**Question - A**

**Dataset:** The dataset is about staff survey and the dataset contains in total 534 rows and 30 columns.

Staff were inquired to complete a brief, anonymous survey (appeared later in this Reference section) containing questions approximately their conclusion of different perspectives of the organisation and the treatment they have received as employees.

**Table

Description automatically generated**

Below table contains SPSS variable name in .sav file for survey details that has been filled by employees across the various locations.

**Table

Description automatically generated**

For the process of statistics tests, the independent variables such as Age of staff member has been chosen and later divided into 4 groups such as ‘1= 18 to 30, ‘2=31 to 40’, ‘3=41 to 50’, ‘4=50+’ for the further analysis. Employment status is denoted as “1” for permanent and “0” for casual. On other hand the dependent variable “total satisfaction level” which represents the level of satisfaction in employees is a continuous variable.

Source of the dataset has been provided below:

<http://spss4.allenandunwin.com.s3-website-ap-southeast-2.amazonaws.com/>

1. **INDEPENDENT-SAMPLES T TEST**

The main purpose of conducting Independent-Samples T Test is to identify if there is any significant difference between two unrelated group. In this method the mean of unrelated groups is compared on the continuous dependent variable.

**OBJECTIVE**:

The purpose of this analysis is to compare the satisfaction level between different age group of employees.

**HYPOTHESIS**:

**Null Hypothesis (H0**) – There is no difference in satisfaction level between employees of different ages. The age group of employees are considered here are 31 to 40 and 41 to 50.

**Alternative Hypothesis (H1)** – There is significant difference between the means satisfaction level between age groups.

We will reject the null Hypothesis if P value is lower than 0.05

**VERIFICATION OF ASSUMPTION:**

* Dependent variable is measured on a Continuous scale.
* Independent variable has two categorical groups,
* Normality of dependent variable:

Table

Description automatically generated

This test of Normality of the dependent variable with respect to different groups is analysed with the help of **Shapiro-Wilk** test. Here, the significance value we got for each group is more than 0.05, which results the presence of Normality in the dependent variable with respect to each individual group in independent variable.

No significant outliers, they dependent variable does not have any outlier that could affect the analysis.

Chart, box and whisker chart

Description automatically generated

**INTERPRETATION OF RESULT:**

After conducting the Independent-Samples T test for the variables considered for this analysis, the following two tables of Group Statistics and Independent sample T test were obtained which are studied below.

**Group Statistics:**

**Table

Description automatically generated**

This table results that there is a slight difference between the sample mean of both individual age groups, where the mean value of group “31 to 40” age stands higher than the other age group ’41 to 50’. And ‘N’ shows the count of observations considered from each group.

**Independent Samples Test:**

The primary purpose of this table is to help us to verify whether our assumption is correct or not, with the help of Significance value obtained. In addition, to this result of Levene’s test helps to verify the assumption of Homogeneity of variance.

Table

Description automatically generated

The result of Levene’s test provided for two conditions, one for Equal variance assumed and other for Equal variance not assumed. The significance value obtained for this Levene’s test is 0.580 which is more than 0.05, this shows that there is equal variance of error of dependent variable. This test also proves the Goodness-of-fit of the model, which is satisfied here.

The significance value of 2 tailed test is 0.661 which is greater than 0.05 alpha value, which infers that the **Null Hypothesis is not rejected**.

**CONCLUSION:**

The Independent Samples T test has been conducted on the staff data, to check whether there is any difference between satisfaction level of two different age group considered in the analysis. And it has been observed that there is **no statistically significant difference** between the satisfaction level of two age group of employees.

1. **Mann – Whitney U Test**

The **Mann–Whitney***U***test** is the non-parametric alternative to the Independent-samples t-test. The test is used to test for differences between two independent groups on ordinal data or non-normal continuous data. Instead of comparing the means of the two groups, as in the case of the t-test, the Mann-Whitney U Test compares medians.

**OBJECTIVE**:

This purpose of this analysis is to identify significant difference between employee satisfaction level among permanent and casual employees.

**HYPOTHESIS:** Null Hypothesis (H0) – There is difference in Satisfaction level among permanent and casual employees.

**Alternate Hypothesis (H1)** – There is no significant difference in Satisfaction level among permanent and casual employees.

**VERIFICATION OF ASSUMPTIONS:**

* Dependent variable (Satisfaction level)
* A Grouping variable – employment status (Permanent and Casual)

Table

Description automatically generated

As you can see above, there is a sizeable difference between the mean ranks of the permanent and casual employment status group. The Mann-Whitney test statistic will tell us whether this difference is big enough to reach significance.

Table

Description automatically generated

As shown in above result table Mann-Whitney U, Z and Asymp.Sig (2-tailed) are key values. In SPSS reporting Z Score is -.573 and 2-tailed p-value is 0.566 which is greater than 0.05 (standard alpha value = 0.05).

This leads to accept the Null hypothesis. Hence, there is no significant difference in Satisfaction level among permanent and casual employees.

**Effect Size Calculation for Mann Whitney U Test**

**Table

Description automatically generated**

Table

Description automatically generated

Effect size cannot be produced in SPSS however value of Z that is reported can be used to derive approximate value of r.

r = Z/√N

for our example Z = -.573 and N (number of cases) = 477. Value of r can be calculated as follows,

r = 0.573 / √477

r = 0.573 / 21.840

r = 0.026

According to Cohen (1988) criteria .1 = Small Effect, .3 = Medium Effect, and .5 = Large Effect. In this scenario, the effect size is small.

**CONCLUSION:**

The Mann Whitney U Test has been conducted on the staff data, to check whether there is any significant difference between satisfaction level of two different employment status group (permanent and casual) considered in the analysis. And it has been observed that there is **no statistically significant difference** between the satisfaction level of two group of employment status group.

1. **CHI-SQUARE TEST OF INDEPENDENCE**

Chi-Square test of Independence, also called as Pearson’s chi-square test, helps to identify the relationship between two variables which are Categorical.

**OBJECTIVE**:

This purpose of this analysis is to identify relation between employment status and Age group.

**HYPOTHESIS:** Null Hypothesis (H0) – The employment status does not have any relationship with age group.

**Alternate Hypothesis (H1)** – The age group has impact on employment status.

**VERIFICATION OF ASSUMPTIONS:**

• Variables considered in this analysis are measured in Ordinal or Nominal level.

• Variables has two or more independent groups, individually.

• The Bar chart shown below, helps to visually read the data. It provides information about one categorical with respect to the frequency of occurrences of the other categorical variable.

*Chart, bar chart

Description automatically generated*

**INTERPRETATION OF THE RESULT:**

Cross Tabulation:

Cross tabulation stands as the important aspect of this interpretation. This table has the information like the observer count and expected count of each categorical variable on each category of other variable. Cross tabulation table for this analysis is shown below.

* Table

  Description automatically generatedIrrespective of age group it is clearly observed that percentage of permanent employees are higher than Casual employees.
* However, there is difference in permanent employment percentage in 18 t0 30 and 50+ age group. 50+ group has higher percentage of permanent employees.
* Overall, the observed count of employees of all age, having permanent job status is higher than count of employees of all age with casual job status. And Expected count obtained is also equal to the observed count, altogether.

**Pearson’s Chi-Square test:**

The result obtained from Pearson’s Chi-square test helps to verify the assumption of this analysis. The Pearson’s chi square significance value obtained in this analysis that is 0.065 is more than 0.05 alpha values which denotes that the **Null hypothesis assumed in this analysis is not rejected**, this value is shown in the figure below.

* The Pearson’s Chi-square test has provided us the value of χ²(1) = 7.220, with the significance value of p-Value = 0.065 which is greater than 0.05, signifies that the Null hypothesis assumption is correct. Table

  Description automatically generated

**CONCLUSION**:

From this study conducted to find the relationship between the age groups of employees and employment status, it is concluded that the age of employees does not have any effect employment status.

**Question-B**

**TWO WAY ANOVA**

Analysis of Variance between two groups, also known as ANOVA. Two-way ANOVA helps to identify the mean difference between the groups of two independent variable (more than 2 groups or variables) and the main purpose of this model is to understand the interaction between the dependent variable and independent variables.

**OBJECTIVE**:

As all the assumptions are confirmed we have performed ANOVA Test which is been illustrated below:

**HYPOTHESIS:** Null Hypothesis (H0) – There is no significant interaction among age groups in permanent and casual employees for Satisfaction level.

**Alternate Hypothesis (H1)** – There is significant interaction among age groups in permanent and casual employees for Satisfaction level.

**VERIFICATION OF ASSUMPTION:**

* Dependent variable considered in this analysis is Continuous variable. (Satisfaction level)
* Two independent variables, Employment status (permanent and casual) and Age group of employees are considered here as two categorical groups each.
* Independence of Observations is satisfied, there is no relationship between values in the different groups. This assumption has been checked with independent sample T test.
* No Significant outliers, no influence of outliers on the output has been observed.

Chart, box and whisker chart

Description automatically generated

Boxplot for Employment Status

Chart, box and whisker chart

Description automatically generated

Boxplot for Age groups

* Verification of normal distribution of dependent variable for every group in independent variables has been conducted with Shapiro-Wilk test and resulted that significance value related from all the group observations are more than 0.05.

1. Employment Status (Permanent and casual)
2. Age group (18 to 30 and 31 to 40 age group are considered here)

Table

Description automatically generated

Table

Description automatically generated

1. **Produce a table giving the sample size, mean, and standard deviation for each group of rats. Is it reasonable to assume homogenous variances?**

**Answer:**

As shown in the below figure, is descriptive statistic for independent variable groups age group and employment status where N denotes for sample size and values for Mean and standard deviation can be seen in respective columns.

**Table

Description automatically generated**

Homogeneity of variances for each combination of two groups in independent variable, will help to identify whether our analysis is correct or not, in other words Goodness of fit. Levene’s test has been used in order to check this assumption, whose result has been shown below in figure.

Table

Description automatically generated

Critical value here is F (7,372) = 0.617 and p-value is 0.742 which is greater than 0.05, shows that there is equal error variance of target variable across other factors. This shows that the assumption of Homogeneity of variance is satisfied.

1. **Run an ANOVA in your software. Report the F statistic with its degrees of freedom and P–value. What do you conclude?**

**Answer:**

Below table helps us to get the details about the significance effect of both the individual independent variable and their interaction on the target/dependent variable.

**Table

Description automatically generated**

We should look at the Sig. Column for the agerecode, employstatus and agerecode\*employstatus.

Firstly, the p-value is 0.586 for agerecode (age groups). So, there is no significant difference in the level of satisfaction different age groups.

Secondly, For Employment status, p is 0.443. Therefore, we fail to reject the null hypothesis and conclude that there is no difference in the level of satisfaction between permanent and casual type of employees.

Finally, P-value is 0.976 for interaction between agerecode and employstatus, p = 0.976. Therefore, there was no statistically significant interaction between agerecode and employstatus on the level of Satisfaction.

The **Degree of Freedom(df)** for agerecode (age\_groups) is 3, for Employment status it is 1 and agerecode\*employstatus group is 3.

**F statistics** for agerecode F (3, 372) = 0.646, p = 0. 586.Therefore, We, fail to reject the null hypothesis that there is no effect of age group on the level of satisfaction.

Secondly, the main effect of employment status was not significant, F (1, 372) = 0.589, p = 0.443, partial eta-squared = 0.002. Therefore, we must fail to reject the null hypothesis that there is no effect of employment status on the level of satisfaction.

Finally, the interaction of age group and employment status was not significant, F (3,372) = 0.069, p = 0.976, partial eta squared = 0.001. Consequently, we fail to reject the null hypothesis that the effect of employment status on the level of satisfaction is the same across all levels of age groups.

1. **Use the Turkey, Bonferroni or other multiple–comparisons post–hoc procedure to determine which pairs of means differ significantly.**

**Answer:**

As shown in below multiple comparison between Turkey HSD and Bonferroni tests significance value for each group in independent variables is higher than p vale that is 0.05. Therefore, **Null hypothesis is not rejected.**

**Table

Description automatically generated**

Means for groups in Homogeneous subsets are displayed below:

**Table

Description automatically generated**

1. **Summaries your results in short report**

**Answer:**

This graph helps to get the information that the Employment status and different age groups among the employees about the satisfaction level does not have significant interaction to present the dependent value.

Chart, line chart

Description automatically generated

**Conclusion**: Since the Significance value is greater than 0.05 which is the P value, therefore the Null hypothesis is true. The results were not statistically significant. Therefore, this analysis results that there is no significant interaction between these factors on the Satisfaction level.