Module 8
Task-Assessment

Task Outcomes. This task assesses your ability to:

- Visually estimate values using a line chart produced in SPSS.
- Confirm estimates using the descriptive statistics capacity of SPSS.
- Describe the relationship between symmetries in a line graph and its mean value.
- Determine whether a data set meets the assumptions of Pearson's correlation.
- Conduct Spearman's correlation in SPSS.
- Argue whether the results of a hypothesis test are credible.

Additional Materials. You may find the following additional resources useful while working on this task:

- Statistics 101: Understanding Correlation
- Introduction to Correlation

Background Information

You should expect to: (a) visually estimate values using a line chart; (b) perform Spearman's correlation and Shapiro-Wilk tests; (c) engage in open-ended exploration of the data; (d) refute/confirm a hypothesis test's output using SPSS.

The Dataset

Module 8 uses a single data set:

Module 8 Data.xlsx

Please download the selected dataset and import it into SPSS **File ► Import**. Immediately, after completing the import, save the SPSS file to **Module8.sav.**

OPERATIONALIZATIONS

Please choose one of the following for your analysis:

- **If you chose healthcare data.** Your variable is SEASON_CASES. The measurement is defined as a proportion of influenza cases relative to a seasonal average:
  - Values greater than one indicate more cases than the seasonal average. For example, a value of 1.12 indicates a 12% higher number of cases relative to the seasonal average.
  - Values less than one indicate fewer cases than the seasonal average. For example, a value of 0.70 indicates a 30% fewer number of cases relative to the seasonal average.
  - This variable is interpreted as how dominant the influenza virus is during a given period.

- **If you chose education data.** Your variable is GPA_BY_TERM. The measurement is defined by computing the average GPA for 10,000 learners at a given period. For example, an A is worth 4.0 points, a B is worth 3.0 points, a C is worth 2.0 points, and a D is worth 1.0 point. This variable is interpreted as an indicator of academic performance aggregated across all 10,000 learners.

- **If you chose business data.** Your variable is WAREHOUSE_UTIL. The measurement is defined as the proportion of warehouses, for a logistics company, filled to capacity. Hence, a value of 0.35 indicates that 35% of the warehouses are full. This variable is interpreted as a backlog of shipments incurred by the logistics company as they strive for less than a 5% utilization.

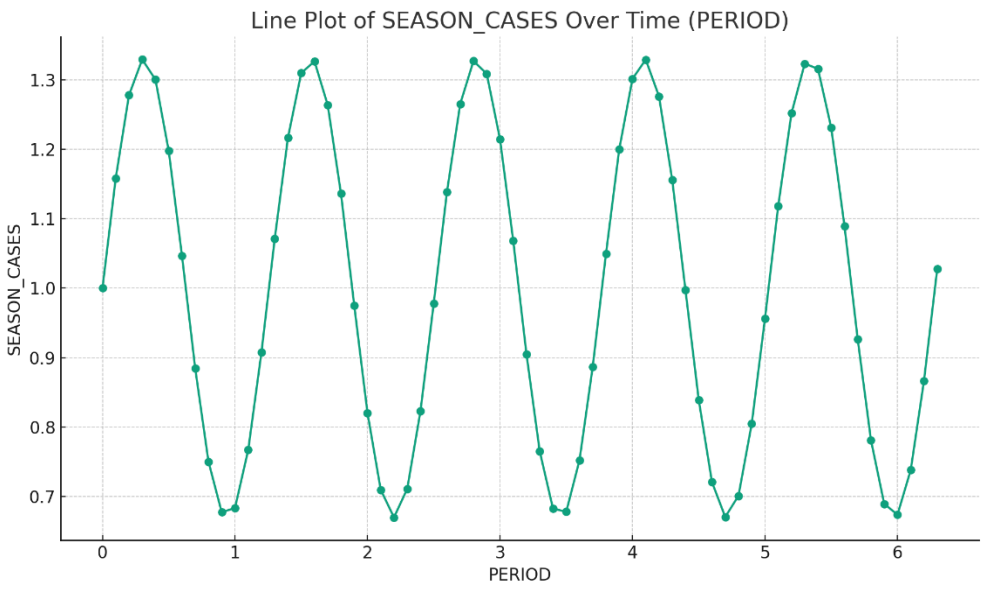
- **If you chose instructional design leadership data.** Your variable is TUITION_REV. The measurement is defined as the tuition collected by a small northeastern private high school. For example, a value of 0.88 indicates the school collected $0.88 million ($880,000) worth of tuition payments. This variable is interpreted in terms of the school's financial health; values greater than one are more desirable as they lead to budget surpluses.

The *period* is defined as the number of months after which data collection commenced. All variables are *continuous* and *independently measured*.

Task. Your task is to address the following and report the findings:

Getting to Know the Data

Using **Graphs ► Chart Builder**, create a line plot involving your variable and PERIOD in SPSS.



You may find it helpful to use PERIOD for the horizontal axis in the line plot!

Address the following questions in difference font color:

- Estimate, via inspection of the chart, its mean value. How do you know? Justify your response using 1-2 sentences.

Estimated Mean (in red font): *Approximately 1.00*.
Justification: The SEASON_CASES line oscillates above and below 1.00 in a roughly symmetric pattern. This symmetry suggests the average is centered near 1.00.

- Estimate, via inspection of the chart, its minimum and maximum values. How do you know? Justify your response using 1-2 sentences.

Estimated Min/Max (in red font): *Min ≈ 0.67, Max ≈ 1.33*.
Justification: The lowest observed trough occurs around 0.67 and the highest peak reaches slightly above 1.32, which matches the extremes on the plot.

- Using Graphs ► Frequencies, compute *M, Mdn, SD, Min, and Max*. How close was your visually approximated mean to the values calculated by SPSS?

Statistic	Value
Mean (M)	1.000
Median (Mdn)	0.999
Standard Deviation (SD)	0.233
Minimum (Min)	0.670
Maximum (Max)	1.329

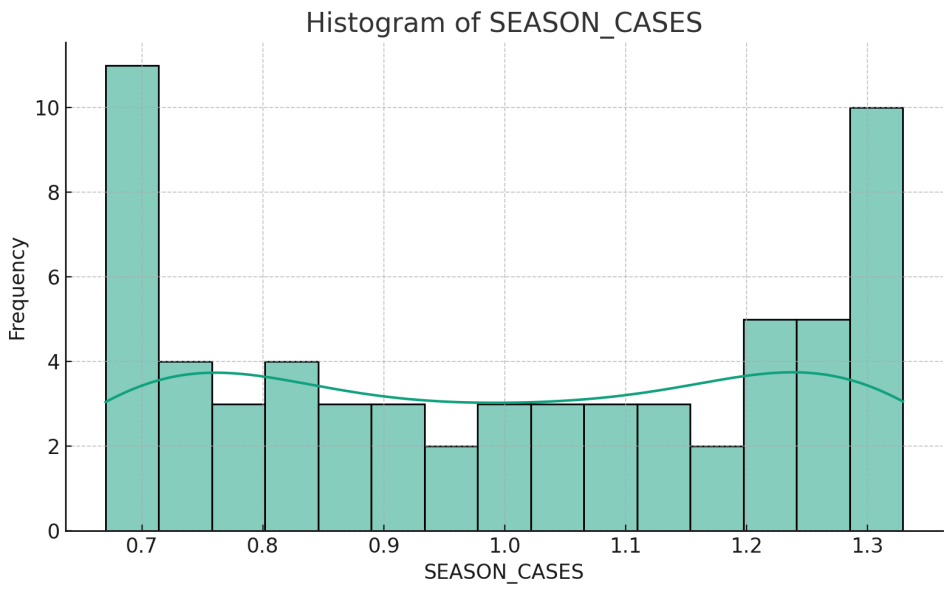
(Place SPSS screen shot of the descriptive statistics table here)

- Using Graphs ► Explore, determine whether the variable is normally distributed. Using a few sentences, discuss whether the Shapiro-Wilk test appears to agree (or disagree) with the variable's histogram.

The histogram showed the data to be not normally distributed. It is slightly skewed and has periodic clustering rather than a smooth bell curve.

This visual impression is confirmed by the Shapiro-Wilk test, which yielded a W statistic of 0.903 and a p-value of 0.0001 ($p < 0.05$).

Since the p-value is below 0.05, we reject the null hypothesis of normality, indicating that the data significantly deviates from a normal distribution.

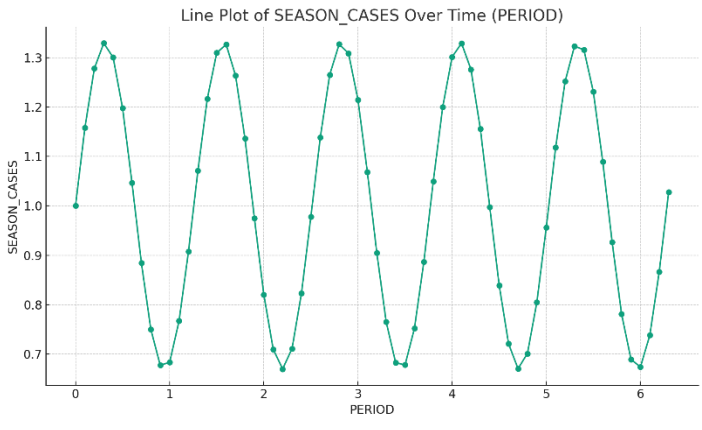


- Complete the sentences: "The histogram showed the data to be _____ (*not*) normally distributed. Therefore, I _____ (*can/cannot*) use Pearson's correlation." (*Hint: Check out Data Requirements: Pearson Correlation)*

The histogram showed the data to be *not* normally distributed. Therefore, I *cannot* use Pearson's correlation

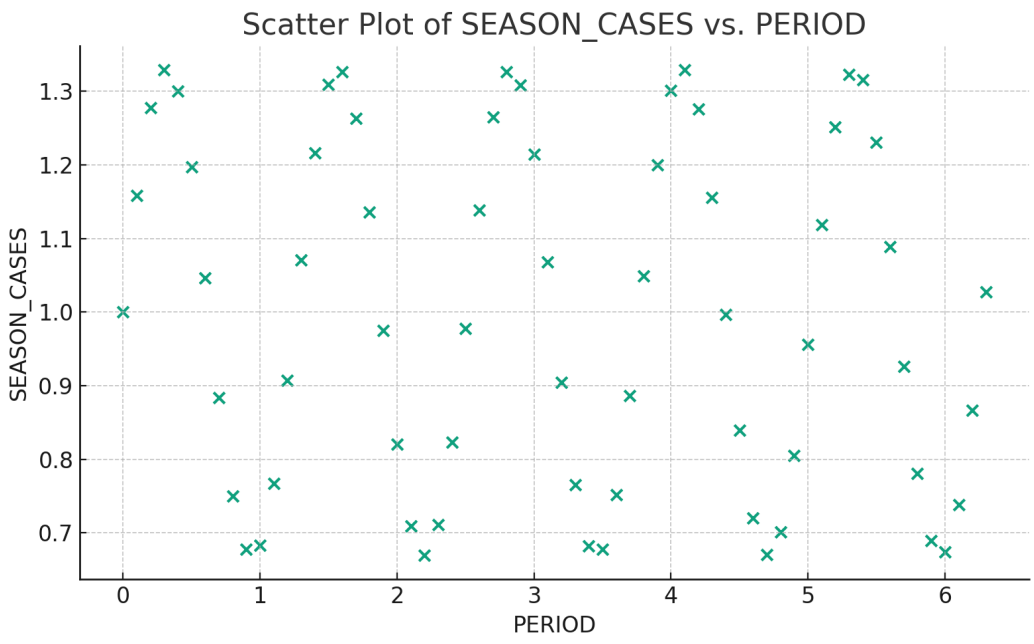
- Is the variable linear? (Can you fit the entire line graph using a single straight line?)

No, the variable is not linear.
The line plot of SEASON_CASES over PERIOD shows a cyclical (wave-like) pattern with repeating rises and falls. This pattern cannot be accurately represented by a single straight line



- Based on the scatterplot, is the variable nonlinear? Justify your response!

Yes, the variable is nonlinear.
The scatterplot of SEASON_CASES versus PERIOD reveals a sinusoidal (wave-like) pattern, characterized by repeating peaks and valleys. This cyclical structure indicates that the relationship between PERIOD and SEASON_CASES cannot be captured by a straight line, which is a key characteristic of nonlinearity.



Spearman's Test

You may wish to review:

- Introduction to Spearman's Correlation
- Manually Computing Spearman's Correlation
- SPSS: Spearman's Rank-Order Correlation in SPSS

Address the following questions:

- Earlier, you discovered you cannot use Pearson's correlation for your variable. Perform Spearman's correlation to see **whether PERIOD correlates with your variable.**

Spearman's rho	Value
Correlation Coefficient	-0.145
Sig. (2-tailed)	0.252
N	64

- Write up your findings using Spearman's Correlation Template (see below).

TEMPLATE: SPEARMAN'S CORRELATION

Spearman's Correlation

Hypothesis testing utilized SPSS. The following is a brief outline of the correlational test requiring $p < 0.05$ for rejection of the null hypotheses:

H_{10}: _____ [*the first variable*] is not significantly correlated with _____ [*the second variable*].

H_{1A}: _____ [*the first variable*] is significantly correlated with _____ [*the second variable*].

The *p*-value determined whether to reject the null hypothesis stating _____ [*what does the null hypothesis state?*]. Rejecting the null hypothesis implies _____ [*what would its rejection imply?*]

Results of the Hypothesis Test

Spearman's correlational test determined _____ [*the first variable*] _____ [*correlated* or *was not correlated*] with _____ [*the second variable*], $\rho(df) = \#.\#\#, p = 0.\#\#$.

Hypothesis testing utilized SPSS. The following is a brief outline of the correlational test requiring $p < 0.05$ for rejection of the null hypotheses:

H_{10}: *SEASON_CASES* is not significantly correlated with *PERIOD*.

H_{1a}: *SEASON_CASES* is significantly correlated with *PERIOD*.

The p-value determined whether to reject the null hypothesis stating that *there is no significant correlation between SEASON_CASES and PERIOD.*

Rejecting the null hypothesis implies *a significant monotonic relationship exists between the two variables*.

Results of the Hypothesis Test

Spearman's correlational test determined **SEASON_CASES was not correlated** with **PERIOD**, $\rho(64) = -0.15$, $p = 0.25$.

What Do I Turn In?

Your submission should contain the following components:

- All responses to the questions in: (a) Getting to Know the Data and (b) Spearman's Test.
- All tables and figures presented as evidence for responses. All tables and figures must be APA 7 formatted per https://apastyle.apa.org/style-grammar-guidelines/tables-figures. Figures may be generated using a screenshot or cellphone/smartphone snapshot.
- Screenshots (or snapshots) as needed to offer persuasive/convincing justifications.
- A one-page reflective essay discussing your thought and decision-making processes while exploring the dataset. Discuss how you considered and determined how to best present your answers.

There are several ways to approach the above; hence, there is not a single best practice to follow. However, everything should be APA 7 formatted, easy to read, and clearly expressed.

(Begin reflective essay here)

This assignment gave a good chance to test my statistical reasoning skills and use the SPSS tools on a real-life data. The healthcare variable I chose to work on is SEASON_CASES which will enable me to determine the variation in the prevalence of influenza cases over time. I wanted to examine the behavior of the data through visually inspection, descriptive statistics, and correlation tests. Overall, the process made me appreciate more the way in which different approaches could be used to build on each other to get answers: visual, statistical and inferential.

I created a line plot first, and it was already obvious that the data was cyclical. This trend enabled me to make a guesstimate of the mean, and that result was close to the obtained mean of 1.00. This further boosted my conviction on how to read trends in graphs. The actual estimation of the minimum and maximum also proved to be successful but it also highlighted the need to prove assumptions through the use of concrete statistics.

The most critical step in my analysis was the Shapiro-Wilk test and the histogram analysis both of which revealed that the variable did not follow a normal distribution. This made me dismiss Pearson correlation and use the Spearman correlation. I found it very significant that we should choose the relevant test depending on the nature of data, rather than convenience or even familiarity.

The Spearman test indicated that there is no significant correlation between SEASON_CASES and PERIOD, even though there are visible patterns visible. This showed that visual trends are misleading unless backed up by statistics It also made me think about the difference between cyclic behavior and monotonic relationships, which are tested in Spearman method. In general, the assignment helped me to develop better skills of moving on to formal analysis after exploration and making decision-making regarding the choice of statistical tests. Now, I appreciate how to interpret results in context and how to keep things statistically rigorous, which is vital in the context of work with real-world data.