

Inferential statistics

The Statistical Test of a Hypothesis

Inferential statistics are the set of statistical tests researchers use to make inferences (decisions) about data. These statistical tests allow researchers to make inferences because they can show whether an observed pattern is due to intervention or chance.

There is a wide range of statistical tests. The choice of which statistical test to utilize relies upon the structure of data, the distribution of the data, and variable type.

For examples,

1-we suspect there might be a non-zero correlation between two variables (e.g. sunspot numbers and annual rainfall totals in Reading).

2-A new medicine you think might work.

3-A way of teaching you think might be better.

4-If I (give exams at noon instead of 7) then (student test scores will improve).

To test these hypothesis, we examine some data and correlation to make true/false decisions about hypotheses based on the evidence provided by the sample of data.

There are many different types of tests in statistics like t-test, Z-test, chi-square test, T-test, ANOVA test, binomial test, one sample median test etc.

In general, if the data is normally distributed, parametric tests should be used. If the data is non-normal, non-parametric tests should be used.

The Parametric tests are used if the data is normally distributed. A parametric statistical test makes an assumption about the population parameters and the distributions that the data came from.

These types of test include T-tests, Z-tests and ANOVA tests, which assume data is from normal distribution.

The decision-making procedure in classical statistical inference proceeds by the following steps:

Z-test

A z-test is a statistical test used to determine whether two population means are different when the variances are known and the sample size is large

In z-test mean of the population is compared. The parameters used are population mean and population standard deviation.

Z-test is used to validate a hypothesis that the sample drawn belongs to the same population.

Ho: Sample mean is same as the population mean (Null hypothesis)

Ha: Sample mean is not same as the population mean (Alternate hypothesis)

$$Z = \frac{(\bar{x} - \mu)}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

where,

\bar{x} =sample mean,

μ =population mean,

σ = population standard deviation.

If *Z value* is less than critical value accepts null hypothesis

while if *Z value* is greater than critical value reject null hypothesis.

Examples (One Sample Z Test)

A principal at a certain school claims that the students in his school are above average intelligence. A random sample of 40 students scores have a **mean** score of 112.

Is there sufficient evidence to support the principal's claim?

The **mean** population is 100 with a **standard deviation** of 15.

solution:

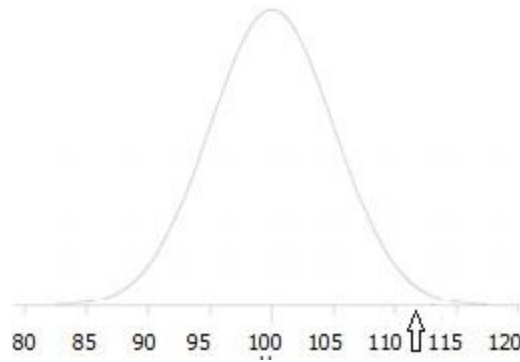
Step 1: State the Null hypothesis. The accepted fact is that the population mean is 100, so: $H_0: \mu=100$.

Step 2: State the Alternate Hypothesis. The claim is that the students have above average scores, so:

$H_1: \mu > 100$.

The fact that we are looking for scores "greater than" a certain point means that this is a one-tailed test.

Step 3: Draw a picture to help you visualize the problem.



Step 4: State the **alpha level**. If you aren't given an alpha level, use 5% (0.05): ($\alpha = 5\%$)

Step 5: Find the rejection region area (given by your alpha level above) from the z-table.

An area of 5% (0.05) is equal to a z-score of 1.96

Step 6: Find the test statistic using this formula:

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

$$Z = \frac{(112-100)}{\left(\frac{15}{\sqrt{40}}\right)}$$

$$Z = 5.05$$

Step 7: If Step 6 is greater than Step 5, reject the null hypothesis.

If it's less than Step 5, you cannot reject the null hypothesis.

In this case, it is greater ($5.05 > 1.96$), so you can reject the null hypothesis.

How to get the value of Z from the statistical tables?

We know that the value of Z corresponding to Confidence level is 95% (0.95) equals 1.96:

1- Confidence Level / 2

In the case of 95 percent, the first step becomes $0.95 / 2 = 0.475$

2: We move to the Z-test table,

3: We find that the corresponding number in the same row is 1.9 and in the same column 0.06, then the value of $Z = 1.96$.

Another example is to assume that the Confidence Level is at 99 percent, as it came in one of the questions:

$0.99 / 2 = 0.495$

Thus, the value of $Z = 2.58$, where it is located in row 2.5 and 0.08 of the column

T-test

In statistics, t-tests are a type of hypothesis test that allows you to compare means.

The one sample t-test is a statistical procedure used to determine whether a sample of observations could have been generated by a process with a specific mean.

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

Hypotheses:

There are two kinds of hypotheses for a one-sample t-test: the null hypothesis and the alternative hypothesis. The alternative hypothesis assumes that some difference exists between the true mean (μ) and the comparison value (μ_0), whereas the null hypothesis assumes that no difference exists.

- The null hypothesis (H_0) assumes that the difference between the true mean (μ) and the comparison value (μ_0) is equal to zero.
- The two-tailed alternative hypothesis (H_1) assumes that the difference between the true mean (μ) and the comparison value (μ_0) is not equal to zero.

Example:

A company wants to improve sales. Past sales data indicate that the average sale was \$100 per transaction. After training the sales force, recent sales data (taken from a sample of 25 salesmen) indicates an average sale of \$130, with a standard deviation of \$15.

Did the training work? Test your hypothesis at a 5% alpha level.

Step 1: Write the null hypothesis statement:

The accepted hypothesis is that there is no difference in sales, so:

$H_0: \mu = \$100.$

Step 2: Write the alternate hypothesis:(there *is* a difference):
(that the mean sales increased), so: $H_1: \mu > \$100.$

Step 3: calculate the test statistic:

1. The sample mean(\bar{x}). This is given in the question as \$130.
2. The population mean(μ). Given as \$100 (from past data).
3. The sample standard deviation(s) = \$15.
4. Number of observations(n) = 25.

Step 4: Insert the items from above into the t score formula.

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

$$t = (130 - 100) / ((15 / \sqrt{25}))$$

$$t = (30 / 3) = 10$$

This is your calculated t-value.

Step 5: Find the t-table value. You need two values to find this:

1. The alpha level: given as 5% in the question.
2. The degrees of freedom, which is the number of items in the sample
(N-1) : (25 - 1 = 24)

Look up 24 degrees of freedom in the left column and 0.05 in the top row:

The intersection is 1.711. This is your one-tailed critical t-value.

What this critical value means is that we would expect most values to fall under 1.711. If our calculated t-value (from Step 4) falls within this range, the null hypothesis is likely true.

Step 5: Compare Step 4 to Step 5. The value from Step 4 does not fall into the range calculated in Step 5, so we can reject the null hypothesis. The value of 10 falls into the rejection region (the left tail).

one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965
18	1.33	1.734	2.101	2.552	2.878	3.61	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.85
21	1.323	1.721	2.08	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.5	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.06	2.485	2.787	3.45	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.66
30	1.31	1.697	2.042	2.457	2.75	3.385	3.646
60	1.296	1.671	2	2.39	2.66	3.232	3.46
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291