**Your name here: Krista Miller**

Once you have completed this assignment, please submit it as a PDF or Word document to the Week 1 Participation Folder.

**Question 1: Give three examples of a population and a sample from that population.**

Example 1 Population: All students currently enrolled in medical school

Example 1 Sample: 1000 medical school students selected at random from 100 programs in the United States.

Example 2 Population: Total sales receipts from a business in the year 2021.

Example 2 Sample: 300 sales receipts, selected at random for each month to audit 2021

Example 3 Population: All boxes of Frosted Flakes produced in April 2021

Example 3 Sample: 20mg of cereal drawn from 500 boxes of Frosted Flakes produced in April 2021.

**Question 2: What measure/s of central tendency - mean, median, or mode - would make sense to compute for the following variables?**

Variable 1: This variable contains the color of a car purchased at a car dealership

Example values: “Blue”, “Red”, “Blue”, “Black”, “Black”, “White”, “Red”...

Measure/s: Mode (car colors are categorical)

Variable 2: This variable contains the outstanding balances of student loan accounts

Example values: $3500; $4320; $510,000; $2190; $3780; $2310...

Measure/s: Median ($510,000 is an outlier and can skew the mean).

Variable 3: This variable contains the number of courses a student is taking in a term

Example values: 1, 2, 2, 1, 4, 3, 3, 3, 2, 1, 1, 3, 4, 2, 5, 3…

Measure/s: Mean (the maximum number of courses a student can take in a term is likely fixed by the university)

Variable 4: This variable contains the number of languages a person speaks proficiently

Example values: 1, 2, 3, 2, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 2, 1, 4, 1, 1....

Measure/s: Mode (The values for spoken languages is an integer and not on a continuous scale)

**Question 3: Match the null/alternative hypothesis statements to the inferential analysis method**

|  |  |
| --- | --- |
| **Hypothesis statements** | **Inferential analysis method** |
| H0: The pre-treatment and post-treatment means are equal  H1: The pre-treatment and post-treatment means are different | Dependent-samples t test (a.k.a. Paired-samples t test) |
| H0: Variable A and B are independent  H1: Variable A and B are not independent  Note: All cells have an expected value of at least 5. | Chi-square test of independence |
| H0: The control group and experimental group means are equal  H1: The control group and experimental group means are different | Independent-samples t test |
| H0: Variable A is consistent with a specified distribution  H1: Variable A is not consistent with a specified distribution | Chi-square test of goodness of fit |
| H0: Variable A and B are independent  H1: Variable A and B are not independent  Note: Some cells have an expected value of less than 5. | Fisher’s exact test |

**Choices of analysis method:**

A. Chi-square test of independence

B. Chi-square test of goodness of fit

C. Fisher’s exact test

D. Independent-samples t test

E. Dependent-samples t test (a.k.a. Paired-samples t test)

**Question 4: P-values**

Let’s say that you’ve conducted an independent-samples t test and obtained a p-value of 0.04. What is the meaning of this p-value? Note that your answer should not include anything about rejecting or failing to reject the null hypothesis.

Your answer here: The independent samples t-test compares the difference in the means from the two groups to a given value. The t-statistic and its associated p-value is calculated under the assumption that the sample comes from an approximately normal distribution. The p-value of 0.04 is evidence that the mean is different from the hypothesized value.

**Question 5: The ANOVA table**

A personal trainer wants to determine the effectiveness of five different running training programs on improvement in VO2 max (a measure of cardiovascular fitness). She recruits 40 healthy women who have no previous experience in running, each of whom completes their training program (i.e., no missing data). The sum of squares for the training factor is 700, and the total sum of squares is 1000. Using this information, do three things:

1. State the null and alternative hypotheses being tested.

Null: The mean VO2 max improvement is the same for each group.

Alternative: The mean VO2 max improvement is different for at least one group.

1. Use the information from the prompt to complete the ANOVA table by hand.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source** | **SS** | **DF** | **MS** | **F** |
| **Training** | 700 (SSA) | 5-1= 4  (k-1) | 700/4= 175  SSA/(k-1) | 175/8.57= 20.42  MSA/MSE |
| **Error** | 300 (SSE) | 40-5= 35  (n-k) | 300/35= 8.57  SSE/(n-k) |  |
| **Total** | 1000 (SST= SSA + SSE) | 40-1 = 39  (n-1) |  |  |

1. Interpret the result in context of the research question.

The F statistic is very large, and the p-value in this case is less than 0.05. The null hypothesis is rejected. It is concluded that the mean improvement in VO2 max is significantly different for at least one running group.