Statistics Concepts

Subscribe for email updates.

statistics × + Add Tag

Sort replies by: Oldest first Newest first Most popular

Cian Michael Moriarty Signature Track · a month ago

Hi, I found that in many videos, particularly in week four, the programming is covered in some detail but the statistical concepts are really glossed over. I'm fairly good at programming in many other languages so I'm having very little problem with the programming side of things. But when statistical things are discussed I have no idea what is being talked about. I passed Stats 1 by the skin of my teeth 21 years ago, but my knowledge of statistics wasn't even very good then.

For example I did the corr.R task from assignment 1 without any problem at all, but I don't really know what it is this correlation thing that I am computing measures. "Correlation" somehow, obviously, but in what way? What does it all mean?

In the lectures R.D. Peng will say things like "of course results of the Poisson distribution will be integers". Without actually even defining what a probability distribution is, let alone a Poisson distribution.

I checked the Wikipedia page for Poisson distribution, but I couldn't make any sense of it. Maybe the text book has more introductory statistics stuff? I might buy it but money is tight at the moment.

I came into the specialisation thinking there would be some statistics taught. Maybe it's taught in later subjects? Or maybe I need to catch up on these concepts somewhere else. Does anyone have any suggestions on resources for learning statistics?

Thanks!

8 · flag

Derek Franks · a month ago

Don't worry about the statistics concepts for now. They'll be taught in later classes. That said, I would brush up on your statistics ahead of time, as going into the inferential statistics class in the Data Science track with no statistics knowledge is like taking this class with no programming experience. You can do it, but it's going to be difficult. I would suggest Duke's Data Analysis and Statistical Inference class as a good background course.

5 · flag

Ian Malone Signature Track · a month ago

Yes, this is the programming bit, the statistics is all the other bits, so there'll probably be no shortage!

As for the Poisson distribution, it's for the number of things that can happen in a given interval (e.g. cars that go past your house in half an hour, trees along a 1 mile stretch of road), so numbers sampled from it have to be integers. The key bit in the wikipedia description, "is a discrete probability distribution that expresses the probability of a given number of events occurring," is the word "discrete". This is as opposed to continuous distributions for things like height. Probability distribution itself just means how likely each outcome is.

For a different discrete distribution (the binomial distribution) that applies to coin tosses, if you have one coin toss the probability of outcome 1 (tails) is 0.5 and the probability of outcome 2 (heads) is 0.5, that's the probability distribution. (There being an infinite number of real numbers, continuous distributions are a little trickier, so you have a probability density function, the same way you have density of materials, a point by itself doesn't weigh anything, but a volume of material does.)

Correlation is how close to a perfect linear relationship two variables are. If the data points form a straight line with non-zero slope then you get 1 (or -1 if one decreases as the other increases). If the points are related but there's noise then the value is closer to zero. On the other hand, even if there's no relationship then noise can lead to a non-zero value.

```
> A<-rnorm(5)
> B<-rnorm(5)
> A
[1] -1.15647437 -1.26917137  0.71595222 -0.02272513  0.96527317
> B
[1] -1.8938202 -0.4046060  0.3795699  0.6925031 -0.6449699
> cor(A,B)
[1] 0.4902643
```

↑ **5** ↓ · flag

free4freedom · a month ago

@Derek Franks - Just for confirmation - is this the Duke's statistics course you are referring to?

Another question - would you suggest to go through the entire material?(it is too much!)

Will those ALL concepts be covered in this Data Science Track?

Perhaps we can learn those concepts from Duke's course that this Data Science Track is not going to touch upon, but are required.

Thanks in advance.

↑ 0 ↓ · flag

Derek Franks · a month ago

Yes that's it. All of the concepts covered in the Duke class (and then some) will be covered in the data science track.

0 · flag

+ Comment

 Pierre Tourigny · a month ago

From EdX, there is Explore Statistics with R.

1 · flag

+ Comment

 Douglas Weathers Signature Track · 14 days ago

Thanks for this thread. One reason I'm taking this certificate is because I don't know any statistics even after a couple of math degrees. I think I'll use at the textbook for the Duke class.

0 · flag

+ Comment

 Rick Henderson Signature Track · 12 days ago

The textbook for this class doesn't have much more in the way of statistical info. I feel the same way, I hope the stats is covered well in other courses in the track. However, the textbook for this course is available free, and if you use it once you get a job you could buy a copy.

0 · flag

 Rick Henderson Signature Track · 12 days ago

Also I came across this piece of info when I was starting to teach descriptive statistics (mean, min, max, st.dev etc). "Don't let the word **statistics** scare you. A statistic is just a number."

- I wish I could remember which teacher told me that. Like the number of people taking this course. Or the average age of people in the room. A statistic is a tool used to summarize a larger set of numbers.

0 · flag

Pradyumna Misra COMMUNITY TA · 11 days ago

If you are not familiar with statistics, the subject not the number, the textbook for the Duke class.referred above is a great resource. It is written for undergraduate courses where as most of material covered in this specialization is graduate level so many people feel rushed or feel the concepts are glossed over. It's a good refresher for those who are comfortable with statistics but do not use it on a regular basis.

0

· flag

+ Comment

Anonymous · 12 days ago

How about the Biostatistics Bootcamp offered by Johns Hopkins?

https://class.coursera.org/biostats-009

0 · flag

 Rick Henderson Signature Track · 11 days ago

Looks interesting, but the requirement for some calculus and set theory may be beyond some people in the course.

0 · flag

Pradyumna Misra COMMUNITY TA · 11 days ago

Data Analysis and Statistical Inference course https://www.coursera.org/course/statistics offered by Duke University via Coursera is a great course to bone up on statistics and some of the material that will be covered in later courses of JHU data science specialization sequence.

0 · flag

Anonymous · 11 days ago

@Rick -- yeah, it may. I found it interesting though that this course also seems to be listed under the "Data Science Specialization" (and indeed, has one of the same instructors). But otherwise it isn't visible on the list of courses in the DSS sequence.

0 · flag

 Rick Henderson Signature Track · 9 days ago

What I find is interesting is that when everyone says "data science" they often mean "statistics".

0 · flag

+ Comment

Anonymous · 11 days ago

Thanks to all of you who suggested references. With no statistical background, I feel that I am unable to go through with the last week's assignments without having a crash course of the basic concepts in record time. I think that this class should have a statistics course as a pre-requisite, and feel that this should be clearly explained in the course information so that students would be well informed *before*

they join the course.

0 · flag

Anonymous · 11 days ago

Hey Anon,

There's zero statistics required in any of the programming assignments in *this* course, don't worry. For example, where PA1 used the concept of correlation, we were not required to understand it -- just that it was a number that could be calculated from `cor()`.

In the final assignment PA3, all we are doing is ordering some data by columns and returning either the 1st index, the last index, or the *n*th index. If there's any concepts here that you need a crash course in, please ask :)

0 · flag

Anonymous · 11 days ago

Thanks for the reassurance :-). I think that a question in the final quiz which may have given me the wrong impression at first sight.

0 · flag

+ Comment

New post

To ensure a positive and productive discussion, please read our forum posting policies before posting.

B | I | Link | <code> | Pic | Math | Edit: Rich | Preview

Make this post anonymous to other students

Subscribe to this thread at the same time

Add post