Epi 3 Check on Learning (COL) Quiz

COL quizzes are *NOT* intended to be extremely challenging. Rather, the goal of COL quizzes are to simply assess students’ basic comprehension of the assigned materials and provide them with feedback early and without high stakes.

**Quiz Instructions:** Please complete this check on learning quiz after reviewing all of the required readings for this module. You may take this quiz as many times as you like.

# 

# Question ideas:

* Placeholder

# Q1. Dependent/Independent Variables

[Multiple Dropdowns]

A group of researchers is performing a case-control study to identify if the physical characteristics of the neighborhood play an important role in the prevalence of obesity in preschool children. In this scenario, the dependent/outcome variable is [Obesity] and the independent/predictor/explanatory variable is [Neighborhood].

Possible answers for [Obesity]

|  |  |
| --- | --- |
| ✅ | Obesity |
|  | Neighborhood Characteristics |
|  | Preschool children |

Feedback:

Correct! In this case, obesity is the variable that we are trying to understand, explain, and/or predict. Therefore it is considered the dependent/outcome variable.

In this case, obesity is the variable that we are trying to understand, explain, and/or predict. Therefore it is considered the dependent/outcome variable.

In this case, obesity is the variable that we are trying to understand, explain, and/or predict. Therefore it is considered the dependent/outcome variable. Preschool children just happens to be the population of interest.

Possible answers for [Neighborhood]

|  |  |
| --- | --- |
|  | Obesity |
| ✅ | Neighborhood Characteristics |
|  | Preschool children |

Feedback:

Not quite. Remember, we suspect that neighborhood characteristics partially determine or explain the outcome - in this case, obesity.

Correct! We suspect that neighborhood characteristics partially determine or explain the outcome - in this case, obesity.

Not quite. Remember, we suspect that neighborhood characteristics partially determine or explain the outcome - in this case, obesity.

# Q2. Relationship between two numeric variables

Which of the following measures can be used to describe the relationship between two numeric variables?

|  |  |
| --- | --- |
| ✅ | Pearson's correlation coefficient |
|  | Sample mean |
|  | Sample median |
|  | All of the above |

Pearson's correlation coefficient can be used to measure the linear relationship between two continuous variables.

A single sample mean can tell us something about an individual continuous variable, but it doesn't tell us anything about its relationship to another variable.

A single sample median can tell us something about an individual continuous variable, but it doesn't tell us anything about its relationship to another variable.

# Q3. Read frequency table

[Nuemerical Answer]

Given the frequency table below, what proportion of participants reported that they are in very good general health (please give the full, exact answer)?

Table

Description automatically generated

0.475

The marginal proportion of respondents who reported that they are in very good was 0.475. The answer comes from the bottom row (margin) of the third column.

# Q4. Read frequency table

[Nuemerical Answer]

Given the frequency table below, what proportion of participants with good general health reported that they do not have any personal doctor? Note that in this table personal doctor is the row variable and general health is the column variable.

Table

Description automatically generated

0.4

8 out of the 20 (0.400) participants with good general health reported that they do not have any personal doctor. This answer is the fourth number down (column percent) in the cell at the intersection of column 4 and row 2 of the summary table.

# Q5. Relationship between two numeric variables

You are interested in understanding if there is a relationship between student GPA's and student's GRE scores. Of the following choices, what is the best measure of the relationship between GPA and GRE scores?

|  |  |
| --- | --- |
| ✅ | Pearson’s Correlation Coefficient |
|  | Comparison of mean GPA and mean GRE score |
|  | 2 X 2 Contingency Table |
|  | Bar chart |

Of these options, Pearson’s Correlation Coefficient is the best measure of the linear relationship between two continuous variables.

Comparison of mean GPA and mean GRE score would give you some information about GPA and GRE individually but would tell you nothing about the relationship between them.

2 X 2 contingency tables are useful for describing the relationship between two categorical variables, but not useful for continuous variables. You can categorize continuous variables and put them into contingency tables, but then they are no longer continuous variables.

Bar charts can be useful for graphically describing categorical variables. They are not useful for graphically describing continuous variables.

# Q6. Interpret correlation coefficcient

[Matching]

Match the following correlation coefficient values with the correct interpretation:

|  |  |
| --- | --- |
| r = 1 | A perfect positive linear relationship |
| r = -1 | A perfect negative linear relationship |
| r = 0 | No linear relationship |
| r = 0.8 | A strong positive linear relationship |
| r = -0.2 | A weak negative linear relationship |
| r = -0.9 | A strong negative linear relationship |
| r = 1.2 | An impossible value |

# Q7. Interpret scatter plot

Please interpret the graph below:

Chart, line chart, scatter chart

Description automatically generated

|  |  |
| --- | --- |
|  | Age and BMI most likely have a positive correlation |
|  | Age and BMI most likely have a negative correlation |
| ✅ | Age and BMI most likely uncorrelated |
|  | It’s impossible to draw a conclusion about the correlation between age and BMI from this graph. |

The random scatter of the points around this chart indicates that there is most likely no correlation between age and BMI in this sample.

# Q8. Code for summary stats

Graphical user interface, application, table

Description automatically generated with medium confidence

Which of the following code chunks below produced the table of summary statistics above?

|  |  |
| --- | --- |
|  | df %>%  group\_by(age\_group) %>%  summarise(  n = n(),  mean = median(score),  sd = sd(score),  min = min(score),  max = max(score)  ) |
|  | df %>%  summarise(  n = n(),  mean = mean(score),  sd = sd(score),  min = min(score),  max = max(score)  ) |
|  | df %>%  group\_by(age\_group) %>%  mutate(  n = n(),  mean = mean(score),  sd = sd(score),  min = min(score),  max = max(score)  ) |
| ✅ | df %>%  group\_by(age\_group) %>%  summarise(  n = n(),  mean = mean(score),  sd = sd(score),  min = min(score),  max = max(score)  ) |

This code is incorrect because it uses the median function instead of the mean function.

This code is incorrect because it is missing group\_by(age\_group).

This code is incorrect because it uses the mutate function instead of the summarise function.

Correct!

# Measures of Association

|  |  |  |
| --- | --- | --- |
| The odds ratio is typically a good approximation of the relative risk when the occurrence of the disease being studied is rare. | | |
| X | True | “In a case-control study, the use of the odds ratio to estimate the relative risk is based on the assumption that the disease under study has a low incidence, thus resulting in a small built-in bias… it follows that when the disease that defines case status in a case-control study is sufficiently rare, the estimated odds ratio will likely be a good approximation to the relative risk.” (Szklo & Nieto, 2019, pg. 107) |
|  | False |  |

|  |  |  |
| --- | --- | --- |
| Which of the following measures of association can be defined as the incidence in the exposed divided by the incidence in the unexposed? | | |
| X | Relative Risk | “The relative risk of developing the disease is expressed as the ratio of the risk (incidence) in exposed individuals to that in unexposed” (Szklo & Nieto, 2019, Kindle Locations 2394-2396) |
|  | Odds Ratio |  |
|  | Relative Odds |  |
|  | Attributable Risk |  |

|  |  |  |
| --- | --- | --- |
| In general, the odds ratio tends to yield an estimate closer to 1.0 than the relative risk. | | |
|  | True |  |
| X | False | “In general, the odds ratio tends to yield an estimate further away from 1.0 than the relative risk on both sides of the scale (above or below 1.0).” (Szklo & Nieto, 2019, Kindle Locations 2471-2472) |

|  |  |  |
| --- | --- | --- |
| The rare disease assumption must be satisfied in order for the odds ratio to approximate the relative risk. | | |
|  | True |  |
| X | False | “The rare-disease assumption is irrelevant in situations in which the control group is a sample of the total reference population, 18 which is the usual strategy in case-control studies within a defined cohort.” (Szklo & Nieto, 2019, Kindle Locations 2860-2861) |

|  |  |  |
| --- | --- | --- |
| Which of the following is the recommended arrangement of the classic 2x2 table used extensively in epidemiology? | | |
|  | |  |  |  | | --- | --- | --- | |  | Diseased | Exposed | | Nondiseased |  |  | | Unexposed |  |  | |  |
|  | |  |  |  | | --- | --- | --- | |  | Exposed | Unexposed | | Diseased |  |  | | Nondiseased |  |  |  |  |  |  | | --- | --- | --- | |  | Nondiseased | Diseased | | Unexposed |  |  | | Exposed |  |  | |  |
| X | |  |  |  | | --- | --- | --- | |  | Diseased | Nondiseased | | Exposed |  |  | | Unexposed |  |  | |  |

What are the most common measures of association in epidemiology?

What does a typical 2x2 table look like?

Red people vs. Blue people

Why can’t we just use the RR all the time? (socrative?)

The OR is a biased estimate of association. T/F socrative

Make a socrative question about language: relative, ratio, etc.

Arrangement of the 2X2 table

Matched pair case-control study