Statistics - handout

-

Andras Vetier

January 13, 2017

Contents

1 Regression in one-dimension **2**

2 Regression in two-dimensions **3**

3 Linear regression **4**

4 Confidence intervals **6**

5 U-tests **10**

6 T-tests **21**

7 Chi-square-test for fitness **23**

8 Chi-test for standard deviation (Chi-square-test for variance) **24**

9 F-test for equality of variances (of standard deviations) **25**

10 Test with ANOVA (Analysis of variance) **27**

1 Regression in one-dimension

Imagine that we work with a random variable X. Let us make N experiments, and let the observed values be $X_1, X_2, \ldots, X_N$. If we replace each observed value by a constant c, then we make an error at each replacement.

Minimizing the expected value of the absolute error. The absolute values of the errors are:

$$|X_1 - c|, |X_2 - c|, \ldots, |X_N - c|$$

The average of the absolute errors is

$$\frac{|X_1 - c| + |X_2 - c| + \ldots + |X_N - c|}{N}$$

For large N, this average is approximated by the expected value of $|X - c|$:

$$\frac{|X_1 - c| + |X_2 - c| + \ldots + |X_N - c|}{N} \approx$$

$$\mathbf{E}(|X - c|) = \int_{-\infty}^{\infty} |x - c| \; f(x)\, dx$$

We learned in Part III that this integral is minimal if c is the median of X.

Minimizing the expected value of the squared error. The squares of the errors are:

$$(X_1 - c)^2, (X_2 - c)^2, \ldots, (X_N - c)^2$$

The average of the squared errors is

$$\frac{(X_1 - c)^2 + (X_2 - c)^2 + \ldots + (X_N - c)^2}{N}$$

For large N, this average is approximated by the expected value of $(X - c)^2$:

$$\frac{(X_1 - c)^2 + (X_2 - c)^2 + \ldots + (X_N - c)^2}{N} \approx$$

$$\mathbf{E}\left((X - c)^2\right) = \int_{-\infty}^{\infty} (x - c)^2 \; f(x)\, dx$$

We learned in Part III that this integral is minimal if c is the expected value of X.

2 Regression in two-dimensions

Imagine that we work with a two-dimensional random variable (X, Y). Let us make N experiments, and let the observed values be $(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)$. If we replace each observed Y-value by a function of the X-value, that is, Y is replaced by $k(X)$, then we make an error at each replacement.

Minimizing the expected value of the absolute error. The absolute values of the errors are:

$$|Y_1 - k(X_1)|, |Y_2 - k(X_2)|, \ldots, |Y_N - k(X_N)|$$

The average of the absolute errors is

$$\frac{|Y_1 - k(X_1)| + |Y_2 - k(X_2)| + \ldots + |Y_N - k(X_N)|}{N} \approx$$

For large N, this average is approximated by the expected value of $|Y - k(X)|$:

$$\frac{|Y_1 - k(X_1)| + |Y_2 - k(X_2)| + \ldots + |Y_N - k(X_N)|}{N} \approx$$

$$\mathbf{E}(|Y - k(X)|) = \iint_{R^2} |y - k(x)| \; f(x, y)\, dx\; dy =$$

$$\iint_{R^2} |y - k(x)| \; f_1(x) \; f_{2|1}(y|x))\, dx\; dy =$$

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |y - k(x)| \; f_{2|1}(y|x)) \; dy \right) f_1(x) \; dx$$

For all x, the inner integral is minimal if $k(x)$ is the conditional median, that is, the median of the conditional distribution associated to the condition $X = x$. The conditional median can be calculated from the equation

$$F_{2|1}(y|x) = \frac{1}{2}$$

so that we express y in term of x to get the function $y = k(x)$.

Minimizing the expected value of the squared error. The squares of the errors are:

$$(Y_1 - k(X_1))^2, (Y_2 - k(X_2))^2, \ldots, (Y_N - k(X_N))^2$$

The average of the squared errors is

$$\frac{(Y_1 - k(X_1))^2 + (Y_2 - k(X_2))^2 + \ldots + (Y_N - k(X_N))^2}{N}$$

For large N, this average is approximated by the expected value of $(Y - k(X))^2$:

$$\frac{(Y_1 - k(X_1))^2 + (Y_2 - k(X_2))^2 + \ldots + (Y_N - k(X_N))^2}{N} \approx$$

$$\mathbf{E}\left((Y - k(X))^2\right) = \iint_{R^2} (y - k(x))^2 \; f(x,y)\,dx\;dy =$$

$$\iint_{R^2} (y - k(x))^2 \; f_1(x)\; f_{2|1}(y|x))\,dx\;dy =$$

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} (y - k(x))^2 \; f_{2|1}(y|x))\;dy \right) f_1(x)\;dx$$

For all x, the inner integral is minimal if $k(x)$ is the conditional expected value, that is, the expected value of the conditional distribution associated to the condition $X = x$. The conditional expected value can be calculated by integration:

$$k(x) = m_1(x) = \int_{-\infty}^{\infty} y\; f_{2|1}(y|x)\;dy$$

3 Linear regression

Imagine that we work with a two-dimensional random variable (X, Y). Let us make N experiments, and let the observed values be $(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)$. If we replace each observed Y-value by a linear function of the X-value, that is, Y is replaced by $aX + b$, then at each of the replacements we make an error. The squares of the errors are:

$$(Y_1 - (aX_1 + b))^2\,,\;\; (Y_2 - (aX_2 + b))^2\,,\;\; \ldots,\;\; (Y_N - (aX_N + b))^2$$

The average of the squared errors is

$$\frac{(Y_1 - (aX_1 + b))^2 + (Y_2 - (aX_2 + b))^2 + \ldots + (Y_N - (aX_N + b))^2}{N}$$

For large N, this average is approximated by the expected value of $(Y - (aX + b))^2$:

$$\frac{(Y_1 - (aX_1 + b))^2 + (Y_2 - (aX_2 + b))^2 + \ldots + (Y_N - (aX_N + b))^2}{N} \approx$$

$$\mathbf{E}\left((Y - (aX + b))^2\right) = \iint_{R^2} (y - (ax + b))^2 \; f(x,y)\,dx\;dy$$

We may be interested in finding the values of a and b so that the expected value of the squared error is minimal.

Solution. Expanding the square $(y-(ax+b))^2$ in the above integral, we get six terms, so the integral is equal to the sum of six integrals as follows:

$$\mathbf{E}\left((Y-(aX+b))^2\right) =$$

$$\iint\limits_{R^2} y^2\ f(x,y)\,dx\ dy + a^2 \iint\limits_{R^2} x^2\ f(x,y)\,dx\ dy + b^2 \iint\limits_{R^2} f(x,y)\,dx\ dy -$$

$$-2a \iint\limits_{R^2} xy\ f(x,y)\,dx\ dy - 2b \iint\limits_{R^2} y\ f(x,y)\,dx\ dy + 2ab \iint\limits_{R^2} x\ f(x,y)\,dx\ dy$$

Each of these six integrals is a constant, so the formula itself is a two-variable quadratic formula. The values of a and b for which this quadratic formula is minimal can be determined by taking the partial derivatives of this quadratic formula with respect to a, and with respect to b, and then solving the arising system of equations for a and b. We omit the details of the calculation, the reader can make it or accept that the solution is

$$a_{\text{opt}} = r\ \frac{\sigma_2}{\sigma_1}$$

$$b_{\text{opt}} = \mu_2 \ - \ a_{\text{opt}}\ \mu_1 = \mu_2 \ - \ r \frac{\sigma_2}{\sigma_1}\ \mu_1$$

Thus, the equation of the line yielding the smallest expected value for the squared error is

$$y = \mu_2 + r\ \frac{\sigma_2}{\sigma_1}\ (x-\mu_1)$$

or, equivalently

$$\frac{y-\mu_2}{\sigma_2} = r\ \frac{x-\mu_1}{\sigma_1}$$

This line is called the **regression line**.

Expected value of the squared error. When we use the regression line, the value of the average of the squared errors is

$$\frac{(Y_1-(a_{\text{opt}}X_1+b_{\text{opt}}))^2 + (Y_2-(a_{\text{opt}}X_2+b_{\text{opt}}))^2 + \ldots + (Y_N-(a_{\text{opt}}X_N+b_{\text{opt}}))^2}{N}$$

For large N, this average is approximated by the expected value of $(Y-(a_{\text{opt}}X+b_{\text{opt}}))^2$, which is equal to

$$\iint\limits_{R^2} (y-(a_{\text{opt}}x+b_{\text{opt}}))^2\ f(x,y)\,dx\ dy$$

It can be shown that this integral is equal to

$$\sigma_2^2 \left(1 - r^2\right)$$

The expression $\sigma_2^2 \left(1 - r^2\right)$ consist of two factors. The first factor is the variance of the random variable Y. The second factor is $\left(1 - r^2\right)$. Since the expected value of the squared error cannot be negative, $\left(1 - r^2\right)$ cannot be negative, so r^2 cannot be larger than 1, that is, r is between -1 and 1, equality permitted. Moreover, $\left(1 - r^2\right)$ is a decreasing function of r^2, so the larger r^2 is, the smaller $\left(1 - r^2\right)$ is, that is, if r^2 is close to 1, then replacing Y by $a_{\text{opt}}X + b_{\text{opt}}$ causes, in most cases, a smaller error, if r^2 is close to 0, then replacing Y by $a_{\text{opt}}X + b_{\text{opt}}$ causes, in most cases, a larger error. This is why we may consider r^2 or $|r|$ as a measure of how well Y can be approximated by a linear function of X, that is, how strong the linear relationship is between X and Y.

Using a calculator. More sophisticated calculators have a key to determine the slope and the intercept of the regression line, as well.

Using Excel. In Excel, the command `SLOPE` (in Hungarian: `MEREDEKSÉG`), and `INTERCEPT` (in Hungarian: `METSZ`) give the slope and the intercept of the regression line.

4 Confidence intervals

Construction of a finite confidence interval when σ is known. Let X be a normally distributed random variable with parameters μ and σ, and let $\overline{X}_n$ be the average of n experimental results. Since $\overline{X}_n$ follows a normal distribution with parameters μ and $\sigma/\sqrt{n}$, the standardized average

$$\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

follows the standard normal distribution. So

$$\mathbf{P}\left(-x \ < \ \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \ < \ x \right) = 2\Phi(x) - 1$$

If, for a given probability value p, we choose x so that

$$2\Phi(x) - 1 = p$$

that is,

$$x = \Phi^{-1}\left(\frac{1+p}{2}\right)$$

then

$$\mathbf{P}\left(-\Phi^{-1}\left(\frac{1+p}{2}\right) \;<\; \frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}} \;<\; \Phi^{-1}\left(\frac{1+p}{2}\right)\right) = p$$

or, equivalently,

$$\mathbf{P}\left(\mu-\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1+p}{2}\right) \;<\; \overline{X}_n \;<\; \mu+\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1+p}{2}\right)\right) = p$$

which means that, with a probability p, both of the following inequalities hold:

$$\mu-\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1+p}{2}\right) \;<\; \overline{X}_n$$

$$\overline{X}_n \;<\; \mu+\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1+p}{2}\right)$$

The first inequality is equivalent to

$$\mu \;<\; \overline{X}_n+\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1+p}{2}\right)$$

The second is equivalent to

$$\overline{X}_n-\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1+p}{2}\right) \;<\; \mu$$

This is how we get that, with a probability p, both of the following inequalities hold:

$$\overline{X}_n-\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1+p}{2}\right) \;<\; \mu \;<\; \overline{X}_n+\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1+p}{2}\right)$$

which means that, with a probability p, the random interval

$$\left(\overline{X}_n-\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1+p}{2}\right)\;,\;\overline{X}_n+\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1+p}{2}\right)\right)$$

called **confidence interval**, contains the parameter μ. The center of the interval is $\overline{X}_n$, the radius (the half of the length) of the interval is $\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1+p}{2}\right)$. Notice that the radius (the half of the length) of the interval is a constant, that is, it does not depend on randomness.

The result we have can be interpreted also like this: the random point $\overline{X}_n$, with a probability p, is an estimation for μ: if μ is not known for us, then we are able to declare a random point based on experimental results, so that, with a probability p, this random point is so close to μ that their difference is less than $\frac{\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{1+p}{2}\right)$.

Construction of an infinitely long confidence interval when σ is known. Let X be again a normally distributed random variable with parameters μ and σ. Let $\overline{X}_n$

be the average if n experimental result. Since $\overline{X}_n$ follows a normal distribution with parameters μ and $\sigma/\sqrt{n}$, the standardized average

$$\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

follows the standard normal distribution. So

$$\mathbf{P}\left(\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \; < \; x\right) = \Phi(x)$$

If, for a given probability value p, we choose x so that $\Phi(x) = p$, that is, $x = \Phi^{-1}(p)$, then

$$\mathbf{P}\left(\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \; < \; \Phi^{-1}(p)\right) = p$$

or, equivalently,

$$\mathbf{P}\left(\overline{X}_n \; < \; \mu + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(p)\right) = p$$

which means that, with a probability p, the following inequality holds:

$$\overline{X}_n \; < \; \mu + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(p)$$

This inequality is equivalent to

$$\overline{X}_n - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(p) \; < \; \mu$$

This is how we get that, with a probability p, the following inequality holds:

$$\overline{X}_n - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(p) \; < \; \mu$$

which means that, with a probability p, the infinitely long random interval with the left end point

$$\overline{X}_n - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(p)$$

called **confidence interval**, contains the parameter μ.

The result we have can be interpreted also like this: the point

$$\overline{X}_n - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(p)$$

with a probability p, is a lower bound for μ: if μ is not known for us, then we are able to declare a random point based on experimental results, so that, with a probability p, this random point is less than μ.

Construction of a finite confidence interval when σ is not known. Let X be a normally distributed random variable with parameters μ and σ, and let $\overline{X}_n$ be the average of the experimental results $X_1, X_2, \ldots, X_n$. If σ is not known for us, then we may replace it by the sample standard deviation, which is

$$s_n^* = \sqrt{\frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \ldots + (X_n - \overline{X})^2}{n-1}}$$

The random variable, which is a modification of the standardized average,

$$\frac{\overline{X}_n - \mu}{\frac{s_n^*}{\sqrt{n}}}$$

follows the t-distribution with degrees of freedom $n-1$. (Accept this fact without a proof.) So, using the distribution function $F(x)$ of the t-distribution with degrees of freedom $n-1$, we get that

$$\mathbf{P}\left(-x \ < \ \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \ < \ x \right) = 2F(x) - 1$$

If, for a given probability value p, we choose x so that

$$2F(x) - 1 = p$$

that is,

$$x = F^{-1}\left(\frac{1+p}{2}\right)$$

then

$$\mathbf{P}\left(-F^{-1}\left(\frac{1+p}{2}\right) \ < \ \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \ < \ F^{-1}\left(\frac{1+p}{2}\right) \right) = p$$

or, equivalently,

$$\mathbf{P}\left(\mu - \frac{s_n^*}{\sqrt{n}} F^{-1}\left(\frac{1+p}{2}\right) \ < \ \overline{X}_n \ < \ \mu + \frac{s_n^*}{\sqrt{n}} F^{-1}\left(\frac{1+p}{2}\right) \right) = p$$

which means that, with a probability p, both of the following inequalities hold:

$$\mu - \frac{s_n^*}{\sqrt{n}} F^{-1}\left(\frac{1+p}{2}\right) \ < \ \overline{X}_n$$

$$\overline{X}_n \; < \; \mu + \frac{s_n^*}{\sqrt{n}} F^{-1}\left(\frac{1+p}{2}\right)$$

The first inequality is equivalent to

$$\mu \; < \; \overline{X}_n + \frac{s_n^*}{\sqrt{n}} F^{-1}\left(\frac{1+p}{2}\right)$$

The second is equivalent to

$$\overline{X}_n - \frac{s_n^*}{\sqrt{n}} F^{-1}\left(\frac{1+p}{2}\right) \; < \; \mu$$

This is how we get that, with a probability p, both of the following inequalities hold:

$$\overline{X}_n - \frac{s_n^*}{\sqrt{n}} F^{-1}\left(\frac{1+p}{2}\right) \; < \; \mu \; < \; \overline{X}_n + \frac{s_n^*}{\sqrt{n}} F^{-1}\left(\frac{1+p}{2}\right)$$

which means that, with a probability p, the random interval

$$\left(\overline{X}_n - \frac{s_n^*}{\sqrt{n}} F^{-1}\left(\frac{1+p}{2}\right) \; , \; \overline{X}_n + \frac{s_n^*}{\sqrt{n}} F^{-1}\left(\frac{1+p}{2}\right)\right)$$

called **confidence interval**, contains the parameter μ. The center of the interval is $\overline{X}_n$, the radius (the half of the length) of the interval is

$$\frac{s_n^*}{\sqrt{n}} F^{-1}\left(\frac{1+p}{2}\right)$$

The radius (the half of the length) of the interval is now not a constant, but it depends on randomness.

The result we have can be interpreted also like this: the random point $\overline{X}_n$, with a probability p, is an estimation for μ: if μ is not known for us, then we are able to declare a random point based on experimental results, so that, with a probability p, this random point is so close to μ that their difference is less than $\frac{s_n^*}{\sqrt{n}} F^{-1}\left(\frac{1+p}{2}\right)$.

5 U-tests

Six tests, called U-tests (also called Z-tests) will be discussed in this chapter. The six cases are:

1. U-test 1: Case of "less than", when n is given
2. U-test 2: Case of "less than", when n is calculated
3. U-test 3: Case of "equality", when n is given
4. U-test 4: Case of "equality", when n is calculated

5. U-test 5: Case of "equality", when an interval is considered instead of the point μ_0
6. U-test 6: Case of "two populations"

In all of these tests, except the last one, which is at the end of this chapter, X is a normally distributed random variable, $\overline{X}_n$ is the average of n experimental results for X.

U-test 1: Case of "less than", when n is given. X is a normally distributed random variable with parameters μ and σ, where σ is known, but μ is not known. μ_0 is a given value. On the basis of n experimental results for X (where n is given), we want to decide whether the hypothesis $\mu \leq \mu_0$ holds or does not hold. For a given p_0 probability value, which is (a little bit) less than 1, we require that

1. if $\mu < \mu_0$, then the probability of accepting the hypothesis is greater than p_0,
2. if $\mu = \mu_0$, then the probability of accepting the hypothesis is equal to p_0,
3. if $\mu > \mu_0$, then the probability of accepting the hypothesis is less than p_0,
4. if μ is "very large", then the probability of accepting the hypothesis is "very small".

Solution. For fixed μ, σ, n and b, the probability

$$\mathbf{P}\left(\overline{X}_n < b\right) = \Phi\left(\frac{b-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

is equal to the area on the left side of b under the graph of the density function of $\overline{X}_n$. For fixed σ, n and b, this expression is a function of μ, which is called the **power function** of the U-test.

Now, for fixed σ, μ_0, p_0 (≈ 1) and n, we look for b so that

$$\Phi\left(\frac{b-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = p_0$$

Here is the solution to this equation:

$$\frac{b-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \Phi^{-1}(p_0)$$

$$b = \mu_0 + \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}(p_0)$$

Since

1. for $\mu = \mu_0$, we have $\mathbf{P}\left(\overline{X}_n < b\right) = \Phi\left(\frac{b-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = p_0$,

2. the power function is a strictly decreasing function of μ, approaching 0 at ∞,

we get that the following test works as required.

U-test 1: Case of "less than", when n is given. *We take the average $\overline{X}_n$ of the experimental results, and compare it to the critical value b, which was determined above in terms of given parameter values. If $\overline{X}_n < b$, then we accept the hypothesis , otherwise we reject the hypothesis.*

U-test 2: Case of "less than", when n is calculated. X is a normally distributed random variable with parameters μ and σ, where σ is known, but μ is not known. $\mu_0 < \mu_1$ are given values. On the basis of n experimental results for X, we want to decide whether the hypothesis $\mu \leq \mu_0$ holds or does not hold. Contrary to U-test 1, now n is not given. In this U-test, we have to determine n so that more requirements will be satisfied. Namely, for a given p_0 probability value, which is (a little bit) less than 1, and for a given p_1 (small) probability value, we require that

1. if $\mu < \mu_0$, then the probability of accepting the hypothesis is greater than p_0,

2. if $\mu = \mu_0$, then the probability of accepting the hypothesis is equal to p_0,

3. if $\mu = \mu_1$, then the probability of accepting the hypothesis is equal to p_1,

4. if $\mu > \mu_1$, then the probability of accepting the hypothesis is less than p_1,

5. if μ is "very large", then the probability of accepting the hypothesis is "very small".

Solution. Now we look for b and n so that

$$\Phi\left(\frac{b-\mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = p_0$$

$$\Phi\left(\frac{b-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right) = p_1$$

Here is the solution to this system of equations:

$$\frac{b-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \Phi^{-1}\left(p_0\right)$$

$$\frac{b-\mu_1}{\frac{\sigma}{\sqrt{n}}} = \Phi^{-1}\left(p_1\right)$$

$$b-\mu_0 = \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}\left(p_0\right)$$

$$b - \mu_1 = \frac{\sigma}{\sqrt{n}}\, \Phi^{-1}(p_1)$$

$$\mu_1 - \mu_0 = \frac{\sigma}{\sqrt{n}}\, \left(\Phi^{-1}(p_0) - \Phi^{-1}(p_1)\right)$$

$$n = \left(\frac{\sigma}{\mu_1 - \mu_0}\, \left(\Phi^{-1}(p_0) - \Phi^{-1}(p_1)\right)\right)^2$$

$$b = \mu_0 + \frac{\sigma}{\sqrt{n}}\, \Phi^{-1}(p_0)$$

$$= \mu_0 + (\mu_1 - \mu_0)\frac{\Phi^{-1}(p_0)}{\Phi^{-1}(p_0) - \Phi^{-1}(p_1)}$$

Since

1. for $\mu = \mu_0$, we have $\mathbf{P}\left(\overline{X}_n < b\right) = p_0$,
2. for $\mu = \mu_1$, we have $\mathbf{P}\left(\overline{X}_n < b\right) = p_1$,
3. the power function is a strictly decreasing function, approaching 0 at ∞,

we get that the following test works as required.

U-test 2: Case of "less than", when n is calculated. *We calculate n and b according to the above formulas, and then we take the average $\overline{X}_n$ of the experimental results, and compare it to b. If $\overline{X}_n < b$, then we accept the hypothesis , otherwise we reject the hypothesis.*

Remark. In both of the above tests, we have to compare the average $\overline{X}_n$ of the experimental results to b, that is, we have to analyze whether the inequality

$$\overline{X}_n < \mu_0 + \frac{\sigma}{\sqrt{n}}\, \Phi^{-1}(p_0)$$

holds or does not hold. This inequality is equivalent to the following inequality:

$$\frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \;<\; \Phi^{-1}(p_0)$$

The expression on the left side of this inequality is called the U-value of the tests:

$$U = \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Using the notion of the U-value, both of the above tests may be performed so that we calculate the U-value, and compare it the so called **standardized critical value**

$$U_{\text{crit}} = \Phi^{-1}(p_0)$$

Notice that the standardized critical value U_{crit} depends only on the probability p_0.

U-test 1 and 2, cases of "less than" - standardized. *We determine U from the experimental results, and we calculate* U_{crit}, *and then compare them. If* $U < U_{crit}$, *then we accept the hypothesis, otherwise we reject the hypothesis.*

U-test 3: Case of "equality", when n is given. X is a normally distributed random variable with parameters μ and σ, where σ is known, but μ is not known. μ_0 is a given value. On the basis of n experimental results for X (where n is given), we want to decide whether the hypothesis $\mu = \mu_0$ holds or does not hold. For a given p_0 probability value, which is (a little bit) less than 1, we require that

1. if $\mu = \mu_0$, then the probability of accepting the hypothesis is equal to p_0,
2. if $\mu \neq \mu_0$, then the probability of accepting the hypothesis is less than p_0,
3. if μ is "farther and farther from μ_0", then the probability of accepting the hypothesis is "smaller and smaller",
4. if μ is "very far from μ_0", then the probability of accepting the hypothesis is "very small".

Solution. For fixed μ, σ, n, a and b, the probability

$$\mathbf{P}\left(a < \overline{X}_n < b\right) = \Phi\left(\frac{b-\mu}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(\frac{a-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

is equal to the area between a and b under the graph of the density function of $\overline{X}_n$. For fixed σ, n, a, b, this expression defines a function of μ, which is called the power function of the U-test.

For a given μ_0, and Δb, the numbers $a = \mu_0 - \Delta b$ and $b = \mu_0 + \Delta b$ define a symmetrical interval around μ_0, and when $\mu = \mu_0$, then

$$\mathbf{P}\left(\mu_0 - \Delta b < \overline{X}_n < \mu_0 + \Delta b\right) = 2\Phi\left(\frac{\Delta b}{\frac{\sigma}{\sqrt{n}}}\right) - 1$$

Now we look for Δb so that

$$2\Phi\left(\frac{\Delta b}{\frac{\sigma}{\sqrt{n}}}\right) - 1 = p_0$$

Here is the solution to this equation:

$$\Delta b = \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}\left(\frac{1+p_0}{2}\right)$$

Since

1. for $\mu = \mu_0$, we have $\mathbf{P}\left(\mu_0 - \Delta b < \overline{X}_n < \mu_0 + \Delta b\right) = p_0$,

2. for$\mu \neq \mu_0$, we have $\mathbf{P}\left(\mu_0 - \Delta b < \overline{X}_n < \mu_0 + \Delta b\right) < p_0$,

3. the power function is strictly increasing on the left side of μ_0, and strictly decreasing on the right side of μ_0, and it is approaching 0 both at $-\infty$ and at ∞,

we get that the following test works as required.

U-test 3: Case of "equality", when n is given. *We take the average $\overline{X}_n$ of the experimental results, and compare it to the critical values $\mu_0 - \Delta b$, $\mu_0 + \Delta b$. If $\overline{X}_n$ is between them, then we accept the hypothesis , otherwise we reject the hypothesis.*

U-test 4: Case of "equality", when n is calculated. X is a normally distributed random variable with parameters μ and σ, where σ is known, but μ is not known. $\mu_0 < \mu_1$ are given values. On the basis of n experimental results for X, we want to decide whether the hypothesis $\mu = \mu_0$ holds or does not hold. Contrary to U-test 3, now n is not given, we have to determine n so that, in this U-test, more requirements will be satisfied. Namely, for a given p_0 probability value, which is (a little bit) less than 1, and for a given p_1 (small) probability value, we require that

1. if $\mu = \mu_0$, then the probability of accepting the hypothesis is equal to p_0,

2. if $\mu = \mu_1$, then the probability of accepting the hypothesis is equal to p_1,

3. if μ is "farther and farther from μ_0", then the probability of accepting the hypothesis is "smaller and smaller",

4. if μ is "very far from μ_0", then the probability of accepting the hypothesis is "very small".

Solution. The probability

$$\mathbf{P}\left(a < \overline{X}_n < b\right) = \Phi\left(\frac{b-\mu}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(\frac{a-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

is equal to the area between a and b under the graph of the density function of $\overline{X}_n$. For fixed σ, a and b and n, this expression is a function of μ, the so called power function of the U-test. The power function obviously increases until the center of the interval $[a, b]$, and then decreases, and approaches 0 both at $-\infty$ and ∞. For a given μ_0, and Δb, the numbers $a = \mu_0 - \Delta b$ and $b = \mu_0 + \Delta b$ define a symmetrical interval around μ_0, and

$$\mathbf{P}\left(\mu_0 - \Delta b < \overline{X}_n < \mu_0 + \Delta b\right) = \Phi\left(\frac{(\mu_0 + \Delta b) - \mu}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(\frac{(\mu_0 - \Delta b) - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

When $\mu = \mu_0$, then

$$\mathbf{P}\left(\mu_0 - \Delta b < \overline{X}_n < b = \mu_0 + \Delta b\right) = 2\Phi\left(\frac{\Delta b}{\frac{\sigma}{\sqrt{n}}}\right) - 1$$

Now we look for Δb and n so that

$$2\Phi\left(\frac{\Delta b}{\frac{\sigma}{\sqrt{n}}}\right) - 1 = p_0$$

$$\Phi\left(\frac{(\mu_0 + \Delta b) - \mu_1}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(\frac{(\mu_0 - \Delta b) - \mu_1}{\frac{\sigma}{\sqrt{n}}}\right) = p_1$$

We can easily handle the first equation, and we get that

$$\Delta b = \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}\left(\frac{1 + p_0}{2}\right)$$

which is a simple linear relation between Δb and $\frac{1}{\sqrt{n}}$. In regards to the second equation, let us notice that $(\mu_0 - \Delta b) - \mu_1$ is a negative number far enough from 0, so

$$\Phi\left(\frac{(\mu_0 - \Delta b) - \mu_1}{\frac{\sigma}{\sqrt{n}}}\right) \approx 0$$

and, omitting this term in the second equation, we get the following approximate equation:

$$\Phi\left(\frac{(\mu_0 + \Delta b) - \mu_1}{\frac{\sigma}{\sqrt{n}}}\right) = p_1$$

From here we get:

$$(\mu_0 + \Delta b) - \mu_1 = \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}(p_1)$$

which is another simple linear relation between Δb and $\frac{1}{\sqrt{n}}$. The two linear equations constitute a system of linear equations for Δb and $\frac{1}{\sqrt{n}}$, which can be solved. From the solution we get n and Δb:

$$n = \left(\frac{\sigma}{\mu_1 - \mu_0}\,\left(\Phi^{-1}\left(\frac{1 + p_0}{2}\right) - \Phi^{-1}(p_1)\right)\right)^2$$

$$\Delta b = (\mu_1 - \mu_0)\frac{\Phi^{-1}\left(\frac{1+p_0}{2}\right)}{\Phi^{-1}\left(\frac{1+p_0}{2}\right) - \Phi^{-1}(p_1)}$$

So the following test works as required.

U-test 4: Case of "equality", when n is calculated. *We calculate the value of n from the above formula, and round up what we get. We calculate Δb, too. Then we make n experiments for X, take the average $\overline{X}_n$ of the experimental results, and compare it to the critical values $\mu_0 - \Delta b$ and $\mu_0 + \Delta b$. If $\overline{X}_n$ is between the critical values, then we accept the hypothesis , otherwise we reject the hypothesis.*

Remark. In Test 3 and Test 4, we have to compare the average $\overline{X}_n$ of the experimental results to $\mu_0 - \Delta b$ and $\mu_0 + \Delta b$, that is, we have to analyze whether the inequalities

$$\mu_0 - \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}\left(\frac{1+p_0}{2}\right) < \overline{X}_n < \mu_0 + \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}\left(\frac{1+p_0}{2}\right)$$

hold or do not hold. These inequalities are equivalent to the following inequalities:

$$-\Phi^{-1}\left(\frac{1+p_0}{2}\right) \;<\; \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \;<\; \Phi^{-1}\left(\frac{1+p_0}{2}\right)$$

The expression in the middle of these inequalities is called the U-value of the tests:

$$U = \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

and the expression on the right side is the so called **standardized critical value**:

$$U_{\text{crit}} = \Phi^{-1}\left(\frac{1+p_0}{2}\right)$$

Notice that the standardized critical value U_{crit} depends only on the probability p_0, but U_{crit} for Test 3 and Test 4 is not the same as U_{crit} for Test 1 and Test 2. Using the notions of U and U_{crit}, we may write the above inequalities like this:

$$-U_{\text{crit}} < U < U_{\text{crit}}$$

or, equivalently,

$$|U| < U_{\text{crit}}$$

We see that Test 3 and Test 4 may be performed so that we calculate the absolute value of the U-value, and compare it the standardized critical value.

U-test 3 and 4, cases of "equality" - standardized. *We determine U from the experimental results, and we calculate U_{crit}, and then we compare the absolute value of U to U_{crit}. If $|U| < U_{crit}$, then we accept the hypothesis , otherwise we reject the hypothesis.*

U-test 5: Case of "equality", when an interval is considered instead of a point. X is a normally distributed random variable with parameters μ and σ, where σ is known, but μ is not known. $\mu_0 < \mu_1 < \mu_2$ are given values. Let μ_1^T and μ_2^T mean the points on the left side of μ_0 which we get if μ_1 and μ_2 are reflected about μ_0. On the basis of n experimental results for X, we want to decide whether the hypothesis $\mu = \mu_0$ holds or does not hold. Similar to U-test 4, n is not given. We have to determine n so that, in this U-test, more requirements will be satisfied. Namely, for a given p_1 probability value, which is (a little bit) less than 1, and for a given p_2 (small) probability value, we require that

1. if $\mu_1^T < \mu < \mu_1$, then the probability of accepting the hypothesis is greater than p_1,
2. if $\mu < \mu_2^T$ or $\mu > \mu_2$, then the probability of accepting the hypothesis is smaller than p_2,
3. if μ is "farther and farther from μ_0", then the probability of accepting the hypothesis is "smaller and smaller",
4. if μ is "very far from μ_0", then the probability of accepting the hypothesis is "very small".

Remark. The first item of the above list of requirements may serve as an explanation for the name of this U-test, since the interval $[\mu_1^T, \mu_1]$ is considered instead of the point μ_0.

Solution. We will try to determine b and its reflection b^T about μ_0, and n so that

$$\Phi\left(\frac{b-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(\frac{b^T-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right) = p_1$$

$$\Phi\left(\frac{b-\mu_2}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(\frac{b^T-\mu_2}{\frac{\sigma}{\sqrt{n}}}\right) = p_2$$

Since $b^T - \mu_1$, $b^T - \mu_2$ are negative numbers far enough from 0,

$$\Phi\left(\frac{b^T-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right) \approx 0$$

$$\Phi\left(\frac{b^T-\mu_2}{\frac{\sigma}{\sqrt{n}}}\right) \approx 0$$

and we get the approximate equations:

$$\Phi\left(\frac{b-\mu_1}{\frac{\sigma}{\sqrt{n}}}\right) = p_1$$

$$\Phi\left(\frac{b-\mu_2}{\frac{\sigma}{\sqrt{n}}}\right) = p_2$$

Taking the inverse of the function Φ, we get that

$$b - \mu_1 = \frac{\sigma}{\sqrt{n}}\, \Phi^{-1}(p_1)$$

$$b - \mu_2 = \frac{\sigma}{\sqrt{n}}\, \Phi^{-1}(p_2)$$

which is system of linear equations for $\frac{1}{\sqrt{n}}$ and b. The solution for n and b is:

$$n = \left(\frac{\sigma}{\mu_2 - \mu_1} \, \left(\Phi^{-1}(p_1) - \Phi^{-1}(p_2) \right) \right)^2$$

$$b = \mu_1 + (\mu_2 - \mu_1) \frac{\Phi^{-1}(p_1)}{\Phi^{-1}(p_1) - \Phi^{-1}(p_2)}$$

So the following test works as required.

U-test 5: Case of "equality", when an interval is considered instead of a point. *We calculate the value of n from the above formula, and round up what we get. We calculate b, too. Then we make n experiments for X, take the average $\overline{X}_n$ of the experimental results, and compare it to the critical values b^T and b. If $\overline{X}_n$ is between the critical values, then we accept the hypothesis , otherwise we reject the hypothesis.*

U-test 6: Case of two populations. X_1 is a normally distributed random variable with parameters μ_1 and σ_1, X_2 is a normally distributed random variable with parameters μ_2 and σ_2. We assume that σ_1 and σ_2 are known for us, but μ_1 and μ_2 are not. On the basis of n_1 experimental results for X_1 and n_2 experimental results for X_2 (n_1 and n_2 are given), we want to decide whether the hypothesis $\mu_1 = \mu_2$ holds or does not hold. For a given p_0 probability value, which is (a little bit) less than 1, we require that

1. if $\mu_1 = \mu_2$, then the probability of accepting the hypothesis is equal to p_0,
2. if $\mu_1 \neq \mu_2$, then the probability of accepting the hypothesis is less than p_0,
3. if μ_1 is "farther and farther from μ_2", then the probability of accepting the hypothesis is "smaller and smaller",
4. if μ_1 is "very far from μ_2", then the probability of accepting the hypothesis is "very small".

Solution. The average of the n_1 experimental results for X_1 is $(\overline{X_1})_{n_1}$, and the average of the n_2 experimental results for X_2 is $(\overline{X_2})_{n_2}$. Let us consider the difference of the averages: $(\overline{X_1})_{n_1} - (\overline{X_2})_{n_2}$ The expected value of this difference is $\mu_1 - \mu_2$, its standard deviation is

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

If the random variable U is defined like this:

$$U = \frac{(\overline{X_1})_{n_1} - (\overline{X_2})_{n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

then the expected value of U is obviously

$$\frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

and its standard deviation is 1. Thus, if $\mu_1 = \mu_2$, then U follows the standard normal distribution, which means that

$$\mathbf{P}\left(-x < U < x\right) = 2\Phi(x) - 1$$

Let us choose x so that $2\Phi(x) - 1 = p_0$, that is, $x = \Phi^{-1}\left(\frac{1+p_0}{2}\right)$. Let us take this x value as a critical value U_{crit}. Thus, if $\mu_1 = \mu_2$, then we have

$$\mathbf{P}\left(-U_{\text{crit}} < U < U_{\text{crit}}\right) = p_0$$

If $\mu_1 \neq \mu_2$, then the distribution of U differs from the standard normal distribution, and

$$\mathbf{P}\left(-U_{\text{crit}} < U < U_{\text{crit}}\right) < p_0$$

If μ_1 and μ_2 are farther and farther from each other, then the distribution of U becomes more and more different from the standard normal distribution, and

$$\mathbf{P}\left(-U_{\text{crit}} < U < U_{\text{crit}}\right)$$

becomes smaller and smaller. Thus, the following test works as required.

U-test 6: Case of two populations. *We determine the value of U from the experimental results, and compare its absolute value to the critical value U_{crit}. If $|U|$ is less than U_{crit}, then we accept the hypothesis , otherwise we reject the hypothesis.*

Important remark: U-tests for NOT normally distributed random variables. At the beginning of this chapter, we assumed that X was a normally distributed random variable. However, if n, the number of experiments is large enough (say it is at least 25, or so), then $\overline{X}_n$, the average of the experimental results for X is approximately normally distributed even if X is not normally distributed. Since in the above tests we were using the normality of $\overline{X}_n$, the above tests are applicable for not normal random variables, as well, if n, the number of experiments is large enough.

6 T-tests

Tests, called T-tests will be discussed in this chapter. In all of them, X is a normally distributed random variable, $\overline{X}_n$ is the average of experimental results $X_1, X_2, \ldots, X_n$, and

$$s_n^* = \sqrt{\frac{(X_1 - \overline{X}_n)^2 + (X_2 - \overline{X}_n)^2 + \ldots + (X_n - \overline{X}_n)^2}{n-1}}$$

is the sample standard deviation.

T-test 1: Case of "equality". X is a normally distributed random variable with parameters μ and σ, where neither μ nor σ is known for us. μ_0 is a given value. On the basis of n experimental results for X (where n is given), we want to decide whether the hypothesis $\mu = \mu_0$ holds or does not hold. For a given p_0 probability value, which is (a little bit) less than 1, we require that

1. if $\mu = \mu_0$, then the probability of accepting the hypothesis is equal to p_0,
2. if $\mu \neq \mu_0$, then the probability of accepting the hypothesis is less than p_0,
3. if μ is "farther and farther from μ_0", then the probability of accepting the hypothesis is "smaller and smaller",
4. if μ is "very far from μ_0", then the probability of accepting the hypothesis is "very small".

Solution. If we knew σ, we could use a standardized U-test (see the remark after U-test 2), and we could calculate

$$U = \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Since σ is not known for us, we have to replace σ by s_n^*, the sample standard deviation. This why we consider the random variable

$$T = \frac{\overline{X}_n - \mu_0}{\frac{s_n^*}{\sqrt{n}}}$$

Based on experimental results, the value of T can be calculated. If $\mu = \mu_0$, then the random variable T follows t-distribution with degrees of freedom $n-1$, so

$$\mathbf{P}\left(-b < T < b\right) = 2F\left(b\right) - 1$$

where $F(x)$ denotes the distribution function of the t-distribution with degrees of freedom $n-1$. We choose b so that $2F(b) - 1 = p_0$, that is, $b = F^{-1}(\frac{1+p_0}{2})$. This b-value will be called the critical value, and will be denoted by T_{crit}:

$$T_{\text{crit}} = F^{-1}(\frac{1+p_0}{2})$$

Thus, if $\mu = \mu_0$, then we get:

$$\mathbf{P}\left(-T_{\text{crit}} < T < T_{\text{crit}}\right) = p_0$$

If $\mu \neq \mu_0$, then the random variable T follows a distribution different from t-distribution with degrees of freedom $n-1$, so that

$$\mathbf{P}\left(-T_{\text{crit}} < T < T_{\text{crit}}\right) < p_0$$

If μ is farther and farther from μ_0, then the distribution of the the random variable T becomes more and more different from t-distribution with degrees of freedom $n-1$, and

$$\mathbf{P}\left(-T_{\text{crit}} < T < T_{\text{crit}}\right)$$

becomes smaller and smaller. We shall not go into the mathematical details of this test. However, we hope that the simulation files given bellow will help to accept that the following test works as required.

T-test 1: Case of "equality". *We calculate the value of T from the experimental results, and compare its absolute value to the critical value T_{crit}. If $|T|$ is less than T_{crit}, then we accept the hypothesis , otherwise we reject the hypothesis.*

T-test 2: Case of "less than". X is a normally distributed random variable with parameters μ and σ, where neither μ nor σ is known for us. μ_0 is a given value. On the basis of n experimental results for X (where n is given), we want to decide whether the hypothesis $\mu \leq \mu_0$ holds or does not hold. For a given p_0 probability value, which is (a little bit) less than 1, we require that

1. if $\mu < \mu_0$, then the probability of accepting the hypothesis is greater than p_0,
2. if $\mu = \mu_0$, then the probability of accepting the hypothesis is equal to p_0,
3. if $\mu > \mu_0$, then the probability of accepting the hypothesis is less than p_0,
4. if μ is "very large", then the probability of accepting the hypothesis is "very small".

Solution. If we knew σ, we could use a standardized U-test (see the remark after U-test 4), and we could calculate

$$U = \frac{\overline{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Since σ is not known for us, we have to replace σ by s_n^*, the sample standard deviation. This why we consider the random variable

$$T = \frac{\overline{X}_n - \mu_0}{\frac{s_n^*}{\sqrt{n}}}$$

Based on experimental results, the value of T can be calculated. If $\mu = \mu_0$, then the random variable T follows t-distribution with degrees of freedom $n-1$, so

$$\mathbf{P}(T < b) = F(b)$$

where $F(x)$ denotes the distribution function of the t-distribution with degrees of freedom $n-1$. We choose b so that $F(b) = p_0$, that is, $b = F^{-1}(p_0)$. This b-value will be called critical value, and will be denoted by T_{crit}:

$$T_{\text{crit}} = F^{-1}(p_0)$$

Thus, if $\mu = \mu_0$, then we get:

$$\mathbf{P}(T < T_{\text{crit}}) = p_0$$

If $\mu < \mu_0$, then the random variable T follows a distribution different from t-distribution with degrees of freedom $n-1$, so that

$$\mathbf{P}(T < T_{\text{crit}}) > p_0$$

If $\mu > \mu_0$, then the random variable T follows a distribution different from t-distribution with degrees of freedom $n-1$, so that

$$\mathbf{P}(T < T_{\text{crit}}) < p_0$$

If μ is very large, then $\mathbf{P}(T < T_{\text{crit}}) < p_0$ becomes very small. We shall not go into the mathematical details of this test. However, we hope that the simulation files given bellow will help to accept that the following test works as required.

T-test 2: Case of "less than". *We calculate the value of T from the experimental results, and compare its absolute value to the critical value T_{crit}. If T is less than T_{crit}, then we accept the hypothesis , otherwise we reject the hypothesis.*

7 Chi-square-test for fitness

Imagine that you need a fair die, and your friend offers you one. You are glad to get a die, but you want to be convinced that the die is really fair. So you toss the die several times, and you get experimental results. How can you decide based on the experimental results whether to accept the hypothesis that the die is fair or to reject it? We will describe a test to decide whether to accept the hypothesis or to reject it. Obviously, since the test is based on experimental results, the decision may be wrong: even if the die is fair, randomness may cause us to reject the hypothesis, so even if the die is fair, the probability of accepting the hypotheses must be less than 1. Let this probability be denoted by p_0. This is why, for a given p_0 probability value, which is (a little bit) less than 1, we require that

1. if the hypothesis holds, then the test, with a probability p_0, will suggest to accept the hypothesis,

2. if the hypothesis does not hold, then the test will suggest to accept the hypothesis with a probability smaller than p_0,
3. if the die differs from a fair die only a "little bit", then the test will suggest to accept the hypothesis with a probability only a "little bit" smaller than p_0,
4. if the die is "far" from a fair die, then the test will suggest to accept the hypothesis with a "small" probability.

Solution. When you toss the die, and observe the number on the top, then the possible values are 1, 2, 3, 4, 5, 6. Make n tosses. Observe the count for each possible value, that is, how many times you have got a 1, how many times you have got a 2, and so on. Then calculate the so called expected counts, too, which means that the hypothetical probability of each possible value is multiplied by n. Then, for each possible value, take the difference between the observed count and the expected count, take the square of the difference, and then divide by the expected count. For each possible value you get a number. Now add these numbers. This sum will be denoted by K^2. Since the value of K^2 is affected by the experimental results, K^2 is a random variable. Suppose that the number of tosses was large enough to guarantee that all the expected counts are greater than 10. If this condition is not fulfilled, then this test is not applicable. If this condition is fulfilled, then the random variable K^2 approximately follows the distribution called chi-square distribution with degrees of freedom $r - 1$, where r is the number of possible values for the die. For a usual die, $r = 6$. If $F(x)$ denotes the distribution function of the chi-square distribution with degrees of freedom $r - 1$, then $\mathbf{P}(K^2 < x) = F(x)$. Let us choose x so that $F(x) = p_0$, that is, $x = F^{-1}(p_0)$. This x value will be denoted by K^2_{crit}, and will be called the critical value of the test. We do not go into the mathematical details, just state that the test given here works properly. Playing with the Excel files, given below, you may learn the algorithm of the test again, and you may have an experience that the test really works as desired.

Chi-square-test for fitness. *We calculate the value of K^2 from the experimental results, and compare it to the critical value K^2_{crit}. If K^2 is less than K^2_{crit}, then we accept the hypothesis , otherwise we reject the hypothesis.*

8 Chi-test for standard deviation (Chi-square-test for variance)

X is a normally distributed random variable with parameters μ and σ, where neither μ nor σ is known for us. σ_0 is a given value. On the basis of n experimental results for X (where n is given), we want to decide whether the hypothesis $\sigma = \sigma_0$ holds or does not hold. For a given p_0 probability value, which is (a little bit) less than 1, we require that

1. if $\sigma = \sigma_0$, then the probability of accepting the hypothesis is equal to p_0,
2. if $\sigma \neq \sigma_0$, then the probability of accepting the hypothesis is less than p_0,
3. if σ is "farther and farther from σ_0", then the probability of accepting the hypothesis is "smaller and smaller",
4. if σ is "very far from σ_0", then the probability of accepting the hypothesis is "very small".

Solution. Let us take the sample standard deviation, which is

$$s_n^* = \sqrt{\frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \ldots + (X_n - \overline{X})^2}{n-1}}$$

and let the random variable K be the sample standard deviation divided by the hypothetical standard deviation σ_0:

$$K = \frac{s_n^*}{\sigma_0}$$

If the hypothesis holds, then this random variable follows the distribution called chi-distribution with degrees of freedom $n-1$. If $F(x)$ denotes the distribution function of the chi-distribution with degrees of freedom $n-1$, then $\mathbf{P}(K < x) = F(x)$. Let us choose x so that $F(x) = p_0$, that is, $x = F^{-1}(p_0)$. This x value will be denoted by K_{crit}, and will be called the critical value of the test. We do not go into the mathematical details, just state that the test given here works properly. Playing with the Excel files, given below, you may learn the algorithm of the test again, and you may have an experience that the test really works as desired.

Chi-test for standard deviation (Chi-square-test for variance). *We calculate the value of K from the experimental results, and compare it to the critical value K_{crit}. If K is less than K_{crit}, then we accept the hypothesis , otherwise we reject the hypothesis.*

Remark. If a random variable follows a chi-distribution with degrees of freedom d, then its square follows chi-square distribution with degrees of freedom d. This is why this test is applicable to test not only the standard deviation, but the variance. If we test the variance and use the sample variance $(s_n^*)^2$, then the critical value K_{crit}^2 for $(s_n^*)^2$ should be chosen using the distribution function of the chi-square distribution.

9 F-test for equality of variances (of standard deviations)

X_1 is a normally distributed random variable with parameters μ_1 and σ_1, X_2 is a normally distributed random variable with parameters μ_2 and σ_2. The parameters μ_1, μ_2,

μ_1, μ_2 are not known for us. On the basis of n_1 experimental results for X_1 and n_2 experimental results for X_2 (n_1 and n_2 are given), we want to decide whether the hypothesis $\sigma_1 = \sigma_2$ holds or does not hold. For a given p_0 probability value, which is (a little bit) less than 1, we require that

1. if $\sigma_1 = \sigma_2$, then the probability of accepting the hypothesis is equal to p_0,
2. if $\sigma_1 \neq \sigma_2$, then the probability of accepting the hypothesis is less than p_0,
3. if σ_1 is "farther and farther from σ_2", then the probability of accepting the hypothesis is "smaller and smaller",
4. if σ_1 is "very far from σ_2", then the probability of accepting the hypothesis is "very small".

Solution. Let the sample standard deviation for X_1 be s_1^*, and the sample standard deviation for X_2 be s_2^*. Their squares are the so called sample variances: $(s_1^*)^2$ and $(s_2^*)^2$. Let us take the quotient of the sample variances:

$$F = \frac{(s_1^*)^2}{(s_2^*)^2}$$

If the hypothesis holds, then the random variable F follows the distribution called F-distribution with degrees of freedom $n_1 - 1, n_2 - 1$. Using (the inverse of) the distribution function of the F-distribution with degrees of freedom $n_1 - 1, n_2 - 1$, we choose two critical values: $F_{\text{crit(lower)}}$ and $F_{\text{crit(upper)}}$ so that

$$\mathbf{P}(F < F_{\text{crit(lower)}}) = \frac{1 - p_0}{2}$$

$$\mathbf{P}(F < F_{\text{crit(upper)}}) = \frac{1 + p_0}{2}$$

With this choice of $F_{\text{crit(lower)}}$ and $F_{\text{crit(upper)}}$, we achieve that

$$\mathbf{P}(F_{\text{crit(lower)}} < F < F_{\text{crit(upper)}}) = p_0$$

We do not go into the mathematical details, just state that the test given here works properly. Playing with the Excel files, given below, you may learn the algorithm of the test again, and you may have an experience that the test really works as desired.

F-test for equality of standard deviations (F-test for equality of variances). *We calculate the value of F from the experimental results, and compare it to the critical values $F_{\text{crit(lower)}}$ and $F_{\text{crit(upper)}}$. If F is between them, then we accept the hypothesis , otherwise we reject the hypothesis.*

10 Test with ANOVA (Analysis of variance)

We have r normally distributed random variables: $X_1, X_2, \ldots, X_r$, which have a common standard deviation σ, and possibly different expected values: $\mu_1, \mu_2, \ldots, \mu_r$. We make a certain number of experiments for each: n_1 experiments for X_1, n_2 experiments for X_2, and so on, n_r experiments for X_r. On the basis of these experimental results, we want to decide whether the hypothesis $\mu_1 = \mu_2 = \ldots = \mu_r$ holds or does not hold. For a given p_0 probability value, which is (a little bit) less than 1, we require that

1. if $\mu_1 = \mu_2 = \ldots = \mu_r$, then the probability of accepting the hypothesis is equal to p_0,
2. if $\mu_1 = \mu_2 = \ldots = \mu_r$ is not true, then the probability of accepting the hypothesis is less than p_0,
3. if $\mu_1, \mu_2, \ldots, \mu_r$ are farther and farther from each other, then the probability of accepting the hypothesis is "smaller and smaller",
4. if $\mu_1, \mu_2, \ldots, \mu_r$ are "very far from each other", then the probability of accepting the hypothesis is "very small".

Solution. We remind the reader that, for a data-set $z_1, z_2, \ldots, z_N$, the average is

$$\overline{z} = \frac{z_1 + z_2 + \ldots + z_N}{N}$$

and the variance is

$$\frac{(z_1 - \overline{z})^2 + (z_2 - \overline{z})^2 + \ldots + (z_N - \overline{z})^2}{N}$$

The n_1 experiments for X_1 constitute a data-set. Its average will be denoted by Ave_1, its variance will be denoted by Var_1. Related to the random variable X_i, we get similarly the average Ave_i and the variance Var_i $(i = 1, 2, \ldots, r)$. The number of all experiments is $n = \sum_i n_i$. The proportion of the ith data-set is $p_i = n_i/n$ $(i = 1, 2, \ldots, r)$. The quantity

$$\text{AVE of Ave} = \sum_i \text{Ave}_i \ p_i$$

will be called the (weighted) average of the averages, and

$$\text{AVE of Var} = \sum_i \text{Var}_i \ p_i$$

will be called the (weighted) average of the variances, and

$$\text{VAR of Ave} = \sum_i \left(\text{Ave}_i - \text{AVE}\right)^2 \ p_i$$

will be called the (weighted) variance of the averages.The random variable F is now defined by

$$F = \frac{\frac{\text{VAR of Ave}}{r-1}}{\frac{\text{AVE of Var}}{n-1}}$$

If the hypothesis holds, then the random variable F follows the distribution called F-distribution with degrees of freedom $r-1, n-1$. Using (the inverse of) the distribution function of the F-distribution with degrees of freedom $r-1, n-1$, we choose the critical value F_{crit} so that

$$\mathbf{P}(F < F_{\text{crit}}) = p_0$$

We do not go into the mathematical details, just state that the test given below works properly. Playing with the Excel file, given below, you may learn the algorithm of the test again, and you may have an experience that the test really works as desired.

Test with ANOVA (Analysis of variance). *We calculate the value of F from the experimental results, and compare F to the critical value F_{crit}. If F is less than F_{crit}, then we accept the hypothesis , otherwise we reject the hypothesis.*