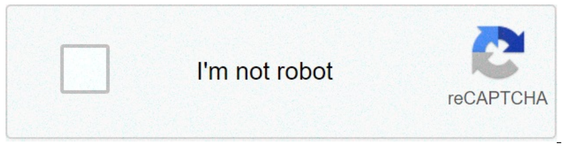

Continue

Bootstrap statistics example

DescriptionR is a programming language and software environment for statistical computing and graphics that is widely used among statisticians and data miners for data analysis. In this course, you'll get a thorough run-through of how R works and how it's applied to data science. Before you know it, you'll be crunching numbers like a pro, and be better qualified for many lucrative careers. Access 82 lectures & 9 hours of content 24/7 Cover basic statistical principles like mean, median, range, etc. Learn theoretical aspects of statistical concepts Discover datatypes & data structures in R, vectors, arrays, matrices & more Understand Linear Regression Visualize data in R using a variety of charts & graphs Delve into descriptive & inferential statistics All featured courses are designed for educational purposes only and do not reflect our views or recommendations. Please note that all course purchasers invest at their own risk. Loonycorn is comprised of four individuals--Janani Ravi, Vitthal Srinivasan, Swetha Kolalapudi and Navdeep Singh--who have honed their tech expertises at Google and Flipkart. The team believes it has distilled the instruction of complicated tech concepts into funny, practical, engaging courses, and is excited to be sharing its content with eager students. Details & RequirementsLength of time users can access this course: lifetime access Access options: web streaming, mobile streaming Certification of completion not included Redemption deadline: redeem your code within 30 days of purchase Experience level required: all levelsCompatibility Terms Unredeemed licenses can be returned for store credit within 30 days of purchase. Once your license is redeemed, all sales are final. Statistics Approximately 34 million children and adults have diabetes in the United States. The numbers associated with diabetes make a strong case for devoting more resources to finding a cure. Read more The national cost of diabetes in the U.S. in 2017 was more than $327 billion, up from $245 billion in 2012. Diabetes is growing at an epidemic rate in the United States. And what's true nationwide is also true in each state. Lisa Nguyen is the co-founder of Baubles + Soles, a line of interchangeable footwear for kids. After all five 'Shark Tank' judges rejected her company's pitch, Nguyen still walked away with a deal. Here's how she did it. Last Updated on November 18, 2020 It’s okay, you can finally admit it. It’s been two months since you’ve seen the inside of the gym. Getting sick, family crisis, overtime at work and school papers that needed to get finished all kept you for exercising. Now, the question is: how do you start again? Once you have an exercise habit, it becomes automatic. You just go to the gym, there is no force involved. But after a month, two months or possibly a year off, it can be hard to get started again. Here are some tips to climb back on that treadmill after you’ve fallen off. Don’t Break the Habit – The easiest way to keep things going is simply not to stop. Avoid long breaks in exercising or rebuilding the habit will take some effort. This may be advice a little too late for some people. But if you have an exercise habit going, don’t drop it at the first sign of trouble. Reward Showing Up – Woody Allen once said that, “Half of life is showing up.” I’d argue that 90% of making a habit is just making the effort to get there. You can worry about your weight, amount of laps you run or the amount you can bench press later. Commit for Thirty Days – Make a commitment to go every day (even just for 20 minutes) for one month. This will solidify the exercise habit. By making a commitment you also take pressure off yourself in the first weeks back of deciding whether to go. Make it Fun – If you don’t enjoy yourself at the gym, it is going to be hard to keep it a habit. There are thousands of ways you can move your body and exercise, so don’t give up if you’ve decided lifting weights or doing crunches isn’t for you. Many large fitness centers will offer a range of programs that can suit your tastes. Schedule During Quiet Hours – Don’t put exercise time in a place where it will easily be pushed aside by something more important. Right after work or first thing in the morning are often good places to put it. Lunch-hour workouts might be too easy to skip if work demands start mounting. Get a Buddy – Grab a friend to join you. Having a social aspect to exercising can boost your commitment to the exercise habit. X Your Calendar – One person I know has the habit of drawing a red “X” through any day on the calendar he goes to the gym. The benefit of this is it quickly shows how long it has been since you’ve gone to the gym. Keeping a steady amount of X’s on your calendar is an easy way to motivate yourself. Enjoyment Before Effort – After you finish any work out, ask yourself what parts you enjoyed and what parts you did not. As a rule, the enjoyable aspects of your workout will get done and the rest will be avoided. By focusing on how you can make workouts more enjoyable, you can make sure you want to keep going to the gym. Create a Ritual – Your workout routine should become so ingrained that it becomes a ritual. This means that the time of day, place or cue automatically starts you towards grabbing your bag and heading out. If your workout times are completely random, it will be harder to benefit from the momentum of a ritual. Stress Relief – What do you do when your stressed? Chances are it isn’t running. But exercise can be a great way to relieve stress, releasing endorphin which will improve your mood. The next time you feel stressed or tired, try doing an exercise you enjoy. When stress relief is linked to exercise, it is easy to regain the habit even after a leave of absence. Measure Fitness – Weight isn’t always the best number to track. Increase in muscle can offset decreases in fat so the scale doesn’t change even if your body is. But fitness improvements are a great way to stay motivated. Recording simple numbers such as the number of push-ups, sit-ups or speed you can run can help you see that the exercise is making you stronger and faster. Habits First, Equipment Later – Fancy equipment doesn’t create a habit for exercise. Despite this, some people still believe that buying a thousand dollar machine will make up for their inactivity. It won’t. Start building the exercise habit first, only afterwards should you worry about having a personal gym. Isolate Your Weakness – If falling off the exercise wagon is a common occurrence for you, find out why. Do you not enjoy exercising? Is it a lack of time? Is it feeling self-conscious at the gym? Is it a lack of fitness know-how? As soon as you can isolate your weakness, you can make steps to improve the situation. Start Small – Trying to run fifteen miles your first workout isn’t a good way to build a habit. Work below your capacity for the first few weeks to build the habit. Otherwise you might scare yourself off after a brutal workout. Go for Yourself, Not to Impress – Going to the gym with the only goal of looking great is like starting a business with only the goal to make money. The effort can’t justify the results. But if you go to the gym to push yourself, gain energy and have a good time, then you can keep going even when results are slow. Inferential statistics look at the relationship between several variables present in a sample. These statistics will predict the future of variables. Sometimes they generalize about larger groups of people. They tell us what is happening. These statistics interpret the data for us. This allows social scientists to view patterns. They can make sense of the information. They also use complex mathematics. This is the core difference between inferential and descriptive statistics. How to Use Inferential Statistics Inferential statistics examine relationships between variables in a sample. The statistics help people make predictions, or inferences, about a larger population. Scientists may use these kinds of statistics as a more affordable way to measure groups based on small samples so that it can later be applied to a large population. For example, if you wanted to know the exact age at which each person in the country went on their first date, you probably wouldn't be able to ask everybody. Instead, you would need to find a sample size and draw conclusions based on the sample. Inferential statistics is all about relationships and quantitative analysis. You can use inferential statistics to create logistic regression analysis and linear regression analysis. Descriptive Statistics Descriptive statistics describe and summarize data. Examples include numerical measures, like averages and correlation. Standard deviation is another descriptive statistic. Descriptive statistics explain only the population you are studying. Scientists cannot use the information to generalize other groups. There are two types of descriptive statistics: measures of spread and measures of central tendency. Measures of Spread A measure of spread shows the distribution of a data set. The measure of spread also shows the relationship between each data point. A measure of spread includes the range, quartiles, variance, frequency distribution and mean absolute deviation. We show measures of spread in different ways. For example, you can show a measure of spread on a bar chart, table or histogram. These charts help people interpret trends in data. Measures of Central Tendency Measures of central tendency are another form of descriptive statistics. The measure of central tendency reveals data trends. It includes the mean, median and mode. Each of these figures tell us something about the data. For example, the mode is the most common value the data shows. The mode can tell you the age at which most people graduate from high school, for instance. The media is the middle range of a data set. It can give us information about the set of ages in which people typically get their first job. Finally, the mean is the average of the data. You can add up each piece of data and then divide that figure by the number of data pieces. You can use the mean to determine the average age at which people begin college, for instance. Bootstrapping is a powerful statistical technique. It is especially useful when the sample size that we are working with is small. Under usual circumstances, sample sizes of less than 40 cannot be dealt with by assuming a normal distribution or a t distribution. Bootstrap techniques work quite well with samples that have less than 40 elements. The reason for this is that bootstrapping involves resampling. These kinds of techniques assume nothing about the distribution of our data. Bootstrapping has become more popular as computing resources have become more readily available. This is because in order for bootstrapping to be practical a computer must be used. We will see how this works in the following example of bootstrapping. We begin with a statistical sample from a population that we know nothing about. Our goal will be a 90% confidence interval about the mean of the sample. Although other statistical techniques used to determine confidence intervals assume that we know the mean or standard deviation of our population, bootstrapping does not require anything other than the sample. For purposes of our example, we will assume that the sample is 1, 2, 4, 4, 10. We now resample with replacement from our sample to form what are known as bootstrap samples. Each bootstrap sample will have a size of five, just like our original sample. Since we are randomly selecting and then are replacing each value, the bootstrap samples may be different from the original sample and from each other. For examples that we would run into in the real world, we would do this resampling hundreds if not thousands of times. In what follows below, we will see an example of 20 bootstrap samples: 2, 1, 10, 4, 2 4, 10, 10, 2, 4 1, 4, 1, 4, 4 4, 1, 1, 4, 10 4, 4, 1, 4, 2 4, 10, 10, 10, 4 2, 4, 4, 2, 1 2, 4, 1, 10, 4 1, 10, 2, 10, 10 4, 1, 10, 1, 10 4, 4, 4, 4, 1 1, 2, 4, 4, 2 4, 4, 10, 10, 2 4, 2, 1, 4, 4 4, 4, 4, 4, 4 4, 2, 4, 1, 1 4, 4, 4, 2, 4 10, 4, 1, 4, 4 4, 2, 1, 1, 2 10, 2, 2, 1, 1 Since we are using bootstrapping to calculate a confidence interval for the population mean, we now calculate the means of each of our bootstrap samples. These means, arranged in ascending order are: 2, 2.4, 2.6, 2.6, 2.8, 3, 3, 3.2, 3.4, 3.6, 3.8, 4, 4, 4.2, 4.6, 5.2, 6, 6, 6.6, 7.6. We now obtain from our list of bootstrap sample means a confidence interval. Since we want a 90% confidence interval, we use the 95th and 5th percentiles as the endpoints of the intervals. The reason for this is that we split 100% - 90% = 10% in half so that we will have the middle 90% of all of the bootstrap sample means. For our example above we have a confidence interval of 2.4 to 6.6. Bootstrapping is a statistical technique that falls under the broader heading of resampling. This technique involves a relatively simple procedure but repeated so many times that it is heavily dependent upon computer calculations. Bootstrapping provides a method other than confidence intervals to estimate a population parameter. Bootstrapping very much seems to work like magic. Read on to see how it obtains its interesting name. One goal of inferential statistics is to determine the value of a parameter of a population. It is typically too expensive or even impossible to measure this directly. So we use statistical sampling. We sample a population, measure a statistic of this sample, and then use this statistic to say something about the corresponding parameter of the population. For example, in a chocolate factory, we might want to guarantee that candy bars have a particular mean weight. It’s not feasible to weigh every candy bar that is produced, so we use sampling techniques to randomly choose 100 candy bars. We calculate the mean of these 100 candy bars and say that the population mean falls within a margin of error from what the mean of our sample is. Suppose that a few months later we want to know with greater accuracy -- or less of a margin of error -- what the mean candy bar weight was on the day that we sampled the production line. We cannot use today’s candy bars, as too many variables have entered the picture (different batches of milk, sugar and cocoa beans, different atmospheric conditions, different employees on the line, etc.). All that we have from the day that we are curious about are the 100 weights. Without a time machine back to that day, it would seem that the initial margin of error is the best that we can hope for. Fortunately, we can use the technique of bootstrapping. In this situation, we randomly sample with replacement from the 100 known weights. We then call this a bootstrap sample. Since we allow for replacement, this bootstrap sample most likely not identical to our initial sample. Some data points may be duplicated, and others data points from the initial 100 may be omitted in a bootstrap sample. With the help of a computer, thousands of bootstrap samples can be constructed in a relatively short time. As mentioned, to truly use bootstrap techniques we need to use a computer. The following numerical example will help to demonstrate how the process works. If we begin with the sample 2, 4, 5, 6, 6, then all of the following are possible bootstrap samples: 2 ,5, 5, 6, 6 4, 5, 6, 6, 6 2, 2, 4, 5, 5 2, 2, 2, 4, 6 2, 2, 2, 2, 2 4,6, 6, 6, 6 Bootstrap techniques are relatively new to the field of statistics. The first use was published in a 1979 paper by Bradley Efron. As computing power has increased and becomes less expensive, bootstrap techniques have become more widespread. The name “bootstrapping” comes from the phrase, “To lift himself up by his bootstraps.” This refers to something that is preposterous and impossible. Try as hard as you can, you cannot lift yourself into the air by tugging at pieces of leather on your boots. There is some mathematical theory that justifies bootstrapping techniques. However, the use of bootstrapping does feel like you are doing the impossible. Although it does not seem like you would be able to improve upon the estimate of a population statistic by reusing the same sample over and over again, bootstrapping can, in fact, do this.