

Hypothesis testing

- Elements of a hypothesis test:

1. Null hypothesis, H_0 (always =).
2. Alternative hypothesis, H_a ($>$, $<$, \neq).
3. Test statistic.
4. Significance level α .

- Hypothesis testing for μ :

$$H_0 : \mu = \mu_0$$

$$H_a : \mu > \mu_0, \mu < \mu_0, \mu \neq \mu_0 \text{ (use only one of these!)}$$

- When σ is known:

Test statistic

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- When σ is unknown:

Test statistic

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

- If σ is known:

Reject H_0 if Z falls in the rejection region. The rejection region is based on the significance level α we choose.

- If σ is unknown:

Reject H_0 if t falls in the rejection region. The rejection region is based on the significance level α we choose and the degrees of freedom $n - 1$.

- What is a p -value? It is the probability of seeing the test statistic or a more extreme value (extreme is towards the direction of the alternative). If p -value $< \alpha$ the H_0 is rejected. This is another way of testing a hypothesis (it should always agree with testing using Z or t).

- Hypothesis testing for p :

$$H_0 : p = p_0$$

$$H_a : p > p_0, p < p_0, p \neq p_0 \text{ (use only one of these 3!)}$$

Test statistic:

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

- Reject H_0 if Z falls in the rejection region. The rejection region is based on the significance level α we choose.
- What is a p -value? As always, it is the probability of seeing the test statistic or a more extreme value (extreme is towards the direction of the alternative). If p -value $< \alpha$ the H_0 is rejected.

Hypothesis Test For Population Mean μ

Hypothesis Test When σ Is Known:

$$H_0 : \mu = \mu_0$$

Alternative Hypothesis H_a	Reject H_0 If
$\mu < \mu_0$	$Z < -Z_\alpha$
$\mu > \mu_0$	$Z > Z_\alpha$
$\mu \neq \mu_0$	$Z < -Z_{\alpha/2}$ or $Z > Z_{\alpha/2}$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Hypothesis Test When σ Is Not Known:

$$H_0 : \mu = \mu_0$$

Alternative Hypothesis H_a	Reject H_0 If
$\mu < \mu_0$	$t < -t_{\alpha; n-1}$
$\mu > \mu_0$	$t > t_{\alpha; n-1}$
$\mu \neq \mu_0$	$t < -t_{\alpha/2; n-1}$ or $t > t_{\alpha/2; n-1}$

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Hypothesis Test For Proportion:

$$H_0 : p = p_0$$

Alternative Hypothesis H_a	Reject H_0 If
$p < p_0$	$Z < -Z_\alpha$
$p > p_0$	$Z > Z_\alpha$
$p \neq p_0$	$Z < -Z_{\alpha/2}$ or $Z > Z_{\alpha/2}$

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

Examples - Hypothesis testing

Example 1

A manufacturer of chocolates claims that the mean weight of a certain box of chocolates is 368 grams. The standard deviation of the box's weight is known to be $\sigma = 10$ grams. If a sample of 49 boxes has sample mean $\bar{x} = 364$ grams, test the hypothesis that the mean weight of the boxes is less than 368 grams. Use $\alpha = 0.05$ level of significance.

Example 2

A large retailer wants to determine whether the mean income of families living within 2 miles of a proposed building site exceeds \$24400. What can we conclude at the 0.05 level of significance if the sample mean income of 60 families is $\bar{x} = \$24524$? Use $\sigma = \$763$.

Example 3

It is claimed that the mean mileage of a certain type of vehicle is 35 miles per gallon of gasoline with population standard deviation $\sigma = 5$ miles. What can be concluded using $\alpha = 0.01$ about the claim if a random sample of 49 such vehicles has sample mean $\bar{x} = 36$ miles?

Example 4

A manufacturer claims that 20% of the public preferred her product. A sample of 100 persons is taken to check her claim. It is found that 8 of these 100 persons preferred her product.

- a. Find the p-value of the test (use a two-tailed test).
- b. Using the 0.05 level of significance test her claim.

Examples - Hypothesis testing Solutions

Example 1

A manufacturer of chocolates claims that the mean weight of a certain box of chocolates is 368 grams. The standard deviation of the box's weight is known to be $\sigma = 10$ grams. If a sample of 49 boxes has sample mean $\bar{x} = 364$ grams, test the hypothesis that the mean weight of the boxes is less than 368 grams. Use $\alpha = 0.05$ level of significance.

Solution:

1.

$$H_0 : \mu = 368$$

$$H_a : \mu < 368$$

2. We compute the test statistic z :

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{364 - 368}{\frac{10}{\sqrt{49}}} \Rightarrow z = -2.8$$

3. We find the rejection region. Here we use significance level $\alpha = 0.05$, therefore the rejection region is when $z < -1.645$.

4. Conclusion: Since $z = -2.8 < -1.645$ we reject H_0 .

Compute the p -value of the test:

$$p\text{-value} = P(\bar{X} < 364) = P(Z < -2.8) = 0.0026.$$

Rule: If $p\text{-value} < \alpha$ then H_0 is rejected. Again, using the p -value we reject H_0 .

Example 2

A large retailer wants to determine whether the mean income of families living within 2 miles of a proposed building site exceeds \$24400. What can we conclude at the 0.05 level of significance if the sample mean income of 60 families is $\bar{x} = \$24524$? Use $\sigma = \$763$.

Solution:

1.

$$H_0 : \mu = 24400$$

$$H_a : \mu > 24400$$

2. We compute the test statistic z :

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{24524 - 24400}{\frac{763}{\sqrt{60}}} \Rightarrow z = 1.26$$

3. We find the rejection region. Here we use significance level $\alpha = 0.05$, therefore the rejection region is when $z > 1.645$.

4. Conclusion: Since $z = 1.26$ does not fall in the R.R. we do not reject H_0 .

Example 3

It is claimed that the mean mileage of a certain type of vehicle is 35 miles per gallon of gasoline with population standard deviation $\sigma = 5$ miles. What can be concluded using $\alpha = 0.01$ about the claim if a random sample of 49 such vehicles has sample mean $\bar{x} = 36$ miles?

Solution:

1.

$$H_0 : \mu = 35$$

$$H_a : \mu \neq 35$$

2. We compute the test statistic z :

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{36 - 35}{\frac{5}{\sqrt{49}}} \Rightarrow z = 1.4$$

3. We find the rejection region. Here we use significance level $\alpha = 0.01$, but because of a two-sided test we have two rejection regions. They are $z < -2.575$ or $z > 2.575$.4. Conclusion: Since $z = 1.4$ does not fall in any of the two rejection regions we do not reject H_0 .

When we have a two-sided test the p -value is computed as follows:

$$p - \text{value} = 2P(\bar{X} > 36) = 2P(Z > 1.4) = 2(1 - 0.9192) = 0.1616.$$

Again, using the p -value H_0 is not rejected.

Example 4

A manufacturer claims that 20% of the public preferred her product. A sample of 100 persons is taken to check her claim. It is found that 8 of these 100 persons preferred her product.

- Find the p -value of the test (use a two-tailed test).
- Using the 0.05 level of significance test her claim.

Solution:

We test the following hypothesis:

$$H_0 : p = 0.20$$

$$H_a : p \neq 0.20$$

We compute the test statistic z :

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.08 - 0.20}{\sqrt{\frac{0.20(1-0.20)}{100}}} = -3.0.$$

Therefore the p -value is:

$$p - \text{value} = 2P(\hat{p} < 0.08) = 2P(Z < -3.0) = 2(0.0013) = 0.0026.$$

We reject H_0 because p -value = 0.0026 < 0.05.

Hypothesis testing - t distribution

Example 1

A tire manufacturer hopes that their newly designed tires will allow a car traveling at 60 mph to come to a complete stop within an average of 125 feet after the brakes are applied. They will adopt the new tires unless there is strong evidence that the tires do not meet this objective. The distances (in feet) for 9 stops on a test track were 129, 128, 130, 132, 135, 123, 125, 128, and 130. These data have $\bar{x} = 128.89$, $s = 3.55$. Test an appropriate hypothesis to conclude whether the company should adopt the new tires. Use $\alpha = 0.05$.

Example 2 (from *Mathematical Statistics and Data Analysis*), by J. Rice, 2nd Edition.

In a study done at the National Institute of Science and Technology (Steel et al. 1980), asbestos fibers on filters were counted as part of a project to develop measurement standards for asbestos concentration. Asbestos dissolved in water was spread on a filter, and punches of 3-mm diameter were taken from the filter and mounted on a transmission electron microscope. An operator counted the number of fibers in each of 23 grid squares, yielding the following counts:

31	29	19	18	31	28
34	27	34	30	16	18
26	27	27	18	24	22
28	24	21	17	24	

Assume normal distribution. These data have $\bar{x} = 24.91$, $s = 5.48$. Using $\alpha = 0.05$ test the following hypothesis:

$$H_0 : \mu = 18$$

$$H_a : \mu \neq 18$$

Hypothesis testing - t distribution

Example 1

A tire manufacturer hopes that their newly designed tires will allow a car traveling at 60 mph to come to a complete stop within an average of 125 feet after the brakes are applied. They will adopt the new tires unless there is strong evidence that the tires do not meet this objective. The distances (in feet) for 9 stops on a test track were 129, 128, 130, 132, 135, 123, 125, 128, and 130. These data have $\bar{x} = 128.89$, $s = 3.55$. Test an appropriate hypothesis to conclude whether the company should adopt the new tires. Use $\alpha = 0.05$.

Solution:

1.

$$H_0 : \mu = 125$$

$$H_a : \mu > 125$$

2. We compute the test statistic t :

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{128.89 - 125}{\frac{3.55}{\sqrt{9}}} \Rightarrow t = 3.29.$$

3. We find the rejection region. Here we use significance level $\alpha = 0.05$ with $n - 1 = 9 - 1 = 8$ degrees of freedom. Therefore the rejection region is when $t > 1.860$.

4. Conclusion: Since $t = 3.29$ falls in any the rejection region we reject H_0 .

The p -value is: p -value = $P(\bar{X} > 128.89) = P(t > 3.29)$. From the t table we can say that the $0.005 < p$ -value < 0.01 . Again, using the p -value H_0 is rejected.

Example 2 (from *Mathematical Statistics and Data Analysis*), by J. Rice, 2nd Edition.

In a study done at the National Institute of Science and Technology (Steel et al. 1980), asbestos fibers on filters were counted as part of a project to develop measurement standards for asbestos concentration. Asbestos dissolved in water was spread on a filter, and punches of 3-mm diameter were taken from the filter and mounted on a transmission electron microscope. An operator counted the number of fibers in each of 23 grid squares, yielding the following counts:

31	29	19	18	31	28
34	27	34	30	16	18
26	27	27	18	24	22
28	24	21	17	24	

Assume normal distribution. These data have $\bar{x} = 24.91$, $s = 5.48$. Using $\alpha = 0.05$ test the following hypothesis:

$$H_0 : \mu = 18$$

$$H_a : \mu \neq 18$$

Solution:

We compute the test statistic t :

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{24.91 - 18}{\frac{5.48}{\sqrt{23}}} \Rightarrow t = 6.05$$

We find the rejection region. Here we use significance level $\alpha = 0.05$ with $n - 1 = 23 - 1 = 22$ degrees of freedom. Therefore the rejection region is when $t < -2.074$ or $t > 2.074$. Conclusion: Since $t = 6.05$ falls in one of the rejection regions we reject H_0 .

Compute the p -value of the test: This is a two-sided test therefore the p -value is p -value = $2P(\bar{X} > 24.91) = 2P(t > 6.05)$. From the t table we can only say that p -value is less than 0.01.

Comparison between confidence intervals and a two-tailed hypothesis test

Two dice are rolled and the sum X of the two numbers that occurred is recorded. The probability distribution of X is as follows:

X	2	3	4	5	6	7	8	9	10	11	12
$P(X)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

This distribution has mean $\mu = 7$ and standard deviation $\sigma = 2.42$. We take 100 samples of size $n = 50$ each from this distribution and compute for each sample the sample mean \bar{x} . Pretend now that we only know that $\sigma = 2.42$, and that μ is unknown. We are going to use these 100 sample means to construct 100 95% confidence intervals for the true population mean μ , and to test using level of significance $\alpha = 0.05$ 100 times the hypothesis:

$$H_0 : \mu = 7$$

$$H_a : \mu \neq 7$$

The results are as follows:

Sample	\bar{x}	95% C.I. for μ	Is $\mu = 7$ included?	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	Reject H_0 ?
1	6.9	$6.23 \leq \mu \leq 7.57$	YES	-0.29	NO
2	6.3	$5.63 \leq \mu \leq 6.97$	NO	-2.05	YES
3	6.58	$5.91 \leq \mu \leq 7.25$	YES	-1.23	NO
4	6.54	$5.87 \leq \mu \leq 7.21$	YES	-1.34	NO
5	6.7	$6.03 \leq \mu \leq 7.37$	YES	-0.88	NO
6	6.58	$5.91 \leq \mu \leq 7.25$	YES	-1.23	NO
7	7.2	$6.53 \leq \mu \leq 7.87$	YES	0.58	NO
8	7.62	$6.95 \leq \mu \leq 8.29$	YES	1.81	NO
9	6.94	$6.27 \leq \mu \leq 7.61$	YES	-0.18	NO
10	7.36	$6.69 \leq \mu \leq 8.03$	YES	1.05	NO
11	7.06	$6.39 \leq \mu \leq 7.73$	YES	0.18	NO
12	7.08	$6.41 \leq \mu \leq 7.75$	YES	0.23	NO
13	7.42	$6.75 \leq \mu \leq 8.09$	YES	1.23	NO
14	7.42	$6.75 \leq \mu \leq 8.09$	YES	1.23	NO
15	6.8	$6.13 \leq \mu \leq 7.47$	YES	-0.58	NO
16	6.94	$6.27 \leq \mu \leq 7.61$	YES	-0.18	NO
17	7.2	$6.53 \leq \mu \leq 7.87$	YES	0.58	NO
18	6.7	$6.03 \leq \mu \leq 7.37$	YES	-0.88	NO
19	7.1	$6.43 \leq \mu \leq 7.77$	YES	0.29	NO
20	7.04	$6.37 \leq \mu \leq 7.71$	YES	0.12	NO
21	6.98	$6.31 \leq \mu \leq 7.65$	YES	-0.06	NO
22	7.18	$6.51 \leq \mu \leq 7.85$	YES	0.53	NO
23	6.8	$6.13 \leq \mu \leq 7.47$	YES	-0.58	NO
24	6.94	$6.27 \leq \mu \leq 7.61$	YES	-0.18	NO
25	8.1	$7.43 \leq \mu \leq 8.77$	NO	3.21	YES
26	7	$6.33 \leq \mu \leq 7.67$	YES	0.00	NO
27	7.06	$6.39 \leq \mu \leq 7.73$	YES	0.18	NO
28	6.82	$6.15 \leq \mu \leq 7.49$	YES	-0.53	NO
29	6.96	$6.29 \leq \mu \leq 7.63$	YES	-0.12	NO
30	7.46	$6.79 \leq \mu \leq 8.13$	YES	1.34	NO
31	7.04	$6.37 \leq \mu \leq 7.71$	YES	0.12	NO
32	7.06	$6.39 \leq \mu \leq 7.73$	YES	0.18	NO
33	7.06	$6.39 \leq \mu \leq 7.73$	YES	0.18	NO
34	6.8	$6.13 \leq \mu \leq 7.47$	YES	-0.58	NO
35	7.12	$6.45 \leq \mu \leq 7.79$	YES	0.35	NO
36	7.18	$6.51 \leq \mu \leq 7.85$	YES	0.53	NO
37	7.08	$6.41 \leq \mu \leq 7.75$	YES	0.23	NO
38	7.24	$6.57 \leq \mu \leq 7.91$	YES	0.70	NO
39	6.82	$6.15 \leq \mu \leq 7.49$	YES	-0.53	NO
40	7.26	$6.59 \leq \mu \leq 7.93$	YES	0.76	NO
41	7.34	$6.67 \leq \mu \leq 8.01$	YES	0.99	NO
42	6.62	$5.95 \leq \mu \leq 7.29$	YES	-1.11	NO
43	7.1	$6.43 \leq \mu \leq 7.77$	YES	0.29	NO
44	6.98	$6.31 \leq \mu \leq 7.65$	YES	-0.06	NO
45	6.98	$6.31 \leq \mu \leq 7.65$	YES	-0.06	NO
46	7.06	$6.39 \leq \mu \leq 7.73$	YES	0.18	NO
47	7.14	$6.47 \leq \mu \leq 7.81$	YES	0.41	NO
48	7.5	$6.83 \leq \mu \leq 8.17$	YES	1.46	NO
49	7.08	$6.41 \leq \mu \leq 7.75$	YES	0.23	NO
50	7.32	$6.65 \leq \mu \leq 7.99$	YES	0.94	NO

Sample	\bar{x}	95% C.I. for μ	Is $\mu = 7$ included?	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	Reject H_0 ?
51	6.54	$5.87 \leq \mu \leq 7.21$	YES	-1.34	NO
52	7.14	$6.47 \leq \mu \leq 7.81$	YES	0.41	NO
53	6.64	$5.97 \leq \mu \leq 7.31$	YES	-1.05	NO
54	7.46	$6.79 \leq \mu \leq 8.13$	YES	1.34	NO
55	7.34	$6.67 \leq \mu \leq 8.01$	YES	0.99	NO
56	7.28	$6.61 \leq \mu \leq 7.95$	YES	0.82	NO
57	6.56	$5.89 \leq \mu \leq 7.23$	YES	-1.29	NO
58	7.72	$7.05 \leq \mu \leq 8.39$	NO	2.10	YES
59	6.66	$5.99 \leq \mu \leq 7.33$	YES	-0.99	NO
60	6.8	$6.13 \leq \mu \leq 7.47$	YES	-0.58	NO
61	7.08	$6.41 \leq \mu \leq 7.75$	YES	0.23	NO
62	6.58	$5.91 \leq \mu \leq 7.25$	YES	-1.23	NO
63	7.3	$6.63 \leq \mu \leq 7.97$	YES	0.88	NO
64	7.1	$6.43 \leq \mu \leq 7.77$	YES	0.29	NO
65	6.68	$6.01 \leq \mu \leq 7.35$	YES	-0.94	NO
66	6.98	$6.31 \leq \mu \leq 7.65$	YES	-0.06	NO
67	6.94	$6.27 \leq \mu \leq 7.61$	YES	-0.18	NO
68	6.78	$6.11 \leq \mu \leq 7.45$	YES	-0.64	NO
69	7.2	$6.53 \leq \mu \leq 7.87$	YES	0.58	NO
70	6.9	$6.23 \leq \mu \leq 7.57$	YES	-0.29	NO
71	6.42	$5.75 \leq \mu \leq 7.09$	YES	-1.69	NO
72	6.48	$5.81 \leq \mu \leq 7.15$	YES	-1.52	NO
73	7.12	$6.45 \leq \mu \leq 7.79$	YES	0.35	NO
74	6.9	$6.23 \leq \mu \leq 7.57$	YES	-0.29	NO
75	7.24	$6.57 \leq \mu \leq 7.91$	YES	0.70	NO
76	6.6	$5.93 \leq \mu \leq 7.27$	YES	-1.17	NO
77	7.28	$6.61 \leq \mu \leq 7.95$	YES	0.82	NO
78	7.18	$6.51 \leq \mu \leq 7.85$	YES	0.53	NO
79	6.76	$6.09 \leq \mu \leq 7.43$	YES	-0.70	NO
80	7.06	$6.39 \leq \mu \leq 7.73$	YES	0.18	NO
81	7	$6.33 \leq \mu \leq 7.67$	YES	0.00	NO
82	7.08	$6.41 \leq \mu \leq 7.75$	YES	0.23	NO
83	7.18	$6.51 \leq \mu \leq 7.85$	YES	0.53	NO
84	7.26	$6.59 \leq \mu \leq 7.93$	YES	0.76	NO
85	6.88	$6.21 \leq \mu \leq 7.55$	YES	-0.35	NO
86	6.28	$5.61 \leq \mu \leq 6.95$	NO	-2.10	YES
87	7.06	$6.39 \leq \mu \leq 7.73$	YES	0.18	NO
88	6.66	$5.99 \leq \mu \leq 7.33$	YES	-0.99	NO
89	7.18	$6.51 \leq \mu \leq 7.85$	YES	0.53	NO
90	6.86	$6.19 \leq \mu \leq 7.53$	YES	-0.41	NO
91	6.96	$6.29 \leq \mu \leq 7.63$	YES	-0.12	NO
92	7.26	$6.59 \leq \mu \leq 7.93$	YES	0.76	NO
93	6.68	$6.01 \leq \mu \leq 7.35$	YES	-0.94	NO
94	6.76	$6.09 \leq \mu \leq 7.43$	YES	-0.70	NO
95	7.3	$6.63 \leq \mu \leq 7.97$	YES	0.88	NO
96	7.04	$6.37 \leq \mu \leq 7.71$	YES	0.12	NO
97	7.34	$6.67 \leq \mu \leq 8.01$	YES	0.99	NO
98	6.72	$6.05 \leq \mu \leq 7.39$	YES	-0.82	NO
99	6.64	$5.97 \leq \mu \leq 7.31$	YES	-1.05	NO
100	7.3	$6.63 \leq \mu \leq 7.97$	YES	0.88	NO

Hypothesis Testing - Type I and Type II error

		ACTUAL SITUATION	
		H_0 IS TRUE	H_0 IS NOT TRUE
STATISTICAL DECISION	DO NOT REJECT H_0	Correct Decision $1 - \alpha$	Type II error β
	REJECT H_0	Type I Error α	Correct Decision $1 - \beta$ (Power)

More examples - Hypothesis testing

Example 1

An experimenter has prepared a drug dosage level that he claims will induce sleep for at least 80% of those people suffering from insomnia. After examining the dosage, we feel that his claims regarding the effectiveness of the dosage are inflated. In an attempt to disprove his claim, we administer his prescribed dosage to 20 insomniacs, and we observe X , the number having sleep induced by the drug dose. We wish to test the hypothesis $H_0 : p = 0.8$ against the alternative $H_a : p < 0.8$. Assume the rejection region $X \leq 12$ is used.

- Find the type I error α .
- Find the type II error β if the true $p = 0.6$.
- Find the type II error β if the true $p = 0.4$.

Example 2

For a certain candidate's political poll $n = 15$ voters are sampled. Assume that this sample is taken from an infinite population of voters. We wish to test $H_0: p = 0.5$ against the alternative $H_a: p < 0.5$. The test statistic is X , which is the number of voters among the 15 sampled favoring this candidate.

- Calculate the probability of a type I error α if we select the rejection region to be $RR = \{x \leq 2\}$.
- Is our test good in protecting us from concluding that this candidate is a winner if, in fact, he will lose? Suppose that he really will win 30% of the vote ($p = 0.30$). What is the probability of a type II error β that the sample will erroneously lead us to conclude that H_0 is true?

Type II error (β) and the power of the test ($1 - \beta$)

Example:

A manufacturer of tires claims that the mean lifetime of these tires is 25000 miles. A random sample of 100 tires will be selected, and assume that the population standard deviation is 3500 miles. Calculate the probability of a Type II error (β) and the power of the test ($1 - \beta$) if the true population mean is 23500 miles using:

a. $\alpha = 0.05$

b. $\alpha = 0.01$

Solution:

a. $\alpha = 0.05$.

We are testing the hypothesis

$$H_0 : \mu = 25000$$

$$H_a : \mu < 25000$$

We find first the values of \bar{X} for which H_0 is rejected. H_0 is rejected when $Z < -1.645$. Or

$$\begin{aligned}\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} &< -1.645 \\ \frac{\bar{X} - 25000}{\frac{3500}{\sqrt{100}}} &< -1.645\end{aligned}$$

Therefore

$$\bar{X} < 25000 - 1.645 \frac{3500}{\sqrt{100}} \Rightarrow \bar{X} < 24424.25$$

If the sample of size $n = 100$ gives a value of \bar{X} less than 24424.25 then H_0 is rejected.

How do we compute the Type II error β ?

$$\begin{aligned}\beta &= P(\text{falsely accepting } H_0) = P(\bar{X} > 24424.25, \text{ when } \mu = 23500) \\ &= P\left(Z > \frac{24424.25 - 23500}{\frac{3500}{\sqrt{100}}}\right) = P(Z > 2.64) = 1 - 0.9959 \Rightarrow \beta = 0.0041.\end{aligned}$$

Therefore, the power of the test is $1 - \beta = 0.9959$.

b. $\alpha = 0.01$

We reject H_0 if $Z < -2.325$ or $\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < -1.645$ or $\bar{X} < 24186.25$.

$$\begin{aligned}\beta &= P(\text{falsely accepting } H_0) = P(\bar{X} > 24186.25, \text{ when } \mu = 23500) \\ &= P\left(Z > \frac{24186.25 - 23500}{\frac{3500}{\sqrt{100}}}\right) = P(Z > 1.96) = 1 - 0.9750 \Rightarrow \beta = 0.025.\end{aligned}$$

Therefore, the power of the test is $1 - \beta = 0.9750$.

Power of the test - example:

Let X be the breaking strength of a steel bar. if the steel bar is manufactured by a certain process, then $X \sim N(50, 6)$. Suppose that a sample of size $n = 16$ will be selected. Find the probability of detecting a shift from $\mu_0 = 50$ to $\mu_a = 55$ if we can accept a Type I error $\alpha = 0.05$.

Power curves

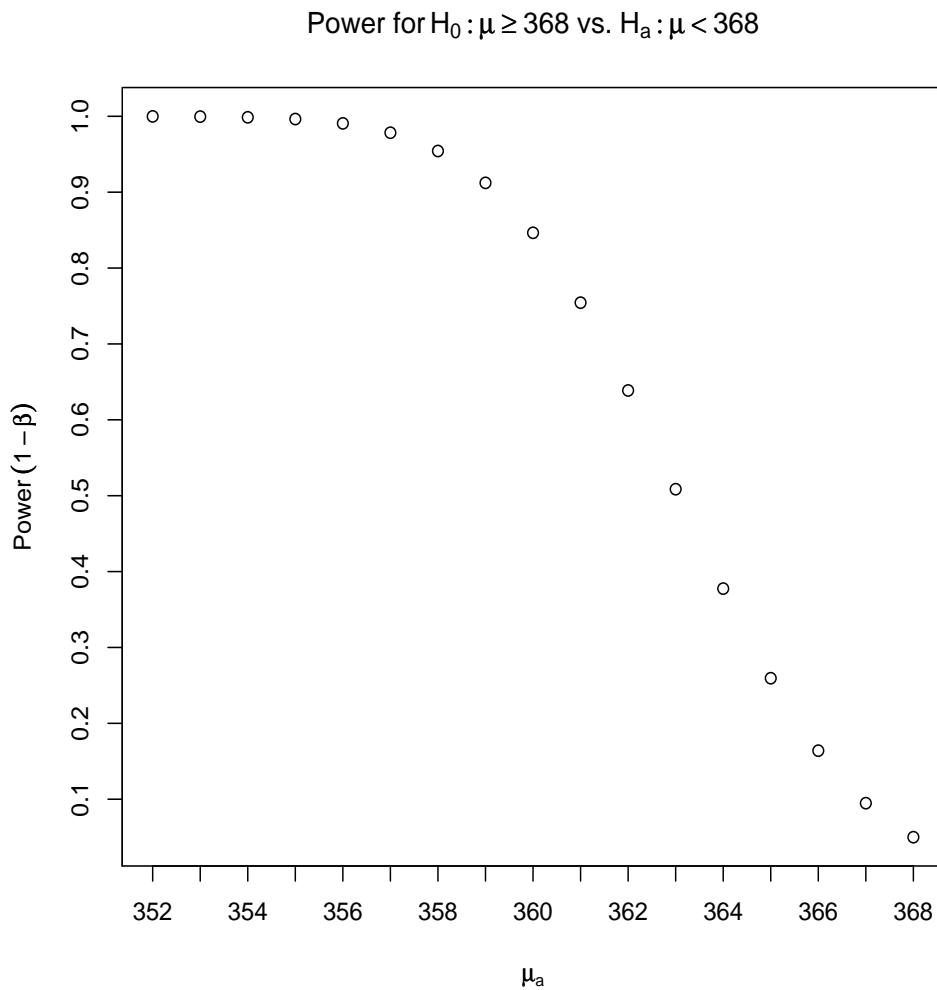
A power curve is a plot of the power $1 - \beta$ against values of the parameter under the alternative hypothesis. Suppose that we want to determine whether or not a cereal box packaging process is in control. Let's assume that the standard deviation of the filling process is known to be $\sigma = 15$ grams and that the weight of the box follows the normal distribution. To test this hypothesis a sample of $n = 25$ boxes of cereal is to be selected. Our goal here is to find the power of the test for different values of μ when we are willing to take a risk of Type I error $\alpha = 0.05$.

a. Suppose our test is:

$$H_0 : \mu \geq 368$$

$$H_a : \mu < 368$$

The plot of the power of the test $1 - \beta$ against values of $\mu < 368$ is the following:

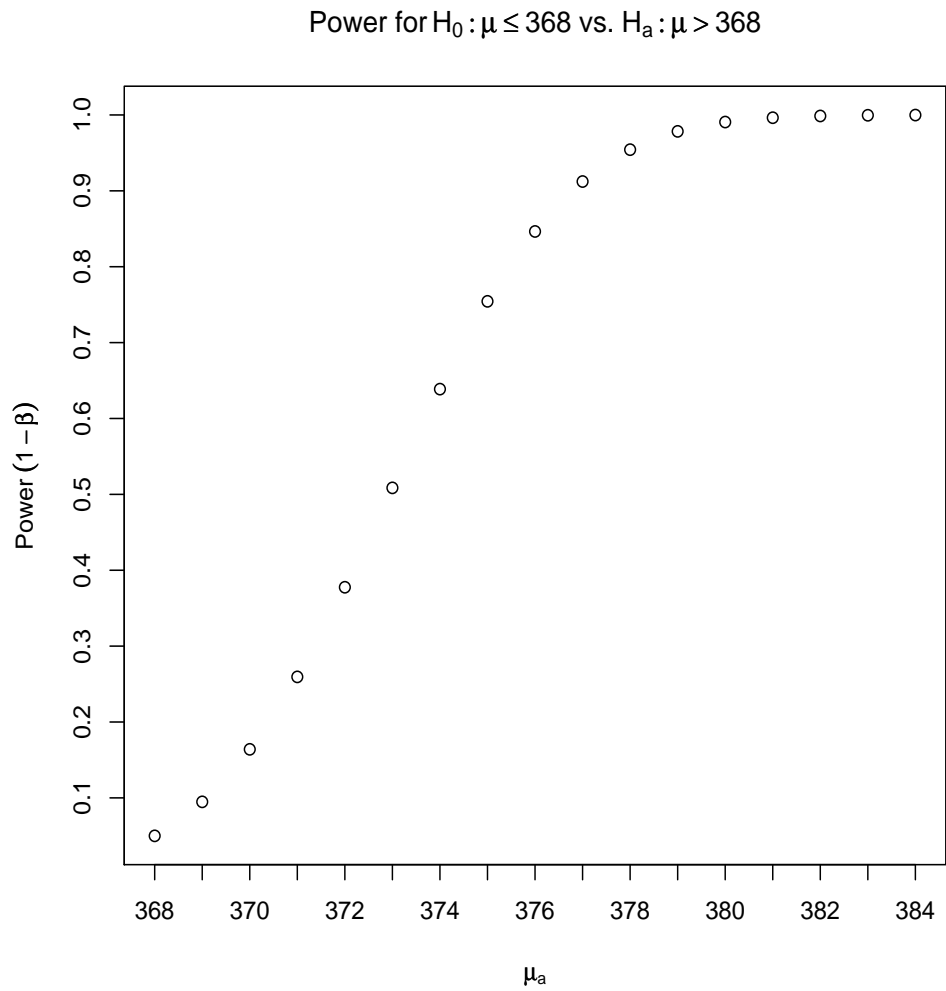


b. Suppose our test is:

$$H_0 : \mu \leq 368$$

$$H_a : \mu > 368$$

The plot of the power of the test $1 - \beta$ against values of $\mu > 368$ is the following:

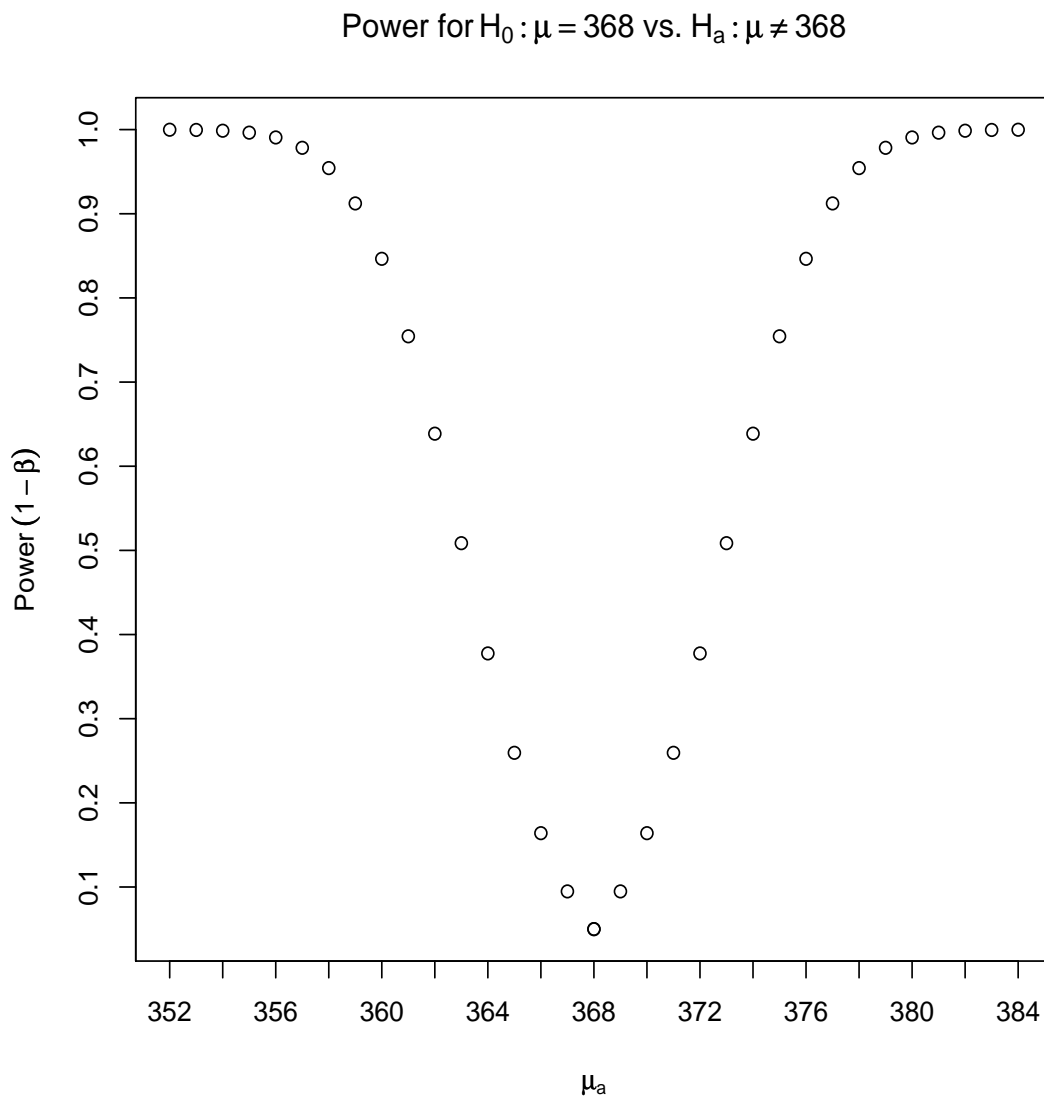


c. Suppose our test is:

$$H_0 : \mu = 368$$

$$H_a : \mu \neq 368$$

The plot of the power of the test $1 - \beta$ against values of $\mu < 368$ or $\mu > 368$ is the following:



Type I and Type II error and finding the sample size - An example

A manufacturer of tires claims that the mean lifetime of these tires is at least 25000 miles when the production process is working properly. Based upon past experience, the standard deviation of the lifetime of the tires is 3500 miles. The production manager will stop the production if there is evidence that the mean lifetime of the tires is below 25000 miles.

- a. If the production manager wishes to have 80% power of detecting a shift in the lifetime mean of the tires from 25000 to 24000 miles and if he is willing to take a 5% risk of committing a Type I error, what sample size must be selected?
- b. If the production manager wishes to have 80% power of detecting a shift in the lifetime mean of the tires from 25000 to 23000 miles and if he is willing to take a 5% risk of committing a Type I error, what sample size must be selected?

Other hypothesis tests

Test for the difference between two population means:

$H_0 : \mu_1 - \mu_2 = \delta$ (δ could be 0, and the test is whether $\mu_1 = \mu_2$).

$H_a : \mu_1 - \mu_2 > \delta$, or $\mu_1 - \mu_2 < \delta$, or $\mu_1 - \mu_2 \neq \delta$

In order to test this hypothesis we select two samples from the two populations. Let the two samples be X_1, X_2, \dots, X_n , and Y_1, Y_2, \dots, Y_n . The test statistic is based on the difference of the two sample means, $\bar{X} - \bar{Y}$, and it depends on whether σ_1^2, σ_2^2 are known, whether $\sigma_1^2 = \sigma_2^2$, whether the sample sizes are small or large. Below we summarize all these different cases.

- a. The two variances, σ_1^2, σ_2^2 are known, and the two populations are normal. Then regardless of the size of the two samples (could be small or large), the test statistics is:

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

If Z falls in the rejection region (based on the significance level α) then H_0 is rejected.

- b. The two variances are unknown and $n_1 \geq 30, n_2 \geq 30$, (large samples). We will estimate the two unknown variances with the sample variances, s_1^2, s_2^2 . Because the two samples are large we can still use the Z test as an approximation.

$$Z \approx \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

If Z falls in the rejection region (based on the significance level α) then H_0 is rejected.

- c. The two variances are unknown but equal ($\sigma_1^2 = \sigma_2^2$) and $n_1 \leq 30$, or $n_2 \leq 30$, (one or both of the samples are small). We will estimate the unknown but common variance with the so called pooled variance s_{pooled}^2 , and the test statistic will be t with $n_1 + n_2 - 2$ degrees of freedom.

$$t = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{s_{\text{pooled}}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where

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

If t falls in the rejection region (based on the significance level α and $df = n_1 + n_2 - 2$) then H_0 is rejected.

Test for the difference between two population proportions:

$H_0 : p_1 - p_2 = 0$ (we test whether the two population proportions are equal).

$H_a : p_1 - p_2 > 0$, or $p_1 - p_2 < 0$, or $p_1 - p_2 \neq 0$

We select two samples of size n_1, n_2 and the number of successes X_1, X_2 are counted in each sample. The test statistic will be based on the the difference between the two sample proportion, $\hat{p}_1 - \hat{p}_2$, where $\hat{p}_1 = \frac{X_1}{n_1}$ and $\hat{p}_2 = \frac{X_2}{n_2}$. The test statistics is:

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where,

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}.$$

Under H_0 the two population proportions are equal and therefore the variance of $\hat{p}_1 - \hat{p}_2$ is:

$$\text{var}(\hat{p}_1 - \hat{p}_2) = \text{var}\left(\frac{X_1}{n_1} - \frac{X_2}{n_2}\right) = \text{var}\left(\frac{X_1}{n_1}\right) + \text{var}\left(\frac{X_2}{n_2}\right) = \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}$$

Because $p_1 = p_2 = p$ (under H_0), the variance of $\hat{p}_1 - \hat{p}_2$ is:

$$\text{var}(\hat{p}_1 - \hat{p}_2) = p(1 - p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right), \quad p \text{ is unknown and it is estimated by } \hat{p}.$$

If Z falls in the rejection region (based on the significance level α) then H_0 is rejected.

The paired t test

In many experiments the same variable is measured under two different conditions. For example, in clinical trials the participants may be evaluated at baseline and then evaluated again at the end of the treatment. For example, the blood pressure is measured for several patients before and after administration of certain drug. The difference between the value at baseline and the value at the end is computed for each participant as follows:

	Value at baseline	Value at the end	Difference
Subject 1	x_{1b}	x_{1a}	$d_1 = x_{1b} - x_{1a}$
Subject 2	x_{2b}	x_{2a}	$d_2 = x_{2b} - x_{2a}$
Subject 3	x_{3b}	x_{3a}	$d_3 = x_{3b} - x_{3a}$
\vdots	\vdots	\vdots	\vdots
\vdots	\vdots	\vdots	\vdots
Subject n	x_{nb}	x_{na}	$d_n = x_{nb} - x_{na}$

We then compute the sample mean and sample standard deviation of the differences.

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n}, \quad s_d^2 = \frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n - 1}$$

The hypothesis we want to test is:

$$H_0 : \mu_d = d_0$$

$$H_a : \mu_d < d_0 \text{ or } \mu_d > d_0 \text{ or } \mu_d \neq d_0$$

If we choose $d_0 = 0$ we are testing whether the before and after treatment are the same.

Test statistic:

$$t = \frac{\bar{d} - d_0}{\frac{s_d}{\sqrt{n}}}$$

Assumption: The differences are treated as a random sample from a normal distribution.

We reject H_0 if the t value falls in the rejection region which is based on the significance level α and $n - 1$ degrees of freedom.

Example:

From *Mathematical Statistics and Data Analysis*, John Rice, Third Edition, Duxbury (2007).

To study the effect of cigarette smoking on platelet aggregation researchers drew blood samples from 11 individuals before and after they smoked a cigarette and measured the percentage of blood platelet aggregation. Platelets are involved in the formation of blood clots, and it is known that smokers suffer more from disorders involving blood clots than do nonsmokers. This study can be found in Levine, P. H. (1973). An acute effect of cigarette smoking on platelet function, *Circulation*, 48, 619-623 (see attached article).

Before	After	Difference
25	27	2
25	29	4
27	37	10
44	56	12
30	46	16
67	82	15
53	57	4
53	80	27
52	61	9
60	59	-1
28	43	15

Test the null hypothesis that the means before and after are the same. Use $\alpha = 0.05$.