**The FOUR Basic Types of Scales/Data**

Before we begin discussing matters about inferential statistics, we examine the 4 types of scales or data:

1. Nominal Scale

Assigning respondents to certain groups or categories e.g. male and female or assigning code number like 1 (males) and 2 (female). There is no overlapping i.e. the categories are mutually exclusive. In other words, nominal scales categorize individuals or objects into mutually exclusive and collectively exhaustive groups. The information generated from nominal scaling is the calculation of the percentage or frequency of males and females in the sample of respondents.

For example if 200 people are interviewed and code 1 is assigned to male respondents and code 2 to female respondents. The results: 98 of the respondents are men and 102 are women. The frequency distribution tells us that 49% of the respondents are men and 51% are women. The nominal data tells only such information and nothing more. Therefore ***nominal scale gives some basic, categorised, gross information.***

1. Ordinal Scale

It categorises the variables to denote differences among the various categories, and it also rank-orders the categories in some meaningful way. The data can be ranked according to categories such as from best to worst and numbered 1, 2, 3 and so on as seen in Likert Scale.

The ordinal scale provides more information than the nominal scale. The ordinal scale goes beyond differentiating categories *to provide information on how respondents distinguish them by rank-ordering them such as excellent, good, satisfactory, bad, terrible*.

In the ordinal scale the persons, objects or events may be ranked such as excellent, good, satisfactory, bad or terrible but their magnitudes are not known. This deficiency is overcome by interval scaling.

1. Interval Scale or Equal Scale

There are equal distances or equal values on the interval scales e.g. a thermometer. The interval scale taps the differences, the order and the equality of the magnitude of the difference in the variable. It is a more powerful scale than the nominal or ordinal scales and has for its measure of central tendency the arithmetic means. It measures dispersion such as the range, the standard deviation and the variance.

1. Ratio Scale

The ratio scale overcomes the disadvantage of the arbitrary origin point of the interval scale, in that it has an absolute zero point. It not only measures the magnitude of the differences between points on the scale but also taps the proportions in the differences.

It is the most powerful of the 4 scales because it has a unique zero origin and subsumes all the properties of the other 3 scales.

The weighing balance is a good example of ratio scale. It has an absolute zero origin calibrated on it, which allows us to calculate the ratio of the weights of 2 individuals e.g. a person weighing 250 pounds is twice as heavy as one who weighs 125 pounds.

The measure of central tendency of the ratio scale may be either the arithmetic or the geometric mean and the measure of dispersion may be either the standard deviation or variance or the coefficient of variation. Examples of ratio scale - actual age, income and number of organisations individuals have worked for.

**1. INFERENTIAL STATISTICS**

**Inferential statistics** is about data analysis techniques for determining how likely it is that results obtained from a sample or samples are the same results that would have been obtained for the entire population.

(Sample values such as means are referred to as statistics. The corresponding values in the population are referred to as parameters.)

Standard Error

The chances of any sample being exactly identical to its population are virtually nil.

The expected random or chance variation among the means is referred to as *sampling error*.

If a difference is found between 2 sample means, the important question is whether the difference is a true or significant one or just the result of sampling error.

A useful characteristic of a sampling error is that they are usually normally distributed in a normal bell-shaped curve.

If a large number of samples of the same size are randomly selected, all the samples will not be the same but that the means of all these samples will form a normal distribution around the mean and yield a good estimate of the population mean. The standard deviation of the sample mean is usually called the standard error of the mean.

The standard error tells us by how much we would expect our sample means to differ if we use other samples from the same population. About 68% of the sample means will fall between ± 1 standard error (±1α) of the mean, 95% will fall between ±2 standard error (±2α) and 99% will fall between ±3 standard errors (±3α).

68%

95%

99%

-3α -2α -1α *x* 1α 2α 3α

Where *x* = sample mean and α = standard error

Standard deviation

Standard error (SE) =

√ N - 1

Standard error of the mean is also affected by the population standard deviation. A large sample size can help to reduce standard error.

However, estimates of standard error can also be computed by other statistics such as measure of variability, relationship and relative position or from differences between 2 or more means (ANOVA i.e. analysis of variance).

To find out whether a difference between sample mean probably represents a true difference or a chance difference (due to sampling error) tests of significance are applied to the data. Many tests of significance are based on an estimate of standard error and typically test a ***null hypothesis.***

Standard deviation (SD) is defined as the positive square root of the mean of the square deviations taken from the arithmetic mean of the data. The standard deviation is commonly denoted by σ and is defined as:

∑( X – X)2

σ =

√ n

Where X is the mean and n is the total number of observations. The standard deviation is in the same units as the units of the original observations. If the original observations are in grams, the value if the standard deviation will also be in grams.)

**The Null Hypothesis**

Hypothesis testing is a process of decision making about the results of a study. If the experimental groups’ mean is 35 and the control group mean is 27, there is a difference of 8 between the 2 means. Is this difference of 8, represents a real significant difference in the treatment or simply sampling error. A true difference is one caused by treatment (the independent variable) and not by chance (i.e. sampling error). Therefore, *to* *test a hypothesis requires both a test of significance and a selected probability level that indicates how much risk you are willing to take that the decision you make is wrong(such as 5% or 1%).*

Test of Significance

At the end of an experimental research study, the researcher has 2 or more group means. These means are very likely to be at least a little different, different enough to conclude that they represent a true difference. In other words, the researcher must make the decision whether to reject the null hypothesis. The researcher does not make the decision based on his own best guess. He has to select and apply an appropriate test of significance, a statistical test used to determine whether or not there is a significant difference between or among two or more means at a selected probability level. If the difference is too large to be attributed to chance, the researcher rejects the null hypothesis because he infers that a real difference exists between A and B.

Normally the selected probability level or level of significance is 5% i.e. 5 out of 100 or 1 out of 100 chances that the observed difference did not occur by chance. The level of significance or probability level, selected determines how large the difference between the means must be to be declared significantly difference.

The usual statistical tests of significance are:

1. The t-test
2. The analysis of variance (ANOVA)

3. The Chi-Square [χ2]

Decision-Making

Levels of significance and Type I and Type II Errors

Four possibilities can arise when testing a hypothesis:

1. If the null hypothesis is really true (i.e. there is no difference) and a researcher agrees with it, the research has made the correct decision.
2. If the null hypothesis is false, (i.e. there is really a difference) and the researcher rejects the null hypothesis (says there is a difference) the researcher has made the correct decision.
3. If the null hypothesis is true (there is no difference) and the researcher makes an incorrect decision by rejecting it - Type 1 error.
4. If the null hypothesis is false (there is really a significant difference between the means) but the researcher concludes that the null hypothesis is true and does not reject it, the researcher makes an incorrect decision - Type 2 error.

Researchers consider Type 1 error to be more serious than Type 2 error.

If the α = 0.05 i.e. 5 out of 100 is due to sampling error (standard error), the test of significance showed a higher figure, the null hypothesis is rejected. It means that the larger difference occurred not solely due to chance but to treatment in the experiment. It also means that 95% of the difference resulted from the independent variable is not random error (sampling error) i.e. ±2 standard deviations. Similarly if the α = 0.01 and the hypothesis is rejected, it means that 99% of the difference is not due to random error (sampling error) i.e. with ± 3 standard deviations.

95% 99%

Region of Region of Region of

rejection rejection rejection

-3 -2 -1 0 1 2 3 -3 -2 -1 0 1 2 3

The chance of probability level (α) is made before the study is done. This is to avoid committing errors. (Students should read further on Type 1 and Type 2 errors to have a better understanding of the importance in choosing a probability level.)

**Two-Tailed and One-Tailed Tests**

Tail refers to the extreme end of the normal distribution bell-shaped curve. One-tailed test assumed that a difference can occur in only one direction. The null hypothesis states that one group is not better than the other group and the one-tailed test assumes that if a difference occurs it will be in favour of that particular group (A > B).

A two-tailed test of significance would allow for the possibility that either the group that received a snack or the group that did not might exhibit better behaviour.

Note: Tests of significance are almost always two-tailed i.e. a 2 directions possibility.

**Degrees of Freedom (df)**

After having determined the significant test to be two-tailed or one-tailed, select the probability level (α) and then computer a test of significance. Next determine the degrees of freedom to determine the significance of the results. The df is dependent on the number of participants and the number of groups. Each test of significance has its own determination of the df.

**Tests of Significance**

Types:

(1) Parametric test - a test of significance that requires certain assumptions to be met in order for it to be valid e.g.

1. The variable measured must be normally distributed in the population (or at least that the form of the distribution is known).
2. The data represented an interval or ratio scale of measurement like the Likert scale.
3. The participants must be independently selected i.e. not subjected to any influence i.e. random selection (equal or at least the ratio of the variances is known).
4. The variances of the population of comparison groups are equal or at least the ratio of the variances is known.

(2) Nonparametric test of significance - a test of significance appropriate when the data

represent an ordinal or nominal scale. No need of a normal distribution of the population.

Note: If the data represent an interval or ratio, the parametric test should be used unless one of the assumptions is greatly violated. Parametric test is more powerful:

1. When small samples are used.
2. It enables tests of a number of hypotheses that could not be done in a nonparametric test.

Methods using Parametric and Nonparametric approaches

|  |  |  |
| --- | --- | --- |
| Method | Purpose | Examples of application |
| Pearson’s product moment correlation (P) | Measuring association between 2 variables | Correlates between age & height of teenagers or between advertising spend & number of sales |
| Spearman’s rank correlation coefficient  (NP) | Measuring association | Comparing 2 managers’ ranked assessment of 10 employees |
| Chi Square test (NP) | Measuring difference | Do some manufacturers produce more faulty goods than others? |
| Student’s t test (P) | Measuring difference | Comparing the sample means of ages of female finance and marketing mangers (independent t test) |
| Simple regression (P) | Assessing the strength of relationship between variables | Strength of relationship between advertising expenditure & sales |
| Multiple regression (P) | Assessing the strength of relationship between variables | Strength of relationships between advertising spend and training spend on sales |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | One Sample Tests | Two Samples Tests | | k Sample Tests | |
| Measurement Scale | One Sample Case | Related Samples | Independent Samples | Related Samples | Independent Samples |
| Nominal | \* Binomial  \* χ2 one sample  test | McNemor | \* Fisher Exact  test  \* χ2 two sample  test | Cochran Q | χ2 for k samples |
| Ordinal | \* Kolmogorov-Smirnov one sample test  \* Runs test | \* Sign test  \*Wilcoxon  matched-pairs  test | \*Median test  \*Mann-Whitney  U  \*Kolmogorov-Smirnov  \*Wald-  Wolfounty | \*Friedman two  way ANOVA | \*Median extension  \*Kruskal-  Wolh’s one  way ANOVA |
| Interval & Ratio  (*Parametric*) | \* t-test  \*Z test | \*t-test for  paired samples | \*t-test  \*Z test | \*Repeated-  measures  ANOVA | \*One way ANOVA  \*n-way  ANOVA |

**Measuring Association**

It means whether 2 variables are associated or not e.g. between age and salary of employees. However a correlation does not mean one variable influences the other. A further test is required to determine this cause-effect. E.g. a correlation is found between umbrella sales and annual rainfall; this does not mean that increased rainfall causes an increase in umbrella sales. Other variables may have impact on sales. These might include pricing, the state of economy or competition in the market.

Correlation Coefficient

This is for bivariate analysis (i.e. 2 variable analysis). It measures the extent the 2 variables are linearly related. The correlation coefficient value can range from -1 to 1. The value of 1 means perfect positive correlation, while a -1 means a perfect native linear association but a 0 value means the 2 variables are perfectly independent of each other i.e. no relationship between the 2 variables.

In essence there are 2 types of correlation coefficient:

a. Pearson’s product moment correlation coefficient (r)

It is a parametric method for measuring the strength of association between 2 variables

e.g. between advertising expenditure and number of sales or between age and height

among teenagers.

*The data used must be an interval or ratio type and be normally distributed.*

The formula is

∑xy - ∑x∑y/n

r =

**√ [∑**x2  - (∑x2/n][∑y2 - (∑y2 - (∑y2)/n]

where n = the number of data pairs, y = the dependent variable and x = the independent

variable.

Next refer the result i.e. the calculated r value with the value of correlation coefficient

table at 5% significant level. If the calculated value of (r) is:

between 0.70 and 0.99 is a strong positive correlation.

between 0.40 and 0.69 is a medium positive correlation

between 0 and 0.39 is a weak positive correlation

between -0.40 and -0.69 is a medium negative correlation

between -0.70 and -0.99 is a strong negative correlation.

b. Spearman’s Rank Correlation Coefficient (rs)

It is used to test the strength and direction of association between 2 ordinal variables e.g. as

seen in the Likert scale. It is used for nonparametric data. It is also used to aid the proving

or disproving of a hypothesis e.g. the profit of a company increases as the size

of the company increases.

The formula is

rs = 1 - 6∑d2/ n(n2 - 1)

where d = the difference between the 2 rankings of one item of data and n = the

number of item of data.

The rs value can arrange from -1 to 1. The closer the calculated value to 1 or -1 means a

strong correlation exists.

**TESTS OF SIGNIFICANCE**

**Student’s t-Test**

It is a parametric technique that is used to determine whether 2 sample means are significantly different at a selected population level.

The strategy of the t-test is to compare the actual mean difference observed (between the 2 samples) with the difference expected by chance. The t-test involves forming the ratio of these 2 values. In other words, the numerator for a t-test is the difference between the sample means X1 and X2  and the denominator is the chance difference (error term) that would be expected if the null hypothesis were true.

This error term, (sampling error) is derived from sample size and group variance. The t-test determines whether the observed difference between the 2 sample means, is significantly larger than a difference that would be expected solely by change.

After the numerator is divided by the denominator, the resulting t value is compared to the appropriate t table value (for the appropriate probability value and df). If the t value is equal to or greater than the table value, the null hypothesis is rejected.

The t-test can also be used to test samples means from two different populations. The t-test establishes the probability that the two populations are the same in relation to the variable that is being tested.

There are 2 different types of t tests; the t-test for independent samples and the t-test for nonindependent samples.

Calculating t-test for Independent Samples

Independent samples are 2 samples that randomly formed without any type of matching. The members of one sample are not related to members of the other sample in any systematic way other than that they are selected from the same population.

If the 2 groups are randomly selected, they are expected to have the same performance at the beginning of the study to the dependent variable and the same is expected of them at the end of the study. Therefore the null hypothesis is to be true. However, if their means are not close at the end of the study, the null hypothesis is probably false and should be rejected.

The question is “How they are significantly different?” The t-test for independent samples is a parametric test of significance that is used to determine whether at a selected probability level a significant difference exists between the means of 2 independent samples. For example, for 2 sets of post-test scores for 2 randomly formed groups are:

Group 1 Group 2

3 2

4 3

5 3

6 3

7 4

Are these 2 sets of scores significantly different? The appropriate test of significance to use in order to answer the question is the t-test for independent samples:

X1 - X2

t =

SS1 + SS2  1 + 1

√ n1 + n2 - 2 n1 n2

where: SS1 = ∑X2 - (∑x1)2 /n1 and SS2 = ∑X2 - (∑X2)2/n2

X1 X2 X2 X2

3 9 2 4

4 16 3 9

5 25 3 9

6 36 3 9

7 49 4 16

∑X1 25 ∑X2135 ∑X2 15 ∑X2  47

X1 = 25/5 = 5 X2  = 15/5 = 3

SS1 = 135 - (25)2/ 5 SS2 = 47 - (15)2/ 5

= 135 - 125 = 47 - 45

= 10 = 2

X1 - X2

t =

SS1 + SS2  1 + 1

√ n1 + n2 - 2 n1 n2

5 - 3

t =

10 + 2  1 + 1

√ 5 + 5 - 2 5 5

2

t =

12 2

√ 8 5

2

t =

√ (1.5) (0.4)

= 2/√ 0.6 = 2/0.774 = **2.58**

Therefore t = 2.58

At α = .05 and df = n1 + n2 - 2 = 5 + 5 - 2 = 8

From t table, the t value = 2.306 = 2.31

The calculated t = 2.58 is greater than the t table value of 2.31, therefore the null hypothesis is rejected.

***Next calculating the t-test for Nonindependent Samples***

Nonindependent samples the members of one group are systematically related to members of a second group (i.e. same group at 2 different times). Therefore the scores on the dependent variable are expected to be correlated with each other and a special t-test for correlated means is used. It is expected that the error term to be smaller and there is a higher probability that the null hypothesis will be rejected.

For example: 2 sets of scores for 2 matched groups are as follows:

X1 X2

2 4

3 5

4 4

5 7

7 10

Are the 2 sets of scores significantly different?

The appropriate formula is:

D

t =

∑D2 - (∑D)2/N

√ N(N - 1)

X1 X2 D D2

2 4 +2 4

3 5 +2 4

4 4 0 0

5 7 +2 4

6 10 +4 16

∑D = 10 ∑D2 = 28

D

t =

∑D2 - (∑D)2/N

√ N(N - 1)

10/5 2

= =

28 - (10)2/5 √ 28-20/20

√ 5(5 - 1)

= 2/ √ 8/20 = 2/√ 0.4 = 2/0.63 = **3.17**

At α = 0.05 and df = N - 1 = 5 - 1 = 4, t value from t table = 2.776 = 2.78

Since the calculated value of t = 3.17 is greater than the t table value of 2.78, the null hypothesis is rejected. It is noted that the group difference 2 is also significant.

(Source: L.R. Gay, G.E. Mills & P. Airasian (2006). *Education Research Competencies for Analysis and Applications* (8th edition) , Pearson Prentice Hall, New Jersey, pp. 337-359.)

**2. Analysis of Variance (ANOVA)**

(A set of techniques used to compare two or more sample means at the same time)

What is a variance?

The variance(v) of a data set is the square of the standard deviation (s). Thus the variance of *n* observations x1, x2, ……xn is defined as

1 1 (∑x)2

variance (v) = s2 = ∑(x - x)2 or s2 = ∑x2 -

n-1 n - 1 n

Where n = number of observations or values

x = observation or value

∑x2 = sum of all the squares of observations

∑x = sum of all observations

The variance and standard deviation are the measures of the average scatter around the mean. The variance possesses certain useful mathematical properties. However, its computation results in squared units such as squared percentages, squared ringgits and squared centimetres. Thus the primary measure of variation is the standard deviation, as the value of this measurement is in the original unit of the data. The standard deviation is the square root of the variance. A large value of standard deviation makes a wide distribution, while a smaller value of standard deviation makes a narrower distribution. The narrower distribution is better because it indicates that there are no extreme data values in the data set.

Coefficient of Variation

When comparing distributions of different means and variances, a useful measure is the coefficient of variation (CV). The coefficient of variation gives us the ratio of the standard deviation to the arithmetic mean expressed as a per cent.

standard deviation

CV = x 100

mean

Examples:

(i) Typist Ani can type 40 words per minute with standard deviation of 5 while typist Jura

can type 160 works per minute with standard deviation of 10. Which typist is more

consistent in her work?

Solution:

The standard deviation of typist Jura is twice that of typist Ani. Jura can type four times

the speed of Ani.

Taking into consideration all the information, the coefficient of variation is used.

CV for Ani = s/ x x 100 = 5/40 x 100 = 12.5%

CV for Jura = 10/160 x 100 = 6.25%

The results show that the typing ability of typist Jura is more consistent than typist Ani.

(ii) The investment of Karu and Kamal are given in the data summary below.

|  |  |  |
| --- | --- | --- |
|  | Karu | Kamal |
| Profit (RM) | 250 | 250 |
| Standard deviation | 8.16 | 238.05 |

Whose investment is considered to be more consistent?

Solution:

Coefficient of variation for Karu = 8.16/250 x 100 = 3.26%

Coefficient of variation for Kamal = 238.05/250 x 100 = 95.2%

It shows that Karu’s profit is more consistent than Kamal’s profit.

Note: When making comparisons, rule of thumb is that the larger the percentage, the greater is the relative variation. A larger relative variation implies less consistency, while a smaller relative variation implies more consistency, respectively.

**One-Way Analysis of Variance (ANOVA)**

If two or more samples are taken from a population, will their sample means be the same?

The one-way analysis of variance (ANOVA) is a *parametric test* of significance used to determine whether a significant difference exists between two or more means of a selected probability level. Thus, for a study involving three groups, ANOVA is the appropriate analysis technique. Whether the differences among the means represent true, significant differences or chance differences due to sampling error. Therefore ANOVA is used and the F ratio is computed to find the answer this question. ANOVA is a more convenient to perform the significant test than the usual t tests (more complicated approach).

The concept underlying ANOVA is that the total variation, or variance, of scores can be divided into two sources - variance between groups (variance caused by the treatment groups) and variance within groups (error variance). A ratio is formed - F ratio - with group differences as the numerator (variance between groups) and error in the denominator (variance within groups). The assumptions are that the randomly formed groups of participants are chosen and are essentially the same at the beginning of a study on a measure of the dependent variable. At the end of the study, the researcher determines whether the between groups (or treatment) variance differs from the within groups (or error) variance by more than what would be expected by chance. This is shown by the F ratio. If the treatment variance is sufficiently larger than the error variance, a significant F ratio results, the null hypothesis is rejected and it is concluded that the treatment had a significant effect on the dependent variable. If, on the other hand, the treatment variance and error variance do not differ by more than what would be expected by chance, the resulting F ratio is not significant, and the null hypothesis is not rejected.

The greater the difference, the larger the F ratio. To determine whether the F ratio is significant, an F ratio table is entered at the place corresponding to the selected probability level and the appropriate degrees of freedom. The degrees of freedom for the F ratio are a function of the number of groups and the number of participants.

Calculating simple analysis of variance (ANOVA):

Three sets of post-test scores from 3 randomly selected groups are given as follows:

|  |  |  |
| --- | --- | --- |
| X1 | X2 | X3 |
| 1  2  2  2  3 | 2  3  4  5  6 | 4  4  4  5  7 |

Are these three sets of data significantly different?

One-way analysis of variance (ANOVA) is used to answer the question.

Total sum of squares = between sum of square + within sum of squares

SStotal = SSbetween  + SSwithin

SStotal = ∑x2 - (∑x)2/N

SSbetween = (∑x1)2/n1 + (∑x2)2/n2 + (∑x3)2/n3 - (∑x)2/N

X1 X2X2 X2X3 X2

1 1 2 4 4 16

2 4 3 9 4 16

2 4 4 16 4 16

2 4 5 25 5 25

3 9 6 36 7 49

Sum 10 22 20 90 24 122

∑x = 10 + 20 + 24 = 54

∑x2  = 22 + 90 + 122 = 234

N = 5 + 5 + 5 = 15

SStotal  = ∑x2 - (∑x)2/N

= 234 - (54)2/15

= 234 - 194.4

= 39.6

SSbetween = (∑x1)2/n1  + (∑x2)2/n2  + (∑x3)2/n3  - (∑x)2/N

= (10)2/5 + (20)2/5 + (24)2/5 - (54)2/15

= 20 + 80 + 115.2 - 194.4

= 215.2 - 194.4

= 20.8

SSwithin = SStotal - SSbetween

= 39.6 - 20.8

= 18.8

Mean squares (MS) = sum of squares/ degree of freedom = SS/df

For between, MS, we get MSB = 20.8/2 = 10.40

For within, MS, we get MSW = 18.8/12 = 1.57

For F ratio:

F = MSB/MSw = 10.40/1.57 = 6.62

**Summary**

Source of variation Sum of squares df Mean square F

Between 20.8 3-1=2 10.40 6.62

Within 18.8 15-3=12 1.57

Total 39.6 15-1=14

Therefore F = 6.62 with 2 and 12 degree of freedom

Looking at the table of distribution for F when α = 0.05 i.e. at 5% probability of significant level, at df =2 (*horizontal* for between) and df = 12 (*vertical* for within) the value = 3.88.

Since 6.62 is greater than 3.88, there is a significant difference among the means of the three groups and therefore the null hypothesis is rejected.

**Differences among the means of the three groups**

The outcome of the analysis (ANOVA) has indicated a significant difference among the three group means, what does it tell about the difference among the three group means? The answer to this question can be found by using the procedures called ***multiple comparisons***.

Of the many multiple comparison techniques available, the ***Scheffe test*** is one of the most widely used. The formula is:

(X1 - X2)2

F = with df = (K - 1), (N - K)

MSW ( 1/n1 + 1/n2)(K - 1)

Using the ANOVA example, the formula can be substituted:

(2.00 - 4.00)2 (-2.00)2  4

F = = =

1.57(1/5 + 1/5)(2) 1.57(2/5)(2) 1.57(.4)(2)

= 4/1.57(.8)

= 4/1.256 = **3.18**

Because the value of F required for significance is 3.88 is α = 0.05 and df = 2 and 12, and because 3.18 is smaller than 3.88, therefore there is no significant difference between X1 and X2.

Calculate the Scheffe test for X1 and X3 and for X2 and for X2 and X3 and determine their significant differences.   
The following results can be found:

Scheffe Tests

Group 1 vs Group 2 3.18 Fail to reject

Group 1 vs Group 3 6.24 Reject

Group 2 vs Group 3 0.51 Fail to reject

(Note: Scheffe test can be used to compare combinations of means. Suppose Group 1 is the control and the mean of Group 1 can be compared to the mean of Groups 2 and 3 combinations. Further detail can be referred to for this requirement.)

**3. Simple Regression and Multiple Regression** (both are parametric methods)

(a) Simple Regression - it determines the strength of relationship between a dependent variable (y) and one independent variable (x). Are they linearly related?

A regression equation is used on a scatter plot by a regression line: y = a + bx, where x = independent variable, y = dependent variable, a = point where the line intersects the y-axis and b = gradient of the line.

E.g. profit (y) against advertising expenditure (x), profit = a + b (advertising expenditure)

(b) Multiple Regression - it is used to find a linear relationship between a dependent variable (y) and several independent variables (x).

The equation is y= a + b1x1 + b2x2 + b3x3 + b4x4 + ……

E.g. Relationship between profit (y) and advertising expenditure, price, bonuses and competition.

Profit = a + (b1 x staff training expenditure) + (b2 x price) + (b3 x bonuses) + (b4 x no. of

competitors)

The value produced is between -1 and +1. A value of +1 indicates the equation is a perfect predictor. Conversely, a value of 0 shows that the equation predicts none of the variation.

**4. Chi Square (X2)**

It is widely used to measure the difference such as testing a *hypothesis* that there is a difference between the observed data and what is expected.

It is a nonparametric test of significance appropriate when the data are in the form of frequency counts or percentages and proportions that can be converted to frequencies. Two or more categories are required. Therefore chi square is appropriate when the data are a nominal scale and fall into either true categories (e.g. male vs female) or artificial categories (e.g. tall vs short). A true category is one in which persons or objects naturally fall, independently of any research study, whereas an artificial category is one that is operationally defined by a researcher.

A chi square test compares the proportions actually observed in a study to the expected proportions to see if they are significantly different. Expected proportions are usually the frequencies that would be expected if the groups were equal, although occasionally they also may be based on past data. The chi square value increases as the difference between observed and expected frequencies increases. Whether the chi square is significant is determined by consulting a chi square table.

*One-Dimensional Chi Square*

It is used to compare frequencies occurring in different categories or groups. For example, you stopped 90 shoppers in a supermarket to taste three unlabelled different brand of peanut butter (X, Y, Z) and to tell you which one tasted best. Suppose that 40 chose Brand X, 30 chose Brand Y and 20 chose Brand Z. If the null hypothesis were true - if there were no difference in taste among the three brands - we would expect an equal number of shippers (30) to select each brand. We can present our data in what is called a contingency table, as shown:

X Y Z

|  |  |  |  |
| --- | --- | --- | --- |
| Observed frequencies | 40 | 30 | 20 |
| Expected frequencies | 30 | 30 | 30 |

To determine whether the observed frequencies (40, 30, 20) are significantly different from the expected frequencies (30, 30, 30) you could carry out a chi square test. If you find that the chi square is significant, you would reject the null hypothesis and conclude that the brands do taste different.

Another example, you might wish to investigate whether college sophomores prefer to study alone or with others. Tabulation, based on a random sample of 100 sophomores might reveal that 45 prefer to study alone and 55 prefer to study with others. The null hypothesis of no preference would suggest a 50-50 split. The corresponding contingency table would look like this:

Alone Others

|  |  |  |
| --- | --- | --- |
| Observed frequencies | 45 | 55 |
| Expected frequencies | 50 | 50 |

To determine whether the groups are significantly different, you would compare the observed frequencies (45, 55) with the expected frequencies (50, 50) using a chi square test of significance.

*Two-Dimensional Chi Square*

This is used when frequencies are categorised along more than one dimension. It is a sort of factorial chi square. In the study example shown above, you may select a stratified sample, comprising 50 males and 50 females.

Responses could then be classified by study preference and by gender, a two-way classification that would allow you to see whether study preference is related to gender. The corresponding contingency table would be set up as follows:

Alone Others

|  |  |  |
| --- | --- | --- |
| Male | Observed  Expected | Observed  Expected |
| Female | Observed  Expected | Observed  Expected |

Although 2 x 2 applications are quite common, contingency tables may be based on any number of categories, for example, 2 x 3, 3 x 3, 2 x 4 and so forth. When a two-way classification is used calculation of expected frequencies is a little more complex, but not difficult.

Calculating Chi Square

X Y Z

|  |  |  |  |
| --- | --- | --- | --- |
| Observed frequencies | 40 | 30 | 20 |
| Expected frequencies | 30 | 30 | 30 |

Take the above peanut butter example:

To determine whether the observed frequencies are significantly different from the expected frequencies, we apply the following formula:

(fo - fe)2

X2 = ∑

fe

X Y Z

(40-30)2 (30-30)2 (20-30)2

X2 = + +

30 30 30

= (10)2/30 + (0)2/30 + (-10)2/30

= 100/30 + 0 + 100/30

= 3.333 + 0 + 3.333

= 6.666 = **6.67**

Thus chi square (X2) is 6.67, df = number of columns - 1 = 3 - 1 = 2 and α = .05. From

the X2 table the value = 5.991 or 5.99. the calculated value = 6.67 is greater than the

table value of 5.99, therefore the null hypothesis is rejected.

If the selected α = .01, the table value would be 9.210 or 9.21, then the calculated

value of 6.67 would be lesser than the table value of 9.21. The null hypothesis

would not be rejected and the conclusion is that the three brands of peanut butter taste

the same. From here, it is seen that the selection of a probability level (α) is important

as different conclusions can be drawn with different probability levels.

For the second case - Study alone or with others, the calculation is as follows:

X2 = (45 -50)2/50 + (55 - 50)2/50

= (-5)2/50 + (5)2/50

= 25/50 + 25/50

= 0.50 + 0.50

= 1

With α = .05 and df = number of columns - 1 = 2 - 1 = 1. the chi square table value is 3.841 or 3.84. Since the table value is greater than the calculated value, the null hypothesis is not rejected. There is no significant difference between observed and expected proportions; sophomores do not prefer to study alone or with others.

Now suppose we want to know if study preference is related to the gender of the students. The 2 x 2 contingency table might look like this:

Study preference

Alone Others

Males O 29 O 21

50

E 22.5 E 27.5

Females O 16 O 34 50

E 22.5 E 27.5

45 55 100

To find the expected frequency for a particular cell or category, we multiply the corresponding row total by the corresponding column total and divide by the overall total. For males who prefer to study alone, the observed frequency is 29. To find the expected frequency, we multiple the total of the male row (50) by the total for the ‘alone’ column (45) and divide by the overall total (100):

Males (alone) = 50 x 45/100 = 2250/100 = 22.5

Similarly for the other cells:

Males (others) = 50 x 55/100 = 2750/100 = 27.5

Females (alone) = 50 x 45/100 = 2250/100 = 22.5

Females (others) = 50 x 55/100 = 2750/100 = 27.5

Chi Square is calculated as follows:

X2 = (29 - 22.5)2/ 22.5 + (21 - 27.5)2/27.5 + (16 - 22.5)2/22.5 + (34 - 27.5)2/27.5

= (6.5)2/22.5 + (-6.5)2/27.5 + (- 6.5)2/22.5 + (6.5)2/27.5

= 42.25/22.5 + 42.25/27.5 + 42.25/22/5 + 42.25/27.5

= 1.88 + 1.54 + 1.88 + 1.54

= 6.84

The df = (Number of rows, R - 1)(Number of columns, C - 1) = (2 - 1)(2 - 1) = 1 x 1 = 1 and

α = .05, the chi square table gives the value of 3.841 or 3.84. Since the calculated value is greater than the table value, the null hypothesis is rejected. Therefore the gender is related to preference in study.

**How do I know which statistical tests to use?**

Brown and Saunder (2008, 103-104) make the following suggestions before choosing a particular test:

* What is the research question I am trying to answer?
* What are the characteristics of the sample? For instance, are you using judgement sampling, etc?
* What types of data do I have?
* How many data variables are there?
* How many groups are there?
* Are the data distributed normally?
* If the data are not distributed normally, will this affect the statistic I want to use?
* Are the samples independent?

In addition if you wish to make inferences about a population:

* Are the data representative of the population?
* Are the groups different?
* Is there a relationship between the variables?

Sources

1. L.R. Gay, G.E. Mills & P. Airasian (2006). Educational Research Competencies for

Analysis and Applications (8th ed.), Pearson Prentice Hall, New Jersey, pp337-375.)

2. Jonathan Wilson (2014), Essentials of Business Research, A Guide to Doing your

Research project, 2nd Edition, Singapore: SAGE, pp. 260-267.

3. Uma Sekaran & Roger Bougie (2016), Research Methods for Business, 7th Edition, UK:

WILEY.