Lesson 1

Quiz, 8 questions

8/8 points (100%)

✔ Congratulations! You passed!

Next Item

✔ 1 / 1 points

1.
Which objective of statistical modeling is best illustrated by the following example?

You fit a linear regression of monthly stock values for your company. You use the estimates and recent stock history to calculate a forecast of the stock's value for the next three months.

- ○ Quantify uncertainty
- ○ Inference
- ○ Hypothesis testing
- ◉ Prediction

Correct
Forecasting is another word for predicting, especially with time series data.

✔ 1 / 1 points

2.

Lesson 1

Quiz, 8 questions

8/8 points (100%)

Which objective of statistical modeling is best illustrated by the following example?

A biologist proposes a treatment to decrease genetic variation in plant size. She conducts an experiment and asks you (the statistician) to analyze the data to conclude whether a 10% decrease in variation has occurred.

- Quantify uncertainty
- Inference
- Hypothesis testing

Correct
The scientist has a specific hypothesis in mind and asks you to evaluate the evidence for or against it.

- Prediction

---

1 / 1 points

3.
Which objective of statistical modeling is best illustrated by the following example?

The same biologist form the previous question asks you how many experiments would be necessary to have a 95% chance at detecting a 10% decrease in plant variation.

- Quantify uncertainty

Correct
Most estimates from data come with uncertainty. Statisticians are often called upon to evaluate how much we can trust the results.

- Inference
- Hypothesis testing
- Prediction

Lesson 1

Quiz, 8 questions

8/8 points (100%)

1 / 1 points

4.
Which of the following scenarios best illustrates the statistical modeling objective of inference?

- [x] A social scientist collects data and detects positive correlation between sleep deprivation and traffic accidents.

Correct
Here the social scientist made an inference about the relationship between two variables.

- [ ] A model inputs academic performance of 1000 students and predicts which student will be valedictorian after another year of school.
- [ ] A natural language processing algorithm analyzes the first four words of a sentence and provides words to complete the sentence.
- [ ] A venture capitalist uses data about several companies to build a model and makes recommendations about which company to invest in next based on growth forecasts.

---

1 / 1 points

5.
Which step in the statistical modeling cycle was **not** followed in the following scenario?

Susan gathers data recording heights of children and fits a linear regression predicting height from age. To her surprise, the model does not predict well the heights for ages 14-17 (because the growth rate changes with age), both for children included in the original data as well as other children outside the model training data.

- [ ] Fit the model
- [ ] Use the model
- [x] Explore the data

8/8 points (100%)

Correct
This is also an issue of understanding the problem. A linear model is inappropriate in this scenario because human growth is not linear. A scatter plot of the data would quickly reveal this nonlinear relationship.

- Plan and properly collect relevant data

---

1 / 1 points

6.
Which of the following is a possible consequence of failure to plan and properly collect relevant data?

- You will not produce enough data to make conclusions with a sufficient degree of confidence.
- You may not be able to visually explore the data.
- Your analysis may produce incomplete or misleading results.

Correct
Unless there are fundamental problems with the data (e.g., invalid numbers, etc.) statistical models will fit to any data and return results. It is the statistician's job to make sure the data properly measure what you intend to measure, that the results are valid, and to interpret them.

- Your selected model will not be able to fit the data.

---

1 / 1 points

7.

For Questions 6 and 7, consider the following:

8/8 points (100%)

Xie operates a bakery and wants to use a statistical model to determine how many loaves of bread he should bake each day in preparation for weekday lunch hours. He decides to fit a Poisson model to count the demand for bread. He selects two weeks which have typical business, and for those two weeks, counts how many loaves are sold during the lunch hour each day. He fits the model, which estimates that the daily demand averages 22.3 loaves.

Over the next month, Xie bakes 23 loaves each day, but is disappointed to find that on most days he has excess bread and on a few days (usually Mondays), he runs out of loaves early.

Which of the following steps of the modeling process did Xie skip?

- Understand the problem
- Postulate a model
- Fit the model
- Check the model and iterate

Correct
Xie skipped directly from collecting data and fitting the model to using the model for future production. He skipped the steps of exploring the data and checking to see if the model fit adequately. Because the model was lacking, his use of the model produced less desirable results.

- Use the model

---

1 / 1 points

8.

What might you recommend Xie do next to fix this omission and improve his predictive performance?

- Abandon his statistical modeling initiative.
- Collect three more weeks of data from his bakery and other bakeries throughout the city. Re-fit the same model to the extra data and follow the results based on more data.

8/8 points (100%)

- [x] Plot daily demand and model predictions against the day of the week to check for patterns that may account for the extra variability. Fit and check a new model which accounts for this.

Correct

Higher demand on Mondays suggests that the time of the week may help explain some of the day-to-day variability in demand. This could be incorporated by fitting different Poisson models for different weekdays, or a Poisson regression incorporating explanatory variables (we will explore these in Lesson 10).

- [ ] Trust the current model and continue to produce 23 loaves daily, since in the long-run average, his error is zero.