**Dilip Sankar**

**Student ID 22225743**

**MS5105 2022-2023**

**Individual Assignment 1**

In submitting this work, I confirm that it is entirely my own. I acknowledge that I may be invited to an online interview if there is any concern in relation to the integrity of my submission, and I am aware that any breach will be subject to the University's Procedures for dealing with breaches of Exam Regulations. I am aware of what the NUI Galway plagiarism policy entails.

# **Question 1: Is the continuous variable Total Life Satisfaction distributed normally?**

**Statistical Test:** Here, to verify the **Normality**, we will use a continuous variable called overall life satisfaction and map it to a label ID and show the results using

**1. Histogram 2. Boxplot 3. Kolmogorov-Smirnov and Q-Q plot.**

***Fig 1.1 Histogram, Normal Q-Q plot, and Box plot of Total like Satisfaction***

Chart, histogram

Description automatically generated

Chart, line chart, scatter chart

Description automatically generated

* Chart, box and whisker chart

  Description automatically generatedThe histogram is shaped nearly like a bell-shaped figure with a few dips in the center.
* The Q-Q plot of total life satisfaction demonstrates that there is no huge difference between the expected value and the observed value, demonstrating that it is normally distributed.
* The box plot shows a low value of around 5 and a high value of around 35 without any IDs present so there is no need to perform further analysis.

Table

Description automatically generated***Fig 1.2 Shows the mean and 5% trimmed mean Fig 1.3 Kolmogorov-Smirnov Test of Normality***

Table

Description automatically generated

* The Kolmogorov-Smirnov test has a significance=.000 that is less than 0.05 which shows this statistic is significant.
* Even though the Kolmogorov-Smirnov test is not satisfied, we still treat this continuous variable as having a normal distribution because the other tests has succeeded.

**From Fig 1.2** we can see that the mean for total life satisfaction=22.38 and the 5% trimmed=22.52 which seems fairly close. The 5% trimmed implied that we cut off 2.5% from the left and right side of the histogram and we have the mean for the remaining 95%. When we extract the very low and high answers, we get an average of 22.52. We can also see Skewness=-.323 which implies it is skewed toward the right and Kurtosis=-.450 which implies it is a flat curve. **Final thoughts:** Given that the majority of the results support this, we are assuming that total life satisfaction is normally distributed.

# Question2: Arethe **Perceived Stress Scale items correlated with each other and are they reliable?**

**Statistical Test:** Here, to test if that items are reliable and can be constructed together, we are using **Reliability Analysis** which should show **Cronbach’s Alpha** value of at least 0.7

***Fig 2.1 Reliability Stats with Cronbach’s Alpha Fig 2.2 Item Stats table***

Table

Description automatically generatedTable

Description automatically generated

The item statistics show all ten items of the perceived stress scale along with mean and Standard deviation and we can see that Cronbach’s alpha=0.360 which is < 0.7 indicating the items don’t have good reliability and correlation.

***Fig 2.3 Correlation matrix shows how the level of total perceived stress are correlated with each other***

Table

Description automatically generated

From Correlation matrix we can see that each item completely corresponds with itself, as seen by the inter-item correlation matrix. When we dig deeper, we find that pss1 and pss2 have a correlation of 0.445, which looks low and suggests that they are not related. According to statistics, if the value is higher than 0.5, we may say that the variables are related, but if it is lower than 0.5, they are not.

From the Item-total statistics when you look at the “Cronbach’s alpha if the item deleted” column we can incur that since none of the item’s values is greater than 0.7 then deleting them will not impact the overall Cronbach’s alpha value.

***Fig 2.4 Item total stats table***

Table

Description automatically generatedFor example, when we delete the value of pps1 then Cronbach’s alpha value will drop down to 0.231 so there will not be any significant change. **Final thoughts:** Cronbach’s alpha <0.7, the items are not correlated and cannot to folded together to form a single construct.

# **Question3: Do more men complete higher education than women and are there a significant difference between gender taking higher education?**

**Statistical Test:** In this scenario, we are utilizing **Chi Square** since we are considering two categorical variables such as higher education completed and gender

***Fig 3.1 Gender Vs Higher education completed summary***

Table

Description automatically generatedThere are no unsolved cases, and all of the questions have been addressed (439). The crosstab displays the proportion of men and women with higher education degrees (from primary – to Postgraduate)

Table

Description automatically generatedTable

Description automatically generated***Fig 3.2 Show Chi-square test with Pearson value as well as Phi and Cramer’s V value***

The value of chi-square=15.361 and the significance=0.009 which is lesser than p=0.05 which means its significant. We do not have yates’ continuity as it’s not a 2x2 table. Since p <0.05 it means the proportion of males who go to higher education is significantly different from the proportion of females who goes to higher education. The Phi and Cramer’s V also support the point that the values are significant as we can see the approximate significance for both the tests are lesser than 0.05.

**Final thoughts:** Based on the significance value in Pearson chi-square and Phi and Cramer’s V we can conclude that there is significant difference in gender taking higher education

***Fig 3.2 Crosstab of Gender Vs Higher education completed***

Table

Description automatically generated

The degree of freedom shows each person could have completed five levels of higher education or vice versa. This means males/females can complete either high school or training or both.

# **Question4: How does two continuous independent variable total life satisfaction and total optimism measure one dependent variable total self-esteem and which IV has more influence on DV?**

**Statistical Test:** We are using **multiple regression** to determine which of the two independent variables, total life satisfaction, and total optimism, contributes more towards the dependent variable total self-esteem.

Based on the correlation table we can see the two IVs correlate with the DV in a very fair manner with 0.488 and 0.565 respectively which is not very high.

***Fig 4.1 Correlation table***

Table

Description automatically generated

In the coefficients table, we are looking for the tolerance and VIF where tolerance is 0.767 which is >0.1 and VIF is 1.304 which is <10 in turn suggests that the variables do not have multi-linearity.

***Fig 4.2 Coefficients table shows the tolerance and VIF of the two independent variables***

Table

Description automatically generated

Chart, line chart

Description automatically generated***Fig 4.3 Expected Vs Observed Probability P-P plot Fig 4.4 Scatterplot***

Chart, scatter chart

Description automatically generated

The normal probability method between the observed cumulative probability vs expected cumulative probability plot demonstrates that it is extremely well aligned in a straight line and the scatterplot demonstrates that there are clusters created around the zero mark.

The amount of variance in the DV predicted by the individual variables can be found in the adjusted R-square table,

***Fig 4.4 Show Adjusted R square value to determine the dependency***

Table

Description automatically generated

The adjusted R-square shows a value of 0.337, indicating that the two independent variables, total life satisfaction, and total optimization, account for 37.7% of the variance in the variable total self-esteem, leaving the remaining 42.3% reliant on other variables. **Final thoughts:** The ANOVA tables below show significance =0.000 which is less than p=0.05 which means the IVs are good predictors. The coefficient table shows the beta values, with total optimism equaling 0.429 and total life satisfaction equaling 0.281. Both values are statistically significant, however, total optimism has a 42.9% contribution that is more favorable to total self-esteem.

# **Question5: How do multiple variables in the perceived stress associated with each other based on correlation?**

**Statistical Test:** Since the perceived stress scale comprises more than five items, **factor analysis** is being used to do the study.

The correlation matrix shows that each item correlates with itself, and we have many correlations>0.3. The Kaiser-Meyer Olkin measure is 0.881 which is> 0.6 and Bartlett’s test shows that these values are significant.

***Fig 5.1 Correlation matrix of perceived stress Fig 5.2 Kaiser-Meyer-Olkin and Bartlett’s Test***

**Table

Description automatically generated**Table

Description automatically generated

Two items with eigenvalues greater than one and percentages of 43.65% and 12.70%, respectively, may be seen when we look at the principal component analysis. In contrast, there is just one primary component when examining the scree plot, and other components are grouped.

***Fig 5.3 Tables shows Eigen values and variance Fig 5.4 Scree plot***

Chart, line chart

Description automatically generated Table

Description automatically generated

***Fig 5.5 Shows stress levels of different components***

Table

Description automatically generated

**Final thoughts:** All of the objects in the component matrix load onto component 1, which has two components. As a result, we may say that there is just one construct for the perceived stress scale.

# **Question6: Is there any significant difference in total life satisfaction based on gender?**

**Statistical Test:** We will use the **T-test** as a statistical tool for our study to determine how total life satisfaction, which is a continuous variable vs. categorical variable, differs for males and females.

Levene’s test in the independent samples test shows a significance of 0.401 which is > 0.05 which shows Levene’s test is insignificant.

***Fig 6.1 Significance shown in the independent sample test table***

Table

Description automatically generated

From the group statistics, we can see that the mean value and standard deviation of females (M=22.90, SD=6.91) are more than males (M=21.67, SD=6.52) but there is no significant difference in life satisfaction. **Final thoughts:** The t-test (434) =-1.881 and t(408.5)=-1.897 shows a significant value of 0.061 and 0.059 respectively which is >0.05 so it means there is no significant difference in total life satisfaction based on gender

***Fig 6.2 Mean and Standard deviation of Male and female***

Table

Description automatically generated

# **Question7: Will there be any significant differences in the total life satisfaction if the person is a smoker or not?**

**Statistical Test:** Here we are analyzing total life satisfaction based on if a person smoke or not which is a categorical variable (<=2 levels) vs a continuous variable. Hence, we will use the **T-test as a statistical tool** for our analysis.

In the independent sample test, Levene's test yields significance = 0.422, which is >0.05 and indicates that it is insignificant

***Fig 7.1 Levene’s test shows the significance***

Table

Description automatically generated

***Fig 7.2 Shows the mean and standard deviation of total life satisfaction***

Table

Description automatically generated

According to the group statistics, there is a considerable difference in the mean total life satisfaction between smokers (M=21.58, SD=6.91) and nonsmokers (M=22.58, SD=6.74). **Final thoughts:** Additionally, the t-significant test's t (431) =-1.252 value is 0.211, which is greater than 0.05, demonstrating that the values are not significantly different.

# **Question8: How does total self-esteem differ based on different age groups?**

**Statistical Test:** We are using **one-way ANOVA** as a statistical method for this research since there are numerous levels of a categorical independent variable (> 2 levels), where independent categorical variable, representing age groups, and a continuous dependent variable, representing self-esteem.

***Fig 8.1 Levene’s significance Fig 8.2 ANOVA table shows significance < 0.05***

Table

Description automatically generatedTable

Description automatically generated

The subjects are divided into 3 age groups: **GROUP1** contains people aged 29 and below, **GROUP2** contains people between the ages of 33 and 44, and finally, **GROUP3** contains people aged greater than 45.

Levene’s statistic is 0.237, which, as we can see across all the rows of mean and median, is not significant. Additionally, the ANOVA's F (2,435) =4.50 and significant value = 0.012 which is < 0.05 demonstrates that there are differences because the value is significant.

***Fig 8.3 Shows the mean and standard deviation of different age groups***

Table

Description automatically generated

The mean and standard deviation between the age groups don't differ much, therefore we will examine post hoc testing to determine where the difference is.

Table

Description automatically generated***Fig 8.4 shows Tukey’s post hoc test Fig 8.5 Shows a measure of significant difference***

Table

Description automatically generated

When you look at the values with \* in the mean difference column we can find the difference. There is a significant difference between GROUP1 (18-29 years) with M=32.60 and SD=5.586 and GROUP3 (45+ years) with M=34.50 and SD=5.15 where mean difference=1.906 and vice versa. But we can also see that GROUP2(33-44 years) doesn’t have any significant difference with both group1 and group3.

To measure how huge, the difference is we can analyze the ANOVA effect sizes and look at the eta squared. The point estimate=0.02 which means even though we have a significant difference based on the eta-squared value it’s very minimum.

# **Question9: How do a person’s marital status and smoking affect total life satisfaction and does a smoker moderate relationship between marital status and life satisfaction**

**Statistical Test:** Here we have considered a dependent continuous variable total life satisfaction vs two independent categorical variables such as smoker and marital status with more than 2 levels. Hence, we are **using 2-way ANOVA** as a statistical tool for this analysis.

**Table

Description automatically generatedTable

Description automatically generated*Fig 9.1 Shows Levene’s significance value Fig 9.2 Test between subjects shows interaction based on the significance***

Levene’s test has a value of 0.870 which proves it is insignificant and the tests of between-subjects effects will show if there is any interaction. The significance of marital status\*smoke F (7,433) =0.508 and p= 0.829 which is not significant hence we can move on to the main factors. Marital status F (7,433) =2.77 and p=0.008 which is significant and smoke F (1,433) =1.06 AND p=0.30 which is not significant. So, life satisfaction doesn’t vary much based on whether the person is smoking or not. But surely there is a difference in life satisfaction based on marital status.

***Fig 9.3 Tukey post hoc test***

Table

Description automatically generated

Table

Description automatically generatedTable

Description automatically generated

The Post-hoc test using Tukey’s HSD shows the below,

1. There is a significant difference in Life satisfaction of **single** and **married first time** with a mean difference of 2.56 where single has M=20.83 & SD=6.721 and married for the first time has M=23.39 & SD = 6.78
2. There is a significant difference in Life satisfaction of **married first time** with **single** & **separated** with a mean difference of 2.56 and 7.69 respectively where separated has M=15.70 & SD=5.79
3. There is a significant difference in Life satisfaction of **remarried** with **separated** and **divorced** with a mean difference of 9.03 and 5.78 respectively where remarried has M=24.73 & SD=6.03 and divorced has M=18.96 & SD=6.93
4. There is a significant difference in Life satisfaction of **separated** with the **married first time, remarried, and widowed** with a mean difference of 7.69, 9.03, and 11.44 respectively where widowed has M=27.14 and SD=5.33.

We were able to see a difference in total life satisfaction for a few groups but not for many hence the total life satisfaction does not have any significant difference based on gender and marital status

# **Question 10: Present graphically how Total life satisfaction differs between age groups and gender.**

Chart, bar chart

Description automatically generated**Statistical Test:** We are going to use a **Bar Chart** to show how total life satisfaction differs between each age group and gender.

The graph shows mean total life satisfaction of male and female between 3 age groups. We can see that females between ages 18-29 and 45+ have the highest life satisfaction. Whereas women between ages of 30-44 shows a slight decrease in total life satisfaction compared to the other 2 age groups. And the men between all the 3 age groups share almost the same number and there is no big difference in them.