-tailed test*.

Z-test - Inference for the mean of a population (σ known)

3. We choose a level of significance $\alpha \in \{1\%, 5\%\}$.
4. We compute the *z-score* (the *statistic* of the test)

$$z = \frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}}$$

5. We determine the critical value corresponding to α:

$z^* = qnorm(\alpha)$ for left asymmetric H_a ($z^* < 0$),

$z^* = qnorm(1-\alpha)$ for right asymmetric H_a ($z^* > 0$),

$z^* = -qnorm(\alpha/2) = qnorm(1-\alpha/2)$

for symmetric H_a ($z^* > 0$).

Z-test - Inference for the mean of a population (σ known)

6. We compare the critical value with the z-score; if the z-score belongs to the *rejection region*, then H_a *is accepted* and H_0 *is rejected*. The rejection regions are:

$$(-\infty, z^*] \text{ for left asymmetric } H_a,$$

$$[z^*, +\infty) \text{ for right asymmetric } H_a,$$

$$(-\infty, -|z^*|] \cup [|z^*|, +\infty) \text{ for symmetric } H_a.$$

If the z-score doesn't belong to the rejection region we say that *there is not sufficient evidence at the α level of significance* to reject the null hypothesis (*we fail to reject H_0*).

Z-test - Inference for the mean of a population - Example

Example.

- We have a very large colony of laboratory mice. Their weight follows a normal distribution with a standard deviation $\sigma = 5g$ and it is believed that their average weight is $30g$.
- For a 25 mice sample we find an average weight $32g$; is this finding significant with 5% level of significance? but with 1% level of significance?

Solution.

- It seems that the real mean of the population is different than that claimed ($\mu_0 = 30g$).
- Since we know that the population follows a normal distribution, and the standard deviation of the population is known we can perform a Z-test for the mean.

Z-test - Inference for the mean of a population - Example

- We gather the data concerning the population and the sample: $\mu_0 = 30$, $\sigma = 5$, $n = 25$, $\overline{x}_n = 32$.

1-2. We can formulate the null hypothesis and a symmetric alternative hypothesis

$$H_0 : \mu = 30 \quad H_a : \mu \neq 30.$$

3. $\alpha = 0.05$.

4. The z-score

$$z = \frac{\overline{x}_n - \mu_0}{\sigma / \sqrt{n}} = \frac{32 - 30}{5/\sqrt{25}} = 2.$$

5. The critical value is $z^* = -qnorm(\alpha/2) = 1.9599$, for $\alpha = 5\%$.

6. Since $|z| > |z^*|$, we can reject the null hypothesis, and accept that the true mean of the population is not $\mu_0 = 30g$.

Z-test - Inference for the mean of a population - Example

- Now we do again the last two steps for the other level of significance:

5'. For $\alpha = 1\%$ the critical value is $z^* = -qnorm(\alpha/2) = 2.5758$.

6'. Since $|z| < |z^*|$, we fail to reject the null hypothesis with 1% level of significance (there is not sufficient evidence at 1% level of significance that the true mean of the population is different from $30g$).

Z-test - Inference for the mean of a population - Example revisited

- If we look at the information from the sample we can observe that the sample mean, $\overline{x}_n$, is greater than the supposed mean of the population.
- In such a case we can formulate a right asymmetric alternative hypothesis.

1-2. The new hypotheses are

$$H_0 : \mu = 30 \quad H_a : \mu > 30.$$

3. $\alpha = 0.05$.
4. The z-score

$$z = \frac{\overline{x}_n - \mu_0}{\sigma / \sqrt{n}} = \frac{32 - 30}{5/\sqrt{25}} = 2.$$

Z-test - Inference for the mean of a population - Example revisited

5. The critical value is $z^* = qnorm(1-\alpha) = 1.6448$, for $\alpha = 5\%$.
6. Since $z > z^*$, we can reject the null hypothesis, and accept that the true mean of the population is greater than $\mu_0 = 30g$.

- For the other level of significance (1%):

5'. For $\alpha = 1\%$ the critical value is $z^* = qnorm(1-\alpha) = 2.3263$.

6'. Since $z < z^*$, we fail to reject the null hypothesis with 1% level of significance (there is not sufficient evidence at 1% level of significance that the true mean of the population is greater than $30g$).

Z-test - Inference for the mean of a population - Remarks

- It is worth noting that with different levels of significance we can have different conclusions: the null hypothesis may be rejected with a level but not with the other.
- If the null hypothesis is rejected with 1% level of significance, then it will be rejected with 5% also; in other words if the null hypothesis cannot be rejected with 5%, then cannot be rejected with 1%.
- The alternative hypothesis is formulated according to the sample mean: if $\mu_0 << \overline{x}_n$ we can formulate a right asymmetric alternative hypothesis, $H_a : \mu > \mu_0$, and if $\mu_0 >> \overline{x}_n$ we can formulate a left asymmetric alternative hypothesis, $H_a : \mu < \mu_0$.
- If the sample mean is not obviously much greater (or much smaller) than the sample mean we may presume that the true mean is just different from the value in the null hypothesis and we can formulate symmetric alternative hypothesis, $H_a : \mu \neq \mu_0$.

Z-test - Exercises

I. Perform again a Z-test for the above exercise with $\overline{x}_n = 27g$ and $\sigma = 6$. (Use both levels of significance).

II. It is claimed that the students at a certain university will score an average of 35 on a given test with $\sigma = 4$. Is the claim reasonable if a random sample of test scores from this university yields $33, 42, 38, 37, 30, 42$? Complete a hypothesis test using $\alpha = 5\%$. Assume test results are normally distributed.

III. According to the National Center for Health Statistics, the average height of females in the US (which is normally distributed) is 63.7 in with a standard deviation $\sigma = 2.75$ in. A random sample of 50 female American health professionals yield a mean of 65.2 in. Test the claim that the mean height of females in the health profession is different from 63.7 in. Use a 5% level of significance.

T-test

- T-test is a statistical hypothesis test for statistics (scores) that follow a Student's distribution.
- T-test is preferred when the standard deviation of the (normally distributed) population is unknown.
- A T-test is also appropriate when we are handling small samples ($n < 30$) for populations that are just approximately normally distributed.
- Following these observations we can say that a T-test (for the mean of a population) is complementary to a Z-test.
- In the next section we will describe a T-test for the mean of a population with unknown standard deviation.

T-test - Inference for the mean of a population (σ unknown)

- We consider a statistical population whose variance (σ^2) is unknown.
- The population is normally distributed and we want to test the true mean of the population.
- The test can be performed even if the population has a very close to a normal distribution, but only when the sample has a small size (otherwise we have to use a Z-test).
- If μ_0 is the mean of the population (known from the null hypothesis), then the following statistic $\dfrac{\overline{x}_n - \mu_0}{s/\sqrt{n}}$ is Student distributed with $(n-1)$ degrees of freedom: $T(n-1)$.
- A difference from a Z-test is the replacement of the population standard deviation, σ, with the sample standard deviation s.
- The test is performed like follows:

T-test - Inference for the mean of a population (σ unknown)

1. We first formulate the *null hypothesis*, which says that the mean of the population has a certain value:

$$\boxed{H_0: \ \mu = \mu_0}$$

2. We formulate the *alternative hypothesis* according to the information gathered from the sample. We can have three different types of alternative hypothesis

$\boxed{H_a: \ \mu < \mu_0}$ (*left asymmetric*) or

$\boxed{H_a: \ \mu > \mu_0}$ (*right asymmetric*) or

$\boxed{H_a: \ \mu \neq \mu_0}$ (*symmetric hypothesis*).

The asymmetric hypothesis corresponds to an *one-tailed test*, while the symmetric one corresponds to a *two-tailed test*.

T-test - Inference for the mean of a population (σ unknown)

3. We choose the level of significance $\alpha \in \{1\%, 5\%\}$.
4. We compute the *t-score* (the *statistic* of the test)

$$t = \frac{\overline{x}_n - \mu_0}{s/\sqrt{n}}$$

5. We determine the corresponding critical value:

$$t^* = qt(\alpha, n-1)$$ for left asymmetric H_a ($t^* < 0$),

$$t^* = qt(1-\alpha, n-1)$$ for right asymmetric H_a ($t^* > 0$),

$$t^* = -qt(\alpha/2, n-1) = qt(1-\alpha/2, n-1)$$

for symmetric H_a ($t^* > 0$).

T-test - Inference for the mean of a population (σ unknown)

6. We compare the the critical value with the t-score; if the t-score belongs to the *rejection region*, then H_a is accepted and H_0 is rejected. The rejection regions are:

$$(-\infty, t^*] \text{ for left asymmetric } H_a,$$

$$[t^*, +\infty) \text{ for right asymmetric } H_a,$$

$$(-\infty, -|t^*|] \cup [|t^*|, +\infty) \text{ for symmetric } H_a.$$

If the t-score doesn't belong to the rejection region we say that *there is not sufficient evidence at the α level of significance* to reject the null hypothesis (*we fail to reject H_0*).

T-test - Inference for the mean of a population - Example

Example

- The concentration of CO (carbon monoxide) is measured with a machine called spectrophotometer that can measure concentrations up to about 100 ppm. These machines must be calibrated every day by measuring CO concentration in a manufactured gas sample which has a controlled concentration of 70 ppm. If the machine reads close to 70 ppm is ready for use otherwise it has to be adjusted.
- We assume that the concentration follows a normal distribution but the standard deviation is unknown. In one particular day five readings give

$$58 \ 71 \ 67 \ 64 \ 62.$$

- Four of these readings are lower than 70; can this be explained on the basis of chance variations? Or does it prove a bias, coming perhaps from improper adjustment of the machine?

T-test - Inference for the mean of a population - Example

Solution

- It is possible that the machine needs an adjustment, hence we will test the assumption that that the real mean is different than $\mu_0 = 70$.
- Since we know that the population follows a normal distribution, and the standard deviation of the population is unknown we can perform a T-test for the mean.
- The data concerning the population and the sample are: $\mu_0 = 70$, $s = 4.9295$, $n = 5$, $\overline{x}_n = 64.4$.

1-2. We can formulate the null hypothesis and a symmetric alternative hypothesis

$$H_0 : \mu = 70 \quad H_a : \mu \neq 70.$$

T-test - Inference for the mean of a population - Example

3. $\alpha = 0.05$.
4. The t-score

$$t = \frac{\overline{x}_n - \mu_0}{s/\sqrt{n}} = \frac{64.4 - 70}{4.9295/\sqrt{5}} = -2.5402.$$

5. The critical value is $t^* = -qt(\alpha/2, 4) = 2.7764$, for $\alpha = 5\%$.
6. Since $|t| < |t^*|$, we fail to reject the null hypothesis (there is not sufficient evidence at 5% level of significance to accept that the true mean of the population is different from 70 ppm).

T-test - Inference for the mean of a population - Example

- Now we do again the last two steps for the other level of significance:

5'. For $\alpha = 1\%$ the critical value is $t^* = -qt(\alpha/2, 4) = 4.6040$.

6'. Since $|t| < |t^*|$, we fail to reject the null hypothesis with 1% level of significance (there is not sufficient evidence at 1% level of significance to accept that the true mean of the population is different from 70 ppm).

- It is worth noting here that the rework of the test with a smaller α is not needed, as we already fail to reject the null hypothesis with 5% level of significance (because $|z^*|$ will increase).

T-test - Inference for the mean of a population - Example revisited

- If we look at the information from the sample we can observe that the sample mean, $\overline{x}_n$ is lower than the supposed mean of the population.
- In such a situation we may formulate a left asymmetric alternative hypothesis.

1-2. The new hypotheses are

$$H_0 : \mu = 70 \quad H_a : \mu < 70.$$

3. $\alpha = 0.05$.
4. The t-score

$$t = \frac{\overline{x}_n - \mu_0}{s/\sqrt{n}} = \frac{64.4 - 70}{4.9295/\sqrt{5}} = -2.5402.$$

T-test - Inference for the mean of a population - Example revisited

5. The critical value is $t^* = qt(\alpha, 4) = -2.1318$, for $\alpha = 5\%$.
6. Since $t < t^*$, we reject the null hypothesis and accept the alternative hypothesis: the true mean is less than 70 ppm.

- For the other level of significance (1%):

5'. For $\alpha = 1\%$ the critical value is $t^* = qt(\alpha, 4) = -3.7469$.

6'. Since $t > t^*$, we fail to reject the null hypothesis with 1% level of significance (there is not sufficient evidence at 1% level of significance to say that the true mean of the population is less than 70 ppm).

T-test - Exercises

I. A student group maintains that each day, the average student must travel at least 25 minutes one way to reach the college. The college admissions office obtained random sample of 32 one-way travel times from students. The sample has a mean of 19.4 minutes and a standard deviation of 9.6 minutes. Does the admissions office have sufficient evidence to reject the students' claim? Use $\alpha = 0.01$. (Assume the travel duration is normally distributed.)

II. It is know that a young adult spends a weekly average of 40$ for fast food. A survey of $1,000$ young adults by Greenfield Online and reported in a USA Today Snapshot finds a weekly average of 35$ on fast food with a sample standard deviation of 14.50$. Assuming fast food weekly expenditures are normally distributed perform an appropriate test on the true mean of weekly spending for fast food.

T-test - Exercises

III. Homes in a nearby college town have a mean value of $88,950\$$. It is assumed that homes in the vicinity of the college have a higher mean value. To test this theory, a random sample of 12 homes is chosen from the college area. Their mean valuation is $92,460\$$, and the standard deviation is $5,200\$$. Complete a hypothesis test using $\alpha = 5\%$. Assume a normal distribution for the prices.

χ-square tests

- Sometimes the sample frequencies do not fit the theoretical ones (according to the expected distribution of the population). For example when a die is tossed we expect each face to have the same chances to occur, but in practical simulations it is not always the case.
- χ^2 test is designed to measure the "difference" between the observed and the expected frequencies for categorical data.
- More precisely a χ^2 test is used to find (if any) a statistically significant difference between the expected and the observed frequencies.
- There are many test which use the χ^2 distribution; all the tests having a statistics which follows a χ^2 distribution are called χ^2-tests (or χ-square tests).
- Another χ^2 test is used for deciding if two categorical variables are dependent.

χ-square test as a test of goodness of fit

- Suppose that in a certain sample the individuals can be grouped in k categories $C_1, C_2, \ldots, C_k$ with observed frequencies $o_1, o_2, \ldots, o_k$. But on the other hand it is expected that the theoretical absolute frequencies are $e_1, e_2, \ldots, e_k$.
- The statistic χ^2 is

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} = \sum_{i=1}^{k} \frac{o_i^2}{e_i} - m, \tag{1}$$

  where m is the total (absolute) frequency: $m = \sum_{i=1}^{k} o_i$.
- Obviously, $\chi^2 = 0$ if and only if the expected and the observed frequencies are pairwise equal.
- The larger the value of χ^2, the greater the difference between the theoretical and the observed frequencies.

χ-square test as a test of goodness of fit

- It is considered that the above statistic follows very close a χ-square distribution if the expected frequencies are at least equal to 5.
- As a test of goodness of fit the χ-square test aims to determine how well a theoretical distribution fits an empirical one (that obtained from the sample). In this case $df = k - 1 - s$, if the expected frequencies can be computed only by estimating s population parameters from sample statistics.
- Note that $df = k - 1$, if the expected frequencies can be computed without having to estimate the population parameters from sample statistics, because if we know $(n - 1)$ of the expected frequencies, the remaining frequency can be determined.

χ-square test as a test of goodness of fit - Example

- Zacariah Labby from University of Chicago repeated in 2009 the following experiment of Walter Weldon (from 1894): he rolled 315, 672 times a die[1].
- The experiment was designed to decide if faces with numbers 5 and 6 have a larger frequency (because an inexpensive die has the pips carved-out on each face).
- The results are given in the following table:

	Outcome					
	1	2	3	4	5	6
Observed	53,222	52,118	52,465	52,338	52,224	52,285
Expected	52,612	52,612	52,612	52,612	52,612	52,612

[1] In fact he rolled twelve dice 26, 306 times for speeding-up the experience.

χ-square test as a test of goodness of fit - Example

- The question is if these results provide enough evidence that the six faces are not equally likely to come up.

$$H_0: \text{ there is no difference between the two distributions}$$

$$H_a: \text{ there is a difference between the two distributions}$$

- Using (??) the χ^2-statistic is

$$\chi^2 = 24.73$$

- The critical value for the required number of degrees of freedom ($df = 6 - 1 = 5$) is:

$$\chi^{2*} = 11.07 \text{ for } \alpha = 0.05 \text{ and}$$

$$\chi^{2*} = 15.08 \text{ for } \alpha = 0.01.$$

χ-square test as a test of goodness of fit - Example

- The null hypothesis is rejected if $\chi^2 > \chi^{2*}$, which is the case for both level of significance. Hence we can conclude that there exists a bias: from the observed data we understand that the faces 1 and 6 (which are opposite) occur with higher frequency (higher than the other two pairs $3-4$ and $2-5$).
- The χ-square test is asymmetric to the right: we reject H_0 if the statistic is larger than the critical value.
- The critical value, for a given level of significance α can be computed in R using $\chi^{2*} = qchisq(1-\alpha, df)$.
- Note that the larger the number of degrees of freedom, the better the approximation of χ^{2*} with the $\chi^2(df)$.
- The rule of the thumb says that df must be at least 5 in order to use this test as a goodness of fit test.

χ-square test as a test of goodness of fit - Multinomial distribution

- In the above example we tested if the six categories are all uniformly distributed (with probability 1/6). More general we can test whether the observations which are random variables belong to a certain class of distribution.
- Consider a random experience which has $k \geqslant 2$ outcomes: $A_1, \ldots, A_k$ with known probabilities, $P(A_i) = p_i$, $i = \overline{1, k}$. We independently perform the experience m times.
- Let X_i be the random variable which counts how many times A_i occur. The variables X_i are not independent, in fact $\sum_{i=1}^{k} X_i = m$.
- The vector $X = (X_1, X_2, \ldots, X_k)$ follows a multinomial distribution with parameters n, $p_1, \ldots, p_k$: $M(m, p_1, \ldots, p_k)$.
- The probability mass function of X is

$$P(X = (n_1, \ldots, n_k)) = \frac{m!}{n_1! \cdot \ldots \cdot n_k!} p_1^{n_1} \cdot \ldots \cdot p_k^{n_k}.$$

χ-square test as a test of goodness of fit - Multinomial test

- The multinomial test is a statistical significance test inferring on the probabilities of a multinomial distribution $p_1, \ldots, p_k$.
- Consider a random experience which has $k \geqslant 2$ outcomes: $A_1, \ldots, A_k$; the unknown probabilities of these outcomes are $\pi_1, \ldots, \pi_k$.
- We sample m individuals in order to get a simple sample; let $o_1, \ldots, o_k$ be the observed frequencies for each possible outcome.
- We want to know if these observations provide enough evidence that the outcome probabilities are different from $(p_1, \ldots, p_k)$.

$$H_0 : (\pi_1, \ldots, \pi_k) = (p_1, \ldots, p_k)$$

$$H_a : (\pi_1, \ldots, \pi_k) \neq (p_1, \ldots, p_k)$$

χ-square test as a test of goodness of fit - Multinomial test

- The variable counting the occurrences of A_i, X_i, is $B(m, \pi_i)$ distributed, hence the expected value for this outcome is $e_i = \mathbb{E}[X_i] = m \cdot \pi_i$, for $i = \overline{1, k}$.
- (??) becomes

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - m\pi_i)^2}{m\pi_i} = \sum_{i=1}^{k} \frac{o_i^2}{m\pi_i} - m, \tag{2}$$

because $m = \sum_{i=1}^{k} m\pi_i$.

- For the remainder of the test we compute the critical value $\chi^{2*} = \chi^{2*}(\alpha)$ for the given level of significance and if $\chi^2 > \chi^{2*}$ we can reject the null hypothesis and accept the alternative one.

χ-square test as a test of goodness of fit - Multinomial test - Example

- The positions on a hypothetical roulette wheel are divided into seven colors: 6 red, 5 black, 7 orange, 10 blue, 8 yellow, 7 brown, and 2 green. For a fair wheel, the chance that the ball lands in any slot is $1/45$.
- In order to test the fairness of such a roulette we spun it 1000 times and we get the following absolute frequencies (we computed also the expected frequencies).

	Slot color						
	red	black	orange	blue	yellow	brown	green
Observed	119	117	140	204	165	136	19
Expected	120	100	140	200	160	140	40

- Do these observed values provide convincing evidence that the roulette is not fair?

χ-square test as a test of goodness of fit - Multinomial test - Example

- We perform a multinomial test using the Pearson's χ-square test:

$$H_0 :$$

$$(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7) = (6/45, 5/45, 7/45, 10/45, 8/45, 7/45, 2/45)$$

$$H_a :$$

$$(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6, \pi_7) \neq (6/45, 5/45, 7/45, 10/45, 8/45, 7/45, 2/45)$$

- The χ^2 statistic computed using (??) is 14.273.
- The critical values for 5% and 1% level of significance are

$$\chi^{2*}_{0.05} = qchisq(0.95, 6) = 12.591,$$

$$\chi^{2*}_{0.01} = qchisq(0.99, 6) = 16.811.$$

- With 5% level of significance we can conclude that the roulette is not fair, but with 1% we cannot reject the null hypothesis.

χ-square test as a test for statistical independence

- As test for statistical independence the χ-square test infers on the independence of two categorical variables. The observed values of these two variables are given in a so called contingency table.

	Y				
X	o_{11}	o_{12}	$\ldots$	o_{1r}	$o_{1,}$
	o_{21}	o_{22}	$\ldots$	o_{2r}	$o_{2,}$
	$\vdots$	$\vdots$	$\ldots$	$\vdots$	
	o_{p1}	o_{p2}	$\ldots$	o_{pr}	$o_{p,}$
	$o_{,1}$	$o_{,2}$	$\ldots$	$o_{,r}$	m

- o_{ij} is the number of observations belonging to the category i of X and j of Y. The expected frequencies under the independence (null) hypothesis are $e_{ij} = \dfrac{o_{i,} \cdot o_{,j}}{m}$,

χ-square test as a test for statistical independence

where

$$p_{i,} = \frac{o_{i,}}{m} = \sum_{j=1}^{r} \frac{o_{ij}}{m}; p_{,j} = \frac{o_{,j}}{m} = \sum_{i=1}^{p} \frac{o_{ij}}{m}.$$

- The statistic of the test is

$$\chi^2 = \sum_{i=1}^{p} \sum_{j=1}^{r} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = m \sum_{i=1}^{p} \sum_{j=1}^{r} \frac{(o_{ij}/m - p_{i,}p_{,j})^2}{p_{i,}p_{,j}} \quad (3)$$

- The number of degrees of freedom is $df = (p-1)(r-1)$, if the expected frequencies can be computed only without estimating other population parameters from sample statistics.

χ-square test as a test for statistical independence - Example

- Suppose that we want to know if the hair and eye color are related. [?] reported the following data

		Eye				
		dark	blue	hazel	green	
Hair	black	68	20	15	5	108
	brown	119	84	54	29	286
	red	26	17	14	14	71
	blond	7	94	10	16	127
		220	215	93	64	592

- We perform a Pearson's χ-square test for statistical independence of the two variables ("Eye color" and "Hair color"):

$$H_0 : \text{ the variables are independent}$$

$$H_a : \text{ the variables are dependent}$$

χ-square test as a test for statistical independence - Example

- Under the null hypothesis the statistic (score) of the test is $\chi^2 = 138.29$.
- The number of degrees of freedom is $(4-1)(4-1) = 9$, hence the critical values for the usual levels of significance are

$$\chi^{2*}_{0.05} = qchisq(0.95, 9) = 16.918,$$

$$\chi^{2*}_{0.01} = qchisq(0.99, 9) = 21.665.$$

- Since the score is greater than the critical value, we can reject the null hypothesis and accept that the two variables are dependent for both levels of significance.
- Note that when we reject the null hypothesis and accept the alternative one, while the dependence exists we cannot say nothing about the "relation" between variables.

Inference about ratio of variances

- The inference procedure to be presented in this section is about standard deviations (or variances) for two normally distributed populations.
- The normality assumptions are very important for this test.
- We choose two simple random independent samples (of size n_1 and n_2, respectively) with sample standard deviations s_1 and s_2.
- Consider that, if the null hypothesis is true, the true standard deviations σ_1 and σ_2 are equal.
- The following statistic is Fisher distributed

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

with $r_1 = n_1 - 1$ and $r_2 = n_2 - 1$ numbers of degrees of freedom.

Fisher distribution $F(r_1, r_2)$

- There exists a family of Fisher distributions. Each such distribution has two numbers of degrees of freedom.
- Properties of Fisher distribution :
  - F has nonegative values;
  - F is asymmetric.
- For the inferences in this section the numbers of degrees of freedom are $r_1 = n_1 - 1$, $r_2 = n_2 - 1$.

F-test - Inference about ratio of variances

- The test is performed like follows:

1. We first formulate the *null hypothesis*:

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

2. We formulate the *alternative hypothesis* according to the information gathered from the sample. We can have two[2] different types of alternative hypothesis

$$H_a : \frac{\sigma_1}{\sigma_2} > 1$$ (*right asymmetric*) for an *one-tailed test*

$$H_a : \frac{\sigma_1}{\sigma_2} \neq 1$$ (*symmetric hypothesis*) for a *two-tailed test*.

[2] It is recommended that the "larger" or "expected to be larger" variance be the numerator.

F-test - Inference about ratio of variances

3. We choose the level of significance $\alpha \in \{1\%, 5\%\}$.
4. We compute the *F-score* (the *statistic* of the test)

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

5. We determine the corresponding critical values

$$F^* = qf(1-\alpha, n_1 - 1, n_2 - 1)$$ for right asymmetric H_a,

$$F_s^* = qf(\alpha/2, n_1 - 1, n_2 - 1), F_d^* = qf(1 - \alpha/2, n_1 - 1, n_2 - 1)$$

for symmetric H_a.

F-test - Inference about ratio of variances

6. We compare the the critical value with the F-score; if the F-score belongs to the *rejection region*, then H_a is accepted and H_0 is rejected. The rejection regions are:

$$[F^*, +\infty) \text{ for right asymmetric } H_a,$$

$$(0, F_s^*] \cup [F_d^*, +\infty) \text{ for symmetric } H_a.$$

If the F-score doesn't belong to the rejection region we say that *there is not sufficient evidence at the α level of significance* to reject the null hypothesis (*we fail to reject H_0*).

F-test - Inference about ratio of variances - Example

- A soft drink bottling company has a machine that fills 16 oz bottles. The company needs to control the standard deviation σ (or variance σ^2) in the amount of soft drink into each bottle. A correct mean amount does not ensure that the filling machine is working correctly. If the variance is too large, many bottles will be overfilled and many underfilled.
- Thus, the bottling company wants to maintain as small a standard deviation (or variance) as possible.
- The company wants to decide whether to install a modern, high-speed bottling machine.
- There are, of course, many concerns in making this decision, and one of them is that the increased speed may result in increased variability in the amount of fill placed into each bottle; such an increase would not be acceptable.

F-test - Inference about ratio of variances - Example

- To this concern, the manufacturer of the new system responded that the variance in fills will be no greater with the new machine than with the old.
- For the new machine a sample of 25 bottles gives a sample variance of 0.0018, and for the present machine a sample of 22 bottles gives a sample variance of 0.0008.

Solution:

- We will test both kind of alternative hypothesis: first that the variances of old and new machines are different, and, then, the fact that the new bottling machine has a smaller variance.
- We gather the data concerning the two populations and the samples: $n_1 = 25$, $s_1^2 = 0.0018$, $n_2 = 22$, and $s_2^2 = 0.0008$.

F-test - Inference about ratio of variances - Example

1-2. We formulate the null hypothesis and a symmetric alternative hypothesis

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \quad H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1.$$

3. We choose $\alpha = 5\%$.
4. We compute the *F-score* of the test

$$F = \frac{s_1^2}{s_2^2} = 2.2500$$

5. The critical values are $F_s^* = qf(\alpha/2, n_1 - 1, n_2 - 1) = 0.4327$, $F_d^* = qf(1 - \alpha/2, n_1 - 1, n_2 - 1) = 2.3675$, for $\alpha = 5\%$.

F-test - Inference about ratio of variances - Example

6. Since $F \in [F_s^*, F_d^*]$, we fail to reject the null hypothesis (there is not sufficient evidence at 5% level of significance that the two variances differ); the difference between the two sample variances is due only to the chance.

- For the other level of significance:

5'. For $\alpha = 1\%$ the critical values will give a larger interval: $F_s^* = qf(\alpha/2, n_1 - 1, n_2 - 1) = 0.3294$, $F_d^* = qf(1 - \alpha/2, n_1 - 1, n_2 - 1) = 3.1473$.

6'. The conclusion must be the same: we cannot reject the null hypothesis.

- As with the Z- and T-tests, if H_0 is not rejected for $\alpha = 5\%$ it will not be rejected for $\alpha = 1\%$ also (the rejection area will be larger).

F-test - Inference about ratio of variances - Example revisited

- Because the first variance is greater than the second we can perform a new test with a right asymmetric alternative hypothesis.

1-2. We formulate the null hypothesis and a symmetric alternative hypothesis

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \quad H_a : \frac{\sigma_1^2}{\sigma_2^2} > 1.$$

3. We choose $\alpha = 5\%$.
4. The *F-score* of the test will be the same

$$F = \frac{s_1^2}{s_2^2} = 2.2500$$

5. The critical value is $F^* = qf(1 - \alpha, n_1 - 1, n_2 - 1) = 2.0540$, for $\alpha = 5\%$.

F-test - Inference about ratio of variances - Example revisited

6. Since $F \in [F^*, +\infty)$, we reject the null hypothesis and accept the alternative hypothesis: the variance of the new machine is greater than that of the old.

5'. For $\alpha = 1\%$ the critical value is: $F^* = qf(1-\alpha, n_1 - 1, n_2 - 1) = 2.8010$.

6'. Since $F < F^*$, we fail to reject the null hypothesis (there is not sufficient evidence at 1% level of significance to assume that the the variance of the new machine is greater).

End.

Bibliography

- Freedman, D., R. Pisani, R. Purves, *Statistics*, W. W. Norton & Company, 4th edition, 2007.
- Johnson, R., P. Kuby, *Elementary Statistics*, Brooks/Cole, Cengage Learning, 11th edition, 2012.
- Spiegel, M. R., Stephens, L. J., *Schaum's Outline of Statistics - Theory and problems of statistics*, McGraw-Hill Education, 4th edition, 2008.
- Shao, J., *Mathematical Statistics*, Springer Verlag, 1998.
- Snee, R. D., *Graphical Display of Two-Way Contingency Tables*, The American Statistician, 28 (1), pp 9–12, 1974.