Statistical Hypothesis Testing

Fatima Batool

Contents

Chapter 1

Test of Significance

Assuming we are interested in estimation of an unknown population parameter. We can do point estimates what, or interval estimates. Hypothesis testing is a formal inferential framework where we can test certain hypothesised value for the parameter of interest. The subsequent section introduces hypothesis testing procedure.

1.1 Hypothesis Testing

Example 1.1.1. Average temperature in a region.

We want to test is the average annual temperature for a weather location differs from its expected average.

Table 1.1: Snippet of the weather data:Average temperature Celsius recorded per hour of a single weather location. Source: NOAA

Data	Location	Temperature
2020-01-01	AD	4.236
2020-01-02	AD	3.875
2020-01-03	AD	4.764
2020-01-04	AD	4.556
2020-01-05	AD	4.764

Example 1.1.2. Test that average BMI of Covid-19 patients aged 40-45 differ from non-Covid patients.

State Hypothesis: formally formulate a question about data. The *null hypothesis* for a single parameter say, population mean μ states that the effect is due to chance alone. It might be stated as:

$$H_0 : \mu = \mu_0.$$

Table 1.2: 0:Male, 1 Female.

No.	Age	Gender	BMI	Covid Death
1	49.28	0	39.36	1
2	41.32	1	41.70	0
3	45.08	0	42.20	0
4	50.04	1	37.44	0
5	44.60	1	38.96	0
6	56.82	0	25.16	0
7	43.26	0	31.62	0
8	45.22	1	26.24	0
9	48.30	0	32.78	0
10	62.90	0	38.72	0
11	50.46	1	40.10	0
12	56.72	0	39.30	0
13	46.18	0	25.66	0
14	57.72	0	36.62	0
15	48.28	0	44.12	0

This is where we assume nothing is going on. Other null hypothesis which one might be interested in testing could be: $H_0 : \mu \leq \mu_0.$, or $H_0 : \mu \geq \mu_0$. The *alternative hypothesis* states a claim to be tested. It might be:

$$H_a : \mu \neq \mu_0,$$
$$H_a : \mu > \mu_0,$$
$$H_a : \mu < \mu_0.$$

Decide Level of Significance: The probability of rejecting a true null hypothesis. There are four possible outcome of a decision process, presented in the table below:

Table 1.3: Decision table for Hypothesis Testing

Decision	Truth	
	H_0 is TRUE	H_0 is FALSE
Accept H_0	correct	Type II Error
Reject H_0	Type I Error	correct

Ideally, we want to reduce probabilities of the incorrect decisions. The probability of committing Type I error is known as level of significance. It representing the maximum error allowed to reject the null hypothesis given that when it is in fact true based on the given data at hand, and is denoted by Greek alphabet α. Common error probabilities for the null are 0.05, and 0.01.

Briefly, the probability of wrongly accepting a false alternative hypothesis is known as Type II error and is denoted by Greek alphabet β. The compliment of Type II error is*power of a test* defined as: (1-β).

Test Statistics under the Null:Decide and calculate the appropriate test statistics. We need to specify the probability distribution of the data. Suppose data is normally distributed with mean μ, variance σ^2.

$$X_i \sim N(\mu, \sigma^2).$$

Since σ is known we use z-statistics:

$$z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1).$$

If σ^2 is not known we use the t-statistics:

$$t = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{st}(n - 1).$$

Construct critical region: a set of values for test statistics for which null is rejected. To make a decision regarding acceptance or rejection of *null hypothesis* of no difference of population mean from μ_0, construct a critical region. If the test statistics falls in the critical region reject null. For a two-tail test reject null:

$$z > |Z_\alpha|, \text{or}(z < -Z_\alpha, z > Z_\alpha).$$

For a Left-tail test reject null:

$$z < -Z_\alpha.$$

For a Right-tail test reject null:

$$z > Z_\alpha.$$

The regions are shown in Figure 1.1.

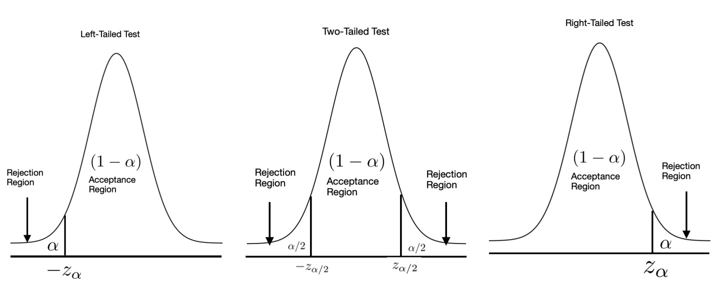

Figure 1.1: The critical region for z test statistics

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Figure 1.2: Cumulative standard normal distribution Table

Data: Suppose we want to test the population mean against $\mu_0 = 1.2$, which sets: $H_0 : \mu = 1.2$. For these data suppose we have sample of 100 observation, and $\sigma^2 = 4$, so that under the null hypothesis z-statistics is calculated as:

$$z = \frac{\sqrt{100}(10 - 1.2)}{\sqrt{4}} = 44.$$

where, for $\alpha = 0.05$, and two-tailed test, the decision rule is: Reject the null hypothesis if $z \leq -1.96$ or $z \geq 1.96$. These critical values are read from Table **??**. Since the z-statistics falls in the rejection region, we reject the null hypothesis. See the R code file for Hypothesis testing for Example 1.1.1.
In the critical region approach the hypothesis is accepted or rejected based on location of test statistics relative to rejection region.

1.2 Significance testing using P-value

The test of significance can also be performed using a P-value without use of a critical value(s).

Definition 1.2.1. The P-value is the probability value of observing the test statistics as or more extreme than the one calculated from the sample data under the assumption that null is true.

The steps for the Hypothesis testing are:

1. State the Null and Alternative hypothesis
2. Decide the Type I error probability
3. Identify the distribution of the data and choose the appropriate test Statistics. Calculate test statistics
4. Choose the P-value formula depending upon alternative hypothesis (one-tail or two tail). Calculate the P-value as appropriate for your hypothesis.
5. Decision

P-value calculation: Let P-value is mathematically denoted by p. Assume a z test statistics is appropriate statistics, the P-value formulae for three alternative hypothesis are given below.

For a Right-tailed rejection region:

$$p = P(Z > z).$$

For a Left-tailed rejection region:

$$p = P(Z < z).$$

For a Two-tailed rejection region:

$$p = 2P(Z > |z|).$$

Decision: The decision whether to accept null hypothesis is made based on a predetermined significance level often set at $\alpha = 0.05$. The table below present decision criterion used to determine the statistical significance of null hypothesis:

Oftentimes, researchers make decision about null hypothesis with out specifying the significance level. The interpretations are:

1. If $P - value$ less than 0.01, strong evidence against null hypothesis.
2. If $P - value$ is between 0.01 and 0.05, some evidence against null hypothesis.
3. If $P - value$ greater than 0.05, insufficient evidence against null hypothesis.

Observed	Decision
P-value $\leq \alpha$	Reject H_o
P-value $> \alpha$	Fail to Reject H_o

1.3 Understanding the P-value

> Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.
>
> *[Wasserstein, 2016]*

1. Small P-value does not equate to practical or scientific significance
2. P-value is not probability of null hypothesis being true or false
3. Significant P-value does not implies cause and effect

1.4 Summary

P-value is the smallest level of significance at which you can reject the null hypothesis. Some researchers define it as exact level of significance and interpret them as they gives total rejection region. However, this dichotomized threshold thinking of significant non-significant should not be focus. The interpretation of p-values should be combined with the sizes of effect estimates, confidence limits, study design, potential biases, and other statistical analysis in light of the context of the research question. Reading on this topic: [Greenland et al., 2016].

Bibliography

Sander Greenland, Stephen J Senn, Kenneth J Rothman, John B Carlin, Charles Poole, Steven N Goodman, and Douglas G Altman. Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology*, 31(4):337–350, 2016.