8. HYPOTHESIS TESTING

Study time: 80 minutes

Aim - you will be able to

- conclude by the pure significance test
- use basic sample and two sample tests
- conclude by paired test and tests for proportions

Explication

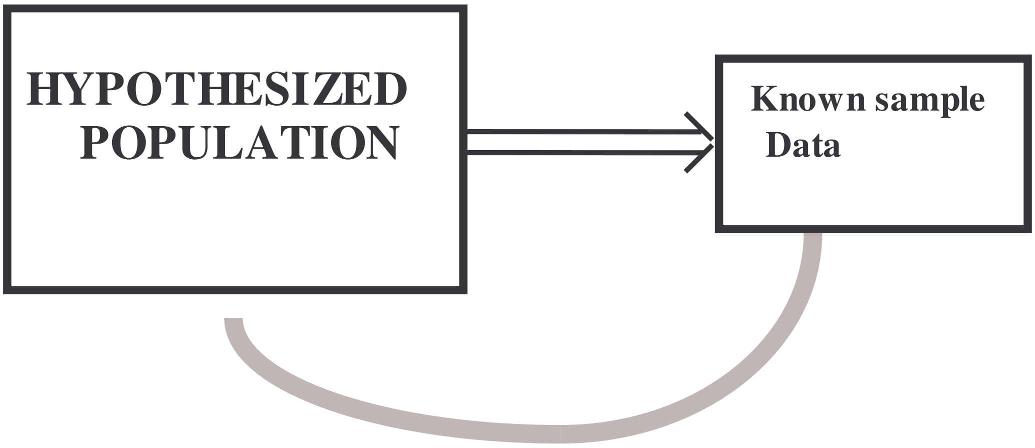
8.1. Introduction

In this chapter we will construct test and with their help we accept or reject some population hypothesis.
The most frequent case is a situation when we can describe a population by some probability distribution which depends on θ parameter. Based on the trial result we can for example want to accept or reject an opinion that θ has some concrete value θ_0. In other situation we can be interested in our hypothesis validity that given population comes from a concrete distribution. Procedures leading to similar decisions are called **significance tests**.
Statistic hypothesis - the assumptions about population whose trueness can be verified by statistic significance tests
Significance tests - procedures which decide if a verified hypothesis should be accepted or rejected based on random sample
Null hypothesis H_0 - verified hypothesis whose rejection is decided by a significance test
Alternative hypothesis H_A - hypothesis which is accepted when we reject null hypothesis

8.2. Pure Significance Tests

The pure significance test asks whether the sample result is extreme with respect to some hypothesized distribution.

If the sample data lies at an extremely high or extremely low percentile of the hypothesized distribution, then the hypothesis is in doubt.



Is data consistent with hypothesized population ?

The pure significance test consists of the following components:

1. Null Hypothesis: H_0
   – The null hypothesis expresses some belief about the nature of the population. It must be specified precisely enough to define a probability measure on the population.

2. Sample Statistic: T(X)
   – The sample statistic is a function of the sample data drawn from the population. The choice of sample statistic is determined by the characteristics of the population's probability distribution with which the null hypothesis is concerned.

3. Null Distribution: $F_0(x)$
   $F_0(x) = P(T(\underline{X}) < x \mid H_0)$
   - The null distribution is the probability distribution of the sample statistic when the null hypothesis is correct. The null hypothesis must be specified precisely enough to determine the null distribution.

4. To determine whether the observed sample statistic $t=x_{OBS}$ is extreme with respect to the null distribution, a statistic known as the p-value is computed. The p-value has 3 definitions depending on the context of the null hypothesis, but in all cases, the interpretation of the p-value is the same.

Definition 1: $P_{VALUE} = F_0(x_{OBS})$

This definition is used when we are concerned that the distribution of the sample statistics may be less than the null distribution.

Definition 2: $P_{VALUE} = 1 - F_0(x_{OBS})$

This definition is used when we are concerned that the distribution of the sample statistics may be greater than the null distribution. Such is the case in our first example.

Definition 3: $p_{VALUE} = 2 \min [F_0 (x_{OBS}), 1 - F_0 (x_{OBS})]$

This definition is used when we are concerned that the distribution of the sample statistics may be either greater or less than the null distribution. Note that this definition is only applied when the null distribution is symmetric.

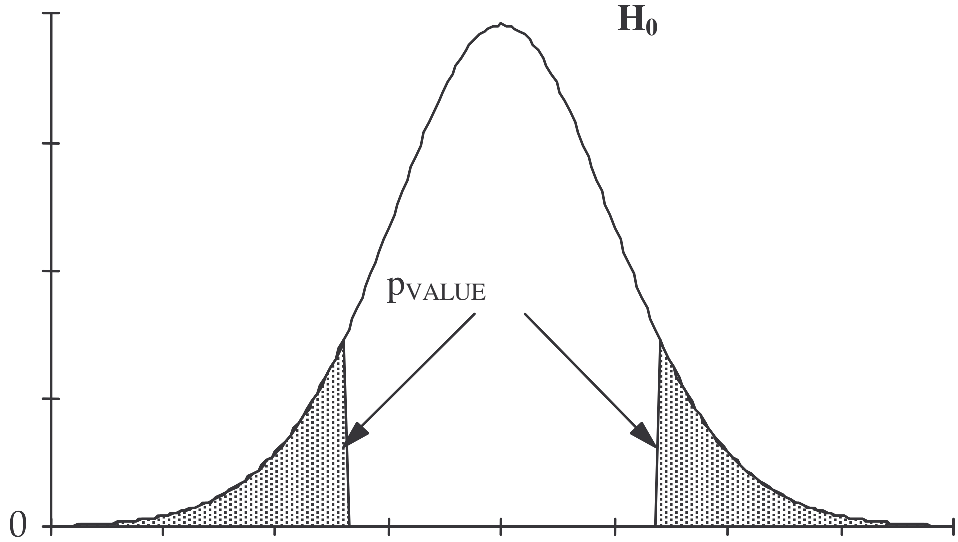


Figure: *Graphical presentation of* p_{VALUE} *for definition 3 by area below spline of density of the null distribution.*

When the null hypothesis is correct, the distribution of the p-value under all three definitions is uniform. That is,

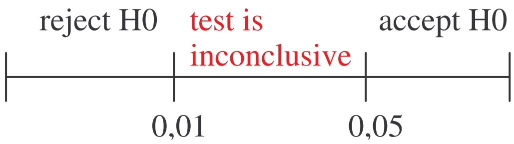
$$P(p_{VALUE}(\underline{X}) < p | H_0) = p$$

Therefore, the p-value has the same interpretation for all null hypotheses independent of the original null distribution. Clearly, smaller p-values are more extreme with respect to null distribution. Therefore, the smaller the p-value, the stronger the evidence of the sample statistic against the null hypothesis. But how small must the p-value be before the evidence is strong enough to reject the null hypothesis? Strictly speaking, this would again depend on the context in which the hypothesis is tested. However, since the weight of evidence against the null hypothesis increases continuously with decreasing p-value, it would be unreasonable to designate a single p-value cut-off point below which the null hypothesis is rejected and above which it is accepted. Rather we should expect an inconclusive region separating the accept and reject p-values.

5. Conclusion in terms of p_{VALUE}

$p_{VALUE} < 0{,}01$	reject H_0
$0{,}01 < p_{VALUE} < 0{,}05$	test is inconclusive

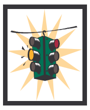
$p_{VALUE} > 0,05$ accept H_0



8.3. Alternate hypothesis

From the definition of the *p-value*, it is clear that the pure significance test procedure for hypothesis testing requires not only a specific null hypothesis but also but notion of which alternative might be correct if the null hypothesis is rejected. The alternate hypothesis need not be specified as precisely as the null hypothesis. To select the appropriate definition of the *p-value*, it is only necessary to know the direction of the alternative with respect to the null. However, the alternate hypothesis will also influence the choice of sample statistic. Those values of the sample statistic which have a small *p-value* under the null hypothesis should tend to have a larger *p-value* for prospective alternatives and vice versa. (Large null *p-values* should have small alternate *p-values*).

8.4. Hypothesis Tests for mean and median



Solved example 1

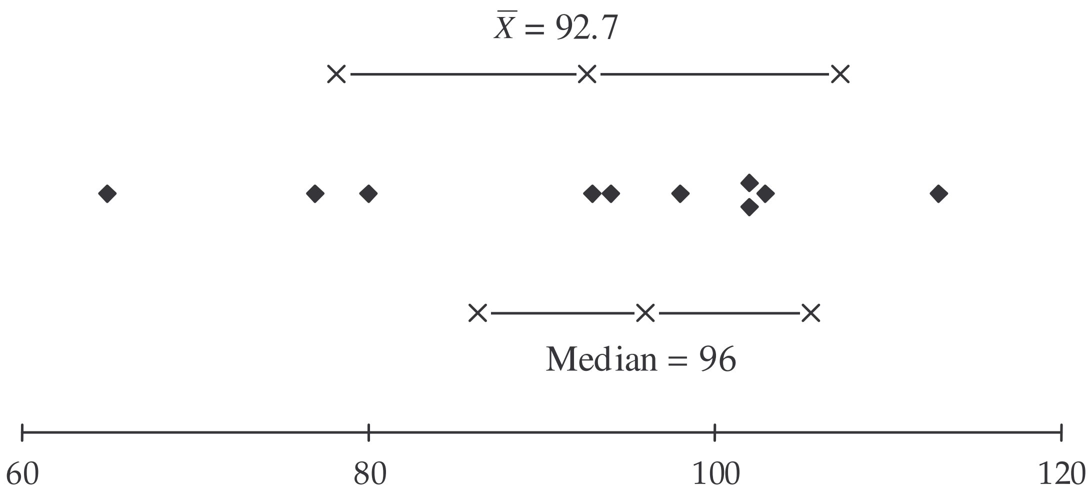
Consider the following ten IQ scores. IQ test scores are scaled to have a mean of 100 and a standard deviation of 15.

65	98	103	77	93
102	102	113	80	94

We wish to test the hypothesis that the mean is 100.

Solution:

We can illustrate this sample:



H_0: X (IQ) has N(100, 15); $\mu_0 = 100$; $\sigma = 15$

$$\text{Under } H_0 \quad \Rightarrow \quad \overline{X} \to N\left(100, \frac{15}{\sqrt{10}}\right)$$

$$\overline{X} = 92.7 ; \qquad s = 14.51$$

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{92.7 - 100}{15 / \sqrt{10}} = -1.54$$

$$p_{VALUE} = \Phi(-1.54) = 0{,}06178$$

Accept H_0: $\mu_0 = 100$

Notes: a) When the sample size n is large, the Central Limit Theorem permits the use of this test when the original population is not normally distributed.

b) If σ is not known and the original population is not normally distributed, the sample standard deviation s may be substituted when the sample size is large.

Solved example 2

- we have same data as with example1

Student's test for mean of small samples

H$_0$: X je N(100, σ); μ_0 = 100 ; σ is uknown

$\bar{X} = 92.7$; $s = 14.51$

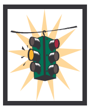
$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{92.7 - 100}{14.51/\sqrt{10}} = -1.59$$

$$t \rightarrow t_{n-1} = t_9$$

$$p_{VALUE} = t_{n-1}(-1.59) = 0{,}073149$$

Accept H$_0$: μ_0 = *100.*

Notes: a) When the sample size n becomes larger than about 30, the t distribution becomes very similar to the normal distribution.



Solved example 3

- we have same data as with example1

Sign test for median

An alternative to testing the hypothesis is that the mean equal to 100 is to test the median equal to 100. If the median is m$_0$, then the probability of any observation exceeding the median is 0.5. Therefore, the number of observations in a sample of n which exceed the hypothesized median will have a binomial distribution with parameters n and 0.5.

H$_0$: X (IQ) has median m0 = 100

Let Y = number of observations > m$_0$

$$Y \rightarrow B\left(n, \frac{1}{2}\right) = B(10, \frac{1}{2})$$

$$Y = 4$$

$$p_{value} = P(Y \leq 4)$$

$$= \sum_{k=0}^{4} \binom{10}{k} \frac{1}{2^{10}} = \frac{386}{1024} = 0{,}377$$

The test result shows no inconsistency between the data and the hypothesis.

Notes: a) The following test makes no assumption about the form of the original distribution and can therefore be applied to any distribution.

b) The sign test has lower power than the t test when the original distribution is normal or the z test when the Central Limit Theorem applies but is not affected by departures from these conditions and is not sensitive to outliers.

Solved example 4

- we have same data as with example1

Wilcoxon signed rank test for medians

A second alternative to tests for sample means based on the normal distribution is to replace the observed values by their ranks and calculate a test statistic from the ranks. To test whether the median is equal to some hypothesized value m_0, we first calculate the absolute difference of each observation from m_0. The absolute differences are then replaced by their ranks or the number of their position. The ranks are then signed -1 if the original observation is less than m_0 and +1 if the original observation is greater than m_0. If the hypothesis that the true median equals m_0 is true, then each rank or integer between 1 and n, the sample size has equal probability of being positive or negative. Therefore, the expected value of the mean of the signed ranks should be zero. Therefore calculating the mean and standard deviation of the signed ranks and forming the z-score as we do for the t test would produce a reasonable test statistic.

H_0: X (IQ) has median $m_0 = 100$

$y_i = |x_i - m_0|$

$r_i = \text{rank}(y_i)$

$r_i^* = \text{sgn}\,(x_i - m_0)\, r_i = \text{signed rank}(y_i)$

For the observations of IQ scores these results are as follows:

IQ score	Absolute difference y_i	Rank of absolute difference r_i	Signed rank r_i^*
93	7	6	-6
94	6	5	-5
77	23	9	-9
80	20	8	-8
103	3	4	4
113	13	7	7
98	2	2	-2
102	2	2	2
65	35	10	-10
102	2	2	2

For the three observations which have the same absolute difference from the hypothesized median, the average of the three ranks has been assigned.

The test statistic of the ranks is calculated as follows. First calculate the mean and standard deviation of the signed ranks.

$$\bar{r} = \frac{\sum_{i=1}^{n} r_i^*}{n} = -2.5; \quad s_r = \sqrt{\frac{\sum_{i=1}^{n}(r_i - \bar{r})^2}{n-1}} = 5.9675$$

Then calculate the z-score for the mean signed rank remembering that the expected value of the mean signed rank is zero under H_0.

$$w = \frac{\bar{r}}{s_r / \sqrt{n}} = -1.325$$

$$p_{VALUE} = W(-1.325) = \Phi(-1.325) = .09257$$

Accept H_0.

Notes: a) Like the sign test, the Wilcoxon signed rank test makes no assumption about the form of the original distribution. If the original distribution is normal, the Wilcoxon test will have less power than the t test, but will be less sensitive to departures from the assumption of normality. As a general rule for small samples it is reasonable to compute both the usual t test and the Wilcoxon test. If the two tests give very different p-values, this would act as a warning that the original distribution may be seriously non-normal.

b) Because the ranks are fixed pre-determined values, the Wilcoxon statistic will not be sensitive to outliers.

c) Computationally simpler formulas for computing the Wilcoxon test statistic which exploit the fact that ranks are fixed values are given in some books.

8.5. Errors through testing

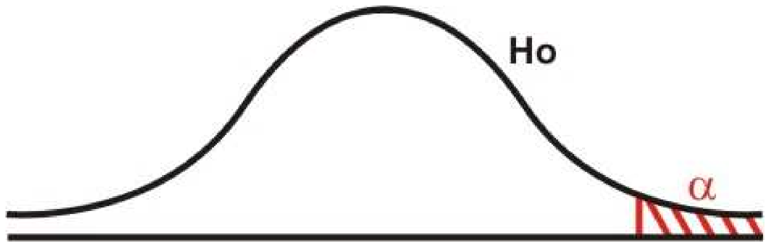
When we make a decision about competing hypotheses, there are two ways of being correct and two ways of making a mistake. This can be depicted by the following table.

		True situation	
		H_0	**H_A**
Decision	**H_0**	OK	Error II
	H_A	Error I	OK

If we make a decision leaning toward H_0 and H_0 is indeed true (true situation), then we did not make an error. If we make a decision leaning toward H_A and H_A is true, then again we did not make an error. These are the probabilities appearing in the upper left and lower right corners. The probabilities in the lower left and upper right are related to the errors made by not making the correct decisions. These probabilities are designated by the Greek letters α ("alpha") and

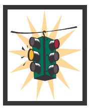
β ("beta").
α is called a Type I Error. It is the probability of falsely rejecting H_0. It is often referred to as the significance level and it represents the risk one is willing to take in rejecting falsely. The user or researcher has complete control over α. Typical (and subjective) values of α are 0.05 and 0 .01. If the consequence of a Type I Error is something in the nature of increased risk of death for a patient or increased risk in financial loses, then one would use a level of significance no greater than 0.01.
β is called a Type II Error. It is the probability of falsely accepting H_0. Unlike a Type I Error, it is difficult to quantify β. More will be said about this later. If the consequence of H_A is extremely attractive and if the results of a Type I Error are not catastrophic, it may be advisable to increase the risk of making a Type I Error and use a level of significance that is 0.05 or higher.
Admittedly it is difficult at this time to fully comprehend these concepts. Hopefully things will make more sense when we go more deeply into hypothesis testing.

$$P(\mathit{Error\ I}) = P(P_{VALUE} < \alpha \mid H_0) = \alpha$$



$$P(\mathit{Error\ II}) = P(P_{VALUE} > \alpha \mid H_A) = \beta$$

8.6. Two sample tests, paired sample tests and tests for proportions



Solved example 5

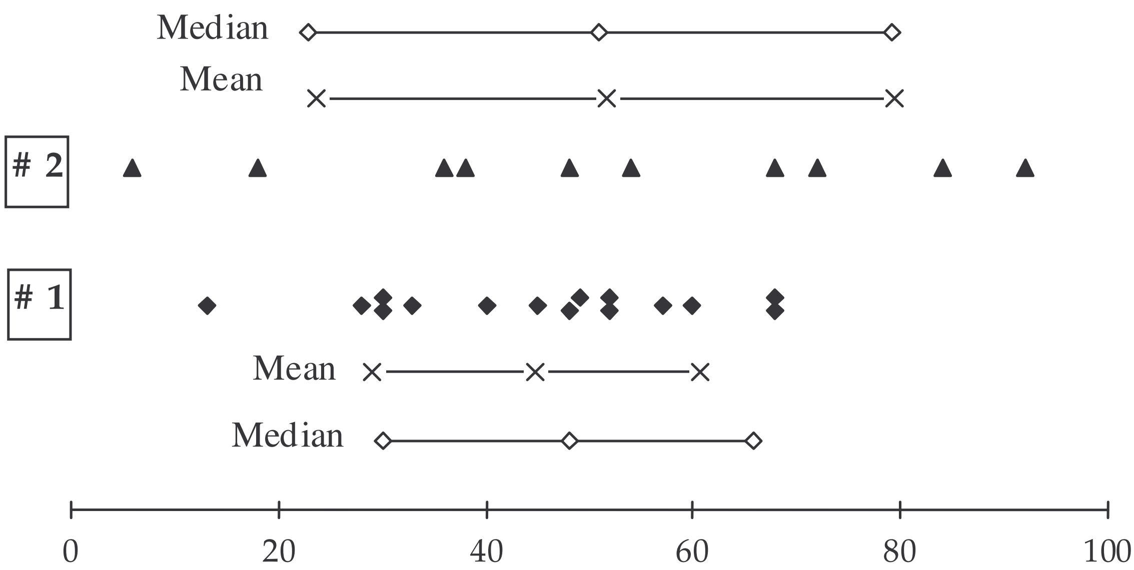
A situation which arises frequently in practice is the two sample test. Two samples have been obtained from difference sources and it is necessary to determine whether the two sources have the same mean or median. One source may be a control group and the other an experimental group. For example, to determine the effectiveness of a new teaching method, a controlled experiment may be conducted in which one group of students, the control group, is taught by traditional methods and a second group by the experimental method. The research question in this case is whether the students taught by the experimental method attained higher results.

Sample from population #1

60	49	52	68	68
45	57	52	13	40
33	30	28	30	48

Sample from population #2

38	18	68	84	72
48	36	92	6	54



The sample means and standard deviations are:

$\#1: \overline{X} = 44.867; \quad s_1 = 15.77$

$\#2: \overline{Y} = 51.6; \quad s_2 = 27.93$

1. We assume that both samples issue from normal distributions:

   $X_i \rightarrow N(\mu_1, \sigma_1^2); \quad i = 1, \ldots, n_1 \quad Y_j \rightarrow N(\mu_2, \sigma_2^2); \quad j = 1, \ldots, n_2,$

   Let:

   $\sigma_1 = 15 \quad \sigma_2 = 25$

   Test **H$_0$:** $\mu_1 = \mu_2$

The test statistics is

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0,1)$$

$$z = \frac{44.867 - 51.6}{\sqrt{\frac{15^2}{15} + \frac{25^2}{10}}} = -0.765$$

$$p_{VALUE} = \Phi(\text{-}0.765) = 0.222$$

2. ***Student's t test for difference of means***

The assumption of equality of variance in both populations requites the computation of a single estimate of standard deviation called the pooled sample standard deviation. The pooled standard deviation is the average of squared deviations of all observations from the sample mean of their respective populations. If x_i is the i^{th} sample observation from population #1, and y_j is the j^{th} sample observation from population #2, then the pooled standard deviation is

$$s_p = \sqrt{\frac{\sum_{i=1}^{n_1} \left(x_i - \bar{x}\right)^2 + \sum_{j=1}^{n_2} \left(y_j - \bar{y}\right)^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}}$$

Under the assumption of equal variance in both populations the estimated standard deviation of the difference of sample means will be

$$s_{\bar{x}-\bar{y}} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Then the two sample t statistic is computed as:

$$t_{n_1+n_2-2} = \frac{\bar{x} - \bar{y}}{s_{\bar{x}-\bar{y}}} = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

and will have a t distribution with ($n_1 + n_2 - 2$) degrees of freedom.

Although the assumption of equal variances is questionable in light of the difference in sample standard deviations, the two sample t test applied to the IQ data above yields the following result.

$$s_p = \sqrt{\frac{14\left(15.77^2\right) + 9\left(27.93^2\right)}{23}} = 21.37$$

$$t_{23} = \frac{44.867 - 51.6}{21.37\sqrt{\frac{1}{15} + \frac{1}{10}}} = -0.772$$

$$p_{VALUE} = t_{n_1+n_2-2}(\text{-}0.772) = 0.224$$

3. ***Mann –Whitney or Wilcoxon rank test for difference of medians***

The two sample rank test is equivalent to ranking the total sample from the two populations and calculating the two sample t test using the ranks rather than the original observations. For the IQ data, this gives the following results.

Ranks for population #1

19	14	15.5	21	21
11	18	15.5	2	10
7	5.5	4	5.5	12.5

Ranks for population #2

9	3	21	24	23
12.5	8	25	1	17

The means and standard deviations of the ranks in each population are

$$\bar{r}_1 = 12.1; \qquad s_{r_1} = 6.29$$
$$\bar{r}_2 = 14.35; \qquad s_{r_2} = 8.89$$

The pooled sample standard deviation of ranks is

$$s_r = \sqrt{\frac{(n_1 - 1)\, s_{r_1}^2 + (n_2 - 1)\, s_{r_2}^2}{(n_1 + n_2 - 2)}}$$
$$= \sqrt{\frac{14\,(6.29^2) + 9\,(8.89^2)}{23}} = 7.42$$

The test statistic is

$$w = \frac{\bar{r}_1 - \bar{r}_2}{s_r\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -0.743$$

$$p_{VALUE} = W(\text{-}0.743) = \Phi(\text{-}0.743) = 0.229$$

Paired sample tests

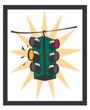
When testing the effect of some experimental condition or comparing the effects of two different conditions, the experimental design often applies either conditions or both experimental and control conditions to the same sampling units or in this case experimental units. The rationale for this design is that variation in experimental results due to differences in sampling units can be eliminated leaving only measurement variation to obscure the effects of the experimental conditions. However, to secure the benefits of reduction in variation offered by this design, the appropriate methods of data analysis and construction of test statistic must be applied.

Suppose two observations under different conditions are taken of n sampling units. For example heart rate before and after exercise. Let X_{i0} be the initial observed value for the i^{th} sampling unit and X_{i1} the subsequent observed value for the same sampling unit. Such a design is called a paired sample design. It is possible to analyze this data and test the hypothesis of no difference between the two experimental conditions using the two sample methods discussed above. However, this approach would failure to take advantage of the opportunity to eliminate variation due to differences in individual sampling units.

A statistically more efficient method to analyze this data is to take advantage of the paired nature of the data and create a single value for each sampling unit. In the simplest data model, this value would be the difference of the two observations for each sampling unit.

$$d_i = X_{i1} - X_{i0}$$

The value d_i is the result only of differences in experimental conditions and experimental error. The methods discussed in the section on one sample tests can then be used to test the hypotheses that the mean or median of d_i is zero which is equivalent to no difference between the two experimental conditions.



Solved example 6

Consider the following example of the heart rates of 12 patients at rest and after ten minutes of exercise.

Resting rate	Rate after Exercise	Difference of Rates	Signed Rank of Difference
42	52	10	3.5
173	175	0	1
113	147	34	11
115	83	-32	-10
69	123	54	12
101	119	20	6
94	69	-25	-7
93	123	30	8.5
112	82	-30	-8.5
67	57	-10	-3.5
104	100	-4	-2
76	89	13	5

The sign test is also applicable to this data. For this data, there are 5 negative signs out of 12 observations. If the true median were 0, the number of negative signs has a binomial distribution with parameters n = 12, and p = 0.5 and the probability if this event is:

$$p_{value} = P(Y \leq 5) = \sum_{k=0}^{5} \binom{12}{k} \frac{1}{2^{12}} = 0.387$$

For such a small sample with unspecified population variance, we assume that the observations are normally distributed and apply the Student's t-test.

The mean and standard deviation of the paired differences are

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = 5$$

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} \left(d_i - \bar{d}\right)^2}{n-1}} = 26.86$$

and the *t*-statistic is

$$t_{11} = \frac{\bar{d}}{s_d / \sqrt{n}} = \frac{5}{26.86 / \sqrt{12}} = 0.645$$

p_{value} = from Definition 2 = $t_{11}(0.645)$ = 0.266

Applying the Wilcoxon signed rank test to these data yields the following results. The mean and standard deviation of the paired differences are

$$\bar{r} = \frac{\sum_{i=1}^{n} r_i}{n} = 1.33$$

$$s_r = \sqrt{\frac{\sum_{i=1}^{n} (r_i - \bar{r})^2}{n-1}} = 7.55$$

and the W statistic is

$$W = \frac{\bar{r}}{s_r/\sqrt{n}} = \frac{1.33}{7.55/\sqrt{12}} = 0.611$$

$$p_{value} = 1 - \Phi(0.611) = 0.271$$

Tests for proportions

When testing hypotheses about the proportion of a population having some attribute, the sample size, n, will be large enough in most cases to use the normal approximation to the distribution of the sample proportion. Under the null hypothesis that the population proportion is equal to some specified value,

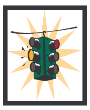
$$H_0: \quad p = p_0$$

The distribution of the sample proportion will be approximately normally distributed for large n.

$$\hat{p} \rightarrow N\left(p_0, \frac{p_0(1-p_0)}{n} \right)$$

and the p-value can be calculated from the z-score of the sample proportion.

$$p_{value} = \Phi\left(\frac{(\hat{p} - p_0)}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right)$$



Solved example 7

If the manufacturer's specifications for the defective rate of an item is not to exceed 3%, and 7 defective items are found in a sample of 95, then the p-value for testing the hypothesis that the sampled population meets the manufacturers specification is

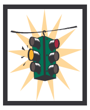
$$z = \frac{(\hat{p} - p_0)}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{7/95 - 0.03}{\sqrt{\frac{(0.03)(0.97)}{95}}} = 2.5$$

$$p_{value} = 1 - \Phi(2.5) = 0.006$$

Reject H_0.

Two sample test for proportions

A two sample test for proportions arises when samples are taken from two populations and the null hypothesis to be tested is that the proportions in both populations are the same. If the samples from each population are large enough, the normal approximation can again to applied to the distribution of the difference of sample proportions. However, since the null hypothesis does not specify a single value for p in each population, the variance is estimated using the total proportion from the samples of both populations which is the maximum likelihood estimate of p under the null hypothesis of equal proportions in both populations.



Solved example 8

Let X_1 be the number of items in a sample of n_1 from population # 1 having the attribute and X_2 be the number of items in a sample of n_2 from population # 2 having the attribute. Then

$$\hat{p}_1 = \frac{X_1}{n_1}; \quad \hat{p}_2 = \frac{X_2}{n_2}; \quad \hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

Then under the null hypothesis that the proportions in the two populations are equal,

H_0: $p_1 = p_2$

the distribution of the difference in sample proportions is:

$$\hat{p}_1 - \hat{p}_2 \rightarrow N\left(0, \hat{p}\,(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

For example, suppose 12 defective items were found in a sample of 88 from one production run and 8 defective items in a sample 92 from a second run. Then,

$$\hat{p}_1 = \frac{12}{88}; \quad \hat{p}_2 = \frac{8}{92}; \quad \hat{p} = \frac{12+8}{88+92}$$

Then the z statistic for testing the hypothesis that the defective rate in both runs is the same is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\,(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.136 - 0.087}{\sqrt{(0.11)\,(0.99)\left(\frac{1}{88} + \frac{1}{92}\right)}} = 1.054$$

$$p_{value} = 1 - \Phi(1.054) = 0.146$$

Accept H_0.

Solved example 9

We have two types of floppy disks - Sony and 3M. In any packet are 20 disks. There were found 24 defective disks into 40 Sony packets and there were found 14 defective disks in 30 3M packets. Does difference in the quality of Sony and 3M disks exist?

Solution:

$\hat{p}_1 = \dfrac{24}{40.20} = 0.030$ (proportion of defective Sony disks)

$\hat{p}_2 = \dfrac{14}{30.20} = 0.023$ (proportion of defective 3M disks)

$$\hat{p} = \frac{24+14}{(40+30).20} = 0.027$$

1. H_0: $p_1 = p_2$
   H_A: $p_1 > p_2$

2. We select test statistic

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}\cdot(1-\hat{p})\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \to N(0,1)$$

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}\cdot(1-\hat{p})\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.030-0.023}{\sqrt{0.027\cdot(1-0.027)(\frac{1}{800}+\frac{1}{600})}} = 0.80$$

3. p-value = 1- Φ (P_2) = 1 - Φ (0,80) = 0,21 Φ (0,80) = 0,79

   /

   (Z has a standard normal distribution)

4. p-value >>> 0.05 $\Rightarrow$ accept H_0

We can't affirm that there exists statistical important difference in quality of Sony and 3M floppy disks.

Summary of notions

The pure significance test answer a question if given random sample $\underline{X}$ (its observed values) is or is not extreme in relation to some tested hypothesis about population. It consists of 5 steps. The last step concludes about acceptation or rejection H_0. The most often we use hypothesis tests for mean and median: **Student's test**, **Wilcoxon test** for median.
Through conclusion of the pure significance test we can commit errors. Cause we don't know a real situation. In case that we reject H_0 but it is true then we have **type 1 error**. If we accept H_A but H_0 holds in fact we have **type 2 error**.
The following tests are the most often used: **Student's test for difference of mean**, Wilcoxon rank test for difference of medians and **paired tests**. There are the most often used the **tests for proportions** in engineering practice

Question

1. How we get P_{VALUE}?
2. What is the alternate hypothesis?
3. Characterize two sample tests for proportions.

Problems

Example 1: Suppose we want to show that only children have an average higher cholesterol level than the national average. It is known that the mean cholesterol level for all Americans is 190. We test 100 only children and find that mean is 198 and standard deviation is 15. Do we have evidence to suggest that only children have an average higher cholesterol level than the national average?

{*Answer:* reject H_0 , therefore, we can conclude that only children have a higher cholesterol level on the average then the national average. }

Example 2: Nine dogs and ten cats were tested to determine if there is a difference in the average number of days that the animal can survive without food. The dogs averaged 11 days with a standard deviation of 2 days while the cats averaged 12 days with a standard deviation of 3 days. What can be concluded?

{*Answer:* We fail to reject the null hypothesis and conclude that there is not sufficient evidence to suggest that there is a difference between the mean starvation time for cats and dogs. }