✓ **Congratulations! You passed!**

| Grade | Latest Submission | To pass 80% or | **Go to next item** |
| received 100% | Grade 100% | higher | |

You have a new skill score!

Great job! Keep learning and making progress in your courses to increase your skill scores.

| Probability & Statistics | 141 |

---

1. Which of the following are tasks in the overall risk management process? (Select three.)   **1 / 1 point**

   ☑ Analysis

   > ✓ **Correct**
   > This is one of the middle tasks in which you assess the nature of a risk, especially its likelihood of occuring and the impact it may have.

   ☑ Mitigation

   > ✓ **Correct**
   > This is one of the later tasks in which you implement some tactic to minimize the impact and/or likelihood of a risk as much as is feasible.

   ☐ Deployment

   ☐ Elimination

   ☑ Identification

   > ✓ **Correct**
   > This is one of the initial tasks in which you discover and learn more about a potential risk.

2. Which of the following risk analysis methods use words like "likely," "unlikely," and "rare" to describe the likelihood of risk, and words like "low," "medium," and "high" to describe the impact of risk? (Select two.)   **1 / 1 point**

   ☐ Semi-qualitative analysis

   ☑ Semi-quantitative analysis

   > ✓ **Correct**
   > This method of risk analysis attempts to incorporate the strengths of both quantitative and qualitative analysis, and still uses qualitative terms like those in the example.

   ☐ Quantitative analysis

☑ Qualitative analysis

> ✓ **Correct**
> This method of risk analysis uses terms like those in the example to assess risk.

3. Which of the following is *not* a valid risk response technique?

1 / 1 point

○ Transfer

○ Avoid

○ Accept

◉ Ignore

> ✓ **Correct**
> Ignoring risk is not a valid response technique because risk still exists whether or not you acknowledge it; therefore, you must address that risk in some way.

4. Which of the following statements are correct about a variable that is normally distributed? (Select two.)

1 / 1 point

☑ The mean, median, and mode of all measurements is the same, and all are located at the center of the distribution.

> ✓ **Correct**
> This is one of the defining characteristics of a normal distribution.

☑ The variable's distribution, when graphed, exhibits a symmetrical bell shape.

> ✓ **Correct**
> This is true of all normal distributions, and occurs due to the symmetrical nature of the distribution of values.

☐ Less than half of all measurements fall within one standard deviation of the mean.

☐ The tails of a normal distribution are denser than the center.

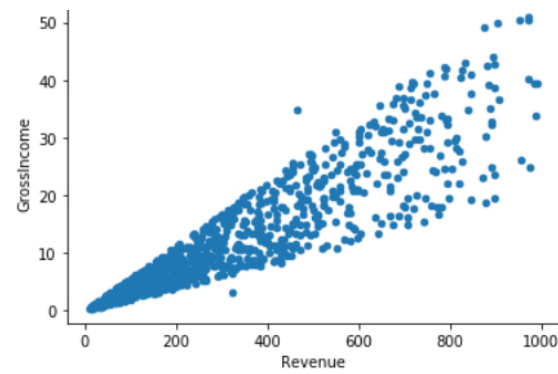**5.** Which of the following statements accurately describes variance?

1 / 1 point

○ Variance measures how much a variable's distribution differs from a normal distribution.

○ Variance measures the shape of the tails in a distribution relative to the center.

◉ Variance measures how far a data example is from the mean.

○ Variance measures the error between predicted values and actual values.

> ✓ **Correct**
> This is the fundamental definition of variance.

6. In the following scatter plot, the **GrossIncome** variable is plotted against the **Revenue** variable. What type of correlation does this plot suggest?

1 / 1 point

○ There is a strong positive correlation between both variables.

○ There is a weak positive correlation between both variables.

○ There is a strong negative correlation between both variables.

○ There is a weak negative correlation between both variables.

> ✓ **Correct**
>
> A strong positive correlation has the data points go from bottom left to top right without much variation, as in the example.

7. How do AI and other data-driven technologies use probability?   1 / 1 point

○ By guaranteeing that some event will occur with 100% likelihood

○ By determining the objective likelihood of some event happening

● By providing a model of belief about the likelihood of some event happening

○ By estimating the likelihood of some event happening without input data

> ✓ **Correct**
>
> AI technologies must build models of belief based on available data, rather than provide definite probabilities of some event happening.

8. You have a dataset of customers that includes each customer's gender, location, and other personal attributes. The label you are trying to predict is how much sales revenue each customer is likely to generate for the business based on these attributes. What type of machine learning outcome is this problem suited for?   1 / 1 point

○ Dimensionality reduction

○ Classification

● Regression

○ Clustering

> ✓ **Correct**
>
> Regression is used to make a prediction about the value of some variable of interest given two or more independent variables. It typically involves predicting numeric values, like how much revenue a customer generates.

9. You're training a model to classify whether or not a bridge is likely to collapse given several factors. You have a dataset of thousands of existing bridges and their attributes, where each bridge is labeled as having collapsed or not collapsed. Only a handful of bridges in the dataset are labeled as having collapsed—the rest are labeled as not collapsed. Given your goal of minimizing bridge collapse and the severe harm it can cause, which of the following metrics would be most useful for evaluating the model?

- ○ Accuracy

- ○ Precision

- ◉ Recall

- ○ Confusion matrix

✓ **Correct**

Recall is useful in evaluating models that are trained on data with a class imbalance, which this example has. However, unlike precision, recall is a good metric for evaluating the presence of false negatives. In this example, a model with high recall will have minimal instances in which it classifies a bridge as not having collapsed, even though it has. Due to the danger involved in the problem the model is trying to solve, the model must be able to avoid these instances as much as possible. In other words, minimizing false negatives is the highest priority, which recall can provide.

10. Which of the following metrics is used to evaluate a linear regression machine learning model?

- ◉ Cost function

- ○ Accuracy

- ○ Receiver operating characteristic (ROC)

- ○ Goodhart's Law

✓ **Correct**

A cost function, of which there are several types, is used to evaluate how poorly a linear regression model performs.