B34.UC2

Numerical Computation and Statistics in Engineering

Unit 5: Hypothesis Testing

Hypothesis Testing

Statistical tests consist of the following elements:

- a *null hypothesis* H_0 about one or more population parameters;
- an *alternative hypothesis* H_1 (or H_a) that replaces H_0 if the test does not support H_0;
- the test statistics;
- acceptance and rejection regions indicating the values of the statistics that will lead to acceptance or rejection of H_0.

The term null hypothesis stems from the fact that we often test for 'something being equal to 0', for example $\mu - 4 = 0$ (i.e., the population mean equals 4), or $\mu_1 - \mu_2 = 0$ (i.e., the two populations have the same mean).

Errors

There are two types of errors we can make when testing a hypothesis:

		H_0 true	H_0 false
Decision:	Reject H_0	Type I error	Correct decision
	Accept H_0	Correct decision	Type II error

Type I errors (rejecting H_0 while it is true) are usually denoted by the symbol α, type II errors (accepting H_0 while it is false) are denoted by the symbol β.

Example. A car retailer believes that more than 20% of his customers are willing to spend extra money for upgrading the stereo equipment of their new car. Before ordering new equipment the retailer wants to ask 10 of his customers whether they would buy the more expensive equipment.

Errors

Here we pick

$$H_0: \; p = 0.2.$$

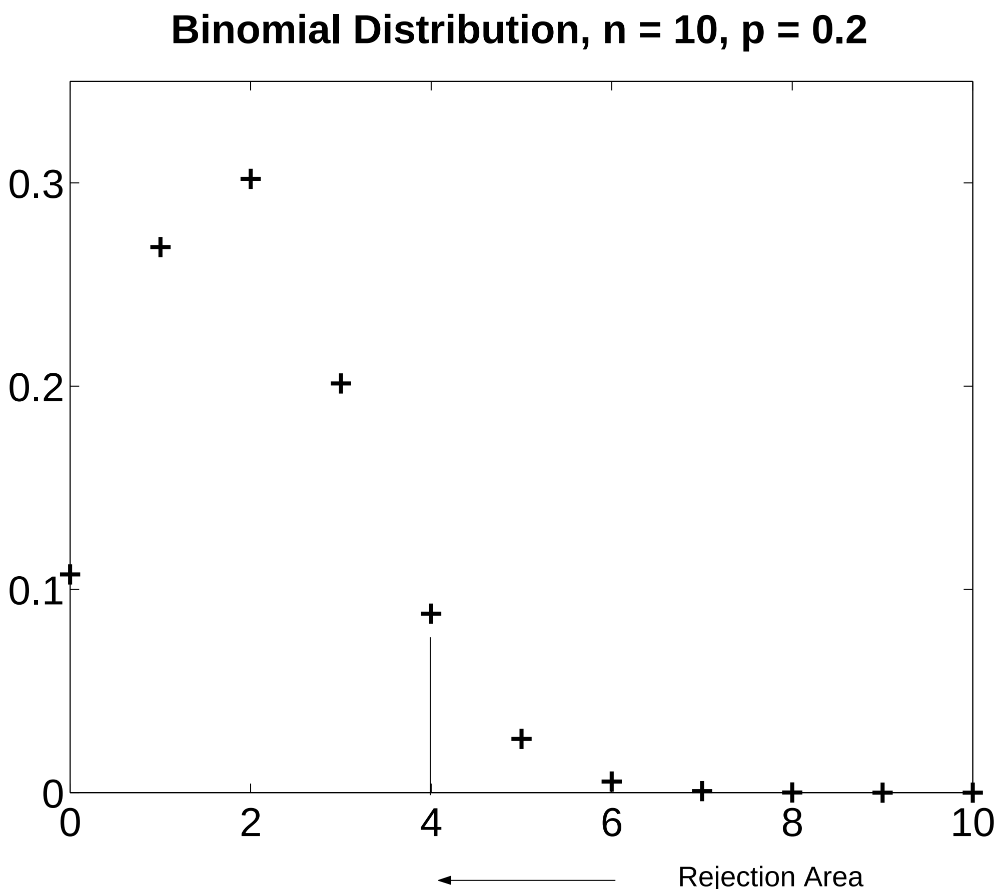
$$H_a: \; p > 0.2.$$

(We do not believe in $p < 0.2$.) The random variable x for this test is the number of people indicating that they would buy better stereo equipment for their cars. If $p = 0.2$ we expect $10 \cdot 0.2 = 2$ people to be in favor of the better product, thus, rejecting H_0 if $x \geq 4$ seems reasonable. For the type II error we find

$$\begin{aligned} \alpha &= P(\text{reject } H_0 \text{ while it is true}) \\ &= P(p = 0.2 \text{ and } \; x \geq 4) \\ &= 1 - P(p = 0.2 \text{ and } \; x \leq 3) \\ &= 1 - \sum_{x=0}^{3} \binom{10}{x} p^x (1-p)^{10-x} \\ &\approx 0.121\,. \end{aligned}$$

Errors

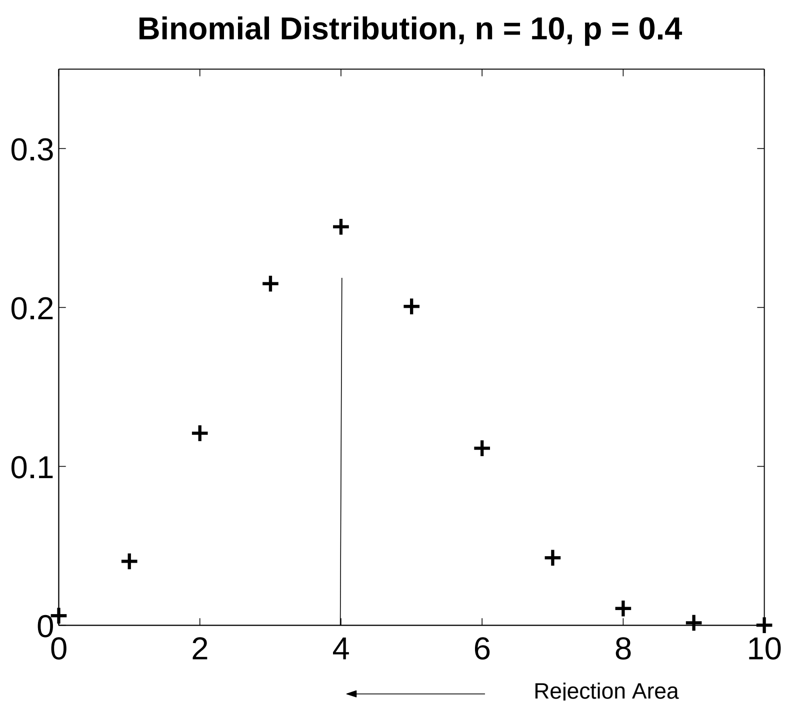
Questioning the customers the retailer finds that 4 out of 10 people are in favor of the better product, thus H_0 is rejected.



Errors

Suppose that the true parameter value is $p = 0.4$. Then

$$\begin{aligned}\beta &= P(\text{accept } H_0 \text{ while } p = 0.4) \;=\; P(x \leq 3 \text{ while } p = 0.4) \\ &= \sum_{x=0}^{3} \binom{10}{x} 0.4^x (1-0.4)^{10-x} \;\approx\; 0.3823\,.\end{aligned}$$



Using the Normal Distribution

In practice we can often use the normal distribution to find acceptance and rejection intervals. Suppose we want to test

$$H_0 : \theta = \theta_0 \qquad H_1 : \theta \neq \theta_0$$

where θ is a parameter of a population (probability, mean, etc.). θ_0 is the value that we think θ has. We assume that the estimator $\hat{\theta}$ that we get from the sample has normal distribution with mean θ_0 and standard deviation $\sigma_{\hat{\theta}}$. Then

$$\text{statistics } z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

has a standard normal distribution. If the rejection region is

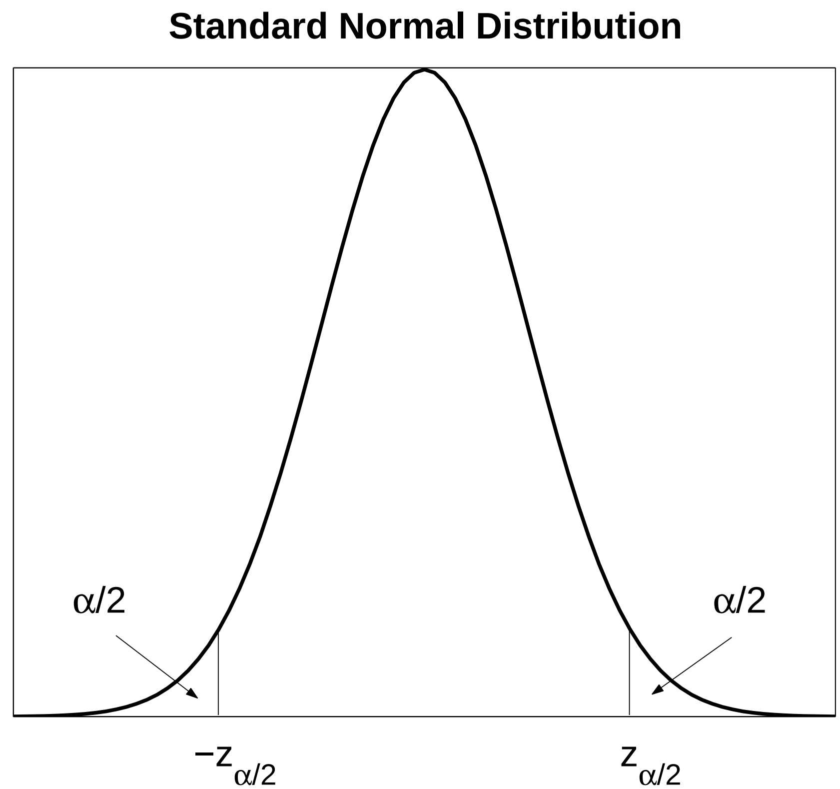
$$z < -z_{\alpha/2}, \qquad z_{\alpha/2} < z$$

then the type I error is α.

Using the Normal Distribution

Indeed,

$$
\begin{aligned}
& \text{type I error} \\
= \; & P(H_0 \text{ holds but is rejected}) \\
= \; & P(z \le -z_{\alpha/2} \text{ or } z_{\alpha/2} \le z) \\
= \; & \alpha \,.
\end{aligned}
$$



Such a test is called a *two-tailed test*.

Using the Normal Distribution

For a *one-tailed test* the data are

- $H_0 : \theta = \theta_0$;
- $H_1 : \theta > \theta_0$ ($\theta < \theta_0$);
- Statistics: $z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$;
- Rejection region: $z > z_\alpha$ ($z < -z_\alpha$);
- Type I error: α.

In practice, we are often given α in advance specifying the type I error probability that we are willing to accept, and we use this to find the acceptance and rejection interval.

Test About a Mean

One-tailed test	Two-tailed test
$H_0\colon \mu = \mu_0$	$H_0\colon \mu = \mu_0$
$H_1\colon \mu > \mu_0$	$H_1\colon \mu \neq \mu_0$
(or $H_1\colon \mu < \mu_0$)	
$z = \frac{\bar{x}-\mu_0}{\sigma_{\bar{x}}}$	$z = \frac{\bar{x}-\mu_0}{\sigma_{\bar{x}}}$
Rejection region:	Rejection region:
$z > z_\alpha$ (or $z < -z_\alpha$)	$z < -z_{\alpha/2}$ or $z_{\alpha/2} < z$

If the sample size is small $(n < 30)$ or if $\sigma_{\bar{x}}$ has to be estimated by $s/\sqrt{n}$ then the normal distribution is replaced by the t-distribution with $n-1$ degrees of freedom.

Test About a Mean

Example. We go back to the machines making wires with diameter approximately 1mm. Data taken from two machines showed the following values:

I:	1.0429	1.0627	0.9203	0.9280	1.0286
	0.9800	1.0345	1.0408	1.0356	1.0645
II:	1.0001	1.0157	0.9439	0.9794	0.9753
	0.9319	0.9877	0.9483	1.0225	0.9558

with $n_1 = n_2 = 10$, $\bar{x}_1 = 1.0138$, $s_1 = 0.0526$, $\bar{x}_2 = 0.9761$, and $s_2 = 0.0309$.

For

$$t_i = \frac{\bar{x}_i - 1.0}{s_i/\sqrt{10}}$$

we find $t_1 = 0.7877$ and $t_2 = -2.4459$. With $\alpha = 0.05$ we consider the following tests:

Test About a Mean

- $H_0 : \bar{x}_1 = 1.0$,
- $H_1 : \bar{x}_1 < 1.0$.

Since $t_1 \not> t_{9,0.05} = 1.8331$ the hypothesis H_0 is accepted.

- $H_0 : \bar{x}_1 = 1.0$,
- $H_1 : \bar{x}_1 \neq 1.0$.

Since $t_{9,0.025} = 2.2622$ and $-2.2622 < t_1 < 2.2622$ the hypothesis is accepted.

In both cases the decision is 'correct'; the data was random data with $\mu = 1.0$ and $\sigma = 0.05$.

Test About a Mean

For the second sample we consider

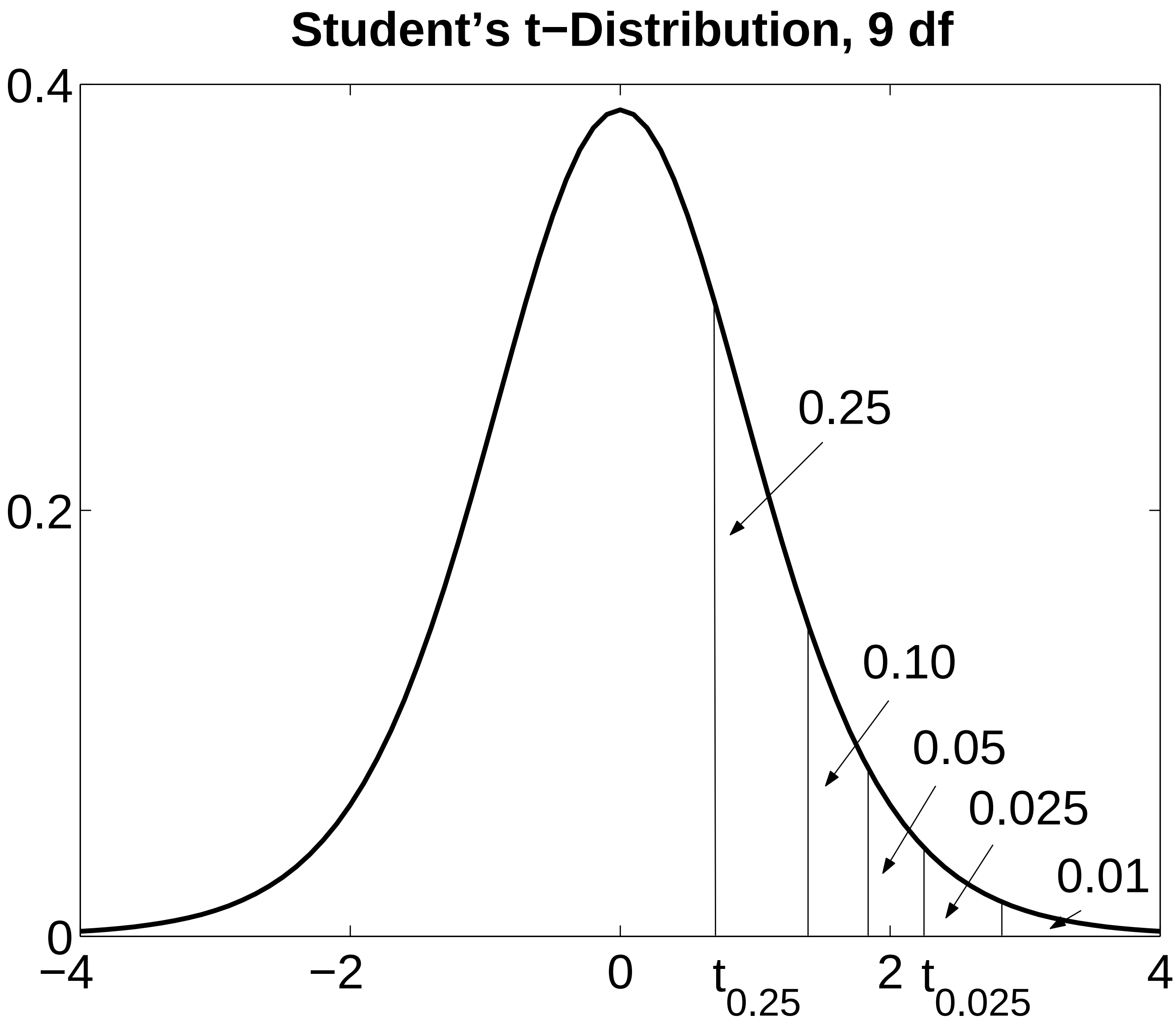
- $H_0: \bar{x}_2 = 1.0$,
- $H_1: \bar{x}_2 < 1.0$.

Since $t_2 = -2.4459 < t_{9,0.05} = -1.833$ the hypothesis is rejected. For the two-tailed test

- $H_0: \bar{x}_2 = 1.0$,
- $H_1: \bar{x}_2 \neq 1.0$

we see that $t_2 < -t_{9,0.025} = -2.262$, and we reject H_0 again. $\square$

Test About a Mean



Testing For the Difference Between Means

One-tailed test	Two tailed test
$H_0: \mu_1 - \mu_2 = d$	$H_0: \mu_1 - \mu_2 = d$
$H_1: \mu_1 - \mu_2 < d$	$H_1: \mu_1 - \mu_2 \neq d$
(or $H_1: \mu_1 - \mu_2 > d$)	
$z = \frac{(\bar{x}_1 - \bar{x}_2) - d}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(\bar{x}_1 - \bar{x}_2) - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$z = \frac{(\bar{x}_1 - \bar{x}_2) - d}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(\bar{x}_1 - \bar{x}_2) - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
Rejection region:	Rejection region:
$z > z_\alpha$ (or $z < z_\alpha$)	$z < -z_{\alpha/2}$ or $z_{\alpha/2} < z$

Testing For the Difference Between Means

If the sample sizes are small and σ_1 and σ_2 are unknown then the t-distribution with $n_1 + n_2 - 2$ degrees of freedom replaces the normal distribution, with

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d}{\sqrt{s_P^2(\frac{1}{n_1} + \frac{1}{n_2})}} \qquad \text{where} \qquad s_P^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$$

provided that the two unknown variances are equal.

Example. The response times of two hard drives are tested. The values found are

Disk 1	Disk 2
$n_1 = 15$	$n_2 = 13$
$\bar{x}_1 = 16$	$\bar{x}_2 = 13$
$s_1 = 5$	$s_2 = 4$

What can be said about the difference between the mean response times?

Testing For the Difference Between Means

Here

$$H_0\colon (\mu_1 - \mu_2) = 0, \qquad H_1\colon (\mu_1 - \mu_2) \neq 0\,.$$

We calculate

$$\begin{aligned} s_P^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \\ &= \frac{14\cdot 5^2 + 12\cdot 4^2}{15+13-2} \\ &= 20.8462 \end{aligned}$$

and

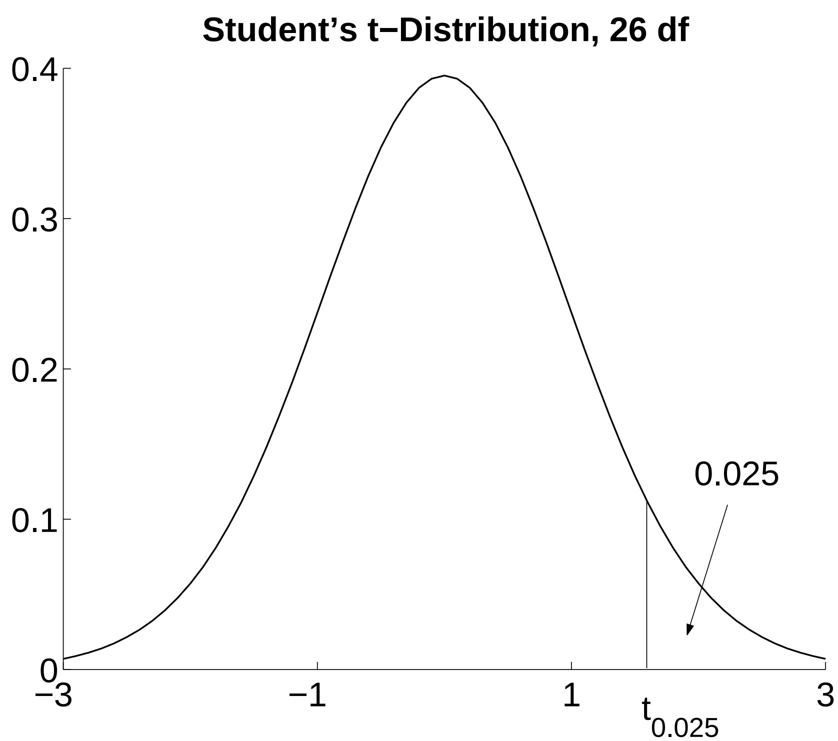
$$\begin{aligned} t &= \frac{(\bar{x}_1 - \bar{x}_2) - d}{\sqrt{s_P^2(\frac{1}{n_1} + \frac{1}{n_2})}} \\ &= \frac{16-13}{\sqrt{20.8462(\frac{1}{15} + \frac{1}{13})}} \\ &= 1.7340\,. \end{aligned}$$

Testing For the Difference Between Means

From the tables we know that for $n_1 + n_2 - 2 = 26$ and $\alpha = 0.1$ (for example) that

$$t_{0.05,26} = 1.706\,.$$

Since $t_{0.05,26} < t$ the hypothesis is rejected.



□

Testing For the Difference Between Means

Example. A study is done in the effectiveness of certain exercises to help weight loss. The following data are collected (in kg):

before	after	before	after
106	99	86	83
90	87	78	77
86	86	92	92
107	104	83	82
91	90	101	100
97	96	90	88
80	81	22	116
91	91	73	71

For the random variables before b and after a we find $\bar{b} = 92.0625$,

Testing For the Difference Between Means

$s_b = 12.3584$, $\bar{a} = 90.0625$, and $s_a = 11.0843$. Our hypotheses are

- $H_0 : \bar{b} - \bar{a} = 0$,
- $H_1 : \bar{b} - \bar{a} > 0$,

and we will test at a 5% level of significance.

Then $t_{30,0.05} = 1.699$, and we reject H_0 if $t \geq 1.699$. To calculate further,

$$
\begin{aligned}
s_P^2 &= \frac{(n_b - 1)s_b^2 + (n_a - 1)s_a^2}{n_b + n_a - 2} \\
&= \frac{15}{30(s_b^2 + s_a^2)} \\
&= \frac{1}{2}(12.3584^2 + 11.0843^2) \\
&= 137.79562\,.
\end{aligned}
$$

Testing For the Difference Between Means

For t we find

$$
\begin{aligned}
t &= \frac{(\bar{b} - \bar{a}) - 0}{\sqrt{s_P^2(\frac{1}{n_b} + \frac{1}{n_a})}} \\
&= \frac{92.0625 - 90.0625}{\sqrt{137.79562 \cdot \frac{1}{5}}} \\
&= 0.38098\,.
\end{aligned}
$$

The hypothesis is accepted, and there is evidence that the exercises help reducing weight. □

Tests Concerning the Variance

Given a random sample from a normal population we will test the null hypothesis $\sigma^2 = \sigma_0^2$ against the alternatives $\sigma^2 \neq \sigma_0^2$ or $\sigma^2 < \sigma_0^2$ ($\sigma^2 > \sigma_0^2$). The random variable

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

has a χ^2 distribution with $n-1$ degrees of freedom. For a two-tailed test the null hypothesis is rejected if

$$\chi^2 \leq \chi^2_{1-\alpha/2,n-1} \qquad \text{or} \qquad \chi^2 \geq \chi^2_{\alpha/2,n-1}\,.$$

For a one-tailed test and the alternative hypothesis $\sigma^2 < \sigma_0^2$ we reject H_0 if $\chi^2 \leq \chi^2_{1-\alpha,n-1}$.

Example. Thickness of a semi-conductor part (in 10^{-5}m) is crucial in a production process. The machine manufacturing these semi-conductors needs to be readjusted if $\sigma^2 \leq 0.36$.

If in a sample of 20 measurements we find $s^2 = 0.74$, what can be said at

Tests Concerning the Variance

a $\alpha = 0.05$ level of significance?

We assume that thickness is normally distributed. Then

- $H_0: \sigma^2 = 0.36$,
- $H_1: \sigma^2 > 0.36$.

We reject the null hypothesis if $\chi^2 \geq \chi^2_{0.05,19} = 30.144$. With $s^2 = 0.74$, $\sigma_0^2 = 0.36$ and $n = 20$ we find

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{19 \cdot 0.74}{0.36} = 34.944\,,$$

and the machine needs to be readjusted.

Note that for $n = 20$, $\sigma_0^2 = 0.36$, and $\alpha = 0.05$ the machine needs readjustment if for a sample of 20 measurements the sample variation s^2 is greater or equal than 0.571. $\square$

Testing for Proportions

Example. Suppose 5 out of 20 transistors are faulty. We test the hypothesis

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

at the 0.05 level of significance.

Instead of determining the rejection and acceptance interval we will find the smallest α which will reject H_0 (note for the calculation that the binomial distribution is symmetric):

$$\begin{aligned} \alpha/2 &= P(x \leq 5) \\ &= \sum_{x=0}^{5} \binom{20}{x} 0.5^x 0.5^{20-x} \\ &= 0.0207\,, \end{aligned}$$

so that $\alpha = 0.0414$. Since $\alpha < 0.05$ we will reject H_0. □

Testing for Proportions

If in a binomial test the size n is large we can use the normal distribution (with or without continuity correction) as an approximation for the random variable x. Then we get

$$H_0: p = p_0$$

$$H_1: p \neq p_0$$

$$z = \frac{\bar{x} - np_0}{\sqrt{np_0(1-p_0)}} \qquad \text{or } z = \frac{(\bar{x} \pm \frac{1}{2}) - np_0}{\sqrt{np_o(1-p_0)}}$$

Rejection region:

$z < -z_{\alpha/2}$ or $z_{\alpha/2} < z$

(If we use the correction factor we use a minus when x exceeds np_0, and a plus when x is less than np_0.)

Testing for Proportions

For the on-tailed test

$$H_0 : p = p_0$$

$$H_1 : p > p_0$$

we use the same statistics z as above with rejection interval $z \geq z_\alpha$.

Example. Suppose $p_0 = 0.2$ and we test

- $H_0 : p = 0.2,$
- $H_1 : p < 0.2,$

at the 0.01 level of significance. Then, using $z_{0.01} = 2.33$ we have the rejection region $z \leq -2.33$. If the test data are $n = 200$, $x = 22$, then

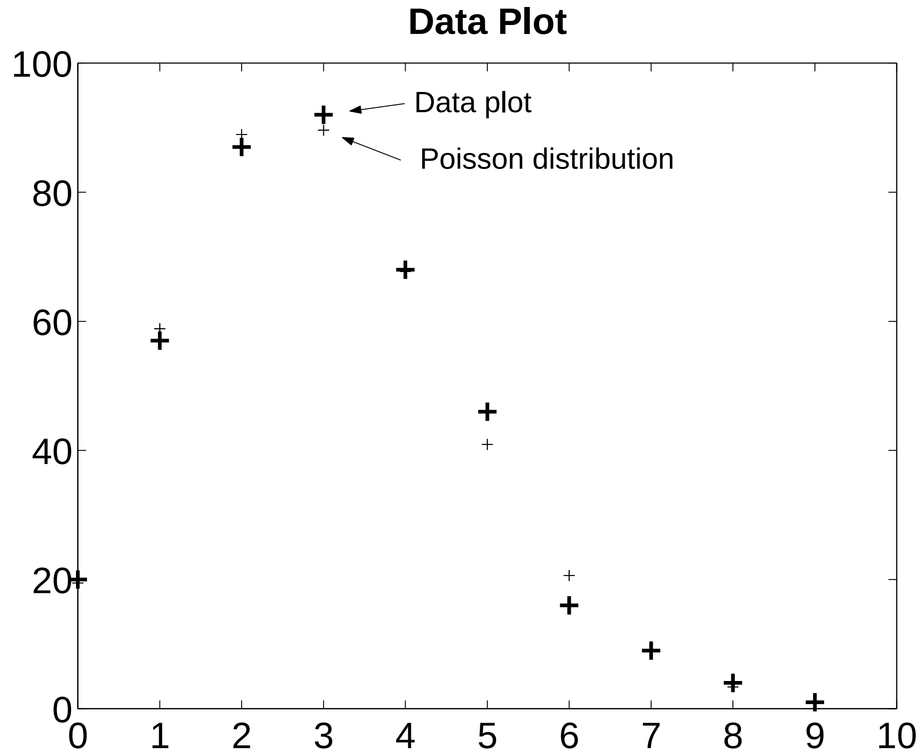
$$z = \frac{\bar{x} - np_0}{\sqrt{np_0(1-p_0)}} = \frac{22 - 200 \cdot 0.2}{\sqrt{200 \cdot 0.2 \cdot 0.8}} \approx -3.18\,,$$

and we reject the null hypothesis. □

Goodness-To-Fit

Goodness to fit tests are applied to test whether a set of data may be looked upon as a random sample from a population having a given distribution.

Suppose we have data from a Poisson distribution with $\lambda = 3$ (see next slide), which gives the following frequency diagram:



Goodness-To-Fit

x	Frequency f_i		Poisson $\lambda = 3$	Expected freq. e_i $(\lambda = 3.0225)$
0	20	0.0500	0.0498	19.4717
1	57	0.1425	0.1494	58.8534
2	87	0.2175	0.2240	88.9421
3	92	0.2300	0.2240	89.6092
4	68	0.1700	0.1680	67.7110
5	46	0.1150	0.1008	40.9313
6	16	0.0400	0.0504	20.6191
7	9	0.0225	0.0216	8.9030
8	4	0.0100	0.0081	3.3637
9	1	0.0025	0.0027	1.1296

Goodness-To-Fit

For the expected frequency we first estimated λ using the third column as

$$\hat{\lambda} = 3.0225\,.$$

The random variable

$$\sum_{i=0}^{m} \frac{(f_i - e_i)^2}{e_i}$$

with m the number of different data (here 10) has a χ^2 distribution with $m - t - 1$ degrees of freedom, where t is the number of parameters estimated from the data (here 1).

Goodness-To-Fit

With the data above we want to test at a 0.05 level of significance whether the data are from a random variable having Poisson distribution. We set

- H_0: The data are from a Poisson random variable.
- H_1: The data are *not* from a Poisson random variable.

We reject H_0 if

$$\chi^2_{\alpha,m-t-1} \le \chi^2 = \sum_{i=0}^{m} \frac{(f_i - e_i)^2}{e_i}$$

Here $m = 10$, $t = 1$, and $\chi^2_{0.05,10-1-1} = 15.507$. With our data, $\chi^2 = 1.9789$, and H_0 is accepted.

Summary

- Statistical tests often consist of a null hypothesis, and an alternative hypothesis. A type I error is made when the null hypothesis is true, but rejected. A type II error is made when the null hypothesis is false, but is accepted.
- We distinguish one-tailed and two-tailed tests.
- Statistical tests are based on sampling and confidence intervals. We thus use the normal distribution, the Student's t-distribution, and the chi-square distribution in standard tests.
- Goodness-To-Fit tests use the chi-square distribution to test whether a set of data fits a given distribution.