Chapter 8

Hypothesis Tests: One Sample

Aim

The aim is the same as for confidence intervals:

Making inferences about a population using a random sample.

Confidence Intervals vs Hypothesis Testing

Confidence Intervals

- Allow us to use sample data to estimate a population value.
- Example: What is the true average amount students spend weekly on coffee? (Find a set of plausible values for this average amount).

Hypothesis Testing

- Allows us to use sample data to test a claim about a population.
- Example: Is the true average amount that students spend weekly on coffee 500 TL?

Example

- **Claim:** The average amount students spend weekly on coffee is 500 TL.
- To test this claim, a survey is conducted with a random sample of 64 students.
- Sample average is 470 TL, with standard deviation 80 TL.
- **Question:** If the true population mean were 500 TL, what is the probability of getting a sample mean of 470 TL or lower?
- By CLT, the sample mean $\overline{X}$ has a normal distribution:

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \approx N\left(500, \frac{80}{\sqrt{64}}\right) = N\left(500, 10\right)$$

  (We use the approximation $\sigma \approx s$ since $n \geq 30$ and σ is unknown.)
- Hence

$$P(\overline{X} < 470) = P\left(Z < \frac{470-500}{10}\right) = P\left(Z < -3\right) = 0.5 - 0.4987 = 0.0013$$

- The probability that sample average is 470 or lower is $0.0013 = 0.13\%$
- **Conclusion:** Such a sample result would be extremely unlikely if the claim were true. So the claim is not reasonable. The population average is less than 500 TL.

Example

- Alice and Bob play backgammon (tavla). After losing too many games, Alice complained about one of the dice. "This die is not fair! It usually gives higher outcomes."
- Bob wanted to test Alice's claim by rolling that die 30 times. The average of these 30 outcomes is 3.9.
- Alice: "You see the average should be 3.5, this is not a fair die!".
- Bob: "No, this is just a chance variation!"
- Who's right here?

Example (cont'd)

- Recall that if X is the outcome of the fair die, then

$$\mu = E(X) = 3.5, \quad \sigma = \text{StDev}(X) = \sqrt{\frac{35}{12}} \approx 1.7078$$

(see question 9 in homework 5.)

- By CLT, if the average of 30 rolls is $\overline{X}$, then

$$\overline{X} \sim N\left(3.5, \frac{1.7078}{\sqrt{30}}\right) \approx (3.5, 0.3118)$$

- Assuming the die is fair, the probability that the average of 30 rolls is 3.9 or more is

$$P(3.9 < \overline{X}) = P\left(\frac{3.9 - 3.5}{0.3118} < Z\right) = P\left(1.28 < Z\right) = 0.5 - 0.3997 = 0.1003$$

Example (cont'd)

- Alice: "You see the chance of this happening is about 10% if it were a fair die!".
- Bob: "I don't think 10% is a low probability. Maybe we should collect more data!"
- Then, Bob rolled that die 50 more times. The average of all 80 outcomes is 3.85.

Example (cont'd)

- By CLT, if the average of 80 rolling is $\overline{X}$, then

$$\overline{X} \sim N\left(3.5, \frac{1.7078}{\sqrt{80}}\right) \approx (3.5, 0.1909)$$

- Assuming the die is fair, the probability that the average of 80 rolls is 3.85 or more is

$$P(3.85 < \overline{X}) = P\left(\frac{3.85 - 3.5}{0.1909} < Z\right) = P\left(1.83 < Z\right) = 0.5 - 0.4664 = 0.0336$$

- Bob: "Ok, I'm convinced now. 3% is low enough for me. Probably, this is not a fair die!"

Example (cont'd)

A note on this example: The correct way to measure whether a die is fair is χ^2-test. But z-test (what we applied using CLT) still gives evidence that this die produces higher values than it should.

χ^2-test is used to measure the difference between observed frequencies and expected frequencies in categories. However, χ^2-test is not in the scope of this course.

We will see z-test (and t-test) which are used to measure the difference between observed mean and expected mean.

The Structure of a Hypothesis Test

- Constructing the hypotheses
- Collecting the Data (Sample)
- Finding P-value
- Making inference

Constructing the hypotheses

We have

$$H_0 \longrightarrow \text{the null hypothesis}$$
$$H_1 \longrightarrow \text{the alternative hypothesis}$$

- H_0 is the status-quo or the statement that we want to test. This is going to be our initial assumption.
- H_1 is the statement we might get by rejecting H_0. Usually, this is the conclusion we want to reach or what the sample suggests.

Constructing the hypotheses

We have

$$H_0 \longrightarrow \text{the null hypothesis}$$
$$H_1 \longrightarrow \text{the alternative hypothesis}$$

Criminal Trial Analogy:

$$H_0 : \text{defendant is not guilty}$$
$$H_1 : \text{defendant is guilty}$$

We assume defendant is not guilty in the beginning.
(A defendant is not guilty, until proven guilty.)

Constructing the hypotheses

In the coffee survey example we want to test whether the population mean is lower than 500 or not:

$$H_0 : \mu = 500$$
$$H_1 : \mu < 500$$

In the die example, we want to test whether the true mean of the die is higher than 3.5 or not:

$$H_0 : \mu = 3.5$$
$$H_1 : \mu > 3.5$$

These are one-tailed hypothesis tests because the alternative hypothesis specifies one direction: either strictly less than or strictly greater than a specified value.

Constructing the hypotheses

Example: The effect of drugs and alcohol on the nervous system has been the subject of considerable research. Suppose a research neurologist is testing the effect of a drug on response time by injecting 100 rats with a unit dose of the drug, subjecting each rat to a neurological stimulus, and recording its response time. The neurologist knows that the mean response time for rats not injected with the drug (the "control" mean) is 1.2 seconds. She wishes to test whether the mean response time for drug-injected rats differs from 1.2 seconds.

$$
\begin{aligned}
H_0 &: \mu = 1.2 \\
H_1 &: \mu \neq 1.2
\end{aligned}
$$

Some statistical investigations seek to show that the population parameter is either larger or smaller than some specified value. Such an alternative hypothesis is called a two-tailed hypothesis.

Constructing the hypotheses

- Select the alternative hypothesis as that which the sampling experiment is intended to establish. The alternative hypothesis will assume one of three forms:
  - One-tailed, upper-tailed. Example: $H_1 : \mu > 100$
  - One-tailed, lower-tailed. Example: $H_1 : \mu < 100$
  - Two-tailed. Example: $H_1 : \mu \neq 100$
- Select the null hypothesis as the status quo, that which will be presumed true unless the sampling experiment conclusively establishes the alternative hypothesis. The null hypothesis will be specified as that parameter value closest to the alternative in one-tailed tests and as the complementary (or only unspecified) value in two-tailed tests.
  - Example: $H_0 : \mu = 100$

Collecting the Data

In the criminal trial analogy, this step corresponds to collecting the evidence (such as fingerprints, blood spots, shoe prints etc).

In statistics, the evidence is data, coming from a sample.

Finding P-value

Now we have H_0, H_1 and a sample.

We calculate the probability that such a sample occurs given that H_0 is true.

This is a conditional probability.

We call this probability P-value.

Making inference

If P is very small, then the chance of getting such a sample is not likely if H_0 were true. But we get this sample, so H_0 is probably not true.

So, if P is very small, we reject H_0.

If P is not that small, we fail to reject H_0. It can still be true or false.

How small should P be? This is our significance level. This depends on how much significant result we want. Usually depends on the situation.

We denote the significance level by α.

Common values of α are 0.10, 0.05, 0.01.

Making inference

If $P < \alpha$: Then we reject H_0. This means that there is enough evidence for H_1 at significance level α.

If $P \geq \alpha$: Then we fail to reject H_0. This means that there is **not** enough evidence for H_1 at significance level α.

Disclaimer

Neither decision entails proving the null hypothesis or alternative hypothesis. We merely state that there is enough evidence to behave one way or the other.

This is always the case in statistics! No matter what decision we make, there is **always a chance that we made an error**!

One Sample z-test for mean

One Sample z-test for mean

Assumptions:

There is a sample of size n randomly selected from a population and either

- $n \geq 30$, or
- $n < 30$ and population is normally distributed with its standard deviation σ known

Example

We have a population of adults and the variable is the body temperature. A sample of 120 has an average of 36.89° with standard deviation 0.418°. Is the average body temperature in the population lower than 37°? Test with $\alpha = 0.01$.

Let μ be the average of the population.

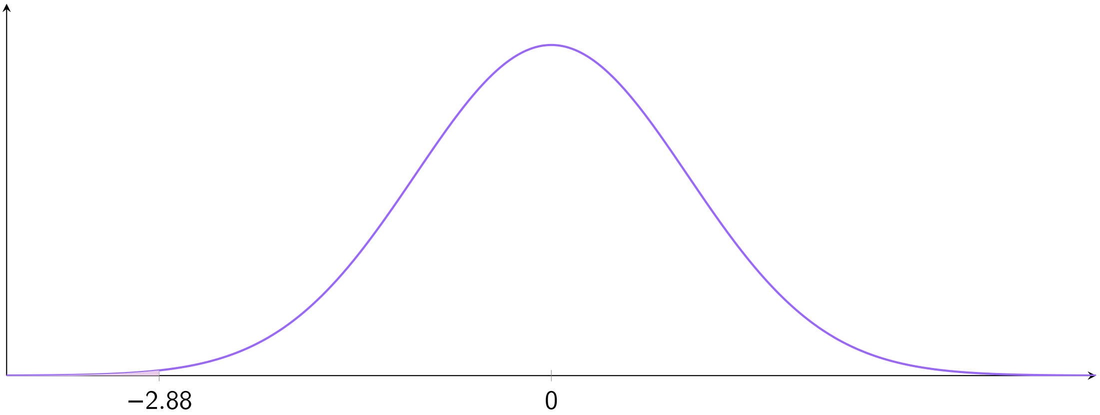
$$H_0 : \mu = 37$$
$$H_1 : \mu < 37$$

$$P = P(\overline{X} < 36.89) = P\left(Z < \frac{36.89 - 37}{0.418/\sqrt{120}}\right) \approx P(Z < -2.88) = 0.002 < 0.01 = \alpha$$

We reject H_0. There is enough statistical evidence at significance level 0.01 to conclude that the average body temperature in the population is lower than 37°.

Example (cont'd)

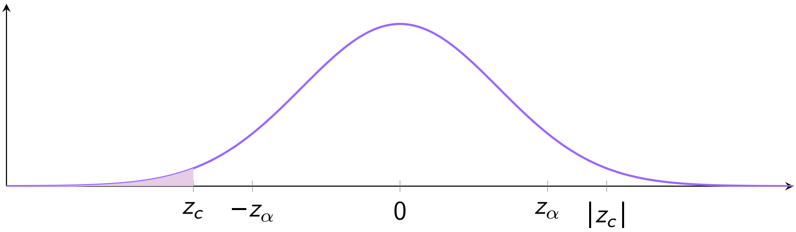
$$z_c = \frac{36.89 - 37}{0.418/\sqrt{120}} = -2.88 \text{ is called } \textbf{test statistic}.$$



$$P = P(Z < z_c)$$

Example (cont'd)

$$P = P(Z < z_c)$$



Checking

$$P < \alpha$$

is equivalent to checking

$$z_c < -z_\alpha$$

or equivalent to checking

$$|z_c| > z_\alpha$$

Example (cont'd)

$$z_c = \frac{36.89 - 37}{0.418/\sqrt{120}} = -2.88$$

$$z_\alpha = z_{0.01} = 2.326$$

Since

$$|z_c| = 2.88 > 2.326 = z_\alpha$$

we reject H_0.

One-tailed (lower-tailed) test

- Hypotheses:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu < \mu_0$$

(This case makes sense only if $\mu_0 > \bar{x}$.)

- Test statistic:

$$z_c = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad \text{if } \sigma \text{ is known}$$
$$z_c = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad \text{if } \sigma \text{ is unknown}$$

(Test statistic is negative.)

- Rejection region:

$$|z_c| > z_\alpha \quad (\text{equivalently, } z_c < -z_\alpha)$$

- P-value:

$$P = P(Z < z_c)$$

Example

The shopping times of 64 randomly selected customers at a local supermarket were recorded. The average and standard deviation of the 64 shopping times were 34 minutes and 16, respectively. Is the average shopping time in the population is 30 minutes or more? Test with $\alpha = 0.05$.

Let μ be the average of the population.

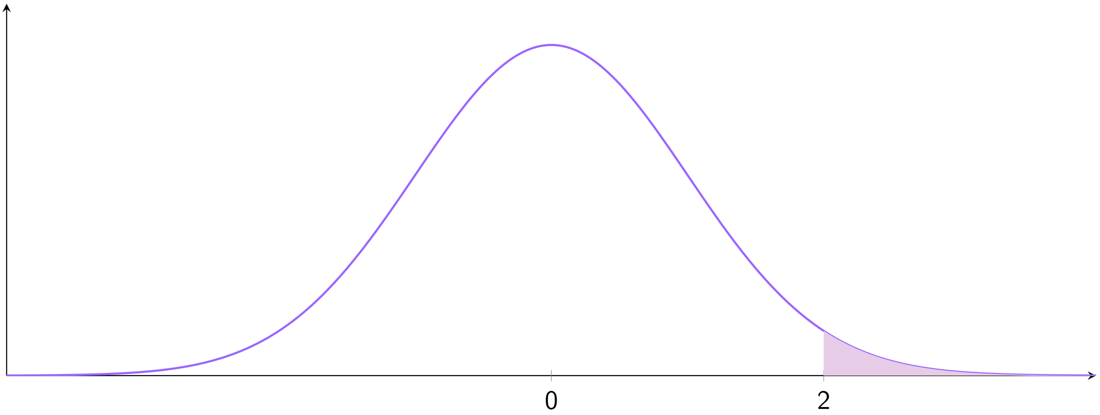
$$H_0 : \mu = 30$$
$$H_1 : \mu > 30$$

$$P = P(\overline{X} > 34) = P\left(Z > \frac{34 - 30}{16/\sqrt{64}}\right) \approx P(Z > 2) = 0.0228 < 0.05 = \alpha$$

We reject H_0. There is enough statistical evidence at significance level 0.05 to conclude that the average shopping time in the population is more than 30 minutes.

Example (cont'd)

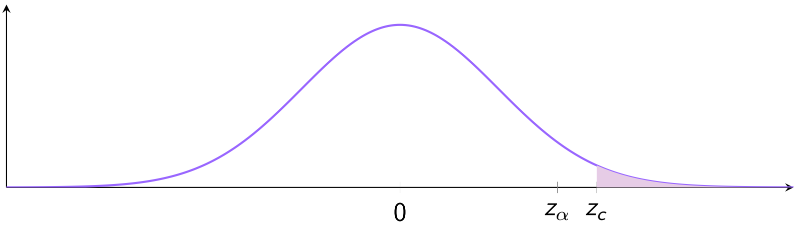
$$z_c = \frac{34 - 30}{16/\sqrt{64}} = 2 \text{ is the test statistic.}$$



$$P = P(Z > z_c)$$

Example (cont'd)

$$P = P(Z > z_c)$$



Checking

$$P < \alpha$$

is equivalent to checking

$$z_c > z_\alpha$$

or equivalent to checking

$$|z_c| > z_\alpha$$

Example (cont'd)

$$z_c = \frac{34 - 30}{16/\sqrt{64}} = 2$$

$$z_\alpha = z_{0.05} = 1.645$$

Since

$$|z_c| = 2 > 1.645 = z_\alpha$$

we reject H_0.

Example (cont'd)

We have

$$0.01 < P = 0.0228 < 0.05$$

So

- At $\alpha = 0.05$, we reject H_0
- At $\alpha = 0.01$, we fail to reject H_0

This means that

- There is enough statistical evidence at significance level 0.05 to conclude that $\mu > 30$ minutes.
- There is NOT enough statistical evidence at significance level 0.01 to conclude that $\mu > 30$ minutes.

One-tailed (upper-tailed) test

- Hypotheses:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

(This case makes sense only if $\mu_0 < \bar{x}$.)

- Test statistic:

$$z_c = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad \text{if } \sigma \text{ is known}$$
$$z_c = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad \text{if } \sigma \text{ is unknown}$$

(Test statistic is positive.)

- Rejection region:

$$|z_c| > z_\alpha \quad (\text{equivalently, } z_c > z_\alpha)$$

- P-value:

$$P = P(Z > z_c)$$

Example

A machine is supposed to fill bottles with an average of 500 ml. From past calibration, the standard deviation is known to be $\sigma = 12$ ml. A random sample of 64 bottles has a sample mean 503.2 ml. Is there significant evidence that the machine's mean fill amount is different from 500 ml? Test at $\alpha = 0.05$.

Let μ be the machine's mean fill amount.

$$H_0 : \mu = 500$$
$$H_1 : \mu \neq 500$$

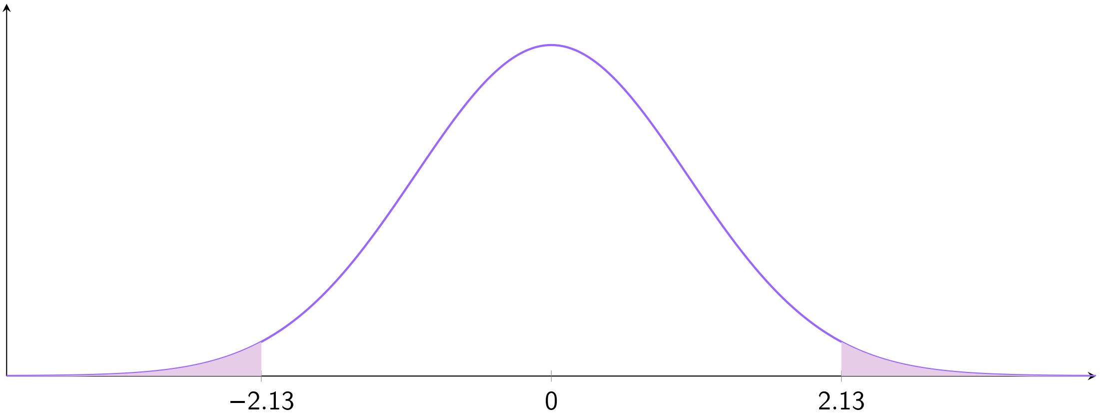
$$z_c = \frac{503.2 - 500}{12/\sqrt{64}} \approx 2.13$$

$$P = P(Z > 2.13) + P(Z < -2.13) = 2P(Z > 2.13) = 2 \cdot 0.0166 = 0.0332 < 0.05 = \alpha$$

We reject H_0. There is enough statistical evidence at significance level 0.05 to conclude that the machine's mean fill amount is different from 500 ml.

Example (cont'd)

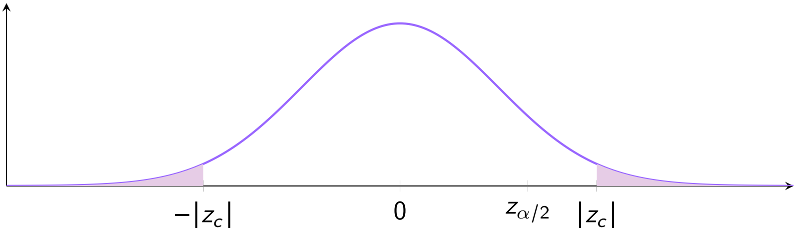
$$z_c = \frac{503.2 - 500}{12/\sqrt{64}} \approx 2.13 \text{ is the test statistic.}$$



$$P = 2P(Z > |z_c|)$$

Example (cont'd)

$$P = 2P(Z > |z_c|)$$



Checking

$$P < \alpha$$

is equivalent to checking

$$P(Z > |z_c|) < \frac{\alpha}{2}$$

or equivalent to checking

$$|z_c| > z_{\alpha/2}$$

Example (cont'd)

$$z_c = \frac{503.2 - 500}{12/\sqrt{64}} \approx 2.13$$

$$z_{\alpha/2} = z_{0.025} = 1.96$$

Since

$$|z_c| = 2.13 > 1.96 = z_{\alpha/2}$$

we reject H_0.

Two-tailed test

- Hypotheses:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

- Test statistic:

$$z_c = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad \text{if } \sigma \text{ is known}$$
$$z_c = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad \text{if } \sigma \text{ is unknown}$$

- Rejection region:

$$|z_c| > z_{\alpha/2}$$

- P-value:

$$P = 2P(Z > |z_c|)$$

Commonly Used Significance Levels

α (significance level)	z_α
0.005	2.575
0.01	2.326
0.025	1.96
0.05	1.645
0.1	1.282

Summary: One Sample z-test for mean

- null hypothesis: $H_0 : \mu = \mu_0$
- alternative hypothesis:
  - One-tailed, upper-tailed: $H_1 : \mu > \mu_0$
  - One-tailed, lower-tailed: $H_1 : \mu < \mu_0$
  - Two-tailed: $H_1 : \mu \neq \mu_0$
- Test statistic:

$$z_c = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad \text{if } \sigma \text{ is known}$$
$$z_c = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad \text{if } \sigma \text{ is unknown}$$

- Decision:

$$\text{One-tailed case: if } |z_c| > z_\alpha, \text{ we reject } H_0$$
$$\text{Two-tailed case: if } |z_c| > z_{\alpha/2}, \text{ we reject } H_0$$

Example

In the International Journal of Obesity (Jan. 2007), researchers at Vanderbilt University investigated the physiological changes that accompany laughter. 90 subjects (18–34 years old) watched film clips designed to evoke laughter. During the laughing period, the researchers measured the heart rate (beats per minute) of each subject, with the following summary results: $\bar{x} = 73.5$, $s = 6$. It is well known that the mean resting heart rate of adults is 71 beats per minute. Test whether the true mean heart rate during laughter exceeds 71 beats per minute with significance $\alpha = 0.05$.

Let μ be the average of the population.

$$H_0 : \mu = 71$$
$$H_1 : \mu > 71$$

$$z_c = \frac{73.5 - 71}{6/\sqrt{90}} \approx 3.95, \quad z_\alpha = z_{0.05} = 1.645 \quad \Longrightarrow \ |z_c| > z_\alpha$$

We reject H_0. There is enough statistical evidence at significance level 0.05 to conclude that mean heart rate during laughter exceeds 71 beats per minute.

Example

Knowledge of the amount of time a patient occupies a hospital bed—called length of stay (LOS)—is important for allocating resources. At one hospital, the mean LOS was determined to be 5 days. A hospital administrator believes the mean LOS may now be less than 5 days due to a newly adopted managed health care system. To check this, the LOSs (in days) for 100 randomly selected hospital patients were recorded and the mean is $\bar{x} = 4.53$. Assuming that $\sigma = 3.68$, test whether the true mean length of stay at the hospital is less than 5 days with significance $\alpha = 0.05$.

Let μ be the average of the population.

$$H_0 : \mu = 5$$
$$H_1 : \mu < 5$$

$$z_c = \frac{4.53 - 5}{3.68/\sqrt{100}} \approx -1.28, \quad z_\alpha = z_{0.05} = 1.645 \quad \implies |z_c| < z_\alpha$$

We fail to reject H_0. There is insufficient statistical evidence at significance level 0.05 to conclude that the true mean LOS at the hospital is less than 5 days.

Example

Suppose a research neurologist is testing the effect of a drug on response time by injecting 100 rats with a unit dose of the drug, subjecting each rat to a neurological stimulus, and recording its response time. She wishes to test whether the mean response time for drug-injected rats differs from 1.2 seconds. The sample of 100 drug-injected rats yielded the results: $\bar{x} = 1.0517$, $s = 0.4982$. Test it with significance $\alpha = 0.01$.

Let μ be the average of the population.

$$H_0 : \mu = 1.2$$
$$H_1 : \mu \neq 1.2$$

$$z_c = \frac{1.0517 - 1.2}{0.4982/\sqrt{100}} \approx -2.98, \quad z_{\alpha/2} = z_{0.005} = 2.575 \quad \Longrightarrow \; |z_c| > z_{\alpha/2}$$

We reject H_0. There is sufficient statistical evidence at significance level 0.01 to conclude that the mean response time for drug-injected rats differs from the control mean of 1.2 seconds.

One Sample t-test for mean

One Sample t-test for mean

Assumptions:

There is a sample of size n randomly selected from a normal population such that

- $n < 30$,
- population is normally distributed and
- population standard deviation σ is not known

One-tailed (lower-tailed) test

- Hypotheses:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu < \mu_0$$

(This case makes sense only if $\mu_0 > \bar{x}$.)

- Test statistic:

$$t_c = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

(Test statistic is negative.)

- Rejection region:

$$|t_c| > t_\alpha \quad (\text{equivalently, } t_c < -t_\alpha)$$

(t-distribution is based on $n - 1$ degrees of freedom.)

- P-value:

$$P = P(T < t_c)$$

One-tailed (upper-tailed) test

- Hypotheses:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

(This case makes sense only if $\mu_0 < \bar{x}$.)

- Test statistic:

$$t_c = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

(Test statistic is positive.)

- Rejection region:

$$|t_c| > t_\alpha \quad (\text{equivalently, } t_c > z_\alpha)$$

(t-distribution is based on $n - 1$ degrees of freedom.)

- P-value:

$$P = P(T > t_c)$$

Two-tailed test

- Hypotheses:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

- Test statistic:

$$t_c = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

- Rejection region:

$$|t_c| > t_{\alpha/2}$$

  (t-distribution is based on $n - 1$ degrees of freedom.)
- P-value:

$$P = 2P(T > |t_c|)$$

Summary: One Sample t-test for mean

- null hypothesis: $H_0 : \mu = \mu_0$
- alternative hypothesis:
  - One-tailed, upper-tailed: $H_1 : \mu > \mu_0$
  - One-tailed, lower-tailed: $H_1 : \mu < \mu_0$
  - Two-tailed: $H_1 : \mu \neq \mu_0$
- Test statistic:

$$t_c = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

- Decision:

$$\text{One-tailed case: if } |t_c| > t_\alpha, \text{ we reject } H_0$$
$$\text{Two-tailed case: if } |t_c| > t_{\alpha/2}, \text{ we reject } H_0$$

where t-curve is based on $n-1$ degrees of freedom.

Example

A smartphone manufacturer claims that the average battery life of their new model is 30 hours. Assume that battery life follows a normal curve. A customer group tests a random sample of 6 phones and finds the following battery lifespans, in hours:

$$31.2,\ 30.4,\ 29.2,\ 28.1,\ 27.9,\ 27.2.$$

At a significance level of $\alpha = 0.1$, can the customer group conclude that the average battery life is less than the manufacturer's claim?

$$H_0 : \mu = 30$$
$$H_1 : \mu < 30$$

$$\bar{x} = \frac{31.2 + 30.4 + 29.2 + 28.1 + 27.9 + 27.2}{6} = 29.$$

$$s = \sqrt{\frac{(2.2)^2 + (1.4)^2 + (0.2)^2 + (-0.9)^2 + (-1.1)^2 + (-1.8)^2}{5}} = \sqrt{2.42}.$$

Example (cont'd)

$$
\begin{aligned}
t_c &= \frac{29 - 30}{\sqrt{2.42}/\sqrt{6}} \approx -1.575 \\
t_\alpha &= t_{0.1} = 1.476 \text{ (degrees of freedom } = 6 - 1 = 5) \\
\Longrightarrow & \left|t_c\right| > t_\alpha
\end{aligned}
$$

We reject H_0. There is enough statistical evidence at significance level 0.1 to conclude that the average battery life is less than the manufacturer's claim.

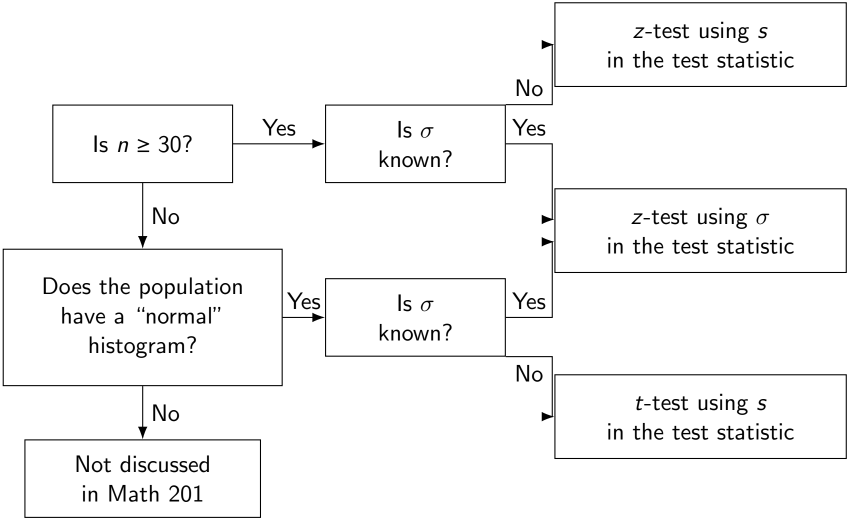
Example (cont'd)

Same question with $\alpha = 0.05$.

$$
\begin{aligned}
t_c &= -1.575 \\
t_\alpha &= t_{0.05} = 2.015 \text{ (degrees of freedom } = 6 - 1 = 5) \\
\Longrightarrow & |t_c| < t_\alpha
\end{aligned}
$$

We fail to reject H_0. There is not enough statistical evidence at significance level 0.05 to conclude that the average battery life is less than the manufacturer's claim.

Hypothesis Test Flow Chart



One Sample z-test for proportion

One Sample z-test for proportion

Assumptions:

There is a sample of size n randomly selected from a population with $np_0 \geq 15$ and $n(1 - p_0) \geq 15$.

One-tailed (lower-tailed) test

- Hypotheses:

$$H_0 : p = p_0$$
$$H_1 : p < p_0$$

  (This case makes sense only if $p_0 > \hat{p}$.)
- Test statistic:

$$z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

  (Test statistic is negative.)
- Rejection region:

$$|z_c| > z_\alpha \quad (\text{equivalently, } z_c < -z_\alpha)$$

- P-value:

$$P = P(Z < z_c)$$

One-tailed (upper-tailed) test

- Hypotheses:

$$H_0 : p = p_0$$
$$H_1 : p > p_0$$

(This case makes sense only if $p_0 < \hat{p}$.)

- Test statistic:

$$z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

(Test statistic is positive.)

- Rejection region:

$$|z_c| > z_\alpha \quad \text{(equivalently, } z_c > z_\alpha\text{)}$$

- P-value:

$$P = P(Z > z_c)$$

Two-tailed test

- Hypotheses:

$$H_0 : p = p_0$$
$$H_1 : p \neq p_0$$

- Test statistic:

$$z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

- Rejection region:

$$|z_c| > z_{\alpha/2}$$

- P-value:

$$P = 2P(Z > |z_c|)$$

Summary: One Sample z-test for proportion

- null hypothesis: $H_0 : p = p_0$
- alternative hypothesis:
  - One-tailed, upper-tailed: $H_1 : p > p_0$
  - One-tailed, lower-tailed: $H_1 : p < p_0$
  - Two-tailed: $H_1 : p \neq p_0$
- Test statistic:

$$z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

- Decision:

$$\text{One-tailed case: if } |z_c| > z_\alpha, \text{ we reject } H_0$$
$$\text{Two-tailed case: if } |z_c| > z_{\alpha/2}, \text{ we reject } H_0$$

Example

Volkan Öge is running in an election. He hires a survey organization, which takes a simple random sample of 1000 voters. In the sample, 200 of 1000 favor him. Will Volkan Öge get more than 17% of the vote in the election? Test it with significance $\alpha = 0.01$.

$$H_0 : p = 0.17$$
$$H_1 : p > 0.17$$

$$\hat{p} = \frac{200}{1000} = 0.20$$

$$z_c = \frac{0.20 - 0.17}{\sqrt{\frac{0.17 \cdot 0.83}{1000}}} \approx 2.53, \quad z_\alpha = z_{0.01} = 2.326 \quad \Longrightarrow \; |z_c| > z_\alpha$$

We reject H_0. There is enough statistical evidence at significance level 0.01 to conclude that Volkan Öge will get more than 17% of the vote in the election.

Example

A random number generator produces either 0 or 1. It is supposed to produce them with equal probabilities. We used it 100 times, and it produced 55 times 1 and 45 times 0. Is this a good random number generator? (Is the true proportion of 1's 0.5?) Test it with significance $\alpha = 0.1$.

$$H_0 : p = 0.5$$
$$H_1 : p \neq 0.5$$

$$\hat{p} = \frac{55}{100} = 0.55$$

$$z_c = \frac{0.55 - 0.50}{\sqrt{\frac{0.50 \cdot 0.50}{100}}} = 1, \quad z_{\alpha/2} = z_{0.05} = 1.645 \quad \Longrightarrow \; |z_c| < z_{\alpha/2}$$

We fail to reject H_0. There is not enough statistical evidence at significance level 0.1 to conclude that this is not a good random number generator.