COMP229: Introduction to Data Science
Lecture 18: statistical significance

Vitaliy Kurlin, vitaliy.kurlin@liverpool.ac.uk
Autumn 2018, Computer Science department
University of Liverpool, United Kingdom

Statistical inference again

In the past lecture confidence intervals were used to estimate a population parameter, e.g. a mean by using a standard deviation and a data sample.

Another common type of inference (tests of significance) aims to test a claim about a population parameter again by using a sample.

Example. Someone claims that they solve on average 70% of problems in any exam and then gets 40% in COMP229. Was the claim disproved? We'll estimate the significance of the mark 40.

Hypothesis: null vs alternative

Definition 18.1. A claim to be checked by a statistical test is called the *null hypothesis* H_0.

Usually, the null hypothesis is a statement of "no new effect", e.g. new data confirms an old trend.

The alternative hypothesis H_a is a claim about the population that we are trying to confirm by data.

Often an alternative hypotehsis H_a claims a new effect and is easier to state before formulating H_0.

Hypotheses: for parameters, not data

Both types of hypotheses (null and alternative) should be stated in terms of parameters of a (very large) population, e.g. the mean of all exam marks, not for a usually small sample outcome.

A good hypothesis: the mean mark is 60 across all outcomes for many students over several years.

A bad hypothesis: the mean mark in January 2019 is 60 (only one data sample for one exam).

Hypotheses: 1-sided vs 2-sided

Let the null hypothesis say that the mean of exam marks (the population parameter) is $\mu = 60$.

An alternative hypothesis H_a may say $\mu \neq 60$ and is called *2-sided* in this case, so H_a is the union of the two hypotheses: 1) $\mu < 60$ and 2) $\mu > 60$.

An alternative hypothesis H_a may not be complementary (exactly opposite) to H_0, e.g.

H_a may say that $\mu > 60$ and is called *1-sided*.

The p-value of a data sample

Assume that a null hypothesis H_0 is true.

A test statistic (a numerical measurement of a sample) estimates how far an actual measurement diverges from an expected value for H_0.

Definition 18.2. The *p-value* is the probability (assuming the null hypothesis) to obtain a result equal to or more extreme than what was observed. If H_0 is $\mu = 60$ and we get $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 40$, then the 1-sided p-value is $P(\bar{x} \leq 40)$, while the 2-sided p-value is $P(|\bar{x} - 60| \geq 20)$.

One sample test for a mean

Let $x_1, \ldots, x_n$ be sample values drawn from a normal distribution with an unknown mean μ and a known standard deviation σ.

To test the null hypothesis H_0 that $\mu = \mu_0$ (a given value), we find the *test statistic* $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$.

The p-value to test $H_a = \{\mu > \mu_0\}$ against H_0 is $P(Z \geq z)$, where $Z \sim N(0, 1)$ is the standard normal variable, the same for $H_a = \{\mu < \mu_0\}$.

The 2-sided p-value against H_0 is $P(|Z| \geq |z|)$.

Reject or not to reject?

Assuming that the standard deviation of exam marks is $\sigma = 15$ and we get $n = 25$ sample marks with the sample mean $\bar{x} = 69$, shall we reject (or not) the hypothesis that the mean mark $\mu = 60$?

Step 1. Compute the z statistic as follows:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{69 - 60}{15/\sqrt{25}} = \frac{9}{3} = 3.$$

Step 2. Compute the 2-sided p-value $P(|Z| \geq 3) = 0.3\%$ by the 68–95–99.7 rule.

$\bar{x} = 69$ seems unlikely for the hypothesis $\mu = 60$.

Significance level α

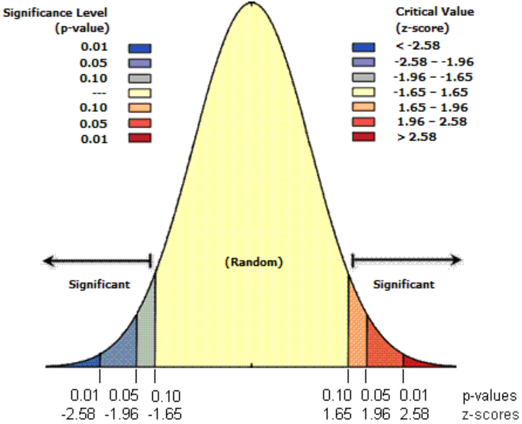
Definition 18.3. If the p–value is (non–strictly) smaller than a specified significance level α, the data are called statistically significant at level α.

The p–value of 0.3% for the null hypothesis $H_0 = \{\mu = 60\}$ means that the actual marks are statistically significant at level 1%, but are not statistically significant at 0.1%. What if $\bar{x} = 66$?

z statistic	1.645	1.96	2.326	2.576
2–sided p–value	10%	5%	2%	1%
1–sided p–value	5%	2.5%	1%	0.5%

Revision: p-levels vs z-scores

For $\bar{x} = 66$, the z-statistic is $z = 2$ and 2-sided p-value is $P(|Z| \geq 2) \approx 5\%$, so the statistic is significant at level 5%, but not at smaller levels.



$$P(|Z| > 1.65) = 10\%,$$
$$P(Z > 1.65) = 5\%,$$
$$P(|Z| > 1.96) = 5\%,$$
$$P(Z > 1.96) = 2.5\%,$$
$$P(|Z| > 2.32) = 2\%,$$
$$P(Z > 2.32) = 1\%,$$
$$P(|Z| > 2.58) = 1\%.$$

Your questions and the quiz

To benefit from the lecture, now you could

- ask or submit your anonymous questions to the COMP229 folder after the lecture;
- write down your summary in 2–3 phrases, e.g. list key concepts you have learned;
- talk to your classmates to revise the lecture.

Question. Let $n = 16$, $\sigma = 20$, $\bar{x} = 47.5$. Should we reject (or not reject) the null hypothesis H_0 that $\mu = 60$ at the significance level 1%?

Answer to the quiz and summary

Answer. $z = \dfrac{47.5 - 60}{20/\sqrt{16}} = -\dfrac{12.5}{5} = -2.5$. The 2-sided p-value $= P(|Z| \geq 2.5)$ is a bit more than 1%, hence we don't reject the null hypothesis H_0.

- The *p-value* is the probability (assuming the null hypothesis) to obtain a result equal to or more extreme than what was observed.
- If the p-value is (non-strictly) smaller than a specified significance level α, the data are called *statistically significant at level* α.