

Title: Running a Manual Z-test to calculate comparisons and probabilities

Target: On completion of this worksheet you should be able to identify key assumptions of, and conduct, a manual Z-test for normal distribution

Key Information to start:

The Z-test is a statistical test which compares the means of two populations.

Assumptions

The Data needs to be normally distributed. (See sheet S9 for more information on this).

Works only for large data sets, this is usually when $n \geq 30$, otherwise if $n < 30$ then a T-test is required.

When finding the probabilities, the bell graph at the mean is symmetrical

General Information

A Z-test uses an assumed value which is within the limits of the data to calculate what is called a Z-score. This value is called our standardised random variable. Our Z-score is calculated by using the following

$$\text{formula: } Z = \frac{X - \bar{X}}{\sigma}$$

Where:

- \bar{X} = Standardised random variable
- \bar{X} = Mean of the data
- σ = Population standard deviation

Mean of the data

The mean of the data is given by

$$\bar{X} = \frac{\sum(X_i)}{n}$$

where

- X_i is the numbers in the data
- n is the data size

Population Standard Deviation

The population standard deviation is given by

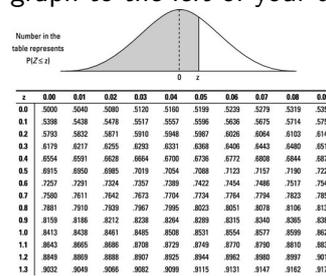
$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$$

where

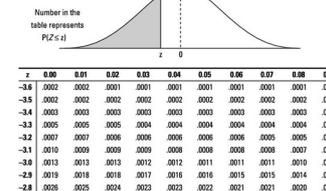
- σ = population standard deviation
- X_i = numbers in the data
- \bar{X} = mean of the data
- n = data size

Using the table

When looking at the Z-table you must check which way the data you are using is represented under the normal distribution. If the values in the table increase to one then you are looking at all the data under the graph to the left of your desired point



While if the data goes to 0, then you are looking at the data to the right of your desired point



Bare in mind, this is only for the positive point, negative will require 1–your value. If we are looking at a two tailed test and finding the value of 95%, we will use the value of 1.96 because that corresponds to 97.5% in the tables.

Calculating probabilities

We can use the reverse to find the value or probability using the confidence value. Firstly find whether its a one tailed or two tailed test. If it's one tailed, the alternative hypothesis, H_1 , will contain either less than or greater than sign to indicate whether we say it has got lower or higher than before, where two tailed will have a non-equal to sign, meaning it can be greater or less than with no direction in mind. Non-equal to, means it cannot be within the acceptance region of the null on either side of its bell graph, and significance level will have to change as a result of it. If its two tailed with 95% confidence level, we split the remainder (which is half of 5), which is 2.5 and add it onto the original confidence (now 97.5) as we are missing on both ends and not just one, and find the value that corresponds to the percentage.

Example 1

In a government organisation, the mean basic salary is £5000. How much will be the Z-score of employees whose salary is £3000, if standard deviation of the population is £850?

Answer

From the question we are given

$$\sigma = 850$$

$$\bar{X} = 5000$$

$$\text{and our } X = 3000$$

$$\text{So by using our formula: } Z = \frac{X - \bar{X}}{\sigma}$$

We get

$$Z = \frac{3000 - 5000}{850} = -2.353$$

This means that our sterling is 2.353 standard deviations below the mean. If you plot this on a normal distribution graph, it may be clearer to understand where the £3000 lies in respect to £5000.

Example 2

A rental car company claims the mean time to rent a car on their website is 60 seconds with a standard deviation of 30 seconds. A random sample of 36 customers attempted to rent a car on the website. The mean time to rent was 75 seconds. Is this enough evidence to contradict the company's claim?

Answer

Firstly we find the standard error of the mean for a sample size of 36: $\frac{30}{\sqrt{36}} = 5$

Now we check how many standard errors are there between the observed sample mean and the population mean $\frac{75 - 60}{5} = 3$

Now using the normal distribution graph/ table, we check how much is between $-\infty$ to 3. This gives us 99.865%.

This shows we would only expect to see a sample mean this extreme less than 1% of the time, therefore we can conclude that this is enough evidence to question the company's claim of 60 seconds.

Example 3

With a mean of 10 and a standard deviation of 1.5, find the values where the 95% confidence level lies.

Answer

Since we are looking for where the intervals lie, we want above and below the mean, therefore we use a two tailed scenario. $1 - 0.95 = 0.05$ and half of 0.05 is 0.025 so in the table we look for the critical value to be 0.975. This value corresponds to 1.96. So $P(a \leq Z \leq b)$ and to calculate a and b we get $\frac{Z - \bar{X}}{\sigma} = \pm \text{critical} \rightarrow \frac{Z - 10}{1.5} = \pm 1.96$

$$Z = 10 \pm (1.96 \times 1.5) \quad Z = 7.06 \text{ and } Z = 12.94$$

Therefore $a = 7.06$ and $b = 12.94$ so we are confident 95% of our data lies between 7.06 and 12.94 when mean is 10 and standard deviation is 1.5

Example 4

Using a standard normal distribution $X \sim N(0,1)$, find the probabilities of less than 0.8 and more than -0.1

Then a factory produces boxes of stock a day on average at 10 and the variance is 9, and we need to find the probabilities that on a random day it will produce more than 16, and another random day less than 7.

Answer

For the first part we know the distribution is $X \sim N(0,1)$ so this does not need to change. We are finding the probability of less than 0.8, so $P(Z < 0.8)$, this is the form we want it in already so we look in our table (for this look at the first one). Look where 0.8 is and the probability is 0.7881

For more than -0.1 we can use the top graph again but need to change how we do it. Since it is a minus and we are looking at greater than the half mark, we need to manipulate it so we can find the probability. $P(Z > -0.1) = P(Z < 0.1)$ and when looked in the table we get the probability to be 0.5398

Now for the factory, we are not in a fully standardised normal distribution, so we have to manipulate our calculations. We are told it is currently $X \sim N(10, 3^2)$. We've put it into 3^2 instead of 9 as it will be easier for the calculation (We want the standard deviation and not the variance). Looking back to our General Information on page 1, we are told the formula to help calculate the probability

$$Z = \frac{X - \bar{X}}{\sigma}$$

Firstly write what our probability will look like. We want more than 16 so $P(Z > 16)$?

$$P(Z > 16) = P\left(Z > \frac{16 - 10}{3}\right) = P(Z > 2) \\ = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$$

So the probability of the factory producing more than 16 on a random day is 0.0228

Finally, we want less than 7, doing the same as before we get

$$P(Z < 7) = P\left(Z < \frac{7 - 10}{3}\right) = P(Z < -1) \\ = 1 - P(Z < 1) = 1 - 0.8643 = 0.1357$$

So the probability of getting less than 7 is 0.1357. Remember to read the direction you want the numbers to fall, which table you are looking at and you want to find the standard deviation if it is not a fully standardised distribution.

Exercises

- Using the Normal distribution given by $X \sim N(174, 92.16)$ for the height of teachers in a particular school, find the Z for teachers over the height of 180cm. Explain what this means.
- A machine is set to produce rods of length L cm, with a mean of 10.6cm and a variance of 0.8cm^2 . Find where 95% of lengths of rods are considered.
- A survey claims that 9 out of 10 doctors recommend aspirin for their patients with headaches. To test this claim, a random sample of 100 doctors is obtained. Of these 100 doctors, 82 indicate that they recommend aspirin. Is this claim accurate? Use $\alpha = 0.05$
- Find the probability that a random student gets higher or the same as 1 more mark than the average of 65 which in this exam has a standard deviation of 4.

- $\frac{180 - 174}{\sqrt{92.16}} = 0.63$
This means that 180cm is 0.63 standard deviations above the mean of 174cm of the distribution.
95% in two tail gives us 1.96
95% of the rods lengths are between (9.032, 12.168)
- Null hypothesis is 0.9, Alternative hypothesis is not equal to 0.9
alpha for 2 tailed test will be 0.025 in each tail, giving a critical value of 1.96.
If Z is greater than 1.96 or less than -1.96, we reject null hypothesis.
Using formula $Z = -2.667$, therefore we reject the null hypothesis as $p > 0.05$
- $X \sim N(65, 4^2)$
 $P(Z < 66) = P\left(Z < \frac{66 - 65}{4}\right) = P(Z < 0.25) = 1 - P(Z > 0.25) = 1 - 0.5987 = 0.4013$

Answers