

Statistik för bioteknik SF1911
Föreläsning 10: Hypotesprövning och statistiska test
Timo Koski

TK

27.11.2017



KTH Matematik

Outline of Lecture 10.

- Statistical Test: What is it?
- Structure and Logic of a Statistical Test (1)
 - hypothesis, null hypothesis, alternative hypothesis, test statistic, significance level, critical value, critical region, rare event rule
 - Decision: Critical Region, p -value, Confidence Interval, Alternative
- Test of Differences of Means
 - Two groups, t-test for difference of means
 - Matched pairs or the paired t-test

Statistical Test: What is it ? Recall Mendel's Peas

Gregor Mendel obtained by crossing peas with green pods and peas with yellow pods the offspring of 580 peas. Among them 428 had green pods, and 152 had yellow pods.

We wanted to infer, e.g., the **proportion of yellow pods that would be obtained in similar experiments**. We understand this proportion as the population parameter denoted p . The best estimate of p based on this data is $\hat{p} = \frac{152}{580} = 26.2\%$ The theory or **hypothesis** of Mendel was that 25 % of the peas would have yellow pods.



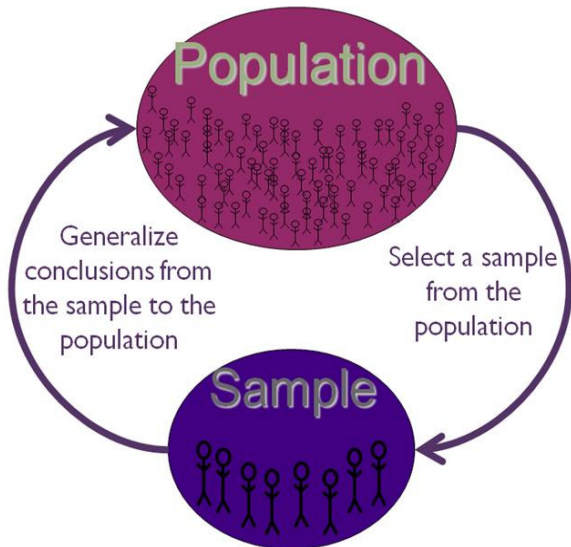
What is it ? Recall the Statement

The CI was found as

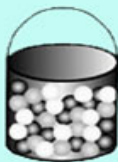
$$0.226 < p < 0.298$$

This is 95 % confidence that the limits 22.6 % and 29.8 % contain the true percentage of offspring peas with yellow pods. This interval includes 25%, so Mendel's expected value of 25% cannot be described as wrong. The results do not appear to provide significant evidence against the 25% rate in Mendel's hypothesis.

This is an example of what is called a **statistical test**. We proceed by finding out in which sense the conclusion above contains the structure and logic of a statistical test.



Statistical Inference: Random sample: no rigging, no selection bias



Probability: Given the information in the pail, what is in your hand?



Statistics: Given the information in your hand, what is in the pail?

Definition

*In statistics, a **hypothesis** is a claim or statement about a property of a population. A **hypothesis test** (or a significance test) is a standard procedure for checking a claim about a property of a population.*

Example of a hypothesis:

- Mendel claims that under certain circumstances, the percentage of the offspring peas with yellow pods is 25 %.

Standard procedure contains: *null hypothesis, alternative hypothesis, test statistic, significance level, critical value, decision rule, proceed assuming null hypothesis is true, with the logic of the rule of rare event.*

The Individual Components of a Statistical Test

- Given a claim identify a statistical model for your population and the **null hypothesis** and the **alternative hypothesis**.
- Given a claim and sample data, compute the value of the **test statistic**
- Given **significance level**, identify the **critical values**.
- Given a value of a test statistic, check against critical value, identify the **p -value**.

We shall now explain what this is.

Null and Alternative Hypothesis

- The **null hypothesis** (denoted by H_0) is a statement that the value of a population parameter (such as proportion, mean, standard deviation) is equal to some claimed value. Examples:

$$H_0 : p = 0.5 \quad H_0 : \mu = 0.86$$

We **assume that H_0 is true and reach a conclusion either to reject H_0 or fail to reject H_0 .**

- The **alternative hypothesis** (denoted by H_1) is a statement that the value of a population parameter is somehow different from the null hypothesis.

$$H_1 : p > 0.5 \quad H_1 : p < 0.5 \quad H_1 : p \neq 0.5$$



The **test statistic** is a value computed from the sample data and it is used to make the decision about the rejection of the null hypothesis. The test statistic is used to for determining whether there is significant evidence against the null hypothesis. Examples:

- Test statistic for proportion: $z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$.
- Test statistic for mean: $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ or $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$.

- The **critical region** (or **rejection region**) is the set of all values of the test statistic that cause us to reject the null hypothesis.
- The **significance level** denoted by α is the probability that the test statistic will fall in the critical region when the null hypothesis is actually true. Common choices of α are 0.10, 0.05, and 0.01.
- A **critical value** is any value that separates the critical region, where we reject the null hypothesis, from the values of the test statistic that do not lead to rejection of the null hypothesis.

Two-tailed, Left-tailed, Right-tailed

The *tails* in a density curve are the regions bounded by critical values. Some hypothesis tests are two-tailed, some are right-tailed and some are left-tailed.

- **Two-tailed test:** the critical region in two parts, significance level is a sum of two areas under the density curve. The significance level is divided equally between the two tails. (Most of the tests in this lecture are of this kind).
- **Left-tailed test:** the critical region in the left tail of the density curve.
- **Right-tailed test:** the critical region in the right tail of the density curve.

Example: Mendel's Peas & Hypothesis Again

$$H_0 : p = 0.25 \quad H_1 : p \neq 0.25$$

We have $\hat{p} = \frac{152}{580} = 26.2\%$. We evaluate the test statistic using the null hypothesis, i.e., assuming that the null hypothesis is true, i.e., $p = 0.25$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.262 - 0.25}{\sqrt{\frac{0.25 \cdot 0.75}{580}}} = 0.67$$

Let us take $\alpha = 0.05$. The critical value is $\lambda_{0.025} = 1.96$ (a two tailed critical region).

Step One: Critical values

Critical values

$P(Z > \lambda_\alpha) = \alpha$ där $Z \in N(0, 1)$

α	λ_α	α	λ_α
0.10	1.2816	0.001	3.0902
0.05	1.6449	0.0005	3.2905
0.025	1.9600	0.0001	3.7190
0.010	2.3263	0.00005	3.8906
0.005	2.5758	0.00001	4.2649

Example: Mendel's Peas & Hypothesis Again

$$H_0 : p = 0.25 \quad H_1 : p \neq 0.25$$

We have, if H_0 is true (c.f. previous lectures)

$$Z = \frac{\hat{p} - 0.25}{\sqrt{\frac{0.25 \cdot 0.75}{n}}} \quad \text{approximately } \sim N(0, 1)$$

Let us take $\alpha = 0.05$. The critical value is $\lambda_{0.025} = 1.96$ (a two tailed critical region). Then

$$P(-1.96 \leq Z \leq 1.96) = 1 - 0.05 = 0.95.$$



Example: Mendel's Peas & Hypothesis Again

The critical region has two parts (tails)

$$z < -1.96 \quad \text{or} \quad z > 1.96$$

Since the computed value of test statistic is $z = 0.67$, it does not lie in the critical region, and we fail to reject H_0 at the significance level 5%. \square

The **p-value** (or **probability value**) is the probability of getting a value of the test statistic which is at least as extreme as the one representing the sample data. The null hypothesis is rejected if the P-value is very small, such as 0.05 or less.

Example

With Mendel's peas,
 $Pr(Z > 0.67) = 1 - Pr(Z \leq 0.67) = 1 - 0.749 = 0.251$, and **P-value**
 $= 2 \cdot 0.251 = 0.502$. We multiply by two because we have a two-tailed test. Hence we fail to reject Mendel's H_0 by P-value.

Recall: Rare Event Rule

If, under a given assumption, the probability of an observed event is very small, we conclude that the assumption is likely not correct.

Rare Event Rule, Hypothesis Testing

If, under $H_0 : p = 0.25$, the probability of the test statistic falling in the critical region is very small, we conclude that $H_0 : p = 0.25$ is likely not correct.

We have a under $H_0 : p = 0.25$ that

$$Z = \frac{\hat{p} - 0.25}{\sqrt{\frac{0.25 \cdot 0.75}{n}}} \quad \text{approximately } \sim N(0, 1)$$

and therefore

$$P(-1.96 \leq Z \leq 1.96) = 1 - 0.05 = 0.95 = 1 - \alpha.$$

or the probability of finding Z in the critical region is

$$P(Z > -1.96) + P(Z \leq 1.96) = 0.025 + 0.025 = 0.05 (= \alpha).$$



- **Basic:** *Reject H_0 if the test statistic falls within the critical region. Fail to reject H_0 if the test statistic does not fall within the critical region.*
- **p-value method:** *Reject H_0 if p-value $\leq \alpha$, where α is the significance level, e.g., 0.05. Fail to reject H_0 if p-value $> \alpha$*
- **Another option:** *Instead of using a significance level like $\alpha = 0.05$ identify the p-value and leave the conclusion to the reader or your boss.*
- **Confidence Intervals:** *Reject a hypothesis if the population parameter has a value under H_0 that is not included in the interval.*

When we used the 95 % confidence interval with the limits 22.6 % and 29.8 % and noted that this interval includes 25%, we failed to reject Mendel's expected value of 25% using the criterion of confidence intervals.

Test of Differences of Means

- Two groups, t-test for difference of means
- Matched pairs or the paired t-test



Two samples & The t-Test

We shall introduce methods for using sample data from two independent samples to test the hypotheses about population means via CI estimates of the difference between two population means.

Definition

*Two samples are **independent** if the sample values selected from one population are not related to or somehow paired or matched with the sample values selected from the other population.*

Two samples

Statistical model:

$$X_1, X_2, \dots, X_{n_1} \sim \mathcal{N}(\mu_1, \sigma_1^2) \quad (\text{sample 1})$$

$$Y_1, Y_2, \dots, Y_{n_2} \sim \mathcal{N}(\mu_2, \sigma_2^2) \quad (\text{sample 2})$$

where all X s and Y s are independent.

Two samples, test statistic for the difference between means: Case (a)

(a) σ_1 and σ_2 **known**

We plan to get a CI for $\mu_1 - \mu_2$, denoted by $I_{\mu_1 - \mu_2}$. A natural estimate of $\mu_1 - \mu_2$ is $\bar{X} - \bar{Y}$. We get the test statistic (Z score)

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

(a) σ_1 and σ_2 **known**

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

implies that the two-tailed critical region with confidence level $1 - \alpha$ is

$$Z < -z_{\alpha/2} \quad \text{or} \quad Z > z_{\alpha/2}$$

Here $z_{\alpha/2}$ is a critical value obtained by the table above.

Two samples, confidence interval for the difference between means: Case (a)

Then, we are in the same situation as in a previous lecture on confidence intervals, we are led to the CI for $\mu_1 - \mu_2$

$$I_{\mu_1 - \mu_2} = \bar{x} - \bar{y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Here $z_{\alpha/2}$ is a critical value obtained by the table above.

Two samples, confidence interval for the difference between means: Case (a)

If $\sigma_1 = \sigma_2 = \sigma$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

and

$$I_{\mu_1 - \mu_2} = \bar{x} - \bar{y} \pm z_{\alpha/2} \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

$z_{\alpha/2}$ is a critical value obtained by the table above.

Two samples, confidence interval for the difference between means: Case (a)

The Confidence Interval Estimate of $\mu_1 - \mu_2$ with $\sigma = \sigma_1 = \sigma_2$ known is

$$\bar{x} - \bar{y} - E < \mu_1 - \mu_2 < \bar{x} - \bar{y} + E$$

where

$$E = z_{\alpha/2} \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

where $z_{\alpha/2}$ is a critical value obtained by the table above. With these we can use the confidence interval criterion of testing a hypothesis on $\mu_1 - \mu_2$ with confidence degree $1 - \alpha$.

Two samples, confidence interval for the difference between means: Confidence interval criterion

$$H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$$

Find

$$(\bar{x} - \bar{y} - E, \bar{x} - \bar{y} + E)$$

If this interval includes zero, we fail to reject H_0 .

Two samples, confidence interval for the difference between means: Case (b)

(b) $\sigma_1 = \sigma_2 = \sigma$ **unknown**

The estimate s of σ is, e.g., the square root of the *pooled* sample variance (s_1^2 and s_2^2 are the sample variances of sample one and two, respectively):

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

and the test statistic is

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

has **t-distribution** with $n_1 + n_2 - 2$ **degrees of freedom**.



Two samples, confidence interval for the difference between means: Case (b)

The Confidence Intervals Estimate of $\mu_1 - \mu_2$ is

$$\bar{x} - \bar{y} - E < \mu_1 - \mu_2 < \bar{x} - \bar{y} + E$$

where

$$E = t_{\alpha/2} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

where $t_{\alpha/2}$ is a critical value of t -distribution with $n_1 + n_2 - 2$ degrees of freedom and

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Two samples, confidence interval for the difference between means: Case (b)

The Simplified Confidence Interval Estimate of $\mu_1 - \mu_2$ is

$$\bar{x} - \bar{y} - E < \mu_1 - \mu_2 < \bar{x} - \bar{y} + E$$

where

$$E = t_{\alpha/2}s$$

where $t_{\alpha/2}$ is a critical value of t -distribution with $n_1 + n_2 - 2$ degrees of freedom and

$$s = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Simplified \rightarrow no pooling in s .

Two samples, confidence interval for the difference between means: Case (b), Example

In an experiment designed to test the effectiveness of paroxetine for treating bipolar depression, subjects were measured using the Hamilton Depression scale with results as follows:

Placebo group	$n_1 = 43$	$\bar{x} = 21.57, s_1 = 3.87$
Paroxetine treatment group	$n_2 = 33$	$\bar{y} = 20.38, s_2 = 3.91$

Two samples, confidence interval for the difference between means: Case (b), Example of t-Test

In an experiment designed to test the effectiveness of paroxetine for treating bipolar depression.

Null hypothesis: there is no difference between treatment and placebo

$$H_0 : \mu_1 = \mu_2$$

Alternative hypothesis: there is a difference between treatment and placebo

$$H_1 : \mu_1 \neq \mu_2$$

We now proceed under H_0 , or $\mu_1 - \mu_2 = 0$. Significance level is $\alpha = 0.05$.

Two samples, confidence interval for the difference between means: Case (b), Example

We now proceed under H_0 , or $\mu_1 - \mu_2 = 0$. We assume that we have independent samples and normal distributions. We must assume this, we have just summaries of data, no chance of looking at histograms or boxplots or normplots. We use the test statistic

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Two samples, confidence interval for the difference between means: Case (b), Example

Insert data and use $H_o : \mu_1 - \mu_2 = 0$ and the simplified computation (no pooling)

$$\begin{aligned} t &= \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\ &= \frac{(21.57 - 20.38) - 0}{\sqrt{\frac{3.87^2}{43} + \frac{3.91^2}{33}}} = 1.321 \end{aligned}$$

Two samples, confidence interval for the difference between means: Case (b), Example

$$t = 1.321$$

Since we are using t-distribution, the critical numbers are found from the t-distribution with $43 + 33 - 2 = 74$ degrees of freedom. This gives us $t_{0.025} = 1.993$ (by `>>tinv(0.975, 74)` in Matlab). Hence the critical region is

$$t < -1.993 \quad \text{or} \quad t > 1.993$$



Two samples, confidence interval for the difference between means: the Statistical Statement

The critical region is

$$t < -1.993 \quad \text{or} \quad t > 1.993$$

Since the test statistic = 1.321 does not fall within the critical region, we fail to reject the null hypothesis $H_0 : \mu_1 - \mu_2 = 0$ at the significance level 5 %.

Based on these numbers paroxetine treatment does not have a significant effect as a treatment for bipolar depression at the significance level 5 %.