Hypothesis Testing

Concepts

- Statistical hypothesis   statement about a set of parameters of a population. It is unknown whether it is true or false

Null hypothesis   a specific statement being tested

Alternative hypothesis   the opposite of the null hypothesis

Null hypothesis : "The drug has no effect"

Alternative hypothesis : "The drug has effect on the disease"

- We speak of either accepting the (null) hypothesis or rejecting the (null) hypothesis

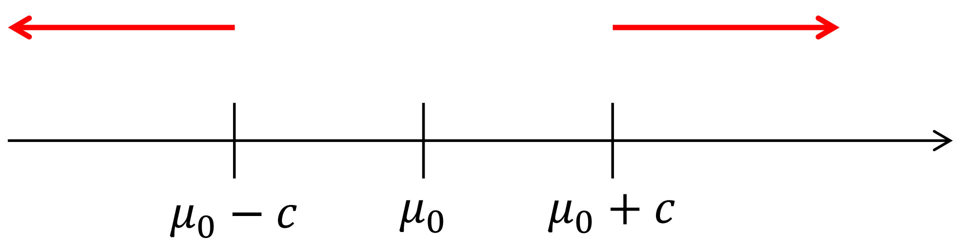
- Accepting the hypothesis (fail to reject the hypothesis) the statistical data is consistent with it. **Does not mean that it is true**

  "The drug has no effect"

- Rejecting the hypothesis the statistical data is unlikely to be consistent with it. **Does not mean that it is false**

  "The drug has effect on the disease"

- A critical region is defined such that if the statistic falls inside the region, the event is regarded as so unlikely to be consistent with the hypothesis that the hypothesis is rejected. On the other hand, if the statistic falls outside the region, the event is consistent with the hypothesis and it is accepted



The critical region is the red part

- Type I error occurs if the test incorrectly rejects the hypothesis when it is actually correct
- Type II error occurs if the test incorrectly accepts the hypothesis when it is actually false

- Level of significance α (also called significant level) is a user defined number

  Usual values for α = 0.1, 0.05, 0.005

  The setting of α is application dependent

Hypothesis Test of the Mean of a Normal Population Assuming Known Variance

For a normal distribution, if the sample size n is large, we can assume that the variance is approximately known.

Suppose that $X_1, \dots, X_n$ is a sample of size n from a normal distribution having an unknown mean μ and a known variance σ^2

Null hypothesis H_0: $\mu = \mu_0$

Alternative hypothesis H_1: $\mu \neq \mu_0$

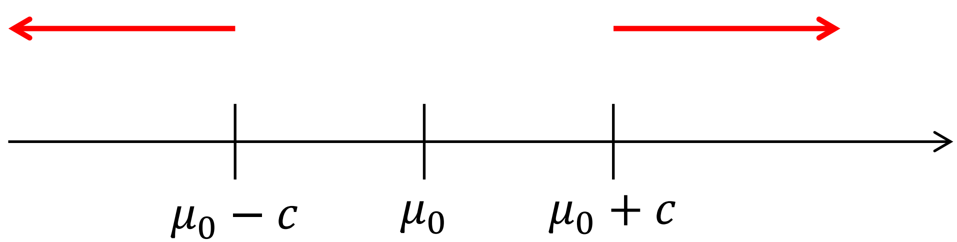
μ_0 is some constant of your choice

$$\bar{X} = \sum_{i=1}^{n} X_i / n$$

is an unbiased estimator of μ

It is reasonable to accept H_0 if $\bar{X}$ is not too far from μ_0. Thus we define the critical region

$$C = \{X_1, \ldots, X_n : |\bar{X} - \mu_0| > c\}$$



We accept H_0 if $|\bar{X} - \mu_0| \leq c$ and reject H_0 if $|\bar{X} - \mu_0| > c$

If we desire that the test has significant level α

$$P(Type\ I\ error) = P\{|\bar{X} - \mu_0| > c\} = \alpha \qquad (1)$$

This probability is computed assuming that $\mu = \mu_0$. Applying the central limit theorem (see the estimation lecture), we know that $\bar{X}$ will be normally distributed with mean μ_0 and variance σ^2/n. Converting to standard normal distribution

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

(1) is rewritten as

$$P\left\{|Z| > \frac{c\sqrt{n}}{\sigma}\right\} = \alpha$$

Since Z is a standard normal random variable,

$$P\{Z > z_{\alpha/2}\} = \frac{\alpha}{2}$$

Thus

$$\frac{c\sqrt{n}}{\sigma} = z_{\alpha/2} \Rightarrow c = \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$$

The critical region $|\bar{X} - \mu_0| > c$ becomes

$$|\bar{X} - \mu_0| > \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$$

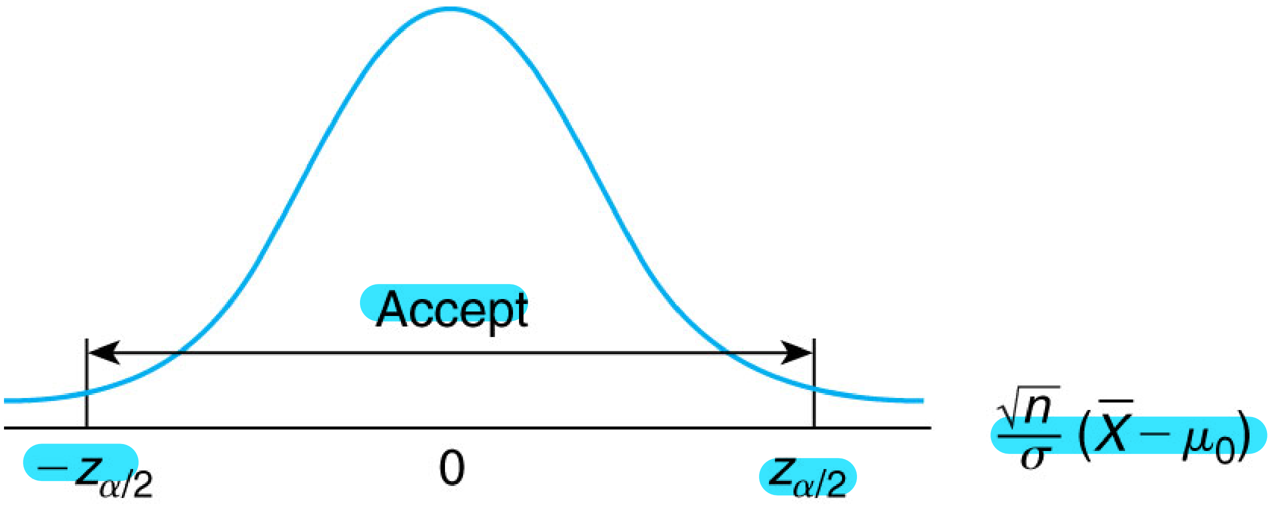
or

$$\frac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} > z_{\alpha/2}$$

The hypothesis test is designed as follows:

Reject H_0 if $\frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| > z_{\alpha/2}$

Accept H_0 if $\frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| \leq z_{\alpha/2}$



Example

Suppose a signal with value μ is sent from A to B and the signal received at B is normally distributed with mean μ and standard deviation 2 due to Gaussian noise. Test the hypothesis $\mu = 8$ if the same signal value is independently sent five times and the average value received is $\bar{X} = 9.5$. If we are testing at 5% level of significance, will we accept or reject the hypothesis? α = 0.05

5 times

= 9.5 - 8

$$Z = \frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| = \frac{\sqrt{5}}{2}(1.5) = 1.68$$

s.d. = 2

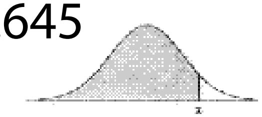
$z_{.025} = 1.96$, the critical region is $|Z| > 1.96$. As 1.68 does not lie on the critical region, the hypothesis is accepted (note: just not enough evidence to conclude otherwise. Not whether it is true or false)

Z Distribution Table

1-0.05/2 = 0.975 -> 1.96

1-0.1/2 = 0.95 -> 1.645

Tables of the Normal Distribution



Probability Content from -oo to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Example

If the level of significance is 10%, will we accept or reject the hypothesis?

Z = 1.68

$z_{.05} = 1.65$. The critical region is $|Z| > 1.65$. As 1.68 lie on the critical region, the hypothesis is rejected

Note that whether it is accepted or rejected depends on the application, in this case, how confident that we are subjectively that the mean is 8. If we are quite confident, we set the level of confidence to a small value (we only reject if things are seriously out of our expectation), and vice versa

Accepting and Rejecting using p value

- Type I error occurs if the test incorrectly rejects the hypothesis when it is actually correct
- P(Type I error) is called the p-value

- To accept the hypothesis, p-value should be larger than the level of significance α
- To reject the hypothesis, p-value should be smaller than α

- It enables us to have a sense of how strong is the hypothesis without deciding on a particular level of significance

Example

In the previous example,

Z = 1.68

$$\frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| = \frac{\sqrt{5}}{2}(1.5) = 1.68$$

p-value $= P\{|Z| > 1.68\} = 2P\{Z > 1.68\} = 2(1 - 0.9535) =$ 0.093

It is the probability of committing a Type I error.

$\alpha = 0.05$

As 0.093 > 0.05, we accept the hypothesis at level of significance 0.05

As 0.093 < 0.1, we reject the hypothesis at level of significance 0.1

$\alpha = 0.1$

Example

Suppose the average of the 5 values received is $\bar{X} = 8.5$, compute the p-value, and hence determine whether the hypothesis will be accepted at level of significance 0.05 and 0.1?

n = 5

8.5 - 8 = 0.5

$$Z = \frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| = \frac{\sqrt{5}}{2}(0.5) = 0.559 \approx 0.56$$

s.d. = 2

$$\text{p-value} = P\{|Z| > 0.56\} = 2P\{Z > 0.56\} = 2(1 - 0.7123)$$
$$= 0.5754$$

$0.5754 > 0.1$. Hence the hypothesis will be accepted at both level of significance 0.05 and 0.1.

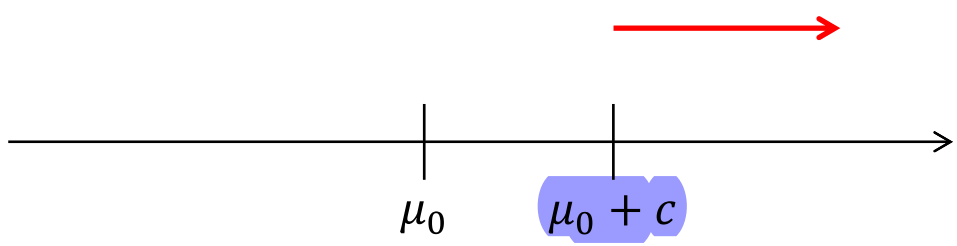
One-sided Test

Sometimes we wish to test the following hypothesis

Null hypothesis	$H_0\colon\ \mu = \mu_0$
Alternative hypothesis	$H_1\colon\ \mu > \mu_0$

Then the critical region is

$$C = \{X_1, \ldots, X_n : \bar{X} - \mu_0 > c\}$$



The hypothesis test is designed as follows:

Reject H_0 if $\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) > z_\alpha$

Accept H_0 if $\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) \leq z_\alpha$

Example

Suppose $\bar{X} = 9.5$ and we are considering whether there is enough evidence that the signal is larger than 8, will we reject the hypothesis at level of significance 0.05?

$$Z = \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) = \frac{\sqrt{5}}{2}(1.5) = 1.68$$

$z_{0.05} = 1.65$. As $Z > 1.65$, the hypothesis will be rejected.

There is enough evidence to conclude that the signal is larger than 8.

Example

Compute the one-sided p-value in the above example. Confirm your conclusion.

p-value $= P(Z > 1.68) = 1 - 0.9535 = 0.0465$

0.0465 < 0.05, hence the hypothesis is rejected.

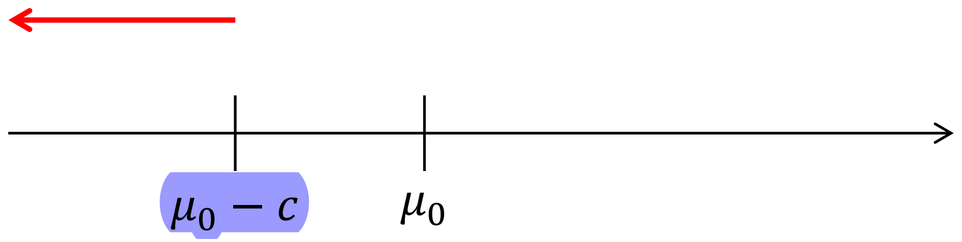
There is enough evidence to conclude that the signal is larger than 8.

Similarly, we may wish to test the following hypothesis

Null hypothesis	$H_0\!: \ \mu = \mu_0$
Alternative hypothesis	$H_1\!: \ \mu < \mu_0$

Then the critical region is

$$C = \{X_1, \ldots, X_n : \bar{X} - \mu_0 < c\}$$



The hypothesis test is designed as follows:

Reject H_0 if $\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) < -z_\alpha$

Accept H_0 if $\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) \geq -z_\alpha$

Example

Suppose $\bar{X} = 9.5$ and we are considering whether there is enough evidence that the signal is smaller than 8, will we reject the hypothesis at level of significance 0.05?

Of course not.

$$Z = \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) = \frac{\sqrt{5}}{2}(1.5) = 1.68$$

As $-z_{0.05} = -1.65$, it is clear that $Z > -1.65$, the hypothesis is accepted (i.e., the hypothesis is not rejected).

p-value= $P(Z < 1.68) = 0.9535 \gg 0.05$

There is very little evidence that the signal is smaller than 8.

Power of a Statistical Test

Type II error occurs if the test incorrectly accepts the hypothesis when it is actually false

$\beta = P(Type\ II\ error) = P(Accept\ H_0\ when\ H_0\ is\ false)$

$Power\ of\ the\ test = 1 - \beta = P(Reject\ H_0\ when\ H_0\ is\ false)$

Power is a number between 0 and 1. It can only be calculated if a specific value for the alternative hypothesis H_1 is specified (e.g. $\mu_w = 2$)

The greater the power, the higher is the ability of the test to reject H_0 if it is false.

Notes in passing

- Note that parameter estimation and hypothesis testing uses the same mathematical model
- In the following, we study other cases of two-sided test. In all these cases, similar procedures may be used for one-sided test.
- We may also consider the power of these tests.

t-test: Hypothesis Test of the Mean of a Normal Distribution Assuming Unknown Variance

This is the famous “t-test”

Since σ^2 is unknown, it seems reasonable to estimate it by

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

and then use the T statistic with $n-1$ d.f. in all the calculations

The hypothesis test is designed as follows:

Reject H_0 if $\frac{\sqrt{n}}{s}|\bar{X} - \mu_0| > t_{\frac{\alpha}{2},n-1}$

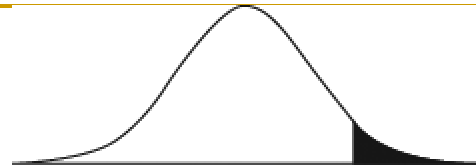
Accept H_0 if $\frac{\sqrt{n}}{s}|\bar{X} - \mu_0| \leq t_{\frac{\alpha}{2},n-1}$

Example

Suppose the standard deviation of the signal received at location B is unknown. Test the hypothesis $\mu = 8$ if the same signal is received 5 times and the recorded signal is

7.5, 10.5, 11, 6.5, 12

Assume a level of significance of 0.05.



t- distribution table

ν	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
	Tail probability									
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
70	0.254	0.678	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435
130	0.254	0.676	1.288	1.657	1.978	2.355	2.614	2.856	3.154	3.367
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

$\bar{X} = 9.5$
$S^2 = 5.625$

$$T = \frac{\sqrt{n}}{S}\left|\bar{X} - \mu_0\right| = \frac{\sqrt{5}}{\sqrt{5.625}}(9.5 - 8) = 1.414213562$$

$$t_{0.025,4} = 2.776$$

As $T < 2.776$, the hypothesis is accepted.

The result in this example is reasonable, as the t test is “more cautious” than the Z test.

Testing the Equality of Means of Two Normal Populations

- This is a useful test to determine whether two different methods have the same performance in the average sense, i.e., whether their mean performance is equal
- Even when the mean is equal, it does not mean that the two method is identical. For example, their variance may be very different!

- We have two cases

  1. Case when the two variances are known
  2. Case when the two variances are unknown

Case of Known Variances

Suppose $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are independent samples from two normal populations having unknown means μ_x and μ_y but known variance ${\sigma_x}^2$ and ${\sigma_y}^2$.

Null hypothesis $\quad H_0\colon\ \mu_x = \mu_y$

Alternative hypothesis $\quad H_1\colon\ \mu_x \neq \mu_y$

It seems reasonable to use the statistic $\bar{X} - \bar{Y}$

Using property of sum of independent normal random variables,

$$\bar{X} - \bar{Y} \sim N\left(\mu_x - \mu_y, \frac{{\sigma_x}^2}{n} + \frac{{\sigma_y}^2}{m}\right)$$

The hypothesis test is designed as follows:

Reject H_0 if $\frac{|\bar{X}-\bar{Y}|}{\sqrt{\frac{{\sigma_x}^2}{n}+\frac{{\sigma_y}^2}{m}}} > z_{\alpha/2}$

Accept H_0 if $\frac{|\bar{X}-\bar{Y}|}{\sqrt{\frac{{\sigma_x}^2}{n}+\frac{{\sigma_y}^2}{m}}} \leq z_{\alpha/2}$

Case of Unknown Variances

Suppose $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ are independent samples from two normal populations having unknown means μ_x and μ_y but unknown variance ${\sigma_x}^2$ and ${\sigma_y}^2$.

Null hypothesis H_0: $\mu_x = \mu_y$

Alternative hypothesis H_1: $\mu_x \neq \mu_y$

At present there is no completely satisfactory solution unless we make the simplifying assumption that

$${\sigma_x}^2 = {\sigma_y}^2$$

Note that this is not very realistic

It is reasonable to use the statistic $\bar{X} - \bar{Y}$ and the two sampled variances ${S_x}^2$ and ${S_y}^2$.

It can be shown that

$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{{S_p}^2(\frac{1}{n} + \frac{1}{m})}} \sim t_{n+m-2}$$

where ${S_p}^2$, the pooled estimator of the common variance σ^2 is

$${S_p}^2 = \frac{(n-1){S_x}^2 + (m-1){S_y}^2}{n+m-2}$$

The hypothesis test is designed as follows:

Reject H_0 if $|T| > t_{\frac{\alpha}{2},n+m-2}$

Accept H_0 if $|T| \leq t_{\frac{\alpha}{2},n+m-2}$

where

$$T = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{{S_p}^2(\frac{1}{n} + \frac{1}{m})}}$$

Paired t-test

For the above, the two sets of data $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are samples from two <u>independent</u> <u>normal populations</u>

In the case when data comes in the form of n pairs (X_i, Y_i), $i = 1, \ldots, n$, then a technique called paired t-test should be used

We compute

$$W_i = X_i - Y_i \qquad \mathrm{i} = 1, \ldots, \mathrm{n}$$

Then for example, we can test

Null hypothesis H_0: $\mu_w = 0$

Alternative hypothesis H_1: $\mu_w \neq 0$

where $W_1, \ldots, W_n$ are assumed to be a sample from a normal population having unknown mean μ_w and unknown variance ${\sigma_w}^2$. This is true due to property of sum of independent normal random variables.

This can be done by using a t-test, which we have studied above.

References

- Text Ch. 8