**A6Wd Calculation of statistical power and probability of a type II error for one and two sample z and t tests**

In this document we include a range of examples to illustrate the calculation of β = P(Type II error) and the corresponding value of statistical power.

α = P(Type I error) = P(rejecting H0 given H0 true)

β = P(Type II error) = P(failing to reject H0 given H0 false)

Power = P(rejecting H0 given H0 is false)



Contents

[Example 1 One tail z test 2](#_Toc9606806)

[Example 2 Two-tail z test 7](#_Toc9606807)

[Example 3 Two-tail z test 8](#_Toc9606808)

[Example 4 Lower one-tail z test 11](#_Toc9606809)

[Example 5 Upper one-tail z-test 12](#_Toc9606810)

[Example 6 Two-tail t test 13](#_Toc9606811)

[Example 7 Two sample independent t-tests 15](#_Toc9606812)

[Example 8 Two sample dependent t-tests 20](#_Toc9606813)

[Example 9 Power of a two sample variance test 22](#_Toc9606814)

[Check your understanding 25](#_Toc9606815)

# Example 1 One tail z test

Consider the problem of estimating the daily revenue on a brand of sweets, where historically, the average revenue is £100 per day. The shop would like to check whether the current spending per day is less than £100 and they have decided to collect a sample on a day. The sample happens to be of size 32, with a population standard deviation of £25.

The shop, after consultation with an analyst, decides to conduct a one sample z-test, but they would like to know how confident they can be in the outcome of applying this test to the data. If the true average spend is £95 (as claimed by the industry experts, for example), what is the probability of rejecting a false null hypothesis? (Test at a significance level of 0.05).

In this example, the hypothesis test is:

H0: population mean μ = £100

H1: population mean μ < £100

The way the alternative hypothesis is stated implies that we need to use a lower one-tail test. The α risk is to be controlled at α = 0.05, when μ = μ0 = 100. Therefore, the rejection probability at μ0 = 100 is P(H1: μ0 = 100) = 0.05. A Type I error occurs if we reject the null hypothesis H0 in favor of the alternative H1 when the null hypothesis H0 is true, where the test significance level α = P(Type I error). If the null hypothesis is true, then we would expect z = 0.

For what values of the sample mean ($\overbar{x}$) do we accept or reject the null hypothesis (H0)? Figure 6.16 represents the Excel solution to calculate the P(Type II error) and the power of the test.



Figure 1 Excel spreadsheet showing the calculation of the power of the test

**Excel solution**

μ0 = Cell B4 Value = 100

n = Cell B5 Value = 32

σ = Cell B6 Value = 25

α = Cell B7 Value = 0.05

zcri = Cell B9 Formula:=NORM.S.INV(B7)

Lower sample mean = Cell B10 Formula:=B4+B9\*B6/SQRT(B5)

if μ1 = 95? Cell B14 Value = 95

zcal = Cell B17 Formula:=(B10-B14)/(B6/SQRT(B5))

β = Area to right of z of - 0.5134... = Cell B18 Formula:=1-NORM.S.DIST(B17,TRUE)

Power = 1 - β = Cell B20 Formula:=1-B18

If the P(Type I error if H0 is true) = α = 0.05, then the lower one tail critical value of z equals – 1.64 (via statistical tables or by using the Excel function = NORM.S.INV(0.05)). As we are using a **one sample z test**, then

$z=\frac{\overbar{x }- μ}{\frac{σ}{\sqrt{n}}}$ (1)

Given the sample size (n = 32) and population standard deviation (σ = £25), we can re-arrange equation (6.6) to calculate the lower and upper critical sample mean values using equation (2).

$\overbar{x}\_{cri} = μ \pm z\_{cri}\frac{σ}{\sqrt{n}}$ (2)

$$Lower critical sample mean value, \overbar{x}\_{cri} = 100- 1.64\*\frac{25}{\sqrt{32}}=92.73$$

We observe the lower mean value is 92.73 (see Excel solution, Figure 1, cell B10). Therefore, if the sample mean lies above this value then the null hypothesis is not violated. If the sample mean is less than 92.73 then we would reject the null hypothesis and accept the alternative hypothesis (H1: μ < £100).

If we have failed to reject the null hypothesis (H0) when the alternative hypothesis (H1) is true, we know that this is a Type II error and it is denoted by the symbol β where, β =P (Type II error). A type II error occurs when we fail to reject the null hypothesis given that the population mean is not £100. We can only calculate the type II error if we assume a new value of the population mean. It is important to note that the type II error value will be different for different population mean values.

In this example, we are told that the population average spend is £95. For the example above, we commit a Type II error if the observed sample mean does not fall in the rejection region, that is, if it is less than £92.73. To calculate the value of β we need to calculate the area on the H1: μ1 = £95 distribution with the sample mean value ≥ 92.73. Figure 2 illustrates graphically the relationship between α, β and the statistical power for this example.



Figure 2 Relationship between α, β, and statistical power

Probability of a Type II error (β) when H1: μ1 = 95

The probability of a Type II error is that we have failed to reject H0 when the alternative hypothesis is true. This can be written as:

β = P(failing to reject H0 given H0 false), or specific in this example:

β = P(failing to reject H0: μ = 100 when H1: μ1 = 95)

$$β=P\left(\overbar{X} \geq 92.73070529 when H\_{1}:μ = 95 \right)$$

Standardizing by substituting values (sample mean = 92.73…, μ1 = 95, σ = 25, n = 32) into equation (1) gives:

$$β = P\left(z \geq \frac{92.73070529-95}{^{25}/\_{\sqrt{32}}}\right)$$

β = P(z ≥ - 0.513482777)

β = 0.696193172

The probability of a Type II error involved in failing to reject the null hypothesis when the true population mean is 95 is 0.696 or approximately 70%.

**Statistical power**

The power of a statistical test is the probability that you correctly reject a false null hypothesis. For this example

Power = P(rejecting H0: μ = 100 when μ = 95)

Power = P(Z < - 0.513482777)

Power = 1 – β

Power = 0.303806828

**>>**

In summary, we have determined that we have a 30% chance of rejecting the null hypothesis H0: μ = 100 in favor of the alternative hypothesis H1: μ < 100 if the true population mean is 95. The relationship between power and β is given by equation (6.5), Statistical power = 1 – β.

What is the relationship between β and different values of the true mean? Figure 6.3 illustrates graphically the relationship between β and values of an alternative mean for the hypothesis test H0: μ = 100 against H1: μ < 100 when assuming the alternative true means range from 86 → 100 (other values are n = 32, σ = 25, and α = 0.05).



Figure 6.3 Variation of β against assumed mean

We observe that when the alternative means are near the value of the null hypothesis, μ = 100, the P(Type II error) is high since it is difficult to discriminate between a distribution with a mean of 100 and a distribution with a mean of 99.9. However, as the assumed mean moves away from the value of the null hypothesis, μ = 100, the β values reduce in size. This reinforces the idea that it is easier to discriminate between a distribution with μ = 100 and a distribution with μ = 90 than between a distribution with μ = 100 and an alternative distribution with μ = 99.9.

What is the relationship between the power of the test and different values of the true mean?

Figure 6.4 illustrates graphically the relationship between the power of the test and various values of the alternative population means. Observe that the power of the test increases as the alternative population mean values move away from the value of the null hypothesis mean, μ = 100. This is what we would expect given that as we move further and further away from the null hypothesis, μ = 100, a correct decision to reject the null hypothesis becomes more likely.



Figure 6.4 The power function

For situations involving one-tail tests in which the actual mean, μ1, exceeds the null hypothesis (H0: μ = 100) the converse would be true. The larger the actual mean, μ1, compared with the null hypothesis (H0: μ = 100), the greater is the power. For two-tail tests, the greater the distance between the actual mean, μ1, and the null hypothesis (H0: μ = 100), the greater the power of the test.

How will a changing sample size affect the power of the test?

If we calculate the power of the test for different samples sizes, n, we find that the relationship is as illustrated in Figure 6.5 where we have calculated the power of the test for a range of assumed alternative population means and superimposed the results on the same graph for two sample sizes, n = 16 and n = 64.



Figure 6.5 Power versus alternative population mean for two different sample sizes

We observe from this graph that if we want to increase our chance of rejecting the null hypothesis when the alternative hypothesis is true, we can do so by increasing our sample size, n.

From the above, the following observations can be made:

1. A one-tail test is more powerful than a two-tail test.
2. An increase in the significance level (α) results in an increase in the power of the test,
3. An increase in sample size (n) results in an increase in the power of the test.

In general, for every hypothesis test that we conduct, we would want to do the following:

1. Minimize the probability of a type II error (typically α ≤ 0.1).
2. Maximize the power of the test at a value of the alternative population mean that is scientifically meaningful (typically power ≥ 0.8) or minimize β (typically β ≤ 0.2).

If you are interested in this topic then you will find want to learn more documents that explore the calculation of β and statistical power for one and two tail z and t-tests.

# Example 2 Two-tail z test

H0: μ = 100

H1: μ ≠ 100

Two-tail test z test

Sample: n = 16, σ = 16

Take significance level 0.05

α = P(Type I error) = P(rejecting H0 given H0 true) = 0.05

Calculate β and power if alternative mean = 102?

Excel solution



We would accept H0 if the sample mean lies between 92.16 → 107.84

β = P(Type II error) = P(failing to reject H0 given H0 false)

Power = P(rejecting H0 given H0 is false)

Power = P(rejecting H0 given H0 is false)

Power = P(sample mean ≤ 92.76 … given μ = 102 or sample mean ≥ 107.84 … given μ = 102)

Power = P(Z ≤ -2.45..) + P(Z ≥ 1.45 …)

Power = 0.079

β = P(92.16 ≤ sample mean ≤ 107.84 given μ = 102)

β = P(sample mean ≤ 107.84 given μ = 102) – P(sample mean ≤ 92.16 given μ = 102)

β = P(Z ≤ 1.45 …) – P(Z ≤ - 2.45 ….)

β = 0.921 (equivalent to Power = 1 - β)



Hence, if the alternative population mean = 102 and we use a sample size of 16 then we have only a 17% chance that we will reject the null hypothesis that μ = 100.

If we continue this process and calculate the value of the statistical power for different alternative population mean values, then we would create the POWER CURVE illustrated below.



# Example 3 Two-tail z test

H0: μ = 1280

H1: μ ≠ 1280

Two-tail test

Sample: n = 80, σ = 110

Two-tail z-test

Take significance level 0.0

α = P(Type I error) = P(rejecting H0 given H0 true) = 0.05

Calculate β and power if alternative mean = 1290?

Excel solution





We would accept H0 if the sample mean lies between 1256 → 1304

β = P(Type II error) = P(failing to reject H0 given H0 false)

Power = P(rejecting H0 given H0 is false)

Power = P(sample mean ≤ 1256 or ≥ 1304) = 0.13

β = P(1256 ≤ sample mean ≤ 1304 given μ = 1290) = 0.87

Hence, if the alternative population mean = 1290 and we use a sample size of 80 then we have only a 13% chance that we will reject the null hypothesis that μ = 1280.

**Example 3 Two-tail z test**

H0: μ = 75

H1: μ ≠ 75

Two-tail test

Sample: n = 16, σ = 8

Two-tail z test

Take significance level 0.05

α = P(Type I error) = P(rejecting H0 given H0 true) = 0.05

Calculate β and power if alternative mean = 76?

Excel solution





We would accept H0 if the sample mean lies between 71.08 → 78.92

β = P(Type II error) = P(failing to reject H0 given H0 false)

Power = P(rejecting H0 given H0 is false)

Power = P(sample mean ≤ 71.08 or ≥ 78.92) = 0.0791

β = P(1256 ≤ sample mean ≤ 1304 given μ = 1290) = 0.9210

Hence, if the alternative population mean = 76 and we use a sample size of 16 then we have only a 8% chance that we will reject the null hypothesis that μ = 76.

# Example 4 Lower one-tail z test

H0: μ = 368

H1: μ < 368

One-tail test

Sample: n = 25, σ = 15

Lower one-tail z test

Take significance level 0.05

α = P(Type I error) = P(rejecting H0 given H0 true) = 0.05

Calculate β and power if alternative mean = 360?

Excel solution





We would accept H0 if the sample mean is ≥ 363.065…..

β = P(Type II error) = P(failing to reject H0 given H0 false)

Power = P(rejecting H0 given H0 is false)

β = P(not rejecting H0: μ = 368 given μ = 360)

β = P(sample mean ≥ 363.065…given μ = 360)

β = 0.1534

Power = P(sample mean ≤ 363.065… given μ = 360)

Power = 0.8466

Hence, if the alternative population mean = 360 and we use a sample size of 25 then we have a 85% chance that we will reject the null hypothesis that μ = 368.

# Example 5 Upper one-tail z-test

H0: μ = 60

H1: μ > 60

One-tail test

Sample: n = 110, σ = 12

Upper one-tail z test

Take significance level 0.05

α = P(Type I error) = P(rejecting H0 given H0 true) = 0.05

Calculate β and power if alternative mean = 62.5?

Excel solution



β = P(Type II error) = P(failing to reject H0 given H0 false)

Power = P(rejecting H0 given H0 is false)

β = 0.2945

Power = 0.7055

Hence, if the alternative population mean = 62.5 and we use a sample size of 110 then we have a 71% chance that we will reject the null hypothesis that μ = 60.

# Example 6 Two-tail t test

H0: μ = 19.44

H1: μ ≠ 19.44

Two-tail test

Sample: n = 32, s = 6.23

Two-tail t test

Take significance level 0.05

α = P(Type I error) = P(rejecting H0 given H0 true) = 0.05

Calculate β and power if alternative mean = 23.4?

Excel solution



β = P(Type II error) = P(failing to reject H0 given H0 false)

Power = P(rejecting H0 given H0 is false)

β = 0.065

Power = 0.9351

Hence, if the alternative population mean = 23.4 and we use a sample size of 32 then we have a 94% chance that we will reject the null hypothesis that μ = 19.44.

**G\*Power solution**



# Example 7 Two sample independent t-tests

Excel solution



Figure 1 Power of a 2-sample 2-tail t-test

Student t statistic

$$t= \frac{\overbar{X}\_{2}- \overbar{X}\_{1}}{S\_{p} \sqrt{\frac{1}{n\_{1}}+ \frac{1}{n\_{2}}}}$$

where sample mean 2 > sample mean 1

$$Pooled variance S\_{p}^{2}= \frac{\left(n\_{1}-1\right) S\_{1}^{2}+ \left(n\_{2}-1\right) S\_{2}^{2}}{n\_{1}+ n\_{2}-2}$$

With standard error (SE) given by the equation

$$SE= \sqrt{\frac{S\_{1}^{2}}{n\_{1}}+ \frac{S\_{2}^{2}}{n\_{2}}}$$

If n1 = n2 = n, then SE is given by the equation

$$SE= \sqrt{\frac{\left(S\_{1}^{2}- S\_{2}^{2}\right)}{n}}$$

H0: μ2 = μ1

H0: μ2 ≠ μ1

Two-tail test

Let us choose a significance level α = 0.05

Now the difference between the two-sample means δ = sample mean 1 – sample mean 2

What is the power of the test: Power = P(rejecting H0 given H0 is false)

From Excel:

Sample mean 2 = 54.2

Sample mean 1 = 43.65

n1 = 20

n2 = 20

sample standard deviation 1 = 26.3504

sample standard deviation 2 = 26.4109

δ = sample mean 2 – sample mean 1 = 54.2 – 43.65 = 10.55

Pooled variance = 695.941

Standard error SE = 8.34231

If H0 true, then what are the critical values of (sample 2 mean – sample 1 mean)?

Given significance level α = 0.05, two-tail

Critical t-value, tcri = t(α/2, df) = t(0.05/2, 38) = T.INV.2T(0.05, 38) = ± 2.02439

Given δ = 10.55 > 0, then calculate the upper value of (sample mean 2 – sample mean 1) using:

Upper critical value of (sample mean 2 – sample mean 1) = $2.02439 × S\_{p} × \sqrt{\frac{1}{n\_{1}}+ \frac{1}{n\_{2}}}$

Therefore, upper critical value of (sample mean 2 – sample mean 1) = 16.88

Calculate the power of this test

Power = P(rejecting H0 given H0 is false)

Power = P(rejecting H0: μ2 - μ1 = 0 given μ2 - μ1 = δ)

Power = P(rejecting H0: μ2 - μ1 = 0 given μ2 - μ1 = 10.55)

Power = P(δ ≥ 16.88)

Power = P(Z ≥ (16.88 – 10.55)/8.34231)

Power = P(Z ≥ 0.759..)

Power = 0.22

The power for this test is 22% which is quite low.

Using Excel > Data > Data Analysis menu





**G\*Power solution**



**Increasing the power**

You can increase the power as follows:

Method 1 Change from 2tail to 1-tail

Change from a 2-tail test to a one-tail test – for example, if we have one-tail and we used a significance level of 0.05, then this 0.05 would be in the right-sided tail. The effect of this would be to reduce the critical value of t. By doing this we observe that the power of the test as increased from 22% to 34%.

A 2-tail test is called a non-directional test and a 1-tail test is called a directional test (given we have an implied direction that we wish to test.

Method 2 Increase the sample size

If we increased the sample size then we can increase the power of this test. For example, we have copied the original data twice, so that we now have 40 observations per sample but kept it at 1-tail with significance level of 0.05. If we do this we observe that the power is now 56%.

Method 3 Conduct a two-sample paired t-test to increase the power of the test (see Example 8 below).

# Example 8 Two sample dependent t-tests

Can increase the power by using a dependent t-test (also called a paired sample t-test).

$$t= \frac{\overbar{X}\_{2}- \overbar{X}\_{1}}{SE}$$

and

$$SE= \sqrt{\frac{\left(S\_{2}^{2}+ S\_{1}^{2}\right)}{n} - 2 × r × \left(\frac{S\_{2}}{\sqrt{n}}\right)\left(\frac{S\_{1}}{\sqrt{n}}\right)}$$

Assuming sample mean 2 > sample mean 1 to keep difference + ve value.

Notice that the SE formula is modified using the correlation co-efficient (r) between sample 2 and sample 1.

Degrees of freedom, df = n – 1

Using the method 1 example for a 1-tail but now using a 2-sample paired t-test.

Excel solution



Power now increased to 77%.

Excel Data > Data Analysis solution





**G\*Power solution**



# Example 9 Power of a two sample variance test

Let s12 and s22 represent the variances of two independent samples of size n1 and n2. If F = s12/s22.

The critical values and beta value for a two-tail test H0: σ12 ≠ σ22 implies F < 1, F > 1}

Lower critical value of F = F.INV (α/2, df1, df2)

Upper critical value of F = F.INV.RT (α/2, df1, df2)

β = F.DIST.RT (lower critical of F/F , df1, df2) - F.DIST.RT (upper critical of F/δ, df1, df2)

Power = 1 - β

The critical value and beta value for the upper one-tailed test H0: σ12 ≤ σ22 {implies F < 1}

Critical value = F.INV.RT (α, df1, df2)

β = F.DIST (critical of F/F , df1, df2)

Power = 1 - β

The critical value and beta value for the lower one-tailed test H0: σ12 ≥ σ22 {implies F > 1}

Critical value = F.INV (α, df1, df2)

β = F.DIST.RT (upper critical of F/F, df1, df2)

Power = 1 – β

Calculate the statistical power:

1. Upper one-tail (F > 1) with α = 0.05, sample sizes 50 and 60 and corresponding variances 1.75 and 2.25.



**G\*Power solution**



1. Lower one-tail (F < 1) with α = 0.05, sample sizes 50 and 60 and corresponding variances 2.25 and 1.20.



**G\*Power solution**



1. Upper one-tail (F ≠ 1) with α = 0.05, sample sizes 50 and 60 and corresponding variances. 2.25 and 1.75.



**G\*Power solution**



# Check your understanding

X1 A tea dispenser is designed to discharge at least 6 ounces of tea per cup (H0: μ ≥ 6), with a standard deviation of 0.24 ounces. If you select a random sample of 25 cups and you are willing to have a significance level of 5% of committing a Type I error, calculate the power of the test when the population mean of tea dispensed is 6.5 ounces.

X2 The lifetime of a tyre is at least 22,300 miles with a standard deviation of 3100 miles. During the production process, a random sample of 80 tyres are selected and their lifetimes measured. What is the probability of a Type II error if the alternative true mean is 23,400 miles? Assume you are happy to have a probability of a Type I error is 5%.