Biomedical Engineering Degree

3. Hypothesis Testing: One-sample Inference

Felipe Alonso Atienza
✉felipe.alonso@urjc.es
@FelipeURJC

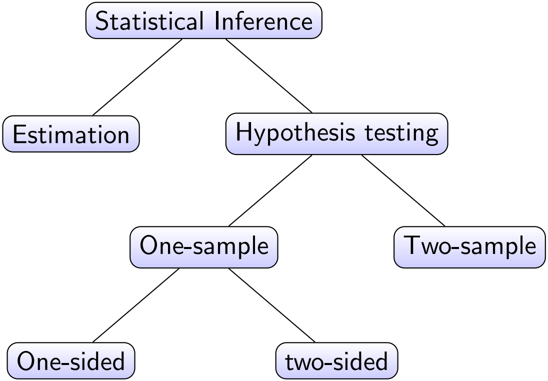
Escuela Técnica Superior de Ingeniería de Telecomunicación
Universidad Rey Juan Carlos

References

1. R. Bernard. *Fundamentals of Biostatistics*. Ed.: Thompson. Chapter 7
2. B. Caffo. *Statistical Inference for Data Science*. Leanpub. Chapters $9-11$
3. D. Díez, M Cetinkaya-Rundel and CD Barr. *OpenIntro Statistics*. Chapter 7.

Outline

Mind map



- Hypothesis testing: deciding between **two decisions**
  - One-sample: hypotheses are specified about a **single distribution**
  - Two-sample: **two different distributions** are compared

Example

Test formulation

We want to test whether mothers with low socioeconomic status deliver babies whose birthweights are lower than the national average.

Data

A list is obtained of birthweights from 100 consecutive, full-term, live-born deliveries from the maternity ward of a hospital in a low socioeconomic status area. The mean birthweight ($\bar{x}$) is found to be $3,2602\,\mathrm{kg}$ with a sample standard deviation of $0,6804\,\mathrm{kg}$. Suppose we know from nationwide surveys based on millions of deliveries that the mean birthweight in the US is $3,4019\,\mathrm{kg}$

Hypothesis

Do these differences arise *purely by chance*?

Outline

Hypothesis testing

We can solve this kind of questions using *hypothesis testing*:

1. **Null hypothesis** (H_0): is the hypothesis that is to be tested and represents the **status quo** (or the *purely by chance* option, or *assume by default*[1])
   - $H_0 =$ there are no differences on mean birthweight
2. **Alternative hypothesis** (H_1): is the hypothesis that in some sense contradicts the null hypothesis which **requires evidence** to conclude
   - $H_1 =$ birthweights in low socioeconomic status area are lower than the national average.

[1] *Everyone is innocent until proven guilty*

Example

Mathematical formulation

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu < \mu_0$$

where μ is the mean birthweight in the low socioeconomic status area hospital, and μ_0 is the national average.

We also assume the **underlying distribution is normal** under either hypothesis

Types of errors

There are four possible outcomes of our statistical decision process

Truth	Decide	Result
H_0	H_0	Correctly accept null
H_0	H_1	Type I error
H_1	H_1	Correctly reject null
H_1	H_0	Type II error

Example

- Type I error would be the probability of deciding that the mean birthweight in the hospital was lower than $3,4019\,\mathrm{kg}$ when in fact it was $3,4019\,\mathrm{kg}$
- Type II error would be the probability of deciding that the mean birthweight was $3,4019\,\mathrm{kg}$ when in fact it was lower than $3,4019\,\mathrm{kg}$

Definitions

Significance level of a test

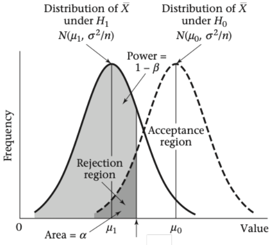
$$\alpha = P(\text{Type I error}) = P(\text{accepting } H_1 | H_0 \text{ is true})^a$$

[a] In the context of detection problems, this corresponds to P_{FA}

Power of a test

$$1 - \beta = 1 - P(\text{Type II error}) = P(\text{rejecting } H_0 | H_1 \text{ is true})^a$$

[a] In the context of detection problems, this corresponds to P_D



Objective

- The general aim in hypothesis testing is to use <u>statistical tests</u> that make α and β **as small as possible**
- However, there's a trade-off between α and β
  - If $\alpha \downarrow \Rightarrow \beta \uparrow \Rightarrow 1 - \beta \downarrow$

Neyman-Pearson approach

We fix α at some specific level (for example, $.10, .05, .01, \ldots$) and to use the test that minimizes β or, equivalently, maximizes the power.

Outline

One-sided

- Or **one-tailed test** is a test in which the values of the parameter being studied under the alternative hypothesis H_1 is allowed to be either *greater than* $>$ or *less than* $<$ the values of the parameter under the null hypothesis, *but not both*.
  - $H_1 : \mu > \mu_0$
  - $H_1 : \mu < \mu_0$

Example

How can we design our test to analyze if birthweight are lower than average? We do not know μ nor σ^2 for low socioeconomic status data.

- Solution: use the sample mean $\bar{x}$.
- Intuition: the higher the difference between $\bar{x}$ and μ_0 the higher the evidence to reject the null hypothesis

One-sample (one-sided) $t-$Test ($\mu < \mu_0$)

- To test the hypothesis

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu < \mu_0$$

with unknown σ (both for H_0 and H_1) with a significance level of α

$t-$Test ($\mu < \mu_0$)

Compute

$$t = \frac{\bar{x} - \mu_0}{s_*/\sqrt{n}} \sim t_{n-1}$$

- if $t < t_{n-1,\alpha}$, the we reject H_0
- if $t \geq t_{n-1,\alpha}$, the we accept H_0

Revisiting our example

Test formulation

We want to test whether mothers with low socioeconomic status deliver babies whose birthweights are lower than the national average.

Data

A list is obtained of birthweights from 100 consecutive, full-term, live-born deliveries from the maternity ward of a hospital in a low socioeconomic status area. The mean birthweight ($\bar{x}$) is found to be $3,2602\,\mathrm{kg}$ with a sample standard deviation of $0,6804\,\mathrm{kg}$. Suppose we know from nationwide surveys based on millions of deliveries that the mean birthweight in the US is $3,4019\,\mathrm{kg}$

Hypothesis

Do these differences arise *purely by chance*?

NOW, YOU CAN ANSWER TO THIS QUESTION![2]

[2] Use a significance level of 0.05

Revisiting our example: solution

- First, compute the statistic

$$t = \frac{\bar{x} - \mu_0}{s_*/\sqrt{n}} = \frac{3.2602 - 3.4019}{0.6804/\sqrt{100}} = -2.08259$$

- Using Python: $t_{n-1,\alpha} =$ `t(df=99).ppf(0.05)` $= -1.66039$
- Since $t < t_{99,0.05}$ then

$$\boxed{\text{We reject the null hypothesis } H_0 \text{ at a significance level of } 0.05}$$

$p-$value

- From the previous slide, it looks like we should be be performing different test at different α values
- Instead, we can calculate the $p-$value for the test

$p-$value

The $p-$value is the α level at which the given value of the test statistic is on the borderline between the acceptance and rejection regions

$$p = P(t_{n-1} \leq t)$$

Thus, p is the **area to the left of t under a t_{n-1} distribution**.

$p-$value

What's is the $p-$value for our previous example?

1. `t(df=99).cdf(-1.6604)` $= 0.05$
2. `t(df=99).cdf(-1.6604)` $= -2.08$
3. `t(df=99).cdf(-2.082)` $= 0.02$
4. None of the above

Guidelines for judging the significance of a $p-$value

- How small should the $p-$value be for results to be considered statistically significant?

- If $.01 \leq p < .05$, then the results are **significant**
- If $.001 \leq p < .01$, then the results are **highly significant**
- If $p < .001$, then the results are **very highly significant**.
- If $p > .05$, then the results are considered **not statistically significant**

Which method should I use?

- Use either:
  1. **Critical-value method** ($t < t_{n-1,\alpha}$) with $\alpha = 0.05$
  2. $p-$**value method** with threshold in $p < 0.05$

One-sample (one-sided) $t-$Test ($\mu > \mu_0$)

- To test the hypothesis

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0$$

with unknown σ (both for H_0 and H_1) with a significance level of α

$t-$Test ($\mu > \mu_0$)

Compute

$$t = \frac{\bar{x} - \mu_0}{s_* / \sqrt{n}} \sim t_{n-1}$$

- if $t > t_{n-1,1-\alpha}$, the we reject H_0
- if $t \leq t_{n-1,1-\alpha}$, the we accept H_0

$p-$value

$$p = P(t_{n-1} \geq t) = 1 - P(t_{n-1} \leq t)$$

Thus, p is the **area to the right of** t **under a** t_{n-1} **distribution**

One-sample (one-sided) $t-$Test ($\mu > \mu_0$)

Example

Suppose the "average" cholesterol level in children is $175\,\mathrm{mg/dL}$. A group of men who have died from heart disease within the past year are identified, and the cholesterol levels of their offspring are measured. Suppose the mean cholesterol level of 10 children whose fathers died from heart disease in is $200\,\mathrm{mg/dL}$ and the sample standard deviation is $50\,\mathrm{mg/dL}$. Test the hypothesis that the mean cholesterol level is higher in this group than in the general population (use $\alpha = 0.05$).

Example solution

- Using the critical-value approach, first we compute the statistic

$$t = \frac{\bar{x} - \mu_0}{s_*/\sqrt{n}} = \frac{200 - 175}{50/\sqrt{10}} = 1.58$$

and then: $t_{n-1,1-\alpha} = \texttt{t(df=9).ppf(0.95)} = 1.833$. Thus, $t \leq t_{9,0.95}$

- On the other hand, using the $p-$value approach, we calculate

$$\begin{aligned} p = P(t_9 > 1.58) &= 1 - P(t_9 \leq 1.58) = \\ &= \texttt{1-t(df=9).cdf(1.58)} = 0.0743 > 0.05 \end{aligned}$$

then, the results are no statistically significant to reject H_0

$$\boxed{\text{We accept the null hypothesis } H_0 \text{ at a significance level of } 0.05}$$

Outline

Two-sided

- We would like to reject if the true mean is **different** $\neq$ than the hypothesized, *not just larger or smaller*.
- In other words, we would reject the null hypothesis if in fact the sample mean was much larger or smaller than the hypothesized mean

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

Two-sided

Example

- Suppose we want to compare fasting serum-cholesterol levels among recent Asian immigrants to the United States with typical levels found in the general U.S. population.
- Suppose we assume cholesterol levels in women ages $21 - 40$ in the U.S. are approximately normally distributed with mean $190\,\mathrm{mg/dL}$. It is unknown whether cholesterol levels among recent Asian immigrants are higher or lower than those in the general U.S. population.
- Let's assume that levels among recent female Asian immigrants are normally distributed with unknown mean μ. Hence we wish to test the null hypothesis $H_0 : \mu = \mu_0 = 190$ vs. the alternative hypothesis $H_1 : \mu \neq \mu_0$
- Blood tests are performed on 100 female Asian immigrants ages $21 - 40$, and the mean level $(\bar{x})$ is $181.52\,\mathrm{mg/dL}$ with standard deviation of $40\,\mathrm{mg/dL}$.

One-sample (two-sided) $t-$Test ($\mu \neq \mu_0$)

- To test the hypothesis

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

with unknown σ (both for H_0 and H_1) with a significance level of α

$t-$Test ($\mu \neq \mu_0$)

Compute

$$t = \frac{\bar{x} - \mu_0}{s_* / \sqrt{n}} \sim t_{n-1}$$

- if $|t| > t_{n-1,1-\alpha/2}$, then we reject H_0
- if $|t| \leq t_{n-1,1-\alpha/2}$, then we accept H_0

Your turn

Solve the previous example

One-sample (two-sided) $t-$Test ($\mu \neq \mu_0$)

$p-$value

What would the $p-$value be?

Connections with confidence intervals

- To test the hypothesis

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

with unknown σ (both for H_0 and H_1) with a significance level of α

confidence interval Test

- if $100\,\%(1-\alpha)$ CI **does not contain** μ_0, then we reject H_0
- if $100\,\%(1-\alpha)$ CI **does contain** μ_0, then we accept H_0

Connections with confidence intervals

Example

Consider our previous cholesterol example, where we had blood tests performed on 100 female Asian immigrants ages $21 - 40$, with mean level $\bar{x} = 181.52\,\mathrm{mg/dL}$, and standard deviation of $40\,\mathrm{mg/dL}$.

- Calculate the $95\,\%$ CI for μ
- Can we accept H_0?

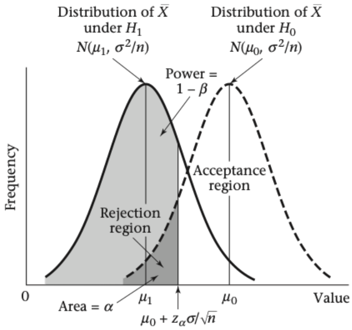
Outline

Power of a test

- Let's calculate the power to a test $(1-\beta)$ for our first example:

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu < \mu_0$$

- Let's assume that σ^2 is know



Power of a test

- The critical value is $\gamma = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$
- So, the power can be calculated as

$$
\begin{aligned}
1-\beta &= P(\overline{X} < \gamma | H_1) \sim \mathcal{N}\left(\mu_1, \frac{\sigma^2}{n}\right) \\
&= P\left(Z < \frac{\gamma - \mu_1}{\sigma/\sqrt{n}}\right) \sim \mathcal{N}(0,1)
\end{aligned}
$$

- Using the value of γ

$$
\begin{aligned}
1-\beta &= P\left(Z < \frac{\mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} - \mu_1}{\sigma/\sqrt{n}}\right) \\
&= P\left(Z < z_\alpha + \frac{\mu_0 - \mu_1}{\sigma}\sqrt{n}\right)
\end{aligned}
$$

Example

- Compute the power of the test for the birthweight data

Data

A list is obtained of birthweights from 100 consecutive, full-term, live-born deliveries from the maternity ward of a hospital in a low socioeconomic status area. The mean birthweight ($\bar{x}$) is found to be $3,2602\,\mathrm{kg}$ with a **true standard deviation** of $\sigma = 0,6804\,\mathrm{kg}$. Suppose we know from nationwide surveys based on millions of deliveries that the mean birthweight in the US is $3,4019\,\mathrm{kg}$

Example solution

- $z_\alpha = z_{0.05} = \texttt{norm().ppf(0.05)} = -1.645$
- $\mu_0 = 3,4019$, $\mu_1 = 3,2602$, $\sigma = 0,6804$, $n = 100$

$$
\begin{aligned}
1-\beta &= P\left(Z < z_\alpha + \frac{\mu_0 - \mu_1}{\sigma}\sqrt{n}\right) = \\
&= P\left(Z < -1.645 + \frac{3,4019 - 3,2602}{0,6804}\sqrt{100}\right) = \\
&= P(Z < 0.4377) = \texttt{norm().cdf(0.4377)} = 0.6692
\end{aligned}
$$

- Therefore, there is about a $67\,\%$ chance of detecting a significant difference using a $5\,\%$ significance level with this sample size.

Factors affecting the power

$$\boxed{1-\beta = P\left(Z < z_{\alpha} + \frac{\mu_0-\mu_1}{\sigma}\sqrt{n}\right)}$$

- if $\alpha \downarrow \Rightarrow (1-\beta) \downarrow$
- if $|\mu_0-\mu_1| \uparrow \Rightarrow (1-\beta) \uparrow$
- if $\sigma \uparrow \Rightarrow (1-\beta) \downarrow$
- if $n \uparrow \Rightarrow (1-\beta) \uparrow$

Sample-size determination

- From previous reasoning, fixing the power $(1-\beta)$, and the values of α, μ_0, μ_1 and σ we could calculate the sample size n
- If we fix the power $1-\beta = P(Z < z_{1-\beta})$, then we can calculate $z_{1-\beta}$ and thus,

$$1-\beta = P(Z < z_{1-\beta}) = P\left(Z < z_\alpha + \frac{\mu_0-\mu_1}{\sigma}\sqrt{n}\right)$$

- Therefore,

$$z_{1-\beta} = z_\alpha + \frac{\mu_0-\mu_1}{\sigma}\sqrt{n}$$

- Taking into account that $-z_\alpha = z_{1-\alpha}$, then

$$\boxed{n = \left[\frac{(z_{1-\alpha}+z_{1-\beta})\sigma}{\mu_0-\mu_1}\right]^2}$$

Example

- Compute the appropriate sample size needed to conduct the test if $\alpha = 0.05$ and $1 - \beta = 0.8$

Data

A list is obtained of birthweights from 100 consecutive, full-term, live-born deliveries from the maternity ward of a hospital in a low socioeconomic status area. The mean birthweight ($\bar{x}$) is found to be $3,2602\,\mathrm{kg}$ with a **true standard deviation** of $\sigma = 0,6804\,\mathrm{kg}$. Suppose we know from nationwide surveys based on millions of deliveries that the mean birthweight in the US is $3,4019\,\mathrm{kg}$

Example solution

- $z_{1-\alpha} = z_{1-0.05} = \texttt{norm().ppf(0.95)} = 1.645$
- $z_{1-\beta} = z_{0.8} = \texttt{norm().ppf(0.8)} = 0.8416$
- $\mu_0 = 3.4019$, $\mu_1 = 3.2602$, $\sigma = 0.6804$

$$
\begin{aligned}
n &= \left[\frac{(z_{1-\alpha} + z_{1-\beta})\sigma}{\mu_0 - \mu_1}\right]^2 \\
&= \left[\frac{(1.645 + 0.8416)0.6804}{3.4019 - 3.2602}\right]^2 = 142.56
\end{aligned}
$$

- Thus, a sample size of 143 is needed to have an $80\,\%$ chance of detecting a significant difference at the $5\,\%$ level if the alternative mean is $3,2602\,\mathrm{kg}$ and a one-sided test is used.

Outline

Point estimate

- Recall that we estimate a proportion p as

$$\hat{p} = \frac{x}{n}$$

  where x is the total number of successes and n is the sample size (that is the sample proportion).

- The standard error is given by

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

Test for $p \neq p_0$

- To test the hypothesis $H_0 : p = p_0 \quad \text{vs} \quad H_1 : p \neq p_0$ with a significance level of α

Test $(p \neq p_0)$[a]

[a]This test is valid under the assumption $np_0(1-p_0) \geq 5$

Compute

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim \mathcal{N}(0,1)$$

- if $|z| > z_{1-\alpha/2}$, then we reject H_0

Confidence interval[a]

[a]This test is valid under the assumption $np_0(1-p_0) \geq 5$

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

- if $100\,\%(1-\alpha)$ CI **does not contain** p_0, then we reject H_0

Example

Example

Suppose that 400 of the 10000 women ages $50-54$ sampled whose mothers had breast cancer, had breast cancer themselves at some time in their lives. Given large studies, assume the prevalence rate of breast cancer for U.S. women in this age group is about $2\,\%$. The question is: How compatible is the sample rate of $4\,\%$ with a population rate of $2\,\%$?

- In other words,

$$H_0 : p = 0.02 \quad \text{vs} \quad H_1 : p \neq 0.02$$