Statistics: Testing Statistical Hypotheses

S. Maset

Dipartimento di Matematica e Geoscienze, Università di Trieste

PEM 2016-2017

Outline

Introduction

In recent years, there have been many discussions about the possible dangers of living near a high-level ElectroMagnetic Field (EMF).

After hearing many anecdotal tales concerning a larger number of cases of cancer, especially among children, in communities living near a high-level EMF, a researcher decided to study the possible dangers.

So, the researcher selected a fairly large community that was located in a high-level EMF area.

By interviewing people in schools, hospitals, and public health facilities, the researcher discovered 32 cases of cancer in children in the previous 3 years.

Then, the researcher consulted a government public health library to learn about the number of cases of childhood cancer that could be expected in a community with size equal to the size of community under consideration.

The researcher learned that the number of cases of childhood cancer over a 3-year period in such a community is a normal random variable with mean 16.2 and standard deviation 4.7.

So, is the discovery of 32 cases of childhood cancers, significantly larger than the mean 16.2, sufficient to conclude that there is some special factor in the community under study (may be a high-level EMF?) that increases the chance for children to contract cancer?

Or is it possible that there is nothing special about the community and that the greater number of cancers is due solely to chance?

We will show how to answer such questions.

Statistical Hypothesis

We recall our mathemathical structure of the Inferential Statistics.

We consider a random variable X, which is a numerical quantity related to an individual randomly selected in a population, with a distribution determined by unknown parameters.

A particular type of inference is involved with the testing of an hypothesis concerning an unknown parameter of the distribution of X.

A **statistical hypothesis** on an unknown parameter $\theta \in \mathbb{R}$ of the distribution of X is a statement of type

$$\theta \in B,$$

where B is a given subset of $\mathbb{R}$.

In the **Testing Statistical Hypotheses** process, there is a statistical hypothesis

$$\theta \in B_0,$$

where B_0 is a given subset of $\mathbb{R}$, to be tested. It is called the **null hypothesis** and is denoted by H_0.

We also assume that, when H_0 is tested, there is another hypothesis

$$\theta \in B_1,$$

where B_1 is another given subset of $\mathbb{R}$ disjoint from B_0, called the **alternative hypothesis** and denoted by H_1.

Aim of the test is to choose one between H_0 and H_1.

- Example. A tobacco firm claims that it has discovered a new treatment for tobacco leaves which makes possible to have cigarettes with mean nicotine content of 1.5 mg or less.

  In this case, X is the nicotine content of a cigarette randomly selected in the population of cigarettes produced in this new way. We assume that X has a normal distribution with unknown mean μ.

  The tobacco firm is claiming that $\mu \leq 1.5\ \mathrm{mg}$.

  Suppose that a researcher is skeptical about this claim and indeed believes that it is not true.

  To disprove the claim of the tobacco firm, the researcher decides to test the null hypothesis

$$H_0 : \mu \leq 1.5\ \mathrm{mg}, \text{ i.e. } \mu \in B_0 = (-\infty, 1.5\ \mathrm{mg}]\,,$$

  versus the alternative hypothesis

$$H_1 : \mu > 1.5\ \mathrm{mg}, \text{ i.e. } \mu \in B_1 = (1.5\ \mathrm{mg}, +\infty)\,.$$

- It is important to remark the following:
  - if the tester is trying to disprove an hypothesis H, then H is the null hypothesis H_0;
  - if the tester is trying to prove an hypothesis H, then H is the alternative hypothesis H_1.
- When we test the null hypothesis H_0 versus the alternative hypothesis H_1, the aim is to reject H_0 in favor of H_1.

  The test has the two possible results: "H_0 is not rejected" and "H_0 is rejected".

  When H_0 is rejected, we have evidence that H_0 is false and H_1 is true.

- Observe that there is not symmetry between H_0 and H_1.

  The falsity of H_0 has to be proved beyond any reasonable doubt. If there is some doubt about the falsity of H_0, we are not ready to reject H_0.

  When we test H_0 versus H_1, we always protect H_0.

  So, the situation where the test rejects H_0 when H_0 is true (called **error of type I**) is considered worse than the situation where the test does not reject H_0 when H_1 is true (**error of type II**).

  This is exactly the same as when a court has to decide

  $$H_0 : \quad \text{the defendant is not guilty}$$

  versus

  $$H_1 : \quad \text{the defendant is guilty.}$$

Test statistic

- Given the sample $\boldsymbol{x}^{\text{obs}} = \left(x_1^{\text{obs}}, \ldots, x_n^{\text{obs}}\right)$ obtained by the Sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ from the distribution of X, we use a test statistic to decide whether reject or not H_0.

  A **test statistic** is a sampling statistic $\tau : \mathbb{R}^n \to \mathbb{R}$ such that the value $\tau\left(\boldsymbol{x}^{\text{obs}}\right)$ enables to decide whether reject or not H_0.

  In the above example of the nicotine content a suitable test statistic is the sample mean, as we will see.

- Given a test statistic τ, we adopt the following rule for rejecting H_0.

  We use a region $CR \subseteq \mathbb{R}$, called **critical region**, or **rejection region**, such that

$$H_0 \text{ is not rejected if } \tau\left(\boldsymbol{x}^{\text{obs}}\right) \notin CR$$
$$H_0 \text{ is rejected if } \tau\left(\boldsymbol{x}^{\text{obs}}\right) \in CR.$$

  The complementary set of CR is called **noncritical region**, or **nonrejection region**, and it is denoted by NCR. So

$$H_0 \text{ is not rejected if } \tau\left(\boldsymbol{x}^{\text{obs}}\right) \in NCR$$
$$H_0 \text{ is rejected if } \tau\left(\boldsymbol{x}^{\text{obs}}\right) \notin NCR.$$

- The **significance level** of the test is

$$\sup_{\theta \in B_0,\ \text{i.e. } H_0 \text{ is true}} \mathbb{P}_\theta \left(H_0 \text{ is rejected} \right),$$

where $\mathbb{P}_\theta$ is the measure of probability for the experiment selection of the sample of individuals in the population, when the unknown parameter of the distribution is equal to θ.

The significance level is the supremum, when H_0 is true, of the probabilities that H_0 is rejected, i.e. the probability of having an error of type I.

The smaller the significance level, more significative is the test. In fact, if we are rejecting H_0 with a small significance level, i.e. if we are rejecting H_0 and there is a small probability to do it when H_0 is true, then we can be quite confident about the fact that H_0 is false.

- Note that we cannot speak about the probability that H_0 is false, i.e. the probability of $\theta \notin B_0$, since the truth value of $\theta \notin B_0$ is determined, not random: B_0 (given) and θ (unknown) do not depend on the outcome of the experiment of selecting a sample from the population.

  On the other hand, we can speak about the probability that H_0 is rejected, since this fact is random: it depends on the outcome of the experiment of selecting a sample from the population, more precisely it depends on the values of the Sample $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$.

- In the test, we proceed as follows: we fix $\alpha \in (0, 1)$ and, then, we choose the critical region (or equivalently the noncritical region) in order to have a significance level equal to α.

  Common used values of α are 5%, 1% and 0.1%. We say the test is **significative**, **very significative**, **extremely significative** for rejecting H_0 if $\alpha \leq 5\%, 1\%, 0.1\%$, respectively.

- By summarizing, a testing statistical hypothesis framework, also said a **statistical test**, is defined by
  - the null hypothesis H_0 and the alternative hypothesis H_1;
  - a test statistic τ;
  - a rule that associates to each $\alpha \in (0, 1)$ a noncritical region NCR_α such that the test

$$
\begin{aligned}
&H_0 \text{ is not rejected if } \tau\left(\boldsymbol{x}^{\mathrm{obs}}\right) \in NCR_\alpha \\
&H_0 \text{ is rejected if } \tau\left(\boldsymbol{x}^{\mathrm{obs}}\right) \notin NCR_\alpha
\end{aligned}
$$

  has significance level α.

The z test

- Consider the situation where the distribution of X is $\mathrm{N}(\mu, \sigma^2)$, with unknown mean μ and known standard deviation σ.

- We begin by considering the case where the null hypothesis has the form
$$H_0 : \mu = \mu_0,$$
where μ_0 is some prefixed number, and the alternative hypothesis has the form
$$H_1 : \mu \neq \mu_0.$$

For this situation, we use the sample mean as test statistic and the intervals
$$\mu_0 \pm k\frac{\sigma}{\sqrt{n}} := \left(\mu_0 - k\frac{\sigma}{\sqrt{n}}, \mu_0 + k\frac{\sigma}{\sqrt{n}}\right), \ k > 0.$$
as noncritical regions.

So

$$H_0 \text{ is not rejected if } \left|\overline{x}^{\mathrm{obs}} - \mu_0\right| < k\frac{\sigma}{\sqrt{n}}$$
$$H_0 \text{ is rejected if } \left|\overline{x}^{\mathrm{obs}} - \mu_0\right| \geq k\frac{\sigma}{\sqrt{n}}.$$

This makes sense. When H_0 is true, i.e. $\mu = \mu_0$, the Sample Mean has distribution

$$\mathrm{N}\left(\mu_0, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$

and so its values are with large probability at a distance from the mean μ_0 smaller than some units of the standard deviation $\frac{\sigma}{\sqrt{n}}$.

Moreover, observe that H_0 is rejected when μ_0 is outside of the interval estimate

$$I\left(\boldsymbol{x}^{\mathrm{obs}}\right) = \overline{x}^{\mathrm{obs}} \pm k\frac{\sigma}{\sqrt{n}}$$

of μ.

If H_0 is true, i.e. $\mu = \mu_0$, then the probability of the event $\mu = \mu_0 \notin I(\boldsymbol{X})$ implying the rejection is the quite small number $1 - C\%$, where $C\%$ is the level of confidence of the interval estimate.

So, when $\mu_0 \notin I\left(\boldsymbol{x}^{\mathrm{obs}}\right)$ happens we can be quite confident, with level of confidence $C\%$, that $\mu \neq \mu_0$.

- The significance level of the test is

$$\sup_{\mu=\mu_0,\text{ i.e. } H_0 \text{ is true}} \mathbb{P}_\mu\left(H_0 \text{ is rejected}\right) = \mathbb{P}_{\mu_0}\left(H_0 \text{ is rejected}\right)$$
$$= \mathbb{P}_{\mu_0}\left(\left|\overline{X}-\mu_0\right| \geq k\frac{\sigma}{\sqrt{n}}\right) = 1 - \mathbb{P}_{\mu_0}\left(\left|\overline{X}-\mu_0\right| < k\frac{\sigma}{\sqrt{n}}\right)$$
$$= 1 - C\%,$$

where $C\% = \mathbb{P}_{\mu_0}\left(\left|\overline{X}-\mu_0\right| < k\frac{\sigma}{\sqrt{n}}\right)$ is the level of confidence of the interval estimate

$$\overline{x}^{\text{obs}} \pm k\frac{\sigma}{\sqrt{n}}$$

of the parameter μ, when $\mu = \mu_0$.

- If a significance level $\alpha \in (0, 1)$ is fixed, the noncritical region NCR_α corresponding to a test with significance level α is

$$\mu_0 \pm k\frac{\sigma}{\sqrt{n}} \text{ where } k = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

In fact, the test has significance level α if the level of confidence of the interval estimate

$$\overline{x}^{\text{obs}} \pm k\frac{\sigma}{\sqrt{n}}$$

of μ, when $\mu = \mu_0$, is $1 - \alpha$ and this is obtained for

$$k = \Phi^{-1}\left(\frac{1 + (1 - \alpha)}{2}\right) = \Phi^{-1}\left(\frac{2 - \alpha}{2}\right) = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

- Example. Suppose that the intensity of the electromagnetic signal emitted from a particular star, measured in some particular unit, is a normal random variable with mean μ unknown and standard deviation $\sigma = 4$.

  In other words, the intensity μ of the emitted signal is altered by random noise, which is normally distributed with mean 0 and standard deviation 4. Since the random noise is always the same for any type of emitted signal, we known its standard deviation.

  A mathematical model provides the value 10 for μ, but it is suspected that this value is not correct.

  So, the null hypothesis

$$H_0 : \mu = 10$$

  is tested versus the alternative hypothesis

$$H_1 : \mu \neq 10.$$

In order to test H_0, the same signal is independently received $n = 20$ times.

Assume that the sample mean is $\overline{x}^{\text{obs}} = 11.6$.

Is H_0 rejected at the significance level 5%?

For $\alpha = 5\%$, we have

$$k = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = \Phi^{-1}(97.5\%) = 1.96.$$

The nonrejection region $NCR_{5\%}$ is

$$\mu_0 \pm k\frac{\sigma}{\sqrt{n}} = 10 \pm 1.96\frac{4}{\sqrt{20}} = 10 \pm 1.75.$$

Since $\overline{x}^{\text{obs}} = 11.6 \in NCR_{5\%}$, H_0 is not rejected.

Exercise. Is H_0 rejected at significance level 10%?

- Exercise. Explain why, if H_0 is not rejected at significance level α, then it is not rejected at any significance level smaller than α.

- It is important to note that the “correct” significance level to be used in a general test depends on the particular situation we are dealing with.

  If the rejection of H_0 in favor of H_1 implies some action requiring a large cost, that would be wasted if H_0 was true, then we have to choose a small significance level.

  An example could be the situation where H_1 is the hypothesis that a new method of production is superior to the one presently in use.

The two-sided z test

- We can see the previous test also in the following way.

  As test statistic, instead of considering the sample mean, we use the test statistic given by

$$z(\boldsymbol{x}) := \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}, \ \boldsymbol{x} \in \mathbb{R}^n,$$

  and as noncritical regions, instead of considering

$$\mu_0 \pm k \frac{\sigma}{\sqrt{n}}, \ k > 0,$$

  we use

$$(-k, k), \ k > 0.$$

So, since

$$\overline{x}^{\text{obs}} \in \left(\mu_0 - k\frac{\sigma}{\sqrt{n}}, \mu_0 + k\frac{\sigma}{\sqrt{n}}\right) \Leftrightarrow z(\boldsymbol{x}^{\text{obs}}) \in (-k, k),$$

the test takes the form

$$\begin{aligned} &H_0 \text{ is not rejected if } |z(\boldsymbol{x}^{\text{obs}})| < k \\ &H_0 \text{ is rejected if } |z(\boldsymbol{x}^{\text{obs}})| \geq k. \end{aligned}$$

- The significance level of the test is

$$\begin{aligned} &\sup_{\mu=\mu_0,\text{ i.e. } H_0 \text{ is true}} (|z(\boldsymbol{X})| \geq k) = \mathbb{P}_{\mu_0}(|z(\boldsymbol{X})| \geq k) = 1 - \mathbb{P}_{\mu_0}(|z(\boldsymbol{X})| < k) \\ &= 1 - (2\Phi(k) - 1) = 2(1 - \Phi(k)), \end{aligned}$$

since the test Statistic $z(\boldsymbol{X}) = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ is a standard normal random variable when $\mu = \mu_0$.

Thus, as we have already seen, the significance level is equal to α for

$$k = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

In Statistics, the value $\Phi^{-1}(1-\beta)$, $\beta \in (0,1)$, is denoted by z_β.

So, the test with level of significance $\alpha \in (0,1)$ becomes

$$H_0 \text{ is not rejected if } |z(\boldsymbol{x}^{\text{obs}})| < z_{\frac{\alpha}{2}}$$
$$H_0 \text{ is rejected if } |z(\boldsymbol{x}^{\text{obs}})| \geq z_{\frac{\alpha}{2}}.$$

In this form the test is called the **two-sided z test** and the test statistic z is called the **z test statistic.**

- Exercise. The level of significance of the two-sided z test is the probability of having an error of type I. On the other hand,

$$\sup_{\mu\neq\mu_0} \mathbb{P}_\mu\left(|z(\boldsymbol{X})| < k\right)$$

is the supremum of the probabilities of having an error of type II. Show that $z(\boldsymbol{X})$ has distribution $\mathrm{N}\left(\frac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}}, 1\right)$. Then show that

$$\begin{aligned}\mathbb{P}_\mu\left(|z(\boldsymbol{X})| < k\right) &= \mathbb{P}_\mu\Big(-k - \frac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}} < z(\boldsymbol{X}) - \frac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}} < k - \frac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}}\Big)\\ &= \Phi_\mu\Big(k - \frac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}}\Big) - \Phi_\mu\Big(-k - \frac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}}\Big).\end{aligned}$$

Now, show that

$$\begin{aligned}&\sup_{\mu\neq\mu_0} \mathbb{P}_\mu\left(|z(\boldsymbol{X})| < k\right) = \sup_{x\neq 0}\left(\Phi(k-x) - \Phi(-k-x)\right)\\ &= \max_{x\in\mathbb{R}}\left(\Phi(k-x) - \Phi(-k-x)\right) = \Phi(k) - \Phi(-k)\end{aligned}$$

by studying the derivative of the function

$$x \mapsto \Phi(k-x) - \Phi(-k-x),\ x \in \mathbb{R}.$$

Finally, conclude that

$$\sup_{\mu\neq\mu_0} \mathbb{P}_\mu\left(|z(\boldsymbol{X})| < k\right) = 1 - \text{significance level}.$$

The previous exercise shows that if the test has a small significance level, i.e. a small probability of having an error of type I, then the supremum of the probabilities of having an error of type II is large.

This can be also understood as follows.

Suppose that $\mu \neq \mu_0$ but $\mu \approx \mu_0$. Then $z(\boldsymbol{X})$ has distribution $\mathrm{N}\left(\frac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}}, 1\right)$ close to $\mathrm{N}(0,1)$ (obtained for $\mu = \mu_0$) and so

$$\begin{aligned}
\mathbb{P}_\mu(\text{error of type II}) &= \mathbb{P}_\mu(|z(\boldsymbol{X})| < k) \\
&\approx \mathbb{P}_{\mu_0}(|z(\boldsymbol{X})| < k) = 1 - \mathbb{P}_{\mu_0}(|z(\boldsymbol{X})| \geq k) \\
&= 1 - \text{significance level.}
\end{aligned}$$

The p value

- Now, we go back to the general situation, where we have a null hypothesis
$$H_0 : \theta \in B_0,$$
versus an alternative hypothesis
$$H_1 : \theta \in B_1.$$
for an unknown parameter θ of the distribution of X, the numerical quantity associated to an individual randomly selected in the population.

- Consider a test of the form
$$H_0 \text{ is not rejected if } f\left(\tau\left(\boldsymbol{x}^{\text{obs}}\right)\right) < k$$
$$H_0 \text{ is rejected if } f\left(\tau\left(\boldsymbol{x}^{\text{obs}}\right)\right) \geq k,$$
where τ is the test statistic, f is a function $\mathbb{R} \to \mathbb{R}$ and $k > 0$ is a number determined by the significance level $\alpha \in (0, 1)$ of the test.

  Observe that the two-sided z test has this form: τ is the z test statistic z, $f(x) = |x|,\ x \in \mathbb{R}$, and $k = z_{\frac{\alpha}{2}}$.

The **p value** (probability value) of the sample $\boldsymbol{x}^{\text{obs}}$, also called the level of significance of the sample $\boldsymbol{x}^{\text{obs}}$, is

$$\sup_{\theta \in B_0,, \text{ i.e } H_0 \text{ is true}} \mathbb{P}_\theta \left(f\left(\tau(\boldsymbol{X})\right) \geq f\left(\tau\left(\boldsymbol{x}^{\text{obs}}\right)\right)\right).$$

The p value is the supremum, when H_0 is true, of the probabilities that the test Statistic $\tau(\boldsymbol{X})$ is at least as unsupportive of H_0 (unsupportivity measured by the difference $f\left(\tau(\boldsymbol{X})\right) - k$) as the observed value $\tau(\boldsymbol{x}^{\text{obs}})$ of $\tau(\boldsymbol{X})$.

The p value is also the level of significance

$$\sup_{\theta \in B_0,, \text{ i.e } H_0 \text{ is true}} \mathbb{P}_\theta \left(f\left(\tau(\boldsymbol{X})\right) \geq k\right)$$

of the test with $k = f\left(\tau\left(\boldsymbol{x}^{\text{obs}}\right)\right)$.

- The p value has the following important property:

$$\text{significance level} < p \text{ value} \quad \Rightarrow \quad H_0 \text{ is not rejected.}$$
$$\text{significance level} > p \text{ value} \quad \Rightarrow \quad H_0 \text{ is rejected.}$$

Here is the proof of such a property.

We prove the two equivalent contrapositive implications

$$H_0 \text{ is rejected} \quad \Rightarrow \quad p \text{ value} \leq \text{significance level}$$
$$H_0 \text{ is not rejected} \quad \Rightarrow \quad \text{significance level} \leq p \text{ value}.$$

If H_0 is rejected, i.e. $f\left(\tau\left(\boldsymbol{x}^{\text{obs}}\right)\right) \geq k$, then

$$f(\tau(\boldsymbol{X})) \geq f\left(\tau\left(\boldsymbol{x}^{\text{obs}}\right)\right) \quad \Rightarrow \quad f(\tau(\boldsymbol{X})) \geq k$$

and so

$$\mathbb{P}_\theta\left(f(\tau(\boldsymbol{X})) \geq f\left(\tau\left(\boldsymbol{x}^{\text{obs}}\right)\right)\right) \quad \leq \quad \mathbb{P}_\theta\left(f(\tau(\boldsymbol{X})) \geq k\right) \quad \text{for any } \theta \in B_0$$

and so

$$\begin{aligned} p \text{ value} &= \sup_{\theta \in B_0} \mathbb{P}_\theta\left(f(\tau(\boldsymbol{X})) \geq f\left(\tau\left(\boldsymbol{x}^{\text{obs}}\right)\right)\right) \\ &\leq \text{significance level} = \sup_{\theta \in B_0} \mathbb{P}_\theta\left(f(\tau(\boldsymbol{X})) \geq k\right). \end{aligned}$$

If H_0 is not rejected, i.e. $f\left(\tau\left(\boldsymbol{x}^{\mathrm{obs}}\right)\right) < k$, then

$$f\left(\tau(\boldsymbol{X})\right) \geq k \quad \Rightarrow \quad f\left(\tau(\boldsymbol{X})\right) \geq f\left(\tau\left(\boldsymbol{x}^{\mathrm{obs}}\right)\right)$$

and so

$$\mathbb{P}_\theta\left(f\left(\tau(\boldsymbol{X})\right) \geq k\right) \quad \leq \quad \mathbb{P}_\theta\left(f\left(\tau(\boldsymbol{X})\right) \geq f\left(\tau\left(\boldsymbol{x}^{\mathrm{obs}}\right)\right)\right) \quad \text{for any } \theta \in B_0$$

and so

$$\begin{aligned} \text{significance level} &= \sup_{\theta \in B_0} \mathbb{P}_\theta\left(f\left(\tau(\boldsymbol{X})\right) \geq k\right) \\ \leq p \text{ value } &= \sup_{\theta \in B_0} \mathbb{P}_\theta\left(f\left(\tau(\boldsymbol{X})\right) \geq f\left(\tau\left(\boldsymbol{x}^{\mathrm{obs}}\right)\right)\right). \end{aligned}$$

- In practice, the significance level is often not fixed in advance. Instead, we determine the p value from the sample $\boldsymbol{x}^{\mathrm{obs}}$ and then we can reject H_0 until a significance level as small as the p value.

  So, the smaller the p value, the greater the evidence that H_0 is false.

  We say that sample $\boldsymbol{x}^{\mathrm{obs}}$ is **significative**, **very significative**, **extremely significative** for rejecting H_0 if the p value is not larger than 5%, 1% and 0.1%, respectively.

- In the two-sided z test case, we have

$$
\begin{aligned}
p \text{ value} &= \sup_{\mu=\mu_0,\text{ i.e. } H_0 \text{ is true}} \mathbb{P}_\mu \left(|z(\boldsymbol{X})| \geq \left|z\left(\boldsymbol{x}^{\text{obs}}\right)\right| \right) \\
&= 2\left(1 - \Phi\left(\left|z\left(\boldsymbol{x}^{\text{obs}}\right)\right|\right)\right),
\end{aligned}
$$

  since the p value is the level of significance of the test with $k = \left|z\left(\boldsymbol{x}^{\text{obs}}\right)\right|$.

- In the example of the signal emitted from the star, the p value is

$$
\begin{aligned}
&2\left(1 - \Phi\left(\left|z\left(\boldsymbol{x}^{\text{obs}}\right)\right|\right)\right) = 2\left(1 - \Phi\left(\frac{11.6 - 10}{\frac{4}{\sqrt{20}}}\right)\right) \\
&= 2\left(1 - \Phi(1.79)\right) = 2\left(1 - 0.9633\right) \\
&= 7.34\%.
\end{aligned}
$$

  Therefore, the null hypothesis that the signal intensity is 10 will be rejected at any significance level larger than 7.34%. The sample is not significative for rejecting H_0.

The one-sided z test

- Now, we consider a null hypothesis of the form

$$H_0 : \mu \leq \mu_0,$$

with alternative hypothesis

$$H_1 : \mu > \mu_0.$$

For this situation, we use the z test statistic

$$z(\boldsymbol{x}) := \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}, \; \boldsymbol{x} \in \mathbb{R}^n,$$

as test statistic and the intervals

$$(-\infty, k), \; k > 0.$$

as noncritical regions.

The test takes the form

$$H_0 \text{ is not rejected if } z(\boldsymbol{x}^{\text{obs}}) < k$$
$$H_0 \text{ is rejected if } z(\boldsymbol{x}^{\text{obs}}) \geq k.$$

- Observe that, in terms of the test statistic sample mean, the noncritical regions are

$$\left(-\infty, \mu_0 + k\frac{\sigma}{\sqrt{n}}\right),\ k > 0,$$

and the test is

$$H_0 \text{ is not rejected if } \overline{x}^{\text{obs}} < \mu_0 + k\frac{\sigma}{\sqrt{n}}$$
$$H_0 \text{ is rejected if } \overline{x}^{\text{obs}} \geq \mu_0 + k\frac{\sigma}{\sqrt{n}}.$$

- The level of significance of the test is

$$\sup_{\mu \leq \mu_0} \mathbb{P}_\mu (z(\boldsymbol{X}) \geq k) = \mathbb{P}_{\mu_0} (z(\boldsymbol{X}) \geq k) = 1 - \Phi(k).$$

In fact, for $\mu \leq \mu_0$, we have

$$z(\boldsymbol{X}) = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

and so

$$z(\boldsymbol{X}) \geq k \implies \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq k$$

and so, since $\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ is a standard normal random variable,

$$\mathbb{P}_\mu (z(\boldsymbol{X}) \geq k) \leq \mathbb{P}_\mu \left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq k \right) = 1 - \Phi(k).$$

Moreover, for $\mu = \mu_0$, $z(\boldsymbol{X})$ is a standard normal random variable and so

$$\mathbb{P}_{\mu_0} (z(\boldsymbol{X}) \geq k) = 1 - \Phi(k).$$

- So, fixed $\alpha \in (0, 1)$, the level of significance of the test is equal to α for
$$k = \Phi^{-1}(1 - \alpha) = z_\alpha.$$

Thus, the test with level of significance $\alpha \in (0, 1)$ is

$$H_0 \text{ is not rejected if } z(\boldsymbol{x}^{\mathrm{obs}}) < z_\alpha$$
$$H_0 \text{ is rejected if } z(\boldsymbol{x}^{\mathrm{obs}}) \geq z_\alpha.$$

This test is called the **one-sided z test**.

- The one-sided z test has the form that we have seen for defining the notion of p value: infact τ is the test statistic z, $f(x) = x,\ x \in \mathbb{R}$, and $k = z_\alpha$.

  The p value of the sample $\boldsymbol{x}^{\mathrm{obs}}$ is

  $$\sup_{\mu \le \mu_0} \mathbb{P}_\mu \left(z(\boldsymbol{X}) \ge z\left(\boldsymbol{x}^{\mathrm{obs}}\right)\right) = 1 - \Phi\left(z\left(\boldsymbol{x}^{\mathrm{obs}}\right)\right)$$

  since the p value is the level of significance of the test with $k = z\left(\boldsymbol{x}^{\mathrm{obs}}\right)$.

- Exercise. Consider the statistical test given by

$$H_0 : \mu = \mu_0 \ \text{ versus } \ H_1 : \mu > \mu_0,$$

the z test statistic and the noncritical regions

$$(-\infty, k)\,,\ k > 0,$$

or, equivalently, the test statistic sample mean and the noncritical regions

$$\left(-\infty, \mu_0 + k\frac{\sigma}{\sqrt{n}}\right),\ k > 0.$$

Determine the noncritical region NCR_α relevant to a given significance level $\alpha \in (0, 1)$ and show that this test is identical to the one-sided z test.

- Example. Consider the previous example of the new type of cigarettes and the claim of the tobacco firm, which is the null hypothesis, that the mean μ of the nicotine content for these cigarettes is 1.5 mg or less.

  Suppose that the skeptical researcher, which wants disprove this claim and so she/he tests

$$H_0 : \mu \leq 1.5 \text{ mg} \;\text{ versus }\; H_1 : \mu > 1.5 \text{ mg},$$

  finds, on a sample of $n = 20$ cigarettes, an observed sample mean $\overline{x}^{\mathrm{obs}} = 1.58$ mg.

  By assuming that the standard deviation of the nicotine content is $\sigma = 0.7$ mg (it is known, since it depends only on the precision of the industrial process producing the cigarettes), we have

$$z\left(\boldsymbol{x}^{\mathrm{obs}}\right) = \frac{\overline{x}^{\mathrm{obs}} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{1.58 - 1.5}{\frac{0.7}{\sqrt{20}}} = 0.511.$$

The p value is

$$1 - \Phi\left(z\left(\boldsymbol{x}^{\mathrm{obs}}\right)\right) = 1 - \Phi\left(0.511\right) = 1 - 0.6950 = 30.5\%.$$

The null hypothesis cannot be rejected. The sample is not significative for rejecting H_0.

- Example. Now we try to answer to what was questioned in the Introduction, i.e. is the larger case of childhood cancer in the community living near an high level EMF due to pure chance?

  Assume that the number of cases of childhood cancer over a 3-year period in the community under study is a normal random variable X with mean μ unknown and standard deviation 4.7 (as the standard deviation reported in the government public health library for a community of the size as that under consideration).

  The researcher wants to prove that $\mu > 16.2$ (16.2 is the mean reported in the government public health library for a community of the size as that under consideration).

  So, she/he tests

  $$H_0 : \mu \leq 16.2 \text{ versus } H_1 : \mu > 16.2.$$

  The observation of the cases of childhood cancer in the last three years constitutes a sample of size $n = 1$ with $\overline{x}^{\text{obs}} = 32$.

We have

$$z\left(\boldsymbol{x}^{\text{obs}}\right) = \frac{\overline{x}^{\text{obs}} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{32 - 16.2}{\frac{4.7}{\sqrt{1}}} = 3.3617.$$

The p value is

$$1 - \Phi\left(z\left(\boldsymbol{x}^{\text{obs}}\right)\right) = 1 - \Phi\left(3.3617\right) = 3.87 \cdot 10^{-4} = 0.0387\%.$$

The null hypothesis can be rejected. The sample is extremely significative for rejecting H_0.

Exercise. On March 28, 1979, a nuclear accident at Three Mile Island (Pennsylvania, US) released low-level nuclear radiation into the areas surrounding it. It has been reported that 11 babies suffering from hyperthyroidism were born in the surrounding areas between the day of the accident and December 28, 1979 (nine months later). Hyperthyroidism, which results when the thyroid gland is malfunctioning, can lead to mental retardation if it is not treated quickly. By examinating the periods before the accident, one can assume the normal number of babies born with hyperthyroidism in the surrounding areas over a 9-month period is normally distributed with mean 3 and a standard deviation 2. Was the period between March 28, 1979, and December 28, 1979, special as for the number of newborns with hyperthyroidism?

Exercise. In the previous examples of the high-level EMF and of the nuclear accident, the number of cases of disease should be a discrete random variable, whose values are natural numbers. In which sense is it considered a continuous random variable, whose values are real numbers?

After the examples of the high-level EMF and of the nuclear accident, please read this.

It is important to note that this test does *not* prove that the nuclear accident was the cause of the increase in hyperthyroidism; and in fact it *does not even prove that there was an increase in this disease*. Indeed, it is hard to know what can be concluded from this test without having a great deal more information. For instance, one difficulty results from our not knowing why the particular hypothesis considered was chosen to be studied. That is, was there some prior scientific reason for believing that a release of nuclear radiation might result in increased hyperthyroidism in newborns, or did someone just check all possible diseases he could think of (and possibly for a variety of age groups) and then test whether there was a significant change in its incidence after the accident? The trouble with such an approach (which is often called *data mining*, or *going on a fishing expedition*) is that even if no real changes resulted from the accident, just by chance some of the many tests might yield a significant result. (For instance, if 20 independent hypothesis tests are run, then even if all the null hypotheses are true, at least one of them will be rejected at the 1 percent level of significance with probability $1 - (0.99)^{20} = 0.18$.)

Another difficulty in interpreting the results of our hypothesis test concerns the confidence we have in the numbers given to us. For instance, can we really be certain that under normal conditions the mean number of newborns suffering from hyperthyroidism is equal to 3? Is it not more likely that whereas on average 3 newborns would normally be diagnosed to be suffering from this disease, other newborn sufferers may go undetected? Would there not be a much smaller chance that a sufferer would fail to be diagnosed as being such in the period following the accident, given that everyone was alert for such increases in that period? Also, perhaps there are degrees of hyperthyroidism, and a newborn diagnosed as being a sufferer in the tense months following the accident would not have been so diagnosed in normal times.

Note that we are not trying to argue that there was not a real increase in hyperthyroidism following the accident at Three Mile Island. Rather, we are trying to make the reader aware of the potential difficulties in correctly evaluating a statistical study.

- Exercise. Consider the example of the signal emitted by the star. It is suspected that $\mu > 10$. What can we say?

- Exercise. Consider the two-sided z test

$$H_0 : \mu = \mu_0 \ \text{ versus } \ H_1 : \mu \neq \mu_0$$

with significance level α and the one-sided z test

$$H_0 : \mu \leq \mu_0 \ \text{ versus } \ H_1 : \mu > \mu_0.$$

with the same significance level α. For a same sample $\boldsymbol{x}^{\text{obs}}$ such that $z(\boldsymbol{x}^{\text{obs}}) > 0$, does the rejection, or the not-rejection, of H_0 in one of the two tests imply the rejection, or the non-rejection, of H_0 in the other test?

- Exercise. Why the two-sided and one-sided z tests are said "two-sided" and "one-sided"?

- Exercise. Determine, for the one-sided z test,

$$\sup_{\mu > \mu_0} \mathbb{P}_\mu \left(z \left(\boldsymbol{X} \right) < k \right),$$

i.e. the supremum of the probabilities of having an error of type II.

- Now, consider

$$H_0 : \mu \geq \mu_0 \quad \text{versus} \quad H_1 : \mu < \mu_0. \tag{1}$$

Here, with respect to the one-sided z test, we have $\geq$ in H_0 rather than $\leq$.

(1) can be equivalently expressed as

$$H_0 : -\mu \leq -\mu_0 \quad \text{versus} \quad H_1 : -\mu > -\mu_0,$$

where now we have $\leq$ in H_0 as in the one-sided z test and $-\mu$ can be seen as the mean of $Y = -X$, whose distribution is $\mathrm{N}(-\mu, \sigma^2)$.

By using the Sample $\boldsymbol{Y} = -\boldsymbol{X} = (-X_1, -X_2, \ldots, -X_n)$ from the distribution of $Y = -X$ with sample $\boldsymbol{y}^{\text{obs}} = -\boldsymbol{x}^{\text{obs}}$, we obtain, since $\overline{y}^{\text{obs}} = -\overline{x}^{\text{obs}}$,

$$z(\boldsymbol{y}^{\text{obs}}) = \frac{\overline{y}^{\text{obs}} - (-\mu_0)}{\frac{\sigma}{\sqrt{n}}} = \frac{\mu_0 - \overline{x}^{\text{obs}}}{\frac{\sigma}{\sqrt{n}}} = -z(\boldsymbol{x}^{\text{obs}}).$$

So, the test is

$$H_0 \text{ is not rejected if } -z\left(\boldsymbol{x}^{\mathrm{obs}}\right) < k$$
$$H_0 \text{ is rejected if } -z\left(\boldsymbol{x}^{\mathrm{obs}}\right) \geq k$$

with $-z\left(\boldsymbol{x}^{\mathrm{obs}}\right) = \frac{\mu_0 - \overline{x}^{\mathrm{obs}}}{\frac{\sigma}{\sqrt{n}}}$.

- Equivalently the test can written as

$$H_0 \text{ is not rejected if } z\left(\boldsymbol{x}^{\mathrm{obs}}\right) > -k$$
$$H_0 \text{ is rejected if } z\left(\boldsymbol{x}^{\mathrm{obs}}\right) \leq -k.$$

Thus, in the one-sided z test for $\geq$ in H_0, we are using the z test statistic and the noncritical regions

$$(-k, +\infty)\,,\ k > 0.$$

- The one-sided z test for $\geq$ in H_0 with significance level $\alpha \in (0, 1)$ is

$$H_0 \text{ is not rejected if } -z\left(\boldsymbol{x}^{\mathrm{obs}}\right) < z_\alpha$$
$$H_0 \text{ is rejected if } -z\left(\boldsymbol{x}^{\mathrm{obs}}\right) \geq z_\alpha.$$

- The p value of the sample $\boldsymbol{x}^{\text{obs}}$ is

$$1 - \Phi\left(z\left(\boldsymbol{y}^{\text{obs}}\right)\right) = 1 - \Phi\left(-z\left(\boldsymbol{x}^{\text{obs}}\right)\right).$$

- Exercise. In the example of the cigarettes, now suppose that the tobacco firm wants to prove that the mean nicotine content μ for the new treatment is less than 1.5 mg. A sample, different from that of the skeptical researcher, of size $n = 25$ gives $\overline{x}^{\text{obs}} = 1.44\ \mathrm{mg}$. What are the conclusions of the test?

- Exercise. For the same sample $\boldsymbol{x}^{\text{obs}}$, find a relation between the p value of the one-sided z test

$$H_0 : \mu \leq \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0$$

and the p value of one-sided z test

$$H_0 : \mu \geq \mu_0 \quad \text{versus} \quad H_1 : \mu < \mu_0.$$

If $\mu \leq \mu_0$ is not rejected in favor of $\mu > \mu_0$, can we reject $\mu \geq \mu_0$ in favor of $\mu < \mu_0$?

The Student t test

- Consider the situation where the distribution of X is $\mathrm{N}(\mu, \sigma^2)$, with unknown mean μ and unknown standard deviation σ.

- For

$$H_0 : \mu = \mu_0 \text{ versus } H_1 : \mu \neq \mu_0,$$

we use the test statistic given by

$$t(\boldsymbol{x}) := \frac{\overline{x} - \mu_0}{\frac{s_x}{\sqrt{n}}}, \; \boldsymbol{x} \in \mathbb{R}^n,$$

where s_x is the sample standard deviation, and the noncritical regions

$$(-k, k), \; k > 0.$$

The test takes the form

$$H_0 \text{ is not rejected if } |t(\boldsymbol{x}^{\text{obs}})| < k$$
$$H_0 \text{ is rejected if } |t(\boldsymbol{x}^{\text{obs}})| \geq k.$$

- We have previously seen that, when $\mu = \mu_0$, the test Statistic

$$t(\boldsymbol{X}) = \frac{\overline{X} - \mu_0}{\frac{S_x}{\sqrt{n}}}$$

has the Student t distribution t_{n-1} with $n - 1$ degrees of freedom.

So, the significance level of the test is

$$\mathbb{P}_{\mu_0}(|t(\boldsymbol{X})| \geq k) = 1 - \mathbb{P}_{\mu_0}(|t(\boldsymbol{X})| < k)$$
$$= 1 - (2\Phi_{n-1}(k) - 1) = 2(1 - \Phi_{n-1}(k)),$$

where Φ_{n-1} is the distribution function of a random variable with distribution t_{n-1}.

- So, fixed $\alpha \in (0,1)$, the significance level of the test is equal to α for
$$k = \Phi_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right) = t_{n-1,\frac{\alpha}{2}}$$
Here we have used the quantities
$$t_{n-1,\beta} := \Phi_{n-1}^{-1}(1-\beta),\ \beta \in (0,1),$$
previously introduced.

  Thus, the test with significance level $\alpha \in (0,1)$ is
$$H_0 \text{ is not rejected if } |t(\boldsymbol{x}^{\text{obs}})| < t_{n-1,\frac{\alpha}{2}}$$
$$H_0 \text{ is rejected if } |t(\boldsymbol{x}^{\text{obs}})| \geq t_{n-1,\frac{\alpha}{2}}.$$

  This test is called the **two-sided Student t test** and t is called the **Student t test statistic.**

- The p value of the sample $\boldsymbol{x}^{\text{obs}}$ is
$$\mathbb{P}_{\mu_0}\left(|t(\boldsymbol{X})| \geq \left|t\left(\boldsymbol{x}^{\text{obs}}\right)\right|\right) = 2\left(1 - \Phi_{n-1}\left(\left|t\left(\boldsymbol{x}^{\text{obs}}\right)\right|\right)\right)$$
since the p value is the significance level with $k = \left|t\left(\boldsymbol{x}^{\text{obs}}\right)\right|$.

- Example. A new drug has to be tested. It is designed to reduce blood cholesterol level in people having high blood cholesterol levels $\geq$ 240 milliliters per deciliter of blood serum.

  To this aim, the drug was given to a group of 40 volunteers with blood cholesterol level $\geq$ 240 for 60 days and the variations in their blood cholesterol levels at the end of this period were noted.

  If the mean of the variations of the blood cholesterol levels is a decrease of 6.8 and the standard deviation of the variations is 12.1, what conclusions can we draw?

  By assuming that the decrease X of the blood cholesterol level at the end of the drug treatment period in a person randomly selected in the population of people having blood cholesterol level $\geq$ 240 is a normal random variable with distribution $\mathrm{N}\left(\mu, \sigma^2\right)$, where both μ and σ are unknown, we test

$$H_0 : \mu = 0 \;\text{ versus }\; H_1 : \mu \neq 0.$$

Exercise. Explain why the assumption of the normal distribution for the decrease of the blood cholesterol level is reasonable.

The significance level 5% is obtained for

$$k = \Phi_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right) = t_{n-1,\frac{\alpha}{2}} = t_{39,2.5\%} = 2.0227.$$

Since

$$\overline{x}^{\mathrm{obs}} = 6.8 \text{ and } s_x^{\mathrm{obs}} = 12.1$$

we have

$$t\left(\boldsymbol{x}^{\mathrm{obs}}\right) = \frac{\overline{x}^{\mathrm{obs}} - \mu_0}{\frac{s_x^{\mathrm{obs}}}{\sqrt{n}}} = \frac{6.8}{\frac{12.1}{\sqrt{40}}} = 3.55.$$

So, H_0 is rejected at the significance level 5%.

Observe that the p value is

$$2\left(1-\Phi_{n-1}\left(t\left(\boldsymbol{x}^{\mathrm{obs}}\right)\right)\right)=2\left(1-\Phi_{39}\left(3.55\right)\right)=0.0010.$$

Thus, we reject H_0 until the significance level 0.1%. The sample is extremely significative.

Exercise. Are we justified at this point to conclude that the variations in the cholesterol levels are due to the used drug?

- Example. Historical data indicate that the mean acidity (pH) level of rain in a certain industrial region in West Virginia is 5.2.

  To see whether there has been any recent change in this value, the acidity levels of 12 rainstorms over the past year have been measured with the following results:

  $$6.1,\ 5.4,\ 4.8,\ 5.8,\ 6.6,\ 5.3,\ 6.1,\ 4.4,\ 3.9,\ 6.8,\ 6.5,\ 6.3.$$

  Are these data strong enough, at the 5 percent level of significance, for concluding that the acidity of the rain has changed from its historical value?

  By assuming that the acidity (pH) level of a rainstorm X in West Virginia is a normal random variable with distribution $\mathrm{N}\left(\mu, \sigma^2\right)$, where both μ and σ are unknown, we test

  $$H_0 : \mu = 5.2 \ \text{ versus } \ H_1 : \mu \neq 5.2.$$

The level of significance 5% is obtained for

$$k = \Phi_{n-1}^{-1}\left(1 - \frac{\alpha}{2}\right) = t_{n-1,\frac{\alpha}{2}} = t_{11,2.5\%} = 2.201.$$

We have

$$\overline{x}^{\text{obs}} = 5.667 \text{ and } s_x^{\text{obs}} = 0.921$$

and then

$$t\left(\boldsymbol{x}^{\text{obs}}\right) = \frac{\overline{x}^{\text{obs}} - \mu_0}{\frac{s^{\text{obs}}}{\sqrt{n}}} = \frac{5.667 - 5.2}{\frac{0.921}{\sqrt{12}}} = 1.76.$$

H_0 is not rejected at the significance level 5%.

The $p-$value is

$$2\left(1 - \Phi_{n-1}\left(t\left(\boldsymbol{x}^{\text{obs}}\right)\right)\right) = 2\left(1 - \Phi_{11}\left(1.76\right)\right) = 0.1061.$$

Therefore, H_0 is rejected only over the significance level 10%. The sample is not significative.

The one-sided Student t test

- For
$$H_0 : \mu \le \mu_0 \;\text{ versus }\; H_1 : \mu > \mu_0,$$
we use the t test statistic t and the noncritical regions
$$(-\infty, k)\,,\; k > 0.$$

  This test has the form
$$H_0 \text{ is not rejected if } t(\boldsymbol{x}^{\mathrm{obs}}) < k$$
$$H_0 \text{ is rejected if } t(\boldsymbol{x}^{\mathrm{obs}}) \ge k.$$

- As in case of the z test, we can prove that the significance level of the test is
$$\sup_{\mu \le \mu_0} \mathbb{P}_\mu\left(t\left(\boldsymbol{X}\right) \ge k\right) = \mathbb{P}_{\mu_0}\left(t\left(\boldsymbol{X}\right) \ge k\right) = 1 - \Phi_{n-1}(k).$$

- So, fixed $\alpha \in (0,1)$, the significance level of the test is α for

$$k = \Phi_{n-1}^{-1}(1-\alpha) = t_{n-1,\alpha}.$$

Thus, the test with significance level $\alpha \in (0,1)$ is

$$H_0 \text{ is not rejected if } t(\boldsymbol{x}^{\text{obs}}) < t_{n-1,\alpha}$$
$$H_0 \text{ is rejected if } t(\boldsymbol{x}^{\text{obs}}) \geq t_{n-1,\alpha}.$$

This test is called the **one-sided Student t test.**

- The p value of the sample $\boldsymbol{x}^{\text{obs}}$ is

$$\sup_{\mu \leq \mu_0} \mathbb{P}_\mu \left(t(\boldsymbol{X}) \geq t\left(\boldsymbol{x}^{\text{obs}}\right)\right) = 1 - \Phi_{n-1}\left(t\left(\boldsymbol{x}^{\text{obs}}\right)\right)$$

since it is the significance level with $k = t\left(\boldsymbol{x}^{\text{obs}}\right)$.

- For

$$H_0 : \mu \geq \mu_0 \quad \text{versus} \quad H_1 : \mu < \mu_0,$$

as in case of the z test, we rewrite it as

$$H_0 : -\mu \leq -\mu_0 \quad \text{versus} \quad H_1 : -\mu > -\mu_0,$$

where now we have $\leq$ in H_0 as in the one-sided Student t test and the mean $-\mu$ of $Y = -X$ is involved.

By using the Sample $\boldsymbol{Y} = -\boldsymbol{X}$ with sample $\boldsymbol{y}^{\mathrm{obs}} = -\boldsymbol{x}^{\mathrm{obs}}$, since $\overline{y}^{\mathrm{obs}} = -\overline{x}^{\mathrm{obs}}$ and

$$s_y^{\mathrm{obs}} = \text{standard deviation of } \boldsymbol{y}^{\mathrm{obs}} = \text{standard deviation of } -\boldsymbol{x}^{\mathrm{obs}}$$
$$= \text{standard deviation of } \boldsymbol{x}^{\mathrm{obs}} = s_x^{\mathrm{obs}}.$$

we obtain

$$t(\boldsymbol{y}^{\mathrm{obs}}) = \frac{\overline{y}^{\mathrm{obs}} - (-\mu_0)}{\frac{s_y^{\mathrm{obs}}}{\sqrt{n}}} = \frac{\mu_0 - \overline{x}^{\mathrm{obs}}}{\frac{s_x^{\mathrm{obs}}}{\sqrt{n}}} = -t(\boldsymbol{x}^{\mathrm{obs}}).$$

- The one-sided Student t test for $\geq$ in H_0 with significance level $\alpha \in (0,1)$ is

$$H_0 \text{ is not rejected if } -t\left(\boldsymbol{x}^{\text{obs}}\right) < t_{n-1,\alpha}$$
$$H_0 \text{ is rejected if } -t\left(\boldsymbol{x}^{\text{obs}}\right) \geq t_{n-1,\alpha}$$

with $-t(\boldsymbol{x}^{\text{obs}}) = \frac{\mu_0 - \overline{x}^{\text{obs}}}{\frac{s_x^{\text{obs}}}{\sqrt{n}}}$.

- The p value of the sample $\boldsymbol{x}^{\text{obs}}$ is

$$1 - \Phi_{n-1}\left(t\left(\boldsymbol{y}^{\text{obs}}\right)\right) = 1 - \Phi_{n-1}\left(-t\left(\boldsymbol{x}^{\text{obs}}\right)\right).$$

- Example. The manufacturer of a new fiberglass tire claims that the average life of a set of its tires is at least $50K$ miles.

  A consumer agency wants to verify this claim and so it selects a sample of 8 sets which are tested on the road.

  By assuming that the life of a set of tires is a normal random variable X with distribution $\mathrm{N}\left(\mu, \sigma^2\right)$, both μ and σ unknown, the consumer agency tests

$$H_0 : \mu \geq 50K \quad \text{versus} \quad H_1 : \mu < 50K.$$

  If the resulting values of the sample mean and sample standard deviation are, respectively, $47.2K$ miles and $3.1K$ miles, what conclusions can be drawn?

Since

$$\overline{x}^{\text{obs}} = 47.2K \;\text{ and }\; s_x^{\text{obs}} = 3.1K,$$

we have

$$-t\left(\boldsymbol{x}^{\text{obs}}\right) = \frac{\mu_0 - \overline{x}^{\text{obs}}}{\frac{s_x^{\text{obs}}}{\sqrt{n}}} = \frac{50K - 47.2K}{\frac{3.1K}{\sqrt{8}}} = 2.55.$$

The $p-$value is

$$1 - \Phi_{n-1}\left(-t\left(\boldsymbol{x}^{\text{obs}}\right)\right) = 1 - \Phi_7\left(2.55\right) = 0.0381.$$

So, H_0 is rejected at the 5% significance level, but not at the 1%. The sample is significative, but not very significative.

- Exercise. Repeat the example of the drug for reducing the blood cholesterol level as a one-sided test.

The Student t test for a general distribution

- When n is large, in the two-sided Student t test for

$$H_0 : \mu = \mu_0 \text{ versus } H_1 : \mu \neq \mu_0$$

or the one-sided Student t test for

$$H_0 : \mu \leq \mu_0 \text{ versus } H_1 : \mu > \mu_0$$

we can use Φ instead of Φ_{n-1} in determining the significance level and the p value. In fact, $\Phi_{n-1}(x) \to \Phi(x),\ n \to \infty$, for any $x \in \mathbb{R}$.

- For n large, we can use the Student t tests also for a general distribution of unknown mean μ and unknown standard deviation σ.

  In fact, for n large, we have with large probability S_X close to σ and so

$$t(\boldsymbol{X}) = \frac{\overline{X} - \mu_0}{\frac{S_X}{\sqrt{n}}} \approx \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = z(\boldsymbol{X}).$$

  Moreover, $\overline{X}$ has distribution close to $\mathrm{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ and so, for $\mu = \mu_0$, $\frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ has distribution close to $\mathrm{N}(0, 1)$.

  This means that the significance level and the p value are computed exactly as in case of the normal distribution.

A test for a change of production

- Suppose that the television tubes produced by a certain manufacturer are known to have a mean lifespan of 3K hours of use.

  An outside consultant claims that a new production method will lead to a greater mean lifespan.

  To check this, a pilot program is designed to produce a sample of tubes by the new suggested method.

  How should the manufacturer use the resulting data?

  By assuming that the lifespan X for a tube produced by the new method is a normal random variable with distribution $\mathrm{N}\left(\mu, \sigma^2\right)$, where μ and σ are both unknown, at first glance it might appear that one has to test

$$H_0 : \mu \leq 3K \quad \text{versus} \quad H_1 : \mu > 3K.$$

A rejection of H_0 means evidence that the new proposed method improves the mean lifetime of the tubes.

- However, if H_1 is true and the sample size is large enough, then H_0 will be rejected even in cases where μ is only a little bit larger than $3K$, for example $\mu = 3.001K$ hours.

In fact, if n large, then with large probability

$$t(\boldsymbol{X}) = \frac{\overline{X} - \mu_0}{\frac{S_x}{\sqrt{n}}} \approx \frac{\mu - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

and so

$$P \text{ Value} = 1 - \Phi_{n-1}(t(\boldsymbol{X})) \approx 1 - \Phi\left(\frac{\mu - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \Phi\left(\frac{\sqrt{n}(\mu - \mu_0)}{\sigma}\right).$$

So, if H_1 is true, i.e. $\mu > \mu_0$, and n is large, then with large probability the sample is extremely significative for rejecting H_0 if

$$1 - \Phi\left(\frac{\sqrt{n}\,(\mu - \mu_0)}{\sigma}\right) \leq 0.1\%,$$

i.e.

$$\sqrt{n}\,(\mu - \mu_0) \geq \sigma z_{0.1\%}.$$

and this happens for n large enough.

- When μ is only a little bit larger than $3K$, it cannot be economically convenient to make the change of the production method because we have only a small increase in the mean lifetime.

  Indeed, we should use

  $$H_0 : \mu \leq 3K + c \;\text{ versus }\; H_1 : \mu > 3K + c,$$

  where c is the smallest increase in the mean lifetime that makes economically convenient the production change.