Hypothesis Testing

1-sample t-test	For the population mean, μ
Paired t-test	For the population mean difference
1-sample z-test	For the population proportion, p

Hypothesis tests

Data are usually collected in order to answer a question about a population of interest.

The mean SBP of men aged 37-44 is 127.2mmHg.

Is the mean SBP of diabetic men the same (i.e. equal to 127.2 mmHg)?

Hypothesis tests

We have a hypothesized value of μ, denoted μ_0,
and we ask the question:

Do the data agree with the hypothesis?

Is the mean SBP of diabetic men equal to 127.2 mmHg?

Hypothesis

Hypothesised value for the mean SBP of diabetic men

Collect data

Do the data agree with the hypothesis?

Hypothesis tests

We have a hypothesized value of μ, denoted μ_0.

Do the data agree with the hypothesis?

Define two hypotheses, which between them cover all possibilities (two-sided test)

Null hypothesis: *mean SBP of diabetic men equal to 127.2 mmHg*

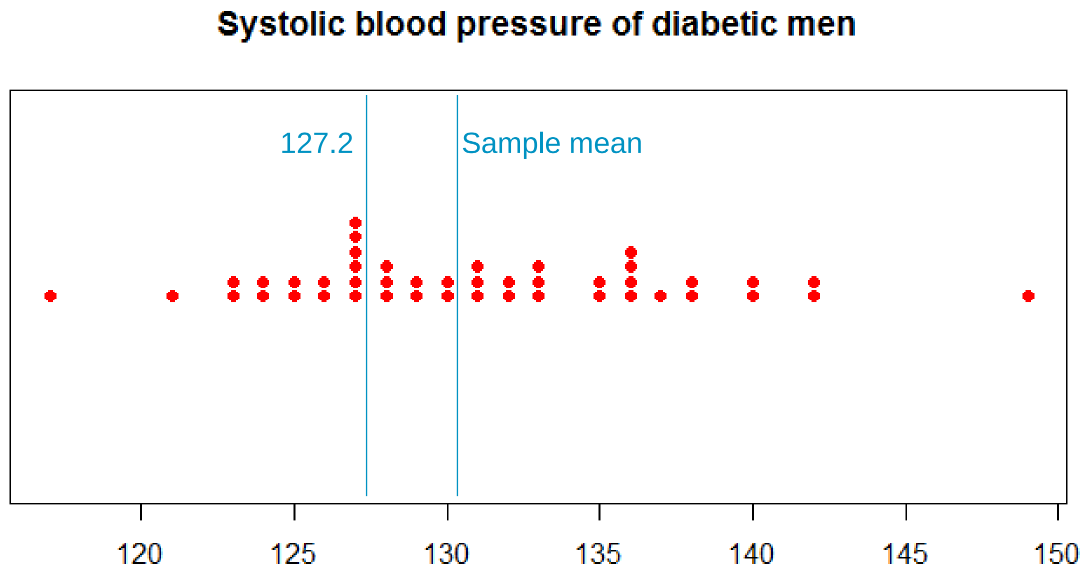
Alternative hypothesis: *mean SBP of diabetic men not equal to 127.2 mmHg*

$$H_0 : \mu = 127.2$$

$$H_1 : \mu \neq 127.2$$

Hypothesis tests

SBP was measured for 45 diabetic men, all aged between 37-44.



If the data do not contradict

$$\mu = 127.2$$

then we "*do not have enough evidence to reject the null hypothesis*"

If the data do contradict

$$\mu = 127.2$$

then we "*have enough evidence to reject the null hypothesis and instead accept the alternative hypothesis*"

General idea

Define the null and the alternative hypothesis
and decide on a level of significance

↓

(Collect data)

↓

Use a statistical test to see whether there is sufficient evidence to reject the null hypothesis

↓

Generate p-value from the test

↓

Interpret the p-value, and hence decide whether to reject the null hypothesis

General idea

Define the null and the alternative hypothesis and decide on a level of significance ✓

↓

(Collect data) ✓

↓

Use a statistical test to see whether there is sufficient evidence to reject the null hypothesis

↓

Generate p-value from the test

↓

Interpret the p-value, and hence decide whether to reject the null hypothesis

The one sample t-test

Data:

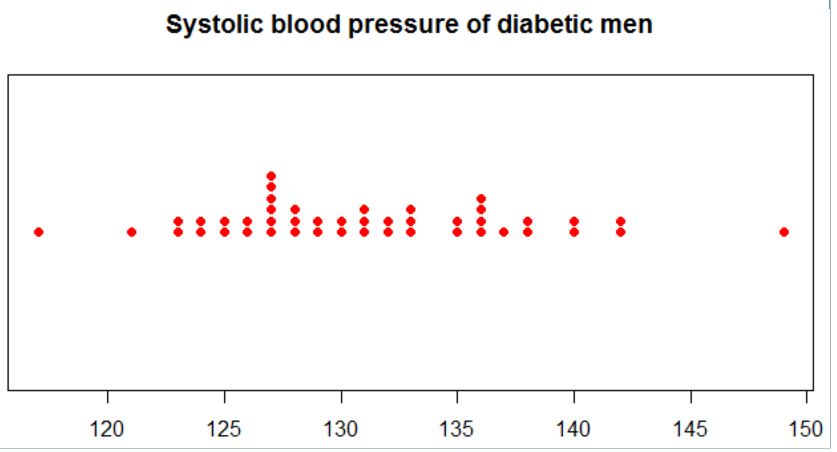
One sample

Testing:

Is there evidence that the (unknown) population mean is equal to some hypothesised value?

Assumption:

Data come from a normally distributed population.



The one sample t-test

Test statistic

$$H_0 : \mu = 127.2$$

$$H_1 : \mu \neq 127.2$$

Using the data and the hypothesised value of the population mean, μ_0 , compute the test statistic:

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

$$\bar{x} = 130.9778$$
$$s = 6.37$$
$$\mu_0 = 127.2$$
$$n = 45$$

The value of our test statistic is:

$$t = \frac{130.9778 - 127.2}{6.37/\sqrt{45}} = 3.98$$

General idea

Define the null and the alternative hypothesis and decide on a level of significance ✓

(Collect data) ✓

"Compute a test statistic" →

Use a statistical test to see whether there is sufficient evidence to reject the null hypothesis

Generate p-value from the test

Interpret the p-value, and hence decide whether to reject the null hypothesis

General idea

The value of our test statistic is:

$$t = \frac{130.9778 - 127.2}{6.37/\sqrt{45}} = 3.98$$

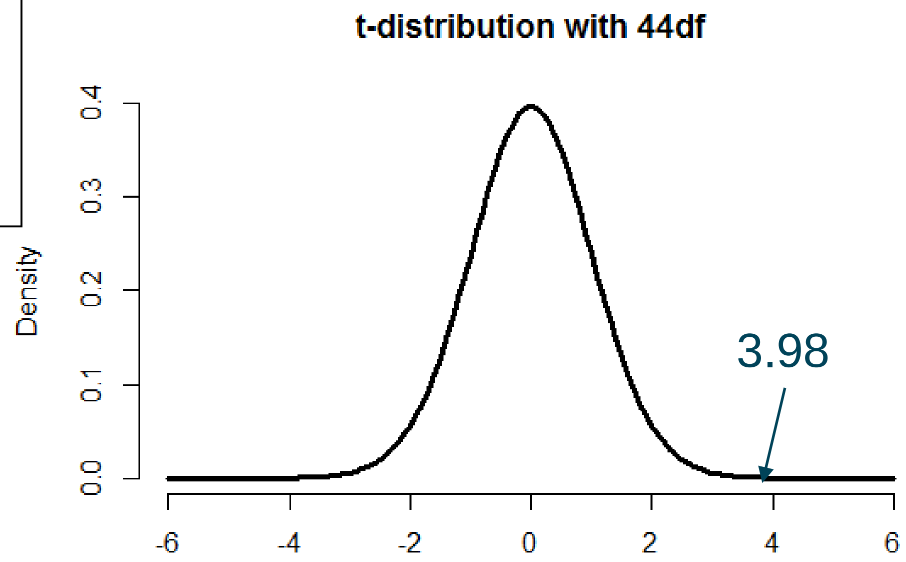
If the data do not contradict the null hypothesis, the test statistic will be:

- a random value from a t-distribution with (n-1) degrees of freedom
- reasonably close to zero...

Should we reject the null hypothesis?

In other words:

Is the value 3.98 extreme?



Measure this using the "p-value"

What is a p-value?

The *probability* that **– if H_0 is true-** we get a test-statistic (data) **as extreme, or more extreme**, than the one observed.

It is a measure of how well the data support the null hypothesis, H_0 .

Small p-value:
Data do not support the null hypothesis

Reject null hypothesis in favour of the alternative hypothesis.

Large p-value
Data do not contradict the null hypothesis

We *cannot reject* the null hypothesis*.

*but we never, ever *accept a null hypothesis*

Computing the p-value

The value of our test statistic is:

$$t = \frac{130.9778 - 127.2}{6.37/\sqrt{45}} = 3.98$$

Is the value 3.98 extreme?

What values are as, or more extreme, than this?

- Values greater than 3.98
- Values less than -3.98

What's the probability of observing this?

$$\text{p-value} = P(T_{44} > 3.98) + P(T_{44} < -3.98)$$
$$= 2\boxed{P(T_{44} > 3.98)}$$

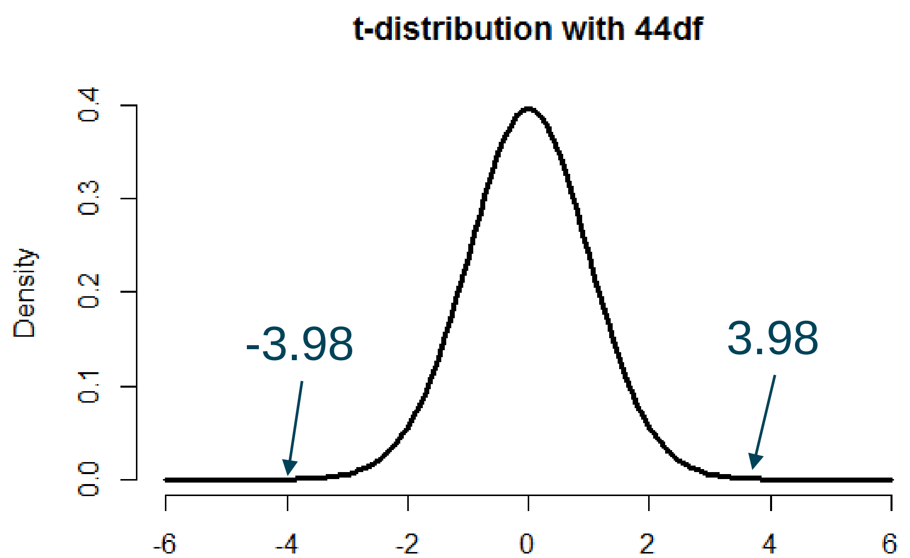
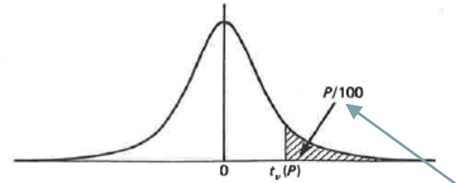


TABLE 10. PERCENTAGE POINTS OF THE t-DISTRIBUTION

This table gives percentage points $t_\nu(P)$ defined by the equation

$$\frac{P}{100} = \frac{1}{\sqrt{\nu\pi}}\frac{\Gamma(\frac{1}{2}\nu+\frac{1}{2})}{\Gamma(\frac{1}{2}\nu)}\int_{t_\nu(P)}^{\infty}\frac{dt}{(1+t^2/\nu)^{\frac{1}{2}(\nu+1)}}.$$

Let X_1 and X_2 be independent random variables having a normal distribution with zero mean and unit variance and a χ^2-distribution with ν degrees of freedom respectively; then $t = X_1/\sqrt{X_2/\nu}$ has Student's t-distribution with ν degrees of freedom, and the probability that $t \geqslant t_\nu(P)$ is $P/100$. The lower percentage points are given by symmetry as $-t_\nu(P)$, and the probability that $|t| \geqslant t_\nu(P)$ is $2P/100$.



The limiting distribution of t as ν tends to infinity is the normal distribution with zero mean and unit variance. When ν is large interpolation in ν should be harmonic.

Percentage area

Degrees of freedom

P	40	30	25	20	15	10	5	2·5	1	0·5	0·1	0·05
ν = 1	0·3249	0·7265	1·0000	1·3764	1·963	3·078	6·314	12·71	31·82	63·66	318·3	636·6
2	0·2887	0·6172	0·8165	1·0607	1·386	1·886	2·920	4·303	6·965	9·925	22·33	31·60
3	0·2767	0·5844	0·7649	0·9785	1·250	1·638	2·353	3·182	4·541	5·841	10·21	12·92
4	0·2707	0·5686	0·7407	0·9410	1·190	1·533	2·132	2·776	3·747	4·604	7·173	8·610
5	0·2672	0·5594	0·7267	0·9195	1·156	1·476	2·015	2·571	3·365	4·032	5·893	6·869
6	0·2648	0·5534	0·7176	0·9057	1·134	1·440	1·943	2·447	3·143	3·707	5·208	5·959
7	0·2632	0·5491	0·7111	0·8960	1·119	1·415	1·895	2·365	2·998	3·499	4·785	5·408
8	0·2619	0·5459	0·7064	0·8889	1·108	1·397	1·860	2·306	2·896	3·355	4·501	5·041
9	0·2610	0·5435	0·7027	0·8834	1·100	1·383	1·833	2·262	2·821	3·250	4·297	4·781
10	0·2602	0·5415	0·6998	0·8791	1·093	1·372	1·812	2·228	2·764	3·169	4·144	4·587
11	0·2596	0·5399	0·6974	0·8755	1·088	1·363	1·796	2·201	2·718	3·106	4·025	4·437
12	0·2590	0·5386	0·6955	0·8726	1·083	1·356	1·782	2·179	2·681	3·055	3·930	4·318
13	0·2586	0·5375	0·6938	0·8702	1·079	1·350	1·771	2·160	2·650	3·012	3·852	4·221
14	0·2582	0·5366	0·6924	0·8681	1·076	1·345	1·761	2·145	2·624	2·977	3·787	4·140
15	0·2579	0·5357	0·6912	0·8662	1·074	1·341	1·753	2·131	2·602	2·947	3·733	4·073
16	0·2576	0·5350	0·6901	0·8647	1·071	1·337	1·746	2·120	2·583	2·921	3·686	4·015
17	0·2573	0·5344	0·6892	0·8633	1·069	1·333	1·740	2·110	2·567	2·898	3·646	3·965
18	0·2571	0·5338	0·6884	0·8620	1·067	1·330	1·734	2·101	2·552	2·878	3·610	3·922
19	0·2569	0·5333	0·6876	0·8610	1·066	1·328	1·729	2·093	2·539	2·861	3·579	3·883
20	0·2567	0·5329	0·6870	0·8600	1·064	1·325	1·725	2·086	2·528	2·845	3·552	3·850
21	0·2566	0·5325	0·6864	0·8591	1·063	1·323	1·721	2·080	2·518	2·831	3·527	3·819
22	0·2564	0·5321	0·6858	0·8583	1·061	1·321	1·717	2·074	2·508	2·819	3·505	3·792
23	0·2563	0·5317	0·6853	0·8575	1·060	1·319	1·714	2·069	2·500	2·807	3·485	3·768
24	0·2562	0·5314	0·6848	0·8569	1·059	1·318	1·711	2·064	2·492	2·797	3·467	3·745
25	0·2561	0·5312	0·6844	0·8562	1·058	1·316	1·708	2·060	2·485	2·787	3·450	3·725
26	0·2560	0·5309	0·6840	0·8557	1·058	1·315	1·706	2·056	2·479	2·779	3·435	3·707
27	0·2559	0·5306	0·6837	0·8551	1·057	1·314	1·703	2·052	2·473	2·771	3·421	3·690
28	0·2558	0·5304	0·6834	0·8546	1·056	1·313	1·701	2·048	2·467	2·763	3·408	3·674
29	0·2557	0·5302	0·6830	0·8542	1·055	1·311	1·699	2·045	2·462	2·756	3·396	3·659
30	0·2556	0·5300	0·6828	0·8538	1·055	1·310	1·697	2·042	2·457	2·750	3·385	3·646
32	0·2555	0·5297	0·6822	0·8530	1·054	1·309	1·694	2·037	2·449	2·738	3·365	3·622
34	0·2553	0·5294	0·6818	0·8523	1·052	1·307	1·691	2·032	2·441	2·728	3·348	3·601
36	0·2552	0·5291	0·6814	0·8517	1·052	1·306	1·688	2·028	2·434	2·719	3·333	3·582
38	0·2551	0·5288	0·6810	0·8512	1·051	1·304	1·686	2·024	2·429	2·712	3·319	3·566
40	0·2550	0·5286	0·6807	0·8507	1·050	1·303	1·684	2·021	2·423	2·704	3·307	3·551
50	0·2547	0·5278	0·6794	0·8489	1·047	1·299	1·676	2·009	2·403	2·678	3·261	3·496
60	0·2545	0·5272	0·6786	0·8477	1·045	1·296	1·671	2·000	2·390	2·660	3·232	3·460
120	0·2539	0·5258	0·6765	0·8446	1·041	1·289	1·658	1·980	2·358	2·617	3·160	3·373
∞	0·2533	0·5244	0·6745	0·8416	1·036	1·282	1·645	1·960	2·326	2·576	3·090	3·291

Don't have 44df in the table, but note that for both 40df and 50df:

$$P(T_{40} > 3.98) < 0.0005$$
$$P(T_{50} > 3.98) < 0.0005$$

Can conclude that

$$P(T_{44} > 3.98) < 0.0005$$

Computing the p-value

The value of our test statistic is:

$$t = \frac{130.9778 - 127.2}{6.37/\sqrt{45}} = 3.98$$

Is the value 3.98 extreme?

What values are as, or more extreme, than this?

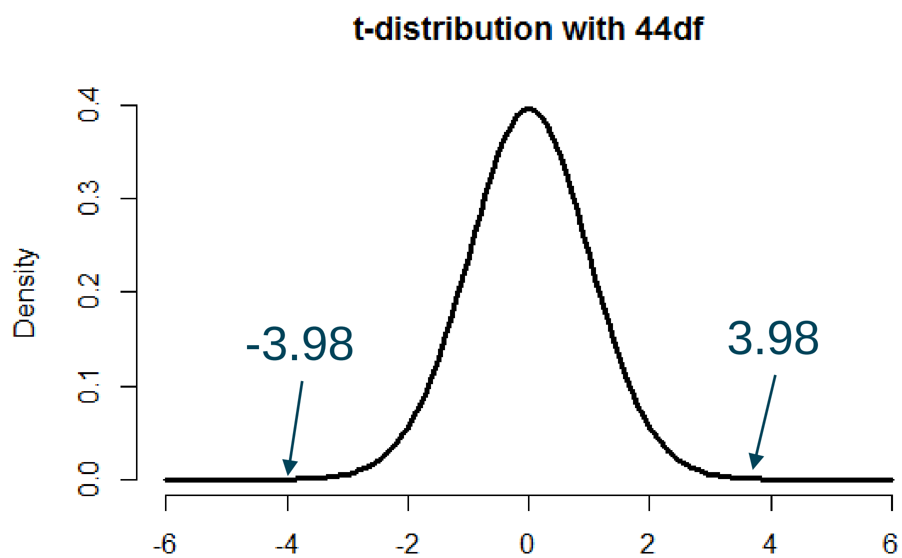
- Values greater than 3.98
- Values less than -3.98

What's the probability of observing this?

$$\text{p-value} = P(T_{44} > 3.98) + P(T_{44} < -3.98)$$

$$= 2P(T_{44} > 3.98)$$

$$\text{p-value} < 0.001$$



Interpretation of the p-value

Recall the definition of a p-value:

The *probability* that **– if H_0 is true-** we get a test-statistic (data) **as extreme, or more extreme**, than the one observed.

p-value	Typical interpretation*
p-value < 0.01	Strong evidence against the null hypothesis
0.01 < p-value < 0.05	Fairly strong evidence against the null hypothesis
p-value >0.05	Little or no evidence against the null hypothesis (i.e. data are consistent with the null)

*This will depend on the context – you may wish to set your threshold much lower than 0.05

Final interpretation

The value of our test statistic is:

$$t = \frac{130.9778 - 127.2}{6.37/\sqrt{45}} = 3.98$$

$$\text{p-value} < 0.001$$

Conclusions?

Since the p-value for the one-sample t-test is less than 0.01, we have **strong evidence against the null hypothesis**.

We therefore conclude that the mean SBP of diabetic men aged 37-44 is not equal to 127.2mmHg.

The mean SBP of diabetic men aged 37-44 differs from the mean SBP of men in the general population aged 37-44.

Hypothesis Testing

✓ **1-sample t-test**	**For the population mean, *μ***
➢ Paired t-test	For the population mean difference
1-sample z-test	For the population proportion, *p*

Paired t-tests

In this case we have paired data, e.g.

- Pain score pre-and post-medication;
- IQ scores from pairs of siblings;
- Blood pressure reading on the same subject using automated and manual readings.

Our interest is whether the mean difference between the pairs is equal to zero

Paired t-tests

Consider, for example, blood pressure readings on the same subject using automated and manual readings

Subject	Manual	Automated	Difference
1	110	112	2
2	125	121	-4
3	109	115	6
4	129	128	-1
5	122	122	0
…	…	…	…

Question:
Is the mean difference equal to zero?

Paired t-tests

Subject	Manual	Automated	Difference
1	110	112	2
2	125	121	-4
3	109	115	6
4	129	128	-1
5	122	122	0
…	…	…	…

Question:
Is the mean difference equal to zero?

Null hypothesis:
mean difference between manual and automated readings is 0

Alternative hypothesis:
mean difference between manual and automated readings is NOT 0

Conduct a one-sample t-test on the differences

$$H_0 : \mu = 0$$

$$H_1 : \mu \neq 0$$

One sample *vs* paired t-tests

One sample t-test

- ✓ One sample
- ✓ Testing:

Is there evidence that the (unknown) population mean is equal to a hypothesized value?

- ✓ Assumption: data come from a normally distributed population

Paired t-test

- ➢ Paired samples
- ➢ Testing:

Is there evidence that the (unknown) population mean difference between the pairs is equal to some hypothesized value (usually 0)?

- ➢ Assumption: the **differences** come from a normally distributed population

Hypothesis Testing

✓ **1-sample t-test**	**For the population mean, μ**
✓ Paired t-test	For the population mean difference
➢ 1-sample z-test	For the population proportion, p

Example

It is known that for the Statistics module STAT6102,

55% of the students had taken a statistics course prior to joining STAT6102.

In a particular year, 21 out of 35 STAT6102 students claimed that they had taken a course in Statistics.

Is there evidence that the proportion of the specific year's cohort of STAT6102 students who have taken a statistics course previously differs from 0.55?

The one-sample z-test

Define the null and the alternative hypothesis
and decide on a level of significance

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

(Collect data)

Attention: We need to make assumptions
Large sample:

$$np_0 > 5 \quad and \quad n(1 - p_0) > 5$$

Use a statistical test to see whether there is sufficient evidence to reject the null hypothesis

Test statistic:

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

Generate p-value from the test

$$2P(Z > |z|)$$

Interpret the p-value, and hence decide whether to reject the null hypothesis

The one-sample z-test

Define the null and the alternative hypothesis
and decide on a level of significance

$$H_0 : p = 0.55 \quad vs \quad H_1 : p \neq 0.55$$

(Collect data)

Attention: We need to make assumptions
Large sample??

$$np_0 = 35 \times 0.55 = 19.25 > 5$$

$$n(1 - p_0) = 35 \times 0.45 = 15.75 > 5$$

Use a statistical test to see whether there is sufficient evidence to reject the null hypothesis

Test statistic :

$$z = \frac{0.6 - 0.55}{\sqrt{0.55 \times 0.45 / 35}} = 0.59$$

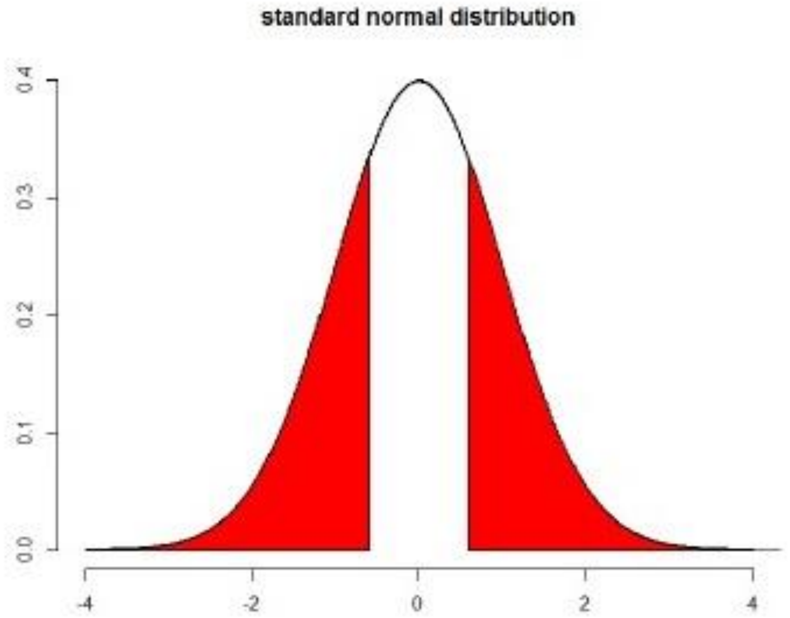
Generate p-value from the test

$$2P(Z > |0.59|)$$

Interpret the p-value, and hence decide whether to reject the null hypothesis

The one-sample z-test

$$H_0 : p = 0.55 \quad vs \quad H_1 : p \neq 0.55$$



$$\text{Test statistic :}$$
$$z = \frac{0.6 - 0.55}{\sqrt{0.55 \times 0.45 / 35}} = 0.59$$

$$2P(Z > |0.59|) = 2 \times P(Z > 0.59) = 2 \times (1 - 0.7224) = 0.5552$$

Conclusions:

- p-value is 0.5552
- Not enough evidence to reject the null hypothesis
- No evidence to suggest that the true proportion of STAT6102 students who have taken a statistics course previously differs from 0.55

What did we do today?

- ✓ What is a hypothesis?
- ✓ How do we do test a hypothesis: general procedure
- ✓ One sample t-test (for the population mean).
- ✓ Paired t-test (mean difference between pairs).
- ✓ One sample z-test (for the population proportion).