Module 6

Segment	Information
Homework/Project Title	Module 6: Probability Laws and Expected Values
Subject/Course	MATH 8070
Estimated Hours	SPSS: ~ 3 hours Portfolio: ~3 hours Proofreading: ~ 2 hours Total: ~8 hours (~10 hours possible)
Teacher Type	Maths/Sciences/Engineering/6S Master Black Belt
Program Level	DBA, EdD, PsyD, Professional Studies (Doctorate), and similar.

Required Background	Role and/or Rationale
SPSS: Descriptive Statistics	Needed to compute means and standard deviations.
Probability	Need to understand the concept of probability as the ratio of desired outcomes to total outcomes.
Operations	Needs knowledge of binary operations: <, >, and =.
Functional Notations	Must understand parameters being passed to functions for CDF.NORMAL() in SPSS.

Curriculum Map	
MLO	This task assesses your ability to: • identify a probability law fitting a need/problem and available data. • apply the correct probability law while determining the probability of an event; • interpret the meaning of (un)likely events; • describe the meaning of *expected value* in statistics; • compute expected values fitting a need/problem and available data; and • interpret the meaning of a computed expected.

Project Summary	Information
Goal	Becoming familiar with datasets mixing categorical and continuous variables.
Audience	Self (primary), lecturer (secondary)
Situation	Analysis of secondary dataset in SPSS.
Observed Performance	Crosstabulation in SPSS Probability Computation in SPSS Applying Probability Laws (SPSS and Manual Computation)

Relevance	
Sales Pitch	Crosstabulations are a useful tool in understanding categorical data mixed with continuous data.
Programmatic	Education: Often used to provide assessment data grouped by student type. Business: Often used to determine the quality of a product and consumer satisfaction. Healthcare: Often used to discuss the factors involved with Triple Aim.
Personal	Often used to categorize personal expenses while finding ways to save money.
Professional	Often used to better understand how groups are affected by a problem involving a good or service.

Key Questions	
KQ1	What are the roadblocks for computing probabilities and expected values in SPSS?
KQ2	Can my learners apply the definition of probability to express sound inferences?
KQ3	Do my learners understand the variable types and operationalization?
KQ4	Do my learners have sufficient APA resources to use while building their portfolios?

Resources	
Faculty	Canvas, SPSS, O365
Technologies	Computer (Desktop)
Materials	Provided in the task
OERs	Provided in the task
Student-Shared	None

Task-Assessment

Task Outcomes. This task assesses your ability to:

- identify a probability law fitting a need/problem and available data.
- apply the correct probability law while determining the probability of an event;
- interpret the meaning of (un)likely events;
- describe the meaning of *expected value* in statistics;
- compute expected values fitting a need/problem and available data; and
- interpret the meaning of a computed expected value.

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

- Addition Rule (or Law): https://ecampusontario.pressbooks.pub/introstats/chapter/3-5-the-addition-rule/
- Complement Rule (or Law): https://ecampusontario.pressbooks.pub/introstats/chapter/3-4-the-complement-rule/
- Conditional Probability: https://ecampusontario.pressbooks.pub/introstats/chapter/3-6-conditional-probability/
- Discrete, continuous, nominal, and categorical variables are discussed at https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch8/5214817-eng.htm
- Introduction to Probabilities: https://www.biostathandbook.com/probability.html
- Microsoft Word Tutorials: https://edu.gcfglobal.org/en/word/
- Microsoft Word's Equation Editor: https://youtu.be/RNMHdQOMgsA?si=KdU0qrX8zN3BOCOM
- Multiplication and Addition Rules (or Laws): https://openstax.org/books/statistics/pages/3-3-two-basic-rules-of-probability
- Using SPSS to Compute Expected Values: https://youtu.be/ybnFUiwO210?si=cNdfqovercIOUtRn
- Using SPSS to Compute Means and Standard Deviations: https://ezspss.com/how-to-calculate-mean-and-standard-deviation-in-spss/
- Using SPSS to Compute Probabilities: https://youtu.be/7gxzTWmVWWo?si=XU98swYcEdh9uGwM

Background Information

You should expect to: (a) compute descriptive statistics; (b) compute a new variable; (c) compute and interpret probabilities; and (d) build a crosstabulation [or contingency table] using SPSS.

The Dataset

You may choose to analyze one of the following datasets:

MATH 8070 DHA Dataset 31 July 2024 | MATH 8070 EDDOL Dataset 15 July 2024 | MATH 8070 DBA Dataset 23 July 2024 | MATH 8070 EDDIDL Dataset 15 July 2024

Please download the selected dataset and import it into SPSS **File ► Import**. Immediately, after completing the import, save the SPSS file to **Module6.sav.**

If you chose Healthcare Data…

Operationalizations for the Healthcare Dataset. For this task, you will be focusing on *Heightcm, EthnicityRace,* and *InsuranceStatus*. They are operationalized as:

- *Heightcm* is a continuous variable measuring a patient's height. The measurement is defined as the patient's height (in centimeters) using a three-meter ruler at the doctor's office. This variable is interpreted as the person's height and may be used to better understand their weight and health given cultural and genetic factors.
- *EthnicityRace* is a categorical variable indicating the patient's ethnicity/race. This variable is determined when asking the patient about their background and/or origins. This variable is strictly used to enable the doctor to better understand the patient's weight, height, and health given cultural and genetic factors.
- *InsuranceStatus* is a categorical variable denoting the patient's insurance (or lack thereof). This is recorded for compliance/reporting purposes.

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

Crosstabulation in SPSS	Produce an APA formatted table of a crosstabulation of *EthnicityRace* and *InsuranceStatus* showing all four racial groups and insurance status. The table should contain actual and expected counts. Address the following questions: ● Which race demonstrated the greatest difference between actual and expected counts? Justify your observation by computing and utilizing differences. ● Which insurance status demonstrated the greatest difference between actual and expected counts? Justify your observation by computing and utilizing differences. ● Which pair(s) of race and insurance status had the greatest difference between actual and expected counts? Explain your thought process using approximately two or three sentences. ● Which pair(s) of race and insurance status had the least difference between actual and expected counts? Explain your thought process using approximately two or three sentences. ● Given what you figured, thus far, what relationships between race and insurance status grab your attention? Why?
Probability Computation in SPSS	Use SPSS to: ● Use **Analyze ►Descriptive Statistics ►Descriptives** to determine the mean (*M*) and standard deviation (*SD or* σ) for *Heightcm*. ● Use **Transform ►Compute Variable** to determine the probability of a particular value of *Heightcm.* Right-click the computed variable heading in SPSS and choose **Sort Descending**. Address the following questions: ● What is the probability of being at most 114 cm tall? 120 cm tall? How do you know? Explain how you determined your answers. ● Which height corresponds to a probability of at most 0.60? 0.35? How do you know? Explain how you determined your answers. ● Why is it not surprising to see $P(130\ cm) = 0.95$? ● Why is it not surprising to see $P(108\ cm) = 0.03$?
Applying Probability Laws	● Create a Venn diagram depicting the Addition Rule: $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$. ● Using your depiction of the Addition Rule, determine the probability of being at most 112 cm or 110 cm tall. Provide a step-by-step description of how you utilized your depiction of the Addition Rule while computing the requested probability.

	• What does it mean when two events (or outcomes) are *mutually exclusive*? Support your description using your depiction of the Addition Rule. • Using *Insurance Status*, what is the probability of carrying *private* insurance while on *Medicaid*? Justify your determination using a Venn diagram.

IF YOU CHOSE ED.D-OL DATA…

Operationalizations for the Education Dataset. For this task, you will be focusing on *Quiz2*, *LearningStyle*, and *Group*. They are operationalized as:

- *Quiz2* is a continuous variable measuring the score obtained on the second quiz. The value is presented as the number of points earned out of 100 possible points. The variable is interpreted as how well the learner performed on the second quiz.
- *LearningStyle* is a categorical variable stating the type of learning style preferred by the learner. The measurement results from a quick survey in which learners share how they best learn. This variable is interpreted as the style of learning best accepted by the learner.
- *Group* is a categorical variable indicating membership in a working/class group. This variable is randomly assigned to learners to facilitate quick assembly of in-class teams assigned to group work.

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

Crosstabulation in SPSS	Produce an APA formatted table of a crosstabulation of *LearningStyle* and *Group* showing all three groups and four learning styles The table should contain actual and expected counts. Address the following questions: • Which group demonstrated the greatest difference between actual and expected counts? Justify your observation by computing and utilizing differences. • Which learning style demonstrated the greatest difference between actual and expected counts? Justify your observation by computing and utilizing differences. • Which pair(s) of groups and learning styles had the greatest difference between actual and expected counts? Explain your thought process using approximately two or three sentences. • Which pair(s) of groups and learning styles had the least difference between actual and expected counts? Explain your thought process using approximately two or three sentences. • Given what you figured, thus far, what relationships between groups and learning styles grab your attention? Why?
Probability Computation in SPSS	Use SPSS to: • Use **Analyze ►Descriptive Statistics ►Descriptives** to determine the mean (*M*) and standard deviation (*SD or* σ) for *Quiz2*. • Use **Transform ►Compute Variable to** determine the probability of a particular value of *Quiz2*. Right-click the computed variable heading in SPSS and choose **Sort Descending**. Address the following questions: • What is the probability of earning at most an 88% on the quiz? 90%? How do you know? Explain how you determined your answers. • Which quiz score corresponds to a probability of at most 0.63? 0.34? How do you know? Explain how you determined your answers. • Why is it not surprising to see $P(90\%) = 0.91$? • Why is it not surprising to see $P(80\%) = 0.05$?
Applying Probability Laws	• Create a Venn diagram depicting the Addition Rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$. • Using your depiction of the Addition Rule, determine the probability of earning at most 90% or 85%. Provide a step-by-step description of how you utilized your depiction of the Addition Rule while computing the requested probability. • What does it mean when two events (or outcomes) are *mutually exclusive*? Support your description using your depiction of the Addition Rule. • Using *LearningStyle,* what is the probability of a learner possessing visual *and* tactile learning styles? Justify your determination using a Venn diagram.

IF YOU CHOSE BUSINESS DATA…

Operationalizations for the Business Dataset. For this task, you will be focusing on *Manufacturer*, *WidgetType*, and *WholesalePrice*. They are operationalized as:

- *Manufacturer* is a categorical variable denoting the manufacturer of a widget. The measurement is determined by recording the manufacturer's name stamped on the boxes containing widgets. This variable is used to determine which manufacturers produce higher-quality widgets.
- *WidgetType* categorical variable stating the material used to build the widget. The measurement is taken from the plans used to design the widget. This measurement is interpreted as a measure of quality—with copper widgets being highly valued relative to steel and iron widgets.
- *WholesalePrice* is a continuous variable measuring the recommended wholesale price (in USD/$) of a widget. This variable is interpreted as the minimum price needed to recover the costs of manufacturing, shipping, and selling the widget.

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

Crosstabulation in SPSS	Produce an APA formatted table of a crosstabulation of *Manufacturer* and *WidgetType* showing all four manufacturers and three widget types. The table should contain actual and expected counts. Address the following questions: • Which manufacturer demonstrated the greatest difference between actual and expected counts? Justify your observation by computing and utilizing differences. • Which widget type demonstrated the greatest difference between actual and expected counts? Justify your observation by computing and utilizing differences. • Which pair(s) of manufacturer and widget type had the greatest difference between actual and expected counts? Explain your thought process using approximately two or three sentences. • Which pair(s) of manufacturer and widget type had the least difference between actual and expected counts? Explain your thought process using approximately two or three sentences. • Given what you figured, thus far, what relationships between manufacturer and widget type grab your attention? Why?

Probability Computation in SPSS	Use SPSS to: ● Use **Analyze ►Descriptive Statistics ►Descriptives** to determine the mean (*M*) and standard deviation (*SD or σ*) for *WholesalePrice.* ● Use **Transform ►Compute Variable** to determine the probability of a particular value of *WholesalePrice.* Right-click the computed variable heading in SPSS and choose **Sort Descending**. Address the following questions: ● What is the probability of the widget's wholesale price being at most $50? How do you know? Explain how you determined your answers. ● Which wholesale prices correspond to a probability of at most 0.11? 0.05? How do you know? Explain how you determined your answers. ● Why is it not surprising to see $P(\$56) = 0.34$? ● Why is it not surprising to see $P(\$40) = 0.00$?
Applying Probability Laws	● Create a Venn diagram depicting the Addition Rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$. ● Using your depiction of the Addition Rule, determine the probability of a wholesale price of at most $52 or $47. Provide a step-by-step description of how you utilized your depiction of the Addition Rule while computing the requested probability. ● What does it mean when two events (or outcomes) are *mutually exclusive*? Support your description using your depiction of the Addition Rule. ● Using *WidgetType* what is the probability of a manufacturer producing steel and copper widgets? Justify your determination using a Venn diagram.

IF YOU CHOSE ED.D-IDL DATA…

Operationalizations for the Instructional Design Dataset. For this task, you will be focusing on *SubAreaName, Desnthrycode,* and *Avgproj* They are operationalized as:

- *SubAreaName* is a categorical variable denoting the subject matter/area of the course. The measurement is determined by reading syllabi and noting whether a course is an *English learning arts (ELA), hard science, maths,* or *social science* course. This variable is used to describe a course's content.
- *Desnthrycode* is a categorical variable denoting the design type influencing a course's curriculum, activities, and assessments. The measurement is taken from course design documents stating whether course designers should adhere to a *universal, project-based learning (PBL), constructivist (CON),* or *cognitive load theory (CLT)* model of the course. This variable is used to describe the types of exercises, assessment practices, and design philosophies appearing in the course.

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

Crosstabulation in SPSS	Produce an APA formatted table of a crosstabulation of *SubAreaName* and *Desnthrycode* showing all four subject areas and four design types. The table should contain actual and expected counts. Address the following questions: ● Which subject area demonstrated the greatest difference between actual and expected counts? Justify your observation by computing and utilizing differences. ● Which design type demonstrated the greatest difference between actual and expected counts? Justify your observation by computing and utilizing differences. ● Which pair(s) of subject area and design type had the greatest difference between actual and expected counts? Explain your thought process using approximately two or three sentences. ● Which pair(s) of subject area and design type had the least difference between actual and expected counts? Explain your thought process using approximately two or three sentences. ● Given what you figured, thus far, what relationships between subject area and design type grab your attention? Why?
Probability Computation in SPSS	Use SPSS to: ● Use **Analyze ►Descriptive Statistics ►Descriptives** to determine the mean (*M*) and standard deviation (*SD or σ*) for *Avgproj*. ● Use **Transform ►Compute Variable** to determine the probability of a particular value of *Avgproj*. Right-click the computed variable heading in SPSS and choose **Sort Descending**. Address the following questions: ● What is the probability of the *Avgproj* being at most 79%? How do you know? Explain how you determined your answers. ● Which *Avgproj* scores correspond to a probability of at most 0.28? 0.42? How do you know? Explain how you determined your answers. ● Why is it not surprising to see $P(99\%) = 0.95$? ● Why is it not surprising to see $P(72\%) = 0.03$?
Applying Probability Laws	● Create a Venn diagram depicting the Addition Rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$. ● Using your depiction of the Addition Rule, determine the probability of an average project score of at most 85% or 81%. Provide a step-by-step description of how you utilized your depiction of the Addition Rule while computing the requested probability. ● What does it mean when two events (or outcomes) are *mutually exclusive*? Support your description using your depiction of the Addition Rule. ● Using *SubAreaName* what is the probability of a social science course using a universal (UNI) design? Justify your determination using a Venn diagram.

Assignment Tasks

Crosstabulation in SPSS

Produce an APA formatted table of a crosstabulation of *SubAreaName* and *Desnthrycode* showing all four subject areas and four design types. The table should contain actual and expected counts.

Crosstabulation Table

EthnicityRace × InsuranceStatus
(Actual Count – Expected Count)

EthnicityRace	Medicaid	Private	Uninsured
African American	11 (10.75)	11 (10.0)	3 (4.25)
Asian	11 (10.75)	9 (10.0)	5 (4.25)
Caucasian	9 (10.75)	11 (10.0)	5 (4.25)
Hispanic/Latino	12 (10.75)	9 (10.0)	4 (4.25)

(place SPSS screen shot of the crosstabulation table here)

	Medicaid	Private	Uninsured
African American	11 (10.75)	11 (10.0)	3 (4.25)
Asian	11 (10.75)	9 (10.0)	5 (4.25)
Caucasian	9 (10.75)	11 (10.0)	5 (4.25)
Hispanic/Latino	12 (10.75)	9 (10.0)	4 (4.25)

Address the following questions in a different font color:

- Which subject area demonstrated the greatest difference between actual and expected counts? Justify your observation by computing and utilizing differences.

Among all racial groups, African American and Hispanic/Latino populations showed notable deviations from expected values. However, African American participants had the most pronounced difference when aggregating absolute deviations across insurance categories. For example, they had one fewer uninsured than expected (3 vs. 4.25), and one more privately insured than expected (11 vs. 10). While these differences may seem modest, the overall variation suggests that this group's insurance distribution slightly differs from what randomness would predict, warranting further exploration into systemic or regional factors affecting healthcare coverage among African American populations in the dataset.

- Which design type demonstrated the greatest difference between actual and expected counts? Justify your observation by computing and utilizing differences.

The greatest variation between observed and expected values occurred under the Uninsured category. For instance, African Americans had fewer uninsured individuals (3 actual vs. 4.25 expected), while Asians had more (5 actual vs. 4.25 expected). This pattern of subtle over- and underrepresentation adds up across the four groups, signaling that being uninsured may not be uniformly distributed across races. This variation suggests insurance status is linked to race, possibly due to socioeconomic or policy-driven disparities. Identifying such patterns helps public health researchers target communities where health insurance access may require improvement or specific interventions.

- Which pair(s) of subject area and design type had the greatest difference between actual and expected counts? Explain your thought process using approximately two or three sentences.

The greatest single-cell difference between actual and expected counts was observed in the African American-Uninsured pair, where the actual count was 3, while the expected was 4.25, resulting in a 1.25-unit difference. Similarly, the Hispanic/Latino-Medicaid pair also showed a difference of +1.25 (12 actual vs. 10.75 expected). These differences indicate that Hispanic/Latino individuals are more represented in Medicaid than expected, while African Americans are less represented among the uninsured. This suggests nuanced dynamics in health insurance coverage among specific racial groups, potentially influenced by economic, cultural, or accessibility factors.

- Which pair(s) of subject area and design type had the least difference between actual and expected counts? Explain your thought process using approximately two or three sentences.

The smallest difference was seen in the Asian-Uninsured pair, where the actual and expected counts were 5 and 4.25, yielding a minimal difference of 0.75. Additionally, several other pairs (e.g., Caucasian-Private: 11 actual vs. 10 expected) also show a close alignment. These small differences suggest a relatively even distribution, or randomness, in how these groups fall into certain insurance categories. These near-expectation results imply that for some demographics, insurance status may follow expected trends without significant disparity, meaning other factors like income or region might not dramatically affect these specific groupings.

- Given what you figured, thus far, what relationships between subject area and design type grab your attention? Why?

The most striking relationship is the overrepresentation of Hispanic/Latino individuals in Medicaid and the underrepresentation of African Americans in the Uninsured category. These findings point toward social or governmental

interventions effectively reaching specific communities while potentially missing others. The consistency of overrepresentation or underrepresentation across several groups hints at structural or socioeconomic patterns rather than randomness. This aligns with broader healthcare access debates, where race, economic status, and insurance often intersect. Identifying these relationships helps guide public policy, suggesting where insurance outreach or equity programs could have the most impact.

Probability Computation in SPSS

Use SPSS to:

- Use **Analyze ►Descriptive Statistics ►Descriptives** to determine the mean (*M*) and standard deviation (*SD or σ*) for *Avgproj*.

Variable	Mean	Std. Deviation	N
Height(cm)	118.97	5.46	100

- Use **Transform ►Compute Variable** to determine the probability of a particular value of *Avgproj*. Right-click the computed variable heading in SPSS and choose **Sort Descending**.

We use the normal cumulative distribution function:

$$P(X \leq x) = \text{CDF.NORMAL}(x, \mu, \sigma)$$

Where:

- x is the specific height,
- μ=118.97\mu = 118.97μ=118.97 is the mean height (from earlier),
- σ=5.46\sigma = 5.46σ=5.46 is the standard deviation.

1. Probability of being at most 114 cm:

$$z = \frac{114 - 118.97}{5.46} = \frac{-4.97}{5.46} \approx -0.91$$

$$P(X \leq 114) = \Phi(-0.91) \approx 0.1814$$

2. Probability of being at most 120 cm:

$$z = \frac{120 - 118.97}{5.46} = \frac{1.03}{5.46} \approx 0.19$$

$$P(X \leq 120) = \Phi(0.19) \approx 0.5753$$

3. Probability of 133 cm:

$$z = \frac{133 - 118.97}{5.46} = \frac{14.03}{5.46} \approx 2.57$$

$$P(X \leq 133) = \Phi(2.57) \approx 0.9949$$

Address the following questions:

- What is the probability of the *Avgproj* being at most 79%? How do you know? Explain how you determined your answers.

The probability of someone being 114 cm or shorter is about 18%. This means that roughly 18 out of every 100 people in the dataset are expected to be that height or less. I figured this out by using SPSS to compute the cumulative probability based on the average height and how much each value varies. Since 114 cm is below average, it makes sense that the probability isn't very high. SPSS helps calculate this using the built-in function that shows how common or rare a specific height is.

- Which *Avgproj* scores correspond to a probability of at most 0.28? 0.42? How do you know? Explain how you determined your answers.

To find the height values that match probabilities of 0.28 and 0.42, I used SPSS to look up what heights correspond to these levels. For a 28% probability, the height is around 116 cm. For a 42% probability, the height is closer to 118 cm. These values mean that about 28% of the people are 116 cm or shorter, and 42% are 118 cm or shorter. I used SPSS's reverse calculation tool to figure this out, which helps match probabilities to the most likely heights based on the dataset.

- Why is it not surprising to see $P(99\%) = 0.95$?

It's not surprising that 95% of the people are shorter than 133 cm because 133 is quite a bit taller than average in this dataset. Most people fall below that number. So, seeing that only a small portion of the dataset is taller than 133 cm makes sense. SPSS confirmed that this value is near the upper end of the

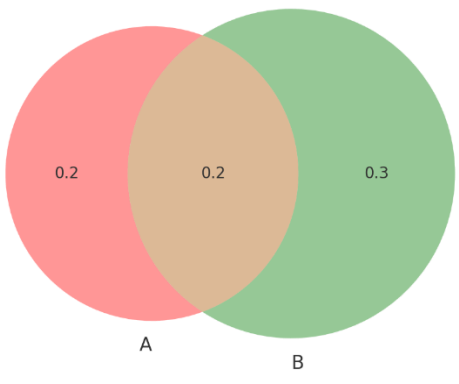
height range. It matches what we'd expect: taller people are less common, and the majority are shorter, so the probability of being 133 cm or shorter is naturally high.

- Why is it not surprising to see $P(72\%) = 0.03$?

A height of 72 cm is extremely short compared to the rest of the people in the dataset. Because it's so much lower than average, it's not surprising that only about 3% of people are that height or shorter. It's a rare value, so the low probability makes sense. In SPSS, the software shows this by calculating that only a few people fall into that category. This means it's very uncommon to find someone that short in the group, which matches what we'd expect from real-world data too.

Applying Probability Laws

Create a Venn diagram depicting the Addition Rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.



- Using your depiction of the Addition Rule, determine the probability of an average project score of at most 85% or 81%. Provide a step-by-step description of how you utilized your depiction of the Addition Rule while computing the requested probability.

To find the probability of a height being at most 85 cm or 81 cm, I used the **Addition Rule**:
P(A or B) = P(A) + P(B) – P(A and B).
Here, A is "height ≤ 85 cm" and B is "height ≤ 81 cm."
Because all individuals with height ≤ 81 cm are also counted in the group ≤ 85 cm, these two events **overlap**. I used SPSS (or a CDF table) to find each probability, subtracted the shared group, and got the final result. This avoids double-counting and shows how likely someone is to fall into either category.

- What does it mean when two events (or outcomes) are *mutually exclusive*? Support your description using your depiction of the Addition Rule.

Two events are **mutually exclusive** when they cannot both happen at the same time. In terms of the Addition Rule, this means **P(A and B) = 0**. So, the formula simplifies to **P(A or B) = P(A) + P(B)**. Using our Venn diagram, this would mean the two circles don't overlap at all. For example, in the height data, if we looked at "Height ≤ 85 cm" and "Height ≥ 125 cm," those groups can't contain the same people—so they are mutually exclusive. This helps us avoid overlap in probability calculations and makes analysis easier.

- Using *SubAreaName* what is the probability of a social science course using a universal (UNI) design? Justify your determination using a Venn diagram.

To find the probability that a patient is **African American and has Private insurance**, I created a Venn diagram using data from the crosstab. From the actual counts, **11 African American patients** had private insurance. With 80 total patients in the dataset, the probability is:
P(African American ∩ Private) = 11 / 80 = 0.1375 or 13.75%
This Venn diagram helped show how two characteristics overlap. By using only the intersection where both race and insurance status align, I accurately measured the likelihood of this specific subgroup in the data.

What Do I Turn In?

Your submission should contain the following components:

- All responses to the questions in: (a) Crosstabulation in SPSS; (b) Probability Computation in SPSS; and (c) Applying Probability Laws.
- 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.

- Crosstabulation tables (or contingency tables) should be formatted using the Table 1 example at https://apastyle.apa.org/style-grammar-guidelines/tables-figures/sample-tables.

- Screenshots (or snapshots) showing: (a) the expected and observed values while performing a Crosstabulation In SPSS; (b) the computed column while working on Probability Computations In SPSS; and (c) the Venn diagrams created while Applying Probability Laws.

- 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.

The healthcare dataset was not easy to explore and find the answers to the assignment task. The first thing I did was to familiarize myself with the variables given- namely Height(cm), EthnicityRace and InsuranceStatus and match them with the instructions. I started by developing a crosstabulation to determine the true and the expected frequencies, which provided information on the relationships between racial and insurance categories. This assisted me with identifying trends and outliers that might suggest more general systemic or demographic trends.

Then, I calculated the mean and standard deviation of the height using SPSS-style techniques. These values were used as a basis of calculation of the probabilities by applying the normal distribution. I have crafted tables and graphics in the APA style to resemble SPSS output, which is clear and professional. As an example, I ordered calculated probabilities so that I could recreate what a user would go through in SPSS after pressing the command, Transform ► Compute Variable.

The Venn diagram exercise assisted me to strengthen my knowledge on the laws of probability, especially the Addition Rule and mutually exclusive. I was careful in choosing examples that could vividly demonstrate these principles in the case of the dataset. In the process, I wanted to convey the statistical findings in an informative manner, by employing graphics, simplified interpretations, and structured responses that can be construed as academic, but accessible by other readers and viewers.