

# Biostatistics

## **Z** statistical

**test** and how to use it in data

analysis and the most important  
conditions for applying the test



**Mohamed Abulibdah**  
**Statistician and Data**  
**Analyst at the NCI -**  
**Cairo University -**  
**Egypt**



# Introduction

## the statistical (z) test used in data analysis

**Z test:** It is one of the types of parametric statistical tests. The z parameter statistical analysis test aims to compare and compare the value of the statistical community with the arithmetic mean of the population sample. This sample is part of it, and the evaluation of the sample is done on the basis of determining the alternative hypothesis and the null hypothesis, calculating the z values, then conducting the comparison, and finally issuing the decision of rejection or acceptance and drawing and showing the result.

## When is the T test and when is the Z test?

There is a slight difference between the time of using the T test and the Z test, as the testing statistical analysis test T is used if the sample size is less than thirty singles, and that the standard deviation is unknown and the community sample is known. T is not used to use in it, that is, if the size of the sample is more than thirty singles, and the standard deviation of society is known, and the society's sample is unknown.

## Conditions for using the Z test

In order to be able to use test 2, some conditions must be met, the most important of which are:

- The population variance is known (often unknown in the field of standard applications).
- The community variance is unknown, but the sample size is large ( $30 \gg n$ ).

To use 2 as a test for significance, it should be known that the arithmetic mean of the normal distribution is  $Z = 0$  and its standard deviation = 1. Then two values must be estimated:

$$Z_{\alpha} = \frac{\hat{\alpha} - \alpha}{S_{\alpha}}$$

$$Z_{\beta} = \frac{\hat{\beta} - \beta}{S_{\beta}}$$

The values of  $S_\alpha$  and  $S_\beta$  are given by the previous two relationships in the standard error part. The values  $\alpha$  and  $\beta$  are estimated by the model. The values of  $\alpha$  and  $\beta$  remain unknown to us and they express the values of the real parameters of the community as a whole. So we have to assume values for the community parameters. Of course, these values can be true or false, so they must be tested. It is customary to use the default value of zero for community parameters. It

Means in the regression parameter that there is no relationship between the variables X and Y in order to test the null hypothesis and the alternative hypothesis, so the two previous relationships become:

$$Z_\alpha = \frac{\hat{\alpha}}{S_\alpha} \quad , \quad Z_\beta = \frac{\hat{\beta}}{S_\beta}$$

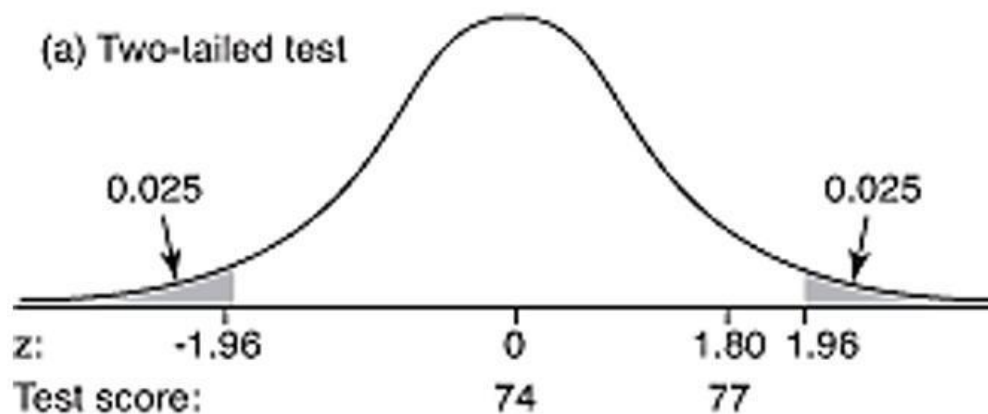
By assuming a normal distribution, and comparing the calculated Z values with their tabular values, we determine the probability of accepting the calculated values. If we test at the 5% level of significance and use assumptions:

$$H_0: \alpha = 0 \quad , \quad \beta = 0$$

$$H_1: \alpha \neq 0 \quad , \quad \beta \neq 0$$

The test here has two parties (positive and negative), so the acceptance of the alternative hypothesis is 95% of the normal distribution, and the rejection area is distributed to occupy an area of 5%, with 2.5% of each side. So the Rejection limits at probability are 0.025 and the tabular value of 2 corresponding to this value is 1.96 as shown in the figure.

Two tests at the same probability level (95%)



When testing at a significant level of 1%, the rejection areas are distributed by 0.005 for each party, and the corresponding tabular value of Z is 2.58, meaning that the rejection area became larger when tested on the basis of a smaller Significance level.

We compare the calculated Z with the tabular and we find two possibilities:

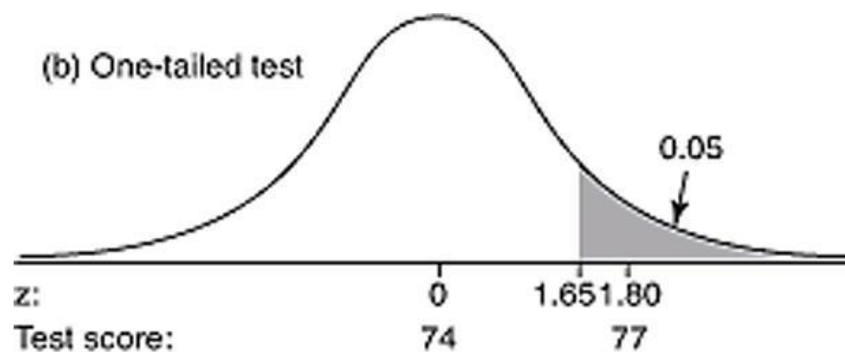
- $Z_{\text{calculated}} < Z_{\text{tabular}}$  (1.96 or 2.58) It means that the probability of imposing nothingness is small and that Because we estimated Z depending on the null hypothesis. So we accept the alternative hypothesis. And we say that the estimated parameter Statistically significant.
- $Z_{\text{calculated}} > \text{Tabular } Z$  (1.96 or 2.58) it means that the probability of imposing nothingness is greater Of 0.025 so we reject the alternative hypothesis. The estimated parameter is statistically insignificant.

.This is in case both parties are tested. But In many cases we have advance Information about the estimated parameter. If the function is a saving or consumption function, for example, we can expect the parameter signal and thus use the one-sided test, so the assumptions in the savings function are as follows:

$$H_0: \alpha = 0 \quad , \quad \beta = 0$$

$$H_1: \alpha < 0 \quad , \quad \beta > 0$$

In this case, the moral values are confined to one party. The tabular values of Z vary, so their value becomes at the level of significance 0.05 equal to 1.645 and at the level of significance 0.001 equal to 2.33 as follows:



## 1- ( one tailed z – test )

### Example

The population of all verbal GRE scores are known to have a standard deviation of 8.5 . The UW Psychology department hopes to receive applicants with a verbal GRE scores over 210 . This year , the mean verbal GRE scores for the 42 applicants was 212.79 . Using a value of  $\alpha = 0.05$  is this new mean significantly greater than the desired mean of 210 ?

For this example , the mean under the null hypothesis is  $\mu_{hyp} = 210$  , the population standard deviation is  $\sigma_x = 8.5$  , and the observed mean  $\bar{x}$  is = 212.79 .

The standard error of the mean is therefore :

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{8.5}{\sqrt{42}} = 1.31$$

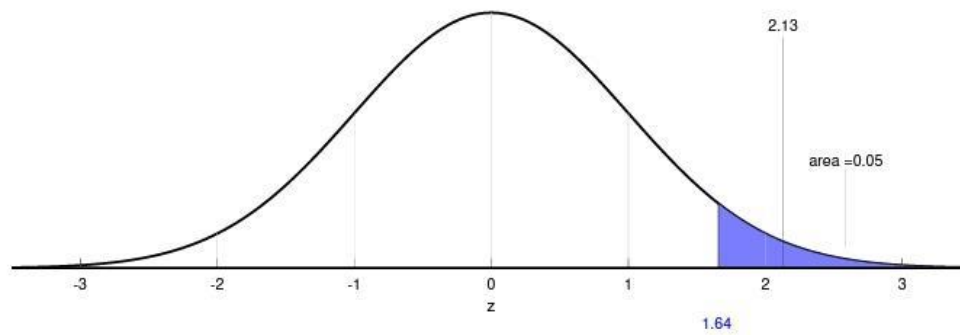
To find the probability of finding a mean above 212.79 we convert our observed mean,  $\bar{x}$  , to A z-score:

$$z = \frac{(\bar{x} - \mu_{hyp})}{\sigma_{\bar{x}}} = \frac{(212.79 - 210)}{1.31} = 2.13$$

This will be a one tailed test because we're only rejecting H0 if our observed mean is Significantly larger than 210. To make our decision we need to find the critical value of z, Which is the z for which the area above is 0.05. Looking at our z-table for  $\alpha=0.05$ :

z	Area between mean and z	Area beyond z
⋮	⋮	⋮
1.62	0.4474	0.0526
1.63	0.4484	0.0516
<b>1.64</b>	<b>0.4495</b>	<b>0.0505</b>
1.65	0.4505	0.0495
1.66	0.4515	0.0485
⋮	⋮	⋮

You can see that the critical value of z is 1.64:



Our observed value of z is 2.13 which is greater than critical value of 1.64. We therefore Reject  $H_0$  - Equivalently, we can calculate the p-value for our observed mean and compare it to alpha.

For this one-tailed test, the p-value is the area under the normal distribution above our Observed value of z. From the z-table:

z	Area between mean and z	Area beyond z
⋮	⋮	⋮
2.11	0.4826	0.0174
2.12	0.4830	0.0170
<b>2.13</b>	<b>0.4834</b>	<b>0.0166</b>
2.14	0.4838	0.0162
2.15	0.4842	0.0158
⋮	⋮	⋮

You can see that our p-value is  $p = 0.0166$ .

Our p-value is less than alpha (0.05). If the null hypothesis is true, then the probability of Obtaining our observed mean or greater is less than 0.05. We therefore reject H0 and state That (in APA format):

The verbal GRE scores of applicants (M = 212.79) is significantly greater than 210,  $z=2.13$ ,  $P= 0.0166$

We could also use the 'pnorm' function R to calculate this p-value. Remember, we need to Divide the population standard deviation (  $\sigma = 8.5$ ) by the square root of n ( $\sqrt{42}$ ):

## 2- (two tailed z-test)

### Example

Suppose you start up a company that has developed a drug that is supposed to increase IQ. You know that the standard deviation of IQ in the general population is 15. You test your drug on 36 patients and obtain a mean IQ of 97.65. Using an alpha value of 0.05, is this IQ significantly different than the population mean of 100?

To solve this, we first calculate the standard error of the mean: 2: (two tailed z-test) Suppose you start up a company that has developed a drug that is supposed to increase IQ. You know that the standard deviation of IQ in the general population is 15. You test your Drug on 36 patients and obtain a mean IQ of 97.65. Using an alpha value of 0.05, is this IQ

Significantly different than the population mean of 100?

To solve this, we first calculate the standard error of the mean:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{36}} = 2.5$$

And then convert our observed mean to a Z-score:

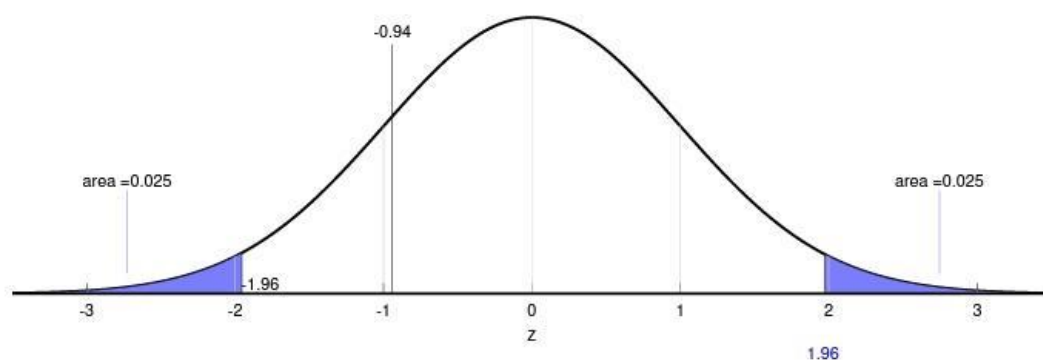
$$z = \frac{(\bar{x} - \mu_{hyp})}{\sigma_{\bar{x}}} = \frac{(97.65 - 100)}{2.5} = -0.94$$



We then compare our observed value of  $z$  to the critical values of  $z$  for  $\alpha = 0.05$ . We Are looking for a significant difference, so this will be a two-tailed test. We reject the Null hypothesis if our observed mean is either significantly larger or smaller than 100. Our Critical values of  $z$  are therefore the two values that span the middle 95% of the area under The standard normal distribution. This means that the areas in each of the two tails is = 0.025:

$z$	Area between mean and $z$	Area beyond $z$
$\vdots$	$\vdots$	$\vdots$
1.94	0.4738	0.0262
1.95	0.4744	0.0256
<b>1.96</b>	<b>0.4750</b>	<b>0.0250</b>
1.97	0.4756	0.0244
1.98	0.4761	0.0239
$\vdots$	$\vdots$	$\vdots$

Which corresponds to a critical value of  $z = 1.96$ .

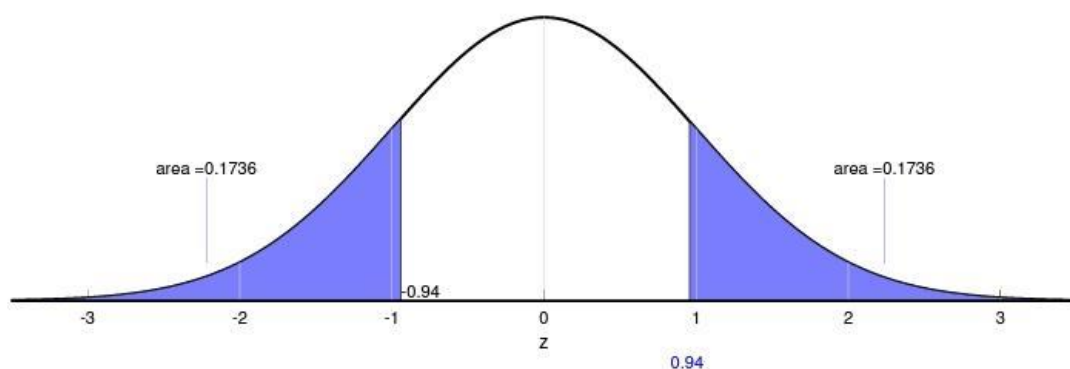


The rejection region contains values of  $z$  less than -1.96 and greater than 1.96. Our observed Value of  $z$  falls outside the rejection region, so we fail to reject  $H_0$  and conclude that our Drug did not have a significant effect on IQ.

To calculate the p-value we need to find the area under the standard normal distribution Beyond our observed value of z and double it. This is because for a two-tailed test we want The probability of obtaining our observed value or more extreme in either direction. This Makes sense if you think about what happens if the observed value of z falls exactly on the Critical value (1.96 in this example). The area beyond the observed value of z in both the Positive direction and the negative direction will add up to alpha (0.05).

z	Area between mean and z	Area beyond z
-0.96	-0.3315	0.8315
-0.95	-0.3289	0.8289
<b>-0.94</b>	<b>-0.3264</b>	<b>0.8264</b>
-0.93	-0.3238	0.8238
-0.92	-0.3212	0.8212
⋮	⋮	⋮

For this example, the area above  $z = 0.94$  plus the area below  $z = -0.94$  is  $0.1736 + 0.1736 = 0.3472$



Since our p-value of 0.3472 is greater than 0.05, we fail to reject  $H_0$  and state that:

The IQ of superb drug patients ( $M = 97.65$ ) is not significantly different than 100,  $z = -0.94$ .  $P = 0.3472$ .

To do this in R, we need to be sure to double our p-value since this is a two-tailed test.



**Mohamed abu libdah**

**Statistician and Data Analyst at the  
NCI - Cairo University - Egypt**