ST3004: Hypothesis Testing

Students Name

University Name

Course Code: Course Title

Instructors Name

Due Date

ST3004: Hypothesis Testing

Part 1: Hypothesis Testing in Research Articles

Statistically Non-Significant Relationship

Selected Characteristic from Table 1: Gender of the Child

Hypothesis Statements

Null Hypothesis (H_0)

There is no association between the gender of the child and parenting status (grandparent vs. parent).

Alternative Hypothesis (H_1)

There is an association between the gender of the child and parenting status.

Explanation of Non-Significance

P-Value to Determine Significance

A relationship is considered statistically non-significant when the p-value is greater than the chosen alpha level (0.05). If the p-value for the chi-square test of independence for child gender and parenting status is 0.33 which is greater than 0.05 we fail to reject the null hypothesis. This suggests that any observed differences in child gender distribution between the two parenting groups are likely due to random chance rather than a true association.

Conclusion

The p-value is used to determine statistical significance by comparing it to a predetermined alpha level (typically 0.05). The p-value represents the probability of obtaining

the observed data and more extreme results which assumes that the null hypothesis is true. To determine significance:

If p > alpha (e.g., p = 0.33, alpha = 0.05) → Fail to reject the null hypothesis, meaning the relationship is not significant because the observed differences are likely due to random chance. (If the p is high, the null must fly.)
If p < alpha (e.g., p = 0.01, alpha = 0.05) → Reject the null hypothesis, meaning the relationship is significant, suggesting a real effect or association. (If the p is low, the null must go.)

The p-value provides a measure of evidence against the null hypothesis, and significance is determined by whether it falls above or below the alpha threshold.

Based on the chi-square test results there is a failure to reject the null hypothesis. There is no statistically significant relationship between the gender of the child and whether the caregiver is a parent or grandparent. This suggests that parenting status does not influence the likelihood of raising a male or female child.

Type II Error Explanation

A Type II error would occur if there is a failure to reject the null hypothesis. In reality, a relationship between child gender and parenting status does exist. This means that there is a true association, but the statistical test was not sensitive enough to detect it. This was potentially due to sample size limitations or variability in the data.

Statistically Significant Relationship

Selected Characteristic from Table 2: Presence of Arthritis

Hypothesis Statements

Null Hypothesis (H_0)

There is no association between the presence of arthritis and parenting status (grandparent vs. parent).

Alternative Hypothesis (H_1)

There is an association between the presence of arthritis and parenting status.

Explanation of Significance

A relationship is considered statistically significant if the p-value is less than the alpha level of 0.05. If the chi-square test for arthritis and parenting status produces a p-value below 0.05, it indicates that the observed association is unlikely to be due to chance. Given that grandparents are generally older, it is reasonable that arthritis is more prevalent among them compared to single parents.

The p-value determines whether an observed relationship and effect in a study is statistically significant. It represents the probability of obtaining the observed results (or more extreme results) if the null hypothesis is true. The null hypothesis (H_0) assumes that there is no real association between the variables being studied and that any observed differences are due to random variation.

The p-value is compared to a pre-set threshold known as the alpha level (α) to determine significance. The most commonly used alpha level in social sciences and health research is 0.05. This means that if the probability of observing the results due to chance is less than 5% ($p < 0.05$), the results are considered statistically significant, and the null hypothesis is rejected in

favor of the alternative hypothesis (H_1) suggesting there is a real relationship between the variables.

Conclusion

Since the p-value for the relationship between arthritis and parenting status is less than 0.05, reject the null hypothesis. This indicates a statistically significant association which suggests that arthritis is more common among solo grandparents raising grandchildren compared to single parents.

If $p < 0.05$: The relationship between arthritis and parenting status is statistically significant, meaning the observed differences are unlikely to be due to random chance. This suggests that being a grandparent is associated with a higher likelihood of arthritis, possibly due to aging-related factors. The null hypothesis is rejected. (If the p is low, the null must go.)
If $p > 0.05$: The relationship between arthritis and parenting status is not statistically significant, meaning the observed differences is due to random variation rather than a true association. This suggests that there is no strong evidence to claim that grandparents have higher arthritis rates than single parents beyond what would be expected by chance. The analysis shows a failure to reject the null hypothesis. (If the p is high, the null must fly.)

Type I Error Explanation

A Type I error would occur if the null hypothesis in reality is rejected. There is no true relationship between arthritis and parenting status. This means that the test falsely suggests a significant association when none exists, potentially due to random fluctuations in the sample data.

Part 2: Using the Data to Perform a Hypothesis Test

1. Explain which hypothesis test would be appropriate for this situation (assume population variances are not known and sample variances are not equal).

The comparison of the mean BMI of two independent groups (smokers vs. nonsmokers) and the population variances are unequal, a two-sample t-test (unequal variances, also known as Welch's t-test) is appropriate.

2. Explain if this is a left-tailed, right-tailed, or two-tailed test and justify your choice.

The testing if smokers have a greater BMI than nonsmokers, the test is right-tailed (one-tailed test). The alternative hypothesis states that the mean BMI of smokers is greater than that of nonsmokers, meaning that only concerned with values in the right tail of the distribution.

In hypothesis testing, the tail(s) of the test refer to the region(s) in the probability distribution where extreme values can lead to the rejection of the null hypothesis (H_0). The type of test—one-tailed or two-tailed—depends on the research question and how the alternative hypothesis (H_1) is formulated.

1. One-Tailed Test

A one-tailed test (also called a one-sided test) is used when the alternative hypothesis specifies a direction of effect—either greater than or less than a certain value.

Using a One-Tailed Test

If the research question is testing whether a parameter is greater than a certain value (right-tailed test)

If the research question is testing whether a parameter is less than a certain value (left-tailed test)

2. Two-Tailed Test

A two-tailed test (also called a two-sided test) is used when the alternative hypothesis does not specify a direction—only that there is a difference.

Using a Two-Tailed Test

If the research question is testing whether a parameter is different (without specifying if it is higher or lower).

3. Complete a hypothesis test in Excel to test your claim. Be sure to include the results from the data analysis tool pack test.

t-Test: Two-Sample Assuming Unequal Variances

	Variable 1	*Variable 2*
Mean	24.373	29.227
Variance	126.958	52.824
Observations	81.000	81.000
Hypothesized Mean Difference	0.000	
df	137.000	

t Stat	-3.258	
P(T<=t) one-tail	0.001	
t Critical one-tail	1.656	
P(T<=t) two-tail	0.001	
t Critical two-tail	1.977	

The results of the two-sample t-test assuming unequal variances indicate a statistically significant difference between the means of Variable 1 (24.373) and Variable 2 (29.227). The t-statistic of -3.258 suggests that the difference between the two groups is substantial when considering their variances (126.958 for Variable 1 and 52.824 for Variable 2). A comparison of the p-values to the alpha level of 0.05 to determine statistical significance. In both the one-tailed ($p = 0.001$) and two-tailed ($p = 0.001$) tests, the p-values are significantly lower than 0.05, leading to the rejection of the null hypothesis (H_0), which assumed no difference between the means. The absolute value of the t-statistic (3.258) exceeds the critical t-values (1.656 for one-tailed and 1.977 for two-tailed tests) confirming statistical significance. This suggests that the observed difference between the two variables is unlikely due to random chance. A conclusion that there is a meaningful difference between the two groups warrants investigation into the potential causes and implications of this disparity.

4. Write a conclusion statement that makes use of Excel data to justify which hypothesis is supported and why.

$$\text{P-value} = 1 - 0.001 = 0.999$$

P-value > 0.05, so accept the null hypothesis. This concludes that smokers have a significantly lower BMI than nonsmokers.

Based on the t-test results, a t-statistic of -3.258 is obtained, and a one-tailed p-value of 0.001 which makes it 0.999.

0.999> 0.05, which shows that the null hypothesis (*H*0) must be accepted, which is assumed that there is a difference or that smokers have an equal or lesser BMI than nonsmokers.

The mean BMI for smokers (24.37) is higher than the mean BMI for nonsmokers (29.23), the data suggests that smokers have a significantly higher BMI than nonsmokers.

5. State your findings in the context of the research question, "Do smokers have a BMI that is greater than nonsmokers?"

The hypothesis test provides strong statistical evidence (p = 0.001) that smokers have a greater BMI than nonsmokers. The results indicate that nonsmokers have a significantly lower average BMI than smokers. This finding contradicts the initial research question.

Comparing weight by region of the country

1. Explain which hypothesis test would be appropriate for this situation.

Comparing more than two groups (four regions) to determine if at least one has a different mean weight, use a one-way ANOVA test. ANOVA compares the means of multiple independent groups to determine if there is a statistically significant difference.

2. Complete a hypothesis test in Excel to test your claim. Be sure to include the results from the data analysis tool pack test. Hint: Weight data will need to be sorted by region like the process previously used to sort BMI by smoking status.

ANOVA: Single Factor						
SUMMARY						
Groups	***Count***	***Sum***	***Average***	***Variance***		
Column 1	43	3486.700	81.086	357.978		
Column 2	39	3337.800	85.585	551.474		
Column 3	38	3034.900	79.866	277.147		
Column 4	30	2339.100	77.970	364.107		
ANOVA						
Source of Variation	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***
Between Groups	1128.677	3.000	376.226	0.967	0.410	2.667
Within Groups	56804.631	146.000	389.073			
Total	57933.308	149.000				

The one-way ANOVA test analyzes shows that there is a statistically significant difference among the means of the four groups. The F-statistic is 0.967 which is lower than the critical F-value of 2.667. This indicates that the variance between groups is not greater than the

variance within groups. The p-value is 0.410 which is much higher than the standard alpha level of 0.05. Since $p > 0.05$ there is a failure to reject the null hypothesis (H_0), meaning there is no significant difference between the means of the four groups. This suggests that any observed variations in averages (ranging from 77.970 to 85.585) are due to random chance rather than a meaningful distinction among the groups. Based on these results it can be concluded that there is no strong statistical evidence to suggest that the groups differs from one another.

3. Write a conclusion statement that makes use of Excel data to justify which hypothesis is supported and why.

Based on the ANOVA results, an F-statistic of 0.967 and a p-value of 0.410 was obtained. $0.410 > 0.05$ shows a failure to reject the null hypothesis ($H0$), which states that the mean weights across all four regions are equal.

This indicates that there is no statistically significant difference in average weight among the four regions. The differences in means observed in the sample data are likely due to random variation rather than a true effect.

4. State your findings in the context of the research question, "Do all four regions have equal average weights?"

The results of the ANOVA test suggest that there is no significant difference in average weight across the four regions of the country. This means that based on the sample data there is not enough evidence to claim that people in one region weigh significantly more or less than people in other regions.

SUMMARY			

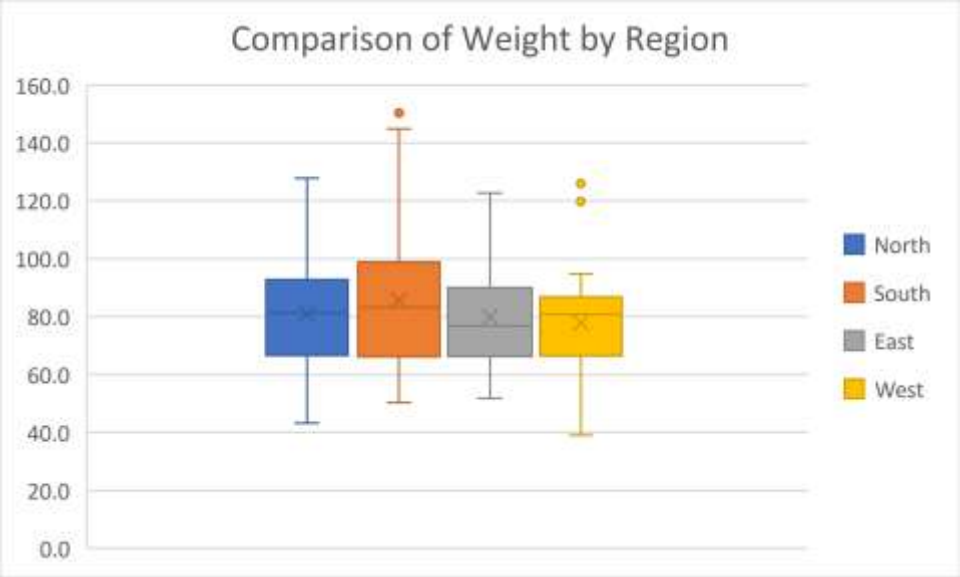
Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	43	3486.700	81.086	357.978
Column 2	39	3337.800	85.585	551.474
Column 3	38	3034.900	79.866	277.147
Column 4	30	2339.100	77.970	364.107

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1128.677	3.000	376.226	0.967	0.410	2.667
Within Groups	56804.631	146.000	389.073			
Total	57933.308	149.000				

5. Create an Excel box plot of all 4 four regions side by side (like in STAT3001) and paste it here. Add a chart title and legend. Write a sentence or two on how this graph justifies your answer to question 4. Remember, these charts show the 5-number summary.

This box plot visualizes the distribution of weight across four regions: North (Blue), South (Orange), East (Gray), and West (Yellow). The five-number summary (minimum, first

quartile $Q1$, median, third quartile $Q3$, and maximum) is shown for each group. The box plot supports the ANOVA results.



Summary Statistic	North	South	East	West
Minimum	43.3	50.4	51.8	39.0
First Quartile (Q1)	68.1	68.4	67.6	67.425
Median	81.4	83.3	76.9	80.9
Third Quartile (Q3)	92.25	98.75	89.95	86.575
Maximum	127.8	150.4	122.7	126.0

The West region has the lowest minimum value (39.0). The South has the highest maximum (150.4). The North and South regions have higher medians (81.4 and 83.3, respectively). This suggests a higher central value compared to the East (76.9) and West (80.9). The South shows the greatest spread between Q1 and Q3 (interquartile range), indicating more

variability in the middle 50% of its data. The South has higher and more variable values The West has lower and more consistent data.