Table of content

Introduction

There are two types of statistical inferences: estimation of population parameters and hypothesis testing. Hypothesis testing is one of the most important tools of application of statistics to real life problems. Most often, decisions are required to be made concerning populations on the basis of sample information. Statistical tests are used in arriving at these decisions.

Statistical Hypotheses:

They are defined as assertion or conjecture about the parameter or parameters of a population, for example the mean or the variance of a normal population. They may also concern the type, nature or probability distribution of the population.

Statistical hypotheses are based on the concept of proof by contradiction. For example, say, we test the mean (δ) of a population to see if an experiment has caused an increase or decrease in δ. We do this by proof of contradiction by formulating a null hypothesis.

Null Hypothesis:

It is a hypothesis which states that there is no difference between the procedures and is denoted by H_0. For the above example the corresponding H_0 would be that there has been no increase or decrease in the mean. Always the null hypothesis is tested, i.e., we want to either accept or reject the null hypothesis because we have information only for the null hypothesis.

Types of Tests

Tests of hypothesis can be carried out on one or two samples. One sample tests are used to test if the population parameter (δ) is different from a specified value. Two sample tests are used to detect the difference between the parameters of two populations (δ_1 and δ_2).

Two sample tests can further be classified as unpaired or paired two sample tests. While in unpaired two sample tests the sample data are not related, in paired two sample tests the sample data are paired according to some identifiable characteristic. For example, when testing hypothesis about the effect of a

treatment on (say) a landfill, we would like to pair th e data taken at different points before and after implementation of the treatment.

Error

When using probability to decide whether a statistical test provides evidence for or against our predictions, there is always a chance of driving the wrong conclusio ns. Even when choosing a probability level of 95%, there is always a 5% chance that one rejects the null hypothesis when it was actually correct. This is called Type I error, represented by the Greek letter δ. It is possible to err in the opposite way if one fails to reject the null hypothesis when it is, in fact, incorrect. This is called Type II error, represented by the Greek letter δ. These two errors are represented in the following chart.

Figure 1. Types of error		
Type of decision	**H_0 true**	**H_0 false**
Reject H_0	Type I error (α)	Correct decision ($1-\beta$)
Accept H_0	Correct decision ($1-\alpha$)	Type II error (β)

Steps in Hypothesis Testing

1

Identify the null hypothesis H_0 and the alternate hypothesis H_A.

2

Choose α. The value should be small, usually less than 10%. It is important to consider the consequences of both types of errors.

3

Select the test statistic and determine its value from the sample data. This value is called the observed value of the test statistic. Remember that statistic is usually appropriate for a small number of samples; for larger number of samples, a z statistic can work well if data are normally distributed.

4

Compare the observed value of the statistic to the critical value obtained for the chosen α.

5

Make a decision.

If the test statistic falls in the critical region:	If the test statistic does not fall in the critical region:
Reject H_0 in favour of H_A.	Conclude that there is not enough evidence to reject H_0.

Practical Examples

The data set was collected from one of the supermarket (local store near my flat) in Cyprus and the amount the customer spends was assumption that I made as they told me that the information regarding the profit and sales is their privacy and thus not allowed to give me. A week before one of my friend and I went to the store and collected the data asking some customer and the store keeper. The objective of this study was to test whether being vegetarian influences the amount spent in Supermarket store. An independent sample t-test was chosen as the methodology to investigate such hypothesis.

The t-test allows to compare two means from different consumer groups and test the null hypothesis that the two means are equal. If the probability of the t-statistics falls below a threshold level (set at 0.05 , i.e. 5%) then the null hypothesis is rejected in favor of the alternative.

The t-test is based upon the normal distribution of the target variable (i.e. amount spent) within each of the groups. If the sample size is reasonably large (>40 -50 units) it is possible to exploit the normal approximation.

The t-test was carried out to check whether the following customer characteristics led to statistically significant differences in the group means:

-vegetarian

-use coupon

-gender

The given below summarized the output of the One Samples T-test.

Figure 2

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Amount Spent	30	320.7333	227.06006	41.45531

Figure 3

One-Sample Test

	Test Value = 400					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Amount Spent	-1.912	29	.066	-79.26667	-164.0523	5.5190

The null hypothesis is not rejected (as the p-value is larger then 0.05).

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Gender	30	.43	.504	.092
Vegetarian	30	.63	.490	.089
Use Coupons	30	1.40	.498	.091
Amount Spent	30	320.7333	227.06006	41.45531

One-Sample Test

	Test Value = 400					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Gender	-4342.235	29	.000	-399.567	-399.75	-399.38
Vegetarian	-4462.918	29	.000	-399.367	-399.55	-399.18
Use Coupons	-4381.579	29	.000	-398.600	-398.79	-398.41
Amount Spent	-1.912	29	.066	-79.26667	-164.0523	5.5190

It shows that the t-statistic for the vegetarian char acteristic has a p-value of 0.66. As this p-value is above 0.05, the null hypothesis of equal means can not be rejected.

This means the vegetarian factor is not influencing the amount spent. However, the mean comparison hypothesis test doesn't not take explicitly onto account the potential influence of other disturbing factors (e.g. store size). Partial correlation and regression analysis could give further information in that direction, but for the objective of this study and given the very high p-value we can confidently assume that the t-test result are reliable.

This study showed that being vegetarian is not an influential factor in determining the amount spent, while there are significant difference in terms of gender and the use of the coupon. More specifically, the average amount spent for male/female and user/non user of coupon. It looks that the amount spent by men is significantly higher, and also the use of coupon lead to a higher expenditure.

A better method for comparing several population means is an analysis of variance, abbreviated as ANOVA.

ANOVA test is based on the variability between the sample means. This variability is measured in relation to the variability of the data values within the samples. These two variances are compared through means of the F ratio test.

If there is a large variability between the sample means, this suggests that not all the population means are equal. When the variability between the samples means is large compared to the variability within the samples, it can be concluded that not all the population means are equal.

The tests used in the testing of hypothesis, viz., t-tests and ANOVA have some fundamental assumptions that need to be met, for the test to work properly and yield good results. The main assumptions for the t-test and ANOVA are listed below.

The primary assumptions underlying the t-test are:

- The samples are drawn randomly from a population in which the data are distributed normally distributed.
- In the case of a two sample t-test, $\delta_1^2 = \delta_2^2$. Therefore it is assumed that s_1^2 and s_2^2 both estimate a common population variance, δ^2. This assumption is called the homogeneity of variances
- In the case of a two sample t-test, the measurements in sample 1 are independent of those in sample 2.

Like the t-test, analysis of variance is based on a model that requires certain assumptions.

Three primary assumptions of ANOVA are that:

- Each group is obtained randomly, with each observation independent of all other observations and the groups independent of each other.
- The samples represent populations in which the data are normally distributed.
- $\delta_1^2 = \delta_2^2 = \delta_3^2 = ... = \delta_k^2$. The assumption of homogeneity of variances is similar to the discussion above under the t-test. The group variances are assumed to be an estimate of a common variance, δ^2.

In actual experimental or sampling situations, the und erlying populations are not likely to be exactly normally distributed with exactly equal variances. Both the t-test and ANOVA are quite robust and yield reliable results when some of the assumptions are not met. For example, if $n_1 = n_2 = ... = n_k$, ANOVA tends to be especially robust with respect to the assumption of homogeneity As the number of groups tested, k, increases there is a greater effect on the value of the F -statistic. It is also seen that a reasonable departure from the assumption of population normality does not have a serious effect on the reliability of the F -statistic or the t-statistic. It is essential however that the assumption of independence be met. The analysis is not *robust* for non-independent measurements. These factors are to be taken into consideration while testing hypotheses.

ANOVA

Amount Spent

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	140058.006	1	140058.006	2.894	.100
Within Groups	1355073.861	28	48395.495		
Total	1495131.867	29			

Limitations for Environmental Sampling

Although hypothesis tests are a very useful tool in general, they are sometimes not appropriate in the environmental field. The following cases illustrate some of the limitations of this type of test:

A) Multiple Comparisons

z and *t* tests are very useful when comparing two population means. However, when it comes to comparing several population means at the same time, this method is not very appropriate.

Suppose we are interested in comparing pollutant concentrations form three different wells with means m_1, m_2 and m_3. We could test the following hypothesis:

H_0: $m_1 = m_2 = m_3$

H_A: not all means are equal

We would need to conduct three different hypothesis tests, which are shown here:

Figure 4.Hypothesis tests needed for testing thr ee different populations

$$\begin{array}{|c|c|c|} \hline \mu_1 = \mu_2 & \mu_2 = \mu_3 & \mu_1 = \mu_3 \\ \mu_1 \neq \mu_2 & \mu_2 \neq \mu_3 & \mu_1 \neq \mu_3 \\ \hline \end{array}$$

For each test, there is always the possibility of committing an error. Since we are conducting three such tests, the overall error probability would exceed the acceptable ranges, and we could not feel very confident about the final conclusion. Table 8 shows the resulting overall a if multiple *t* tests are conducted. Assume that each k value represents the number of populations to be compared.

Summarize

Figure 5

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Amount Spent * Health Food Store * Size of Store	30	93.8%	2	6.3%	32	100.0%

Figure 6

Case Summaries

Amount Spent

Health Food Store	Size of Store	N	Mean	Std. Error of Mean	Maximum	Minimum
yes	small	11	330.7273	41.99980	525.00	100.00
	medium	3	237.3333	14.65530	258.00	209.00
	large	2	875.0000	25.00000	900.00	850.00
	Total	16	381.2500	56.80262	900.00	100.00
no	small	12	266.8333	64.75968	880.00	45.00
	medium	1	265.0000	.	265.00	265.00
	large	1	55.0000	.	55.00	55.00
	Total	14	251.5714	57.18677	880.00	45.00
Total	small	23	297.3913	39.03610	880.00	45.00
	medium	4	244.2500	12.45910	265.00	209.00
	large	3	601.6667	273.71417	900.00	55.00
	Total	30	320.7333	41.45531	900.00	45.00

Frequencies

Figure 7

Statistics

Amount Spent

N	Valid	30
	Missing	2
Mean		320.7333
Std. Error of Mean		41.45531
Median		261.5000[a]
Mode		150.00[b]
Std. Deviation		227.06006
Percentiles	10	77.5000[c]
	20	133.3333
	30	204.5000
	40	247.5000

50	261.5000
60	346.6667
70	360.0000
80	417.5000
90	687.5000

a. Calculated from grouped data.

b. Multiple modes exist. The smallest value is
shown

c. Percentiles are calculated from grouped data.

Figure 8

Amount Spent

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	45.00	1	3.1	3.3	3.3
	50.00	1	3.1	3.3	6.7
	55.00	1	3.1	3.3	10.0
	100.00	1	3.1	3.3	13.3
	120.00	1	3.1	3.3	16.7
	125.00	1	3.1	3.3	20.0
	150.00	2	6.3	6.7	26.7
	200.00	1	3.1	3.3	30.0
	209.00	1	3.1	3.3	33.3
	222.00	1	3.1	3.3	36.7
	245.00	1	3.1	3.3	40.0
	250.00	1	3.1	3.3	43.3
	255.00	1	3.1	3.3	46.7
	258.00	1	3.1	3.3	50.0
	265.00	1	3.1	3.3	53.3
	300.00	1	3.1	3.3	56.7
	345.00	1	3.1	3.3	60.0
	350.00	2	6.3	6.7	66.7
	355.00	1	3.1	3.3	70.0
	365.00	1	3.1	3.3	73.3
	368.00	1	3.1	3.3	76.7

	390.00	1	3.1	3.3	80.0
	445.00	1	3.1	3.3	83.3
	500.00	1	3.1	3.3	86.7
	525.00	1	3.1	3.3	90.0
	850.00	1	3.1	3.3	93.3
	880.00	1	3.1	3.3	96.7
	900.00	1	3.1	3.3	100.0
	Total	30	93.8	100.0	
Missing	System	2	6.3		
Total		32	100.0		

Histogram

Figure 9
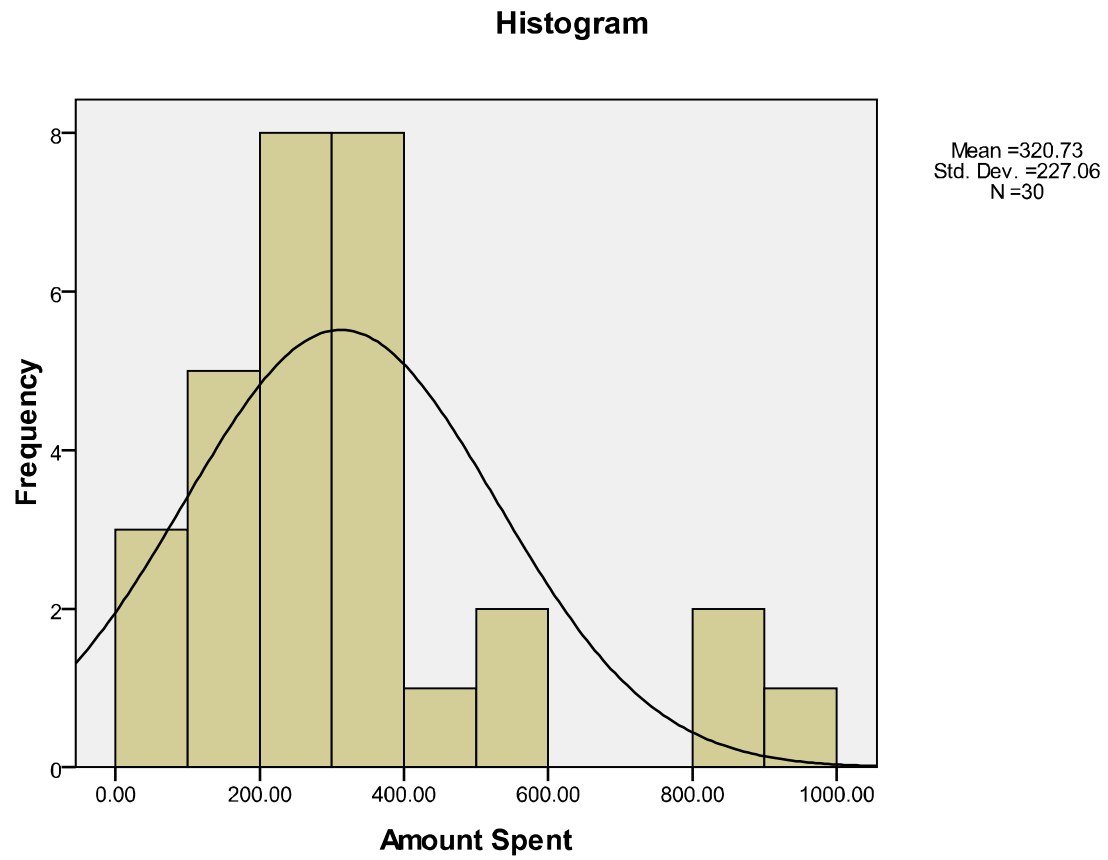
Histogram

Mean =320.73
Std. Dev. =227.06
N =30

Data View From SPSS

	storid	hlthfood	size	org	custid	gender	shopfor	veg	style	usecoup	amtspent
1	4	1	1	1	250.00	0	1	0	1	1	250.00
2	4	1	1	1	115.00	0	1	0	1	1	300.00
3	4	1	1	1	98.00	0	1	0	1	2	350.00
4	4	0	1	1	95.00	0	1	0	1	2	200.00
5	4	0	2	1	88.00	1	1	1	2	2	258.00
6	4	1	2	1	165.00	1	2	1	2	2	265.00
7	4	0	2	1	105.00	0	2	1	2	1	245.00
8	4	0	2	1	125.00	0	2	1	2	1	209.00
9	4	0	1	1	126.00	1	2	1	3	1	150.00
10	4	1	1	1	127.00	0	3	1	3	1	120.00
11	5	1	1	1	140.00	1	3	1	3	1	125.00
12	5	1	1	1	150.00	1	3	1	2	1	50.00
13	5	0	1	1	155.00	1	3	0	2	1	100.00
14	5	0	1	1	160.00	0	2	0	2	2	500.00
15	5	0	3	1	165.00	0	2	0	1	2	900.00
16	5	0	3	1	130.00	0	1	0	1	2	850.00
17	5	0	1	1	240.00	0	1	1	1	2	525.00
18	5	0	1	3	254.00	1	2	1	1	1	445.00
19	5	1	1	3	255.00	1	2	1	1	1	222.00
20	5	1	1	3	265.00	1	1	1	1	1	880.00
21	20	1	1	3	257.00	1	1	1	2	1	150.00
22	20	1	3	3	350.00	1	3	1	2	1	55.00
23	20	1	1	3	355.00	0	3	0	2	2	45.00
24	20	0	1	3	300.00	0	2	0	2	2	255.00
25	20	0	1	3	302.00	0	2	1	1	2	350.00

Variable View From SPSS

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	storid	Numeric	4	0	Store ID	None	None	8	Right	Scale
2	hlthfood	Numeric	4	0	Health Food Store	{0, yes}...	None	7	Right	Scale
3	size	Numeric	4	0	Size of Store	{1, small}...	None	8	Right	Scale
4	org	Numeric	4	0	Store Organization	{1, emphasizes pro...	None	8	Right	Scale
5	custid	Numeric	8	2	Customer ID	None	None	8	Right	Scale
6	gender	Numeric	4	0	Gender	{0, male}...	None	8	Right	Scale
7	shopfor	Numeric	4	0	Who Shopping For	{1, self}...	None	8	Right	Scale
8	veg	Numeric	4	0	Vegetarian	{0, no}...	None	8	Right	Scale
9	style	Numeric	4	0	Shopping Style	{1, weekly}...	None	8	Right	Scale
10	usecoup	Numeric	4	0	Use Coupons	{1, not using coupo...	None	8	Right	Scale
11	amtspent	Numeric	8	2	Amount Spent	None	None	8	Right	Scale

15