

# Data Science Prep Course - Technical Evaluation

This test assesses your mastery of the knowledge and skills taught in the Data Science Bootcamp Prep curriculum. It consists of fifteen questions and should take you between 30 and 60 minutes to complete.

Some of the questions test your knowledge of statistics concepts, while other ones require you to use the data science toolkit to analyze data we provide.

Before starting, download the dataset linked below, which contains data about plane crashes between 1950 and 2009. Fire up a Jupyter notebook and create a dataframe with this data.

This is an open book test, and you're welcome to consult the curriculum to answer questions. Furthermore, you'll have to have to run a Jupyter notebook and use the data science toolkit in order to answer the questions that ask for computations on the dataset.

Download this dataset: [https://tf-curricula-prod.s3.amazonaws.com/assets/plane\\_crashes\\_data.csv](https://tf-curricula-prod.s3.amazonaws.com/assets/plane_crashes_data.csv)

\* Required

Email address \*

Your email

---

In the context of histograms, what is a bin? \*

1 point

- ☐ The part of the chart where outlier values get displayed.
- ☐ Ranges with an upper and lower value that individual data points get put in.
- ☐ The most frequent value in the histogram.
- ☐ Trick question: bins have nothing to do with histograms. It's a different chart entirely called a "bin visualization".

What does it mean to normalize a histogram? \*

1 point

- ☐ Rescaling one or more variables so you can compare their distributions, by making the area of the curve equal to 1.
- ☐ Removing outlier values so that only the interquartile range gets displayed.
- ☐ Displaying only the top-3 most populated bins.

What does "central tendency" mean? \*

1 point

- ☐ The point around which datapoints for a variable cluster.
- ☐ The main conclusion an objective observer should reach by looking at the distribution of a datapoint.
- ☐ The well-known tendency for values to converge around fifty for datapoints whose value can range from 0 to 100.

Which of the following is not a measure of central tendency? \*

1 point

- ☐ mean
- ☐ median
- ☐ variance
- ☐ mode

True or false, for any dataset, there is always one, and only one mode. \*

1 point

- ☐ True
- ☐ False

## What is variance? \*

1 point

- ☐ The difference in central tendency for data from the same survey given at two different points in time (for instance, 2000 vs. 2010).
- ☐ The variance of a variable describes how much values differ from the central tendency, and how much they differ from each other.
- ☐ The range of values (minimum to maximum) of answers to a question with a continuous, numeric answer.

You're on a game show and the host presents you with three doors. Behind two doors are piles of dirty laundry, and behind one is a check for \$10,000. You pick door one, then the host — who you believe knows which door has the prize behind it — opens door three, revealing a pile of dirty laundry. He offers you the chance to switch from door one to door two. If you want to maximize your chance of picking the \$10,000 door, you should: \*

- ☐ Stick with your first choice (door one) because it's statistically more likely to be correct.
- ☐ Flip a coin and decide what to do because the odds of the \$10,000 being behind either of the remaining doors are the same.
- ☐ Switch to the door the host offers you (door two) because you're more likely to win.

You have two six-sided dice. What are the odds of rolling "snake eyes" (aka, two ones), three times in a row? \*

1 point

- ☐ 1:16
- ☐ 1:720
- ☐ 1:32,768
- ☐ 1:46,656
- ☐ 1:58,244

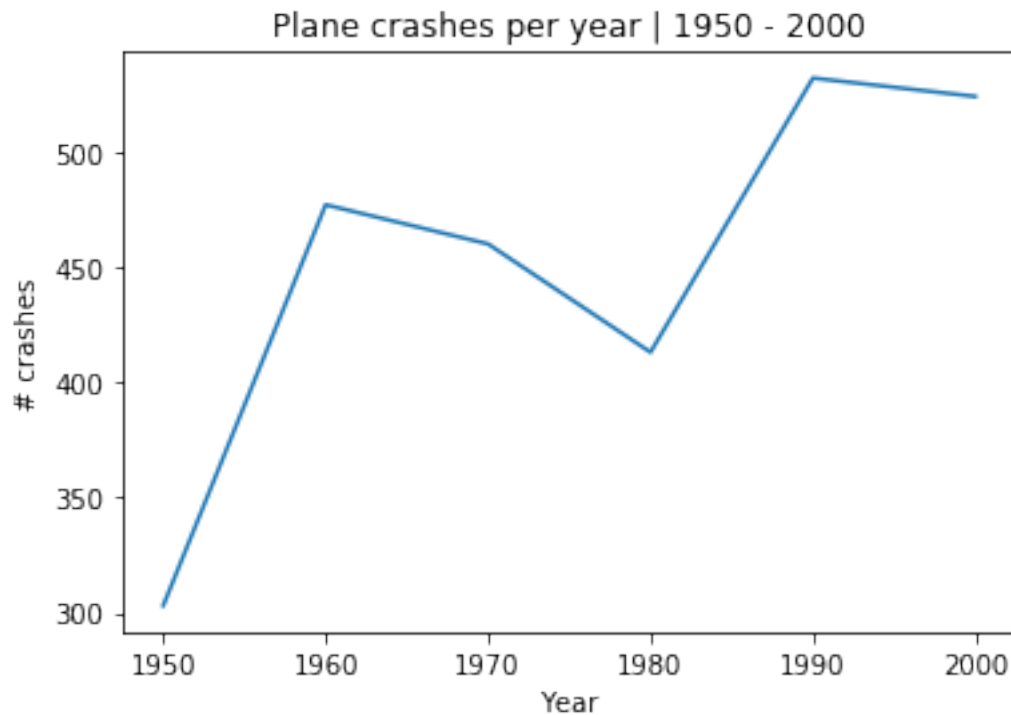
Imagine there's a lottery called "Rainbow ball". In this game, players choose four distinct numbers between 1 and 50, and one number between one 1 and 25 for the "rainbow ball". When the lottery drawing happens, four balls are chosen at random from a container holding 50 balls with values ranging from 1 to 50 (without replacement), and the rainbow ball is chosen from a separate container with 25 balls with values ranging from 1 to 25. To win the jackpot, you have to match the four initial balls, as well as the rainbow ball. Note that order doesn't matter for the four initial balls. Before the drawing, you're given a lotto ticket with the numbers 1, 2, 3, 4 for the "normal" balls and 5 for the "rainbow" ball. What are the odds that this is a winning ticket? \*

1 point

- ☐ 1:5,757,500
- ☐ 1:138,180,000
- ☐ 1:156,250,000

Someone provides the following chart to substantiate the claim that, overall, it's become much more risky to fly since the 1950s. This claim is problematic because: \*

1 point



- ☐ It's not problematic. That's precisely what the chart demonstrates.
- ☐ There are different kinds of risks nowadays. Terrorism was less of a risk in the 50s and 60s.
- ☐ There was a downward trend between 1960 and 1980.
- ☐ The number of flights per year has substantially grown since the 1950s.

NEXT

This form was created inside of Thinkful. [Report Abuse](#) - [Terms of Service](#) - [Additional Terms](#)

Google Forms