**HUYNH DO**

**Module 4A, Part 1/2**

1. **Specify which variable is the dependent and independent variable:**

The analysis is to investigate if there is a difference in the mean age of participants depending on whether they experienced unemployment for the last 10 years (**Yes/No**). Thus, employment status is the factor that categorizing the responders, while age (A continuous variable)is the outcome and being measured.

With the opening variable explanations, we can construct these 2 variables:

* **Independent Variable** – **UNEMP:** Ever unemployed in the last ten years.
* **Dependent Variable – AGE:** The age of the participant.
1. **Using the variables that you’ve chosen, state the null and alternative hypotheses.**

With the 2 declared above variables, we can construct these 2 Hypotheses:

1. **Null Hypothesis (H₀):**
* There is **no** significant difference in the mean age of individuals who have had unemployment in the past ten years compared to those who have not.
* **Mathematical Representation**: H0​:μYES​ = μNO​
1. **Alternative Hypothesis (H₁):**
* There is a difference in the mean age of individuals who have had unemployment in the past ten years compared to those who have not.
* **Mathematical Representation**: H1:μYES​ <> μNO​

1. **Perform the independent samples t-test and interpret the results.**
* **Group Statistic**



1. **Sample Size (N):**
* **“YES”** group: There are 529 people who experienced unemployment for the last 10 years.
* **“NO”** group: There are 1,026 people who reported haven’t experienced unemployment for the last 10 years.
1. **Mean:**
* The average age of respondents who experienced unemployment ("YES") is 41.55 years.
* The average age of those who did not experience unemployment ("NO") is 52.28 years.
* The “YES” and “NO” Means suggest that individuals who have had unemployment tend to be younger (Y: 41.55 < N: 52.28) than those who have not, indicating a possible age-related trend in unemployment over the last 10 years.
1. **Standard Deviation**
* For the "**YES**" group, the standard deviation is 14.847, indicating that the ages in this group are spread out within the range of [41.55 +- 14.847] years.
* For the "**NO**" group, the standard deviation is 18.320, indicating a broader range of [52.28 +- 18.320] years.
1. **Standard Error of the Mean:**
* Since both groups "YES" = 0.646 and "NO" group = 0.572 are fairly low indicating more precise estimates of the mean.
1. **Group Statistic Findings:**

There’s a significant 10.730 years difference (Almost 11 years apart) in average (Mean) age between the 2 groups. This age difference suggests that younger people may be out of work more frequently, while older people may have more stable employment. Additionally, the larger variability in the age of the "NO" group may reflect a broader age range among those who remained employed.

* **Independent Sample Test**

****

* 1. **Levene's Test for Equality of Variances:**
* The extreme small p-value (< .001) indicates that the both variances “**YES**” and “**NO**” groups are significantly different, leading that the assumption of equal variances cannot be used. Therefore, the "**Equal variances not assumed**" results should be used for interpretation.
* Why should we **reject** the **null hypothesis** (**H₀**)?
1. Extreme Low p-value < 0.001: This number suggests a significant difference in the mean age between individuals who experienced unemployment and those who did not. The mean age difference is statistically significant, with younger individuals more likely to have experienced unemployment.
2. Confidence interval lies on the negative side of 0 (but does not include 0 value): The 95% confidence interval for the mean difference in ages is [-12.422, -9.038] further supporting the rejection of the null hypothesis.

	1. **t-test for Equality of Means**
* The magnitude of the t-Value **(-12.411**):

The absolute (**12.411)** indicates how much the sample mean is likely to fluctuate from the 0, the number 12.411in this exercise is considered a quite large number indicating that there’s a substantial difference between the means ages of 2 groups (“Y” and “N). The **negative sign** means the mean age of “YES” group is lower than the mean age of the “NO” group. In another word, the younger age group is a higher likelihood experienced out of work in the last 10 years.

* From the data table, we can see that both results of the t-test show a very significant p-value < .001 which indicating a significant difference in the mean age between the two groups. Moreover, Levene's test also confirmed the inequality of the 2 variances, thus the "**Equal variances not assumed**" results should be used for more accurate interpretation.
	1. **Mean Difference and Confidence Interval Independent**
* The mean difference is 10.73 indicating that, on average, those who experienced unemployment are 10.73 years younger than those who did not.
* 95% Confidence Interval for the Difference: [-12.422, -9.038], but does not include 0 value, indicates that the true mean difference in age between the two groups lies between 9.038 and 12.422 years younger for those who experienced unemployment.
	1. **Independent Sample Test Findings:**

The mean age of the "YES" group being **41.55 years**, compared to **52.28 years** for the "NO" group indicates a significant difference in average age between those who experienced unemployment in the past decade and those who did not. The results suggest that young adults are more likely to be out of works compared to older individuals, possibly due to factors such as job instability or career transitions, while older group prefers a stable job.

1. **Make sure to assess the test assumption**

The following assumptions must be considered for the Independent Samples t-test:

1. **Assumption of Independence**:

If the employment status is collected randomly and each participant’s age is an independent observation, ­then assumption is satisfied.

1. **Assumption of Normality – Special Notes**:

Although both samples are quite large (“YES”:529 and “NO”:1026) which violates t-test’s requirement (n<30) , however, even if the age distributions are not perfectly normal, the Central Limit Theorem suggests that the distribution of the sample means is approximately normal, making the t-test acceptable and reliable. Furthermore, the “Robustness of the t-Test” is also fairly flexible to violations of normality, particularly when the sample size is large.

1. **Assumption of Scale of Measurement**:

In order to derive meaningful differences between the average ages of the 2 groups, the dependent variable (AGE) should be measured on an interval. From the section 1 (Explanation of variables), Age is a ratio variable, so this assumption is satisfied.

1. **Assumption of Homogeneity of Variance**:

To satisfy the “Homogeneity of Variance”, the 2 variances should be almost equaled, however, from the Levene’s test (**F** = 34.727 and **p** < 0.001), suggesting that they are not equal. This violates the assumption of Homogeneity of Variance. Therefore, in this case, the “**Equal variances not assumed**” results should be used.

1. **Using an error histogram and ­bar chart, visually display the results of the independent samples t-test and explain your findings**

****

 **Histogram**

****

**Error Bar Chart**

1. **Mean Age Difference:**

The mean age for the “YES” group (41.55 years) is visually different from the “NO” group (52.28 years) in the Error Mean Bar Chart. The histograms of the 2 groups also confirms that the 2 variances are not distributed equally, indicating a statistically significant difference between the 2 groups.

The differences, with the “**YES**” group is lower than the “**NO**” group, suggests that individuals who have not experienced unemployment in the last decade tend to be older on average than those who have, and specifically it is a **negative** (inverse) relationship.

1. **Error Bar Mean and Overlap**:

The Error mean bars visually, with quite large gap between the means (about 10.73 years) suggest a statistically significant difference in mean ages between the groups, consistent with the t-test results.

1. **Findings:**

The Histogram and the Simple Error Bar Chart show the age difference associated with unemployment status over the past 10 years. The charts also point out that younger individuals are more likely to have experienced unemployment in the past decade compared to older individuals.