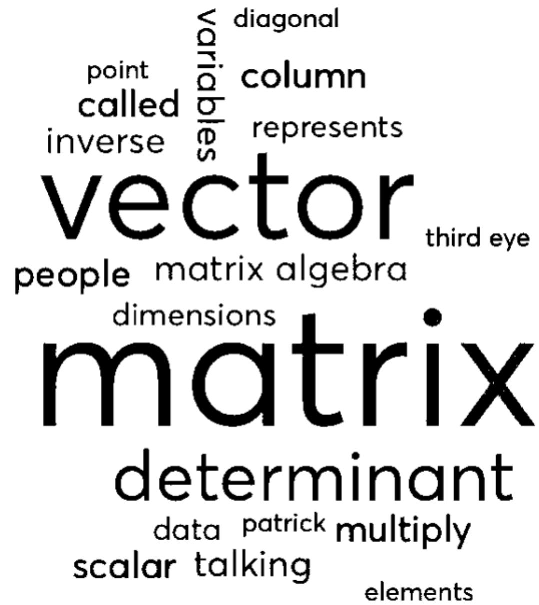
The Podcast *Quantitude*

with Greg Hancock & Patrick Curran

Season 3, Episode 22:

S3E22: The Mättrix Part 1: Defining & Manipulating Matrices



Published Tuesday March 1st, 2022 • 56:39

SUMMARY KEYWORDS
matrix, vector, determinant, matrix algebra, inverse, column, people, multiply, called, scalar, variables, talking, dimensions, data, represents, elements, third eye, diagonal, point, patrick

Greg 00:01
Hi everybody, my name is Greg Hancock and along with my third eyeless friend Patrick Curran we make a quantity food we're a podcast dedicated to all things quantitative ranging from the relevant to the completely irrelevant. In this first of two episodes will lay the foundations of matrix algebra mathematically and geometrically in start connecting these important underlying ideas to statistics. Along the way, we also mentioned STDs, the 110 to the 10 to the 405. Mistakes, skin bags of water, vector Victor tak flashlights on a misty night. Whoa, remembering mnemonics, Stephen Hawking, Bilbo Baggins, and my sultry voice. We hope you enjoy today's episode. All right, I know that we had dug on a couple of episodes ago, but I'm still a little bit gun shy about the whole audio thing.

Patrick 00:53
I still don't know what happened. I sent him an email. We had it all set up. We had a backup recorder, we had the main recorder, and when I got all the audio afterwards, it sounded like he was in the bottom of a trash dumpster. So yeah, I'm a little gun shy as well. We're just gonna hope that the audio works on this one.

Greg 01:14
I will tell people out there that when we post process and edit the episodes, there's often a lot of stuff that we go, oh, that's kind of cute. Wish we could keep that not with Doug's, there was stuff that was like, I don't know what he said. Like he went on this whole thing about STDs, STDs. I didn't know why he was talking about STDs. Turned out, he was talking about s VDS, which stands for singular value decomposition. And he used the words a couple of times at some point to sort of make me understand him. But all of that stuff was just unintelligible and had to be cut.

Patrick 01:49
So wait, you're telling me we weren't talking about STDs? Oh, that makes so much more sense now.

Greg 01:58
Oh, yeah. But you know, it also got me thinking about the fact that in that episode, and in a number of other episodes, we make certain assumptions about people's familiarity with things, one of the things we make assumptions about is matrix algebra and a lot of the language around matrix algebra, manipulations. And you and I do try to keep things at a conceptual level. But every once in a while we get into the weeds to varying degrees. So there's a part of me that thought it might be good to try to go together into the whole matrix algebra world. But then there's this other voice that says that's gonna require even more mind's eye than usual.

Patrick 02:37
Yeah, maybe we should start with something a little easier. You could describe some Monet prints so that people could envision them, maybe some of your favorite Escher prints. I mean, let's start with something easier. That requires visualization before going into matrix algebra. And you give me a hard time about requiring the mind's eye.

Greg 02:59
If I'm not mistaken, you had a Mind's Eye kind of story. Third Eye kind of story that you have not shared. Is that right?

Patrick 03:07
You remember that? Uh huh. Oh, man. So it was 100 years ago, I was in LA. I was on postdoc at UCLA. So this would have been in 95. Maybe I was newly married. I mean, like 15 minutes newly married. We lived in all hombre in the shadow of Dodger Stadium, and I had drive to UCLA every day. Mm hmm. That is the 110 to the 10 to the 405. On a typical day, it's 90 minutes each way. It was six in the morning, there was a stoplight, there was a major intersection where cars were three lanes each way. I got the green light, I was first in line, I started to go forward, and something just didn't seem right. And I slammed on the brakes. And in a flash a car went left to right in front of me through the red light. Wow, probably going 60 miles an hour. That would have been a T bone and the side, your side, my side. And that would not have ended well in so I kind of shook my head and went on. Few years later. My wife and I are pre kid and we're in a bed and breakfast. Right? Which is something that my wife loves and I less love. I love bed and breakfasts if none of the people would be there. Like including the hosts. Yeah, right. If you could just have the house. But we were sitting around in this mandatory kind of sitting in the living room where everyone is chatting. And there was a lovely couple who ran the b&b. But they were mystics. And they had dream catchers hanging and incense and all of this.

And they started talking about third eyes. And my wife said, Oh, well, my husband has a story about that. And so I briefly told it, and I said yeah, it was just a Since I had to not go through, and the woman said, Well, you saw it through your third eye as annoyed and it just didn't feel right in the intersection. And she started to double down and I started to double down on our perspectives, mine turned into differential equations. The first and third lanes were slowing the middle lane was speeding. There was a juxtaposition in velocity and acceleration, and I didn't like that. And she doubled down on my third I saw that was happening from above me. So I doubled down further. And I said, there's no such thing as a third eye. We are skin bags of water governed by random electrical impulses and molecules. And everyone is silent. And at one point, she just had this look of resignation. And she said, I never thought I would ever meet a person who was not born with a third eye. And then she walked out. So yes, is I have been told by a mystic that I was born without a third eye.

Greg 06:08
Wow, yeah, not just Third Eye Blind. You don't even have a third eye.

Patrick 06:20
And she said it with this sadness of just the weight of the world on her. I never thought I would meet someone born without a third story, you can ask me to use my third eye here. Wow.

Greg 06:35
Well, I do want to wait into if I may wait into the matrix with you. It might require some third eye but maybe it'll be my third eye.

Patrick 06:44
If we're gonna do this, and you're gonna force me to do this. I know this is low hanging fruit. This episode is the matrix. Okay, so this like writes itself? Red Pill. blue pill.

Greg 07:04
Hang on.

Patrick 07:05
Dude, we're recording here.

Greg 07:07
Sorry. It's Michael. Oh, crap. Yeah, I know. Hang on.

Patrick 07:10
This happened with Applebee's.

Greg 07:11
Yeah. Hey, Michael, what's up? You're kidding. Really? Um, okay. No, maybe we'll try that. Alright. Thanks. Bye. Michael seems to think that there could be some legal concerns if we use the matrix, because Oh, I know. I know, seriously. He had one suggestion. I don't know how you feel about it. That instead, we refer to it as the mah-trix.

Patrick 07:42
So it's like Simpson is when they did the shining. But to avoid getting sued, they had to do the Shinan. We were already a third rate podcast. And so we're gonna do the mah-trix.

Greg 08:01
That's what he said, When we make explicit reference so as not to irritate Warner Brothers.

Patrick 08:06
It sounds like an adjustable bed I'd buy my mom.

Greg 08:11
All right, well, let's give it a shot. Which pill did we take? The red pill.

MovieClip 08:14
You take the blue pill, the story ends. You wake up in your bed and believe whatever you want to believe. You take the red pill, you stay in Wonderland. And I show you how deep the rabbit hole goes.

Greg 08:28
Let me ask you for starters, do you teach matrix algebra in your classes? Oh, dude,

Patrick 08:33
I wrestle with this so hard. We talked about in Doug's episode, I told a little bit of the story. I used to have a hardcore multivariate stat where we did like eight weeks of matrices. Yeah, I don't do that anymore. And it is missed. It's deeply missed, because it's foundational to everything we do. And so when I teach SEM and longitudinal SEM, I have a section on matrix algebra, but it's really basics. And that's it. And I'm actually struggling with that right now. How about you?

Greg 09:05
Yeah, I'm have such mixed minds on this for a variety of reasons. One is the students who take our courses, at least in my case, are a huge mix of people from all over campus, and then a few people who are in our program. And so there's always that issue of where do you pitch the level of the course. And part of that has to do with the mathematics. So I wrestle with it, I consider it completely foundational, and yet, I keep asking myself, How much depth do people who are applied researchers actually need? And I don't know what the right answer to that is. Two ways that I have tried to resolve this are the following. One is for students in our own program, we have a separate course that includes all of the matrix foundation stuff paired with other content that is quite specific to people who are trying to become quantitative methodologists. So if I don't cover it in some of my classes, then I don't feel like I'm doing a huge disservice because I know some of those connections are going to be made in another class. But the other thing is, I want people to understand what the heck we're talking about. And so what I've tried to do, and I don't know how successful I am, but what I've tried to do is what I consider to be the equivalent of matrix algebra without the matrix algebra, and by that I mean that all of the matrix, okay, the matrix, it's not gonna come naturally, all of the matrix stuff that we do.

Patrick 10:28
This is gonna be a long hour. If we keep doing this, how about is we'll do the matrix, if it's a proper noun, okay? And the matrix is even Warner Brothers can't sue us for saying Cy is a k by k matrix. Alright,

Greg 10:44
so if I say it with a lowercase m, its matrix. If I say it with an uppercase M, its matrix matrix. Good. All right, back on point. Yeah, sorry, sir. So the matrix algebra that we're going to talk about is a

representation of things that go on geometrically. And so what I try to do in my classes is, rather than drag people through all of the matrix algebra, I try and help them to understand what's going on geometrically. And that is the equivalent of matrix algebra, it's just a way to try to visualize it. So that's my compromise to provide the students in our own program, the depth that they need mathematically elsewhere, and then to help people who I think are trying to be good researchers and make good thoughtful choices to understand what's going on visually as much as possible. And I don't know that I'm doing it right.

Patrick 11:30
I'm looking forward to hearing how you talk about that geometrically, because I don't do that. I'm particularly excited if you'll spend a couple of minutes on that because I think there's a nonzero chance, you'll stab yourself in the palm of the hand with a skewer again, since I saw that with an eigenvalue example you gave a while back. Setting that aside, I do see two motivations for understanding matrices from an applied user standpoint, right? If you're quanti, this is peeling back the layers of the onion and really seeing what's happened. If you're an applied user of these. For me, there are two advantages one really is is popping open the hood. And understanding what's happening is there's a level of commitment of saying, Okay, I'm going to take some control over this, right, this is what I'm doing. This is what I'm imposing on my data. And this is why this is happening. So one is a little bit more ethereal. The other is more practical, because you're going to get an error that says Sign non positive definite likely involves elements six, four, well, exactly. You have an NPD sigh What what does that mean? We need to understand the basics of matrices to understand Oh, first sigh represents this particular part of my model. non positive definite means this particular characteristic of the matrix, now I have a really good sense of what the issue is and how I might go about fixing that. So why don't we go through

Greg 13:03
some basic terminology associated with matrix algebra or linear algebra, as it's sometimes called, and talk about the operations that we perform on those and try to make anchors to things that are geometric and try also to connect them to some of the statistical things that we do

Patrick 13:20
good, because like so many other things, we're going to take stuff that you all are completely familiar with. And we're going to rename it and use more occluded terminology, and then try to make you feel bad about not knowing it. So you know, business as usual. Really?

Greg 13:39
Yeah, welcome. Alright, so let's start off small scalar. Give me a definition of a scalar.

Patrick 13:44
What do you got? A scalar? An individual number,

Greg 13:46
it is a number. Do you know why it's called a scalar?

Patrick 13:49
I did not know this was going to be pop quiz.

Greg 13:53
Because when you multiply by that it scales it to a number greater than one, it scales at a number smaller than one it shrinks it. If it's negative, it actually scales it in the opposite direction. So yeah, it's called a scalar. Because it scales things. Look at

Patrick 14:07
you, Mr. high school math teacher, I did not know that.

MovieClip 14:11
I'm trying to free your mind. But this is the very first

Patrick 14:15
example is all of you know what a number is, but we're going to bring it into the garage, sand it down, repaint it, bring it out and call it a scalar. There you go. So a scalar is an individual number. What if I have more than one? How might I organize those in some systematic way? Ooh,

Greg 14:33
well, one systematic way that you could organize them would be in a nice column that we would have. And that's sometimes referred to as a vector, like a data vector, right? If we have measurements for a bunch of people on a particular x variable, and we put them all in a column like we typically do in our data file, then that would be a vector would be a vector that would have just one column, but as many rows as there are people, so there would be n that would be an example of an N by one vector, column vector. After a victory

Patrick 15:00
vector, Roger roger, we have clearance clients, Roger roger, what's our vector Victor, sorry, that was legally bound to say. So you used n by one. So we need a postal system. All right, which means everybody needs an address is where does each scaler reside. So that when we talk about an N by one, there are n rows and one column, the first number we say, are going to be the number of rows. And the second number we say, are going to be the number of columns. So when Greg says an N by one vector, and if we assume n is sample size, which we often do in the kind of work that we do, and n by one vector means that they're n rows and the single column, alright, so it's always row by column, that's going to be our postal address system.

Greg 15:52
And we're going to use that very, very soon as things start to get a little bit messier. I would like to put a geometric spin on this,

MovieClip 15:59
this is your last chance. After this, there is no turning back.

Greg 16:03
Here we go. The word vector has meaning in other contexts, maybe the word vector meant something to you outside of the math here. But a vector is sometimes used to represent an arrow, something that has a particular direction, something that has a particular magnitude. It exists in physics, it exists in a lot of different disciplines. So what I would like you to do is, and this is going to take some visualization, but I think you can do it,

MovieClip 16:26
I can only show you the door, you're the one that has to walk through it.

Greg 16:29
For all of our n people. However many people we have n of them, I would like you to imagine each person has their own axis. So for simplicity, let's imagine there are three people, the first person I will call owl. And the second person I will call Beth, and the third person I will call Carl. So imagine that I have one axis, that is owls axis, and owl could have some score that falls along that axis. And Beth has some axis and has a score that you follow along that axis. And Carl has another axis. Now these people are all independent of each other. So we're going to make their axes go out at right angles, just like you're used to doing when you think about things in three dimensions. Now when I represent the data for owl, Beth and Carl, what I'm going to do is figure out where along the first continuum is owl score, and then Beth score and then Carl score. And I'm going to draw an arrow that goes from the origin out to wherever that point is, that is the corresponding location of ALS score, best score and Carl's score. So that is why it's called a vector because those three numbers tell us where to go. If we're sitting at the origin, how are we going to drive out to where the data point is, and that's it. Now, if we had more people and people, 100 people, 1000 people, however many you have, it's harder to visualize, but we're doing the exact same thing, every person has their own axis. And the data essentially give us a direction to point if we're standing on the origin, we're going to go out to that place that is a location that has everybody's score represented by that ultimate point,

Patrick 18:01
one of the most famous books ever written in factor analysis was by LOL, thirst, stone, and the title is vectors of mind. It's one of my favorite titles of all time, because it's the marrying of the statistics and the subsequent work that we all do. So that is what he's referring to, are these vectors. Exactly.

Greg 18:25
So I would like to take an idea that you had and expand it out. Now you had an idea of giving each element in that vector, that column vector and address so the second score down would be element two, one because it's in the second row and the first column, but we only have one column in this example, what if we had many columns? What if we had p columns, and it would be this array of numbers? Well, that's what we call a matrix. A matrix is like a bunch of column vectors all side by side. And the one that we're most familiar with is a data matrix.

Patrick 18:58
We work with matrices all the time, if you have a calendar on your wall, the weeks are the rows and the days are the columns, and you go to the third week and the second day, and it's Tuesday, you've done

the three two elements of that matrix. So if somebody says, Well, what is the fourth three element of your data matrix, you go to the fourth row, you go to the third column, and there's a little cubby hole that has the scalar that lives in that for three element. The term I sometimes like of a matrix is it is a doubly ordered structure of scalars. And what that means is the W ordered is it's ordered by row and it's ordered by column. Now, as soon as we start thinking about that row by column, we can have rows and columns that are not equal to one another. That would be called a rectangular matrix. But we might have rows and columns that are equal to each other and wait for it says square matrix, ooh, we work with square matrices every day. They're a correlation matrix, their covariance matrix. Yeah, it's p by p. So if n by p is your sample, and is the number of cases, p is the number of variables, P by P is a square matrix. So we can have a column vector, we can have a rectangular matrix, we can have a square matrix,

Greg 20:23
a couple of things, I will add to that, first of all the square matrices that you just talked about, they have a diagonal, right, the elements that go down 112233, all the way down. The thing that's nice about the matrices that you mentioned is that they're also symmetric. The elements above the diagonal, whether they're correlations, or covariances, are the same as the elements below the diagonal correlations or covariances. So for example, if we are in row three, column two, whatever that element is at that address, then that will be the same value as what would be in row two, column three. So everything above is symmetric relative to everything below. Let's

Patrick 21:01
do a little notation because of course, it's a podcast. And we really want to start describing notation, in addition to your vectors moving out into three dimensional space. Alright, we're all going to agree that if is a lowercase letter that is neither italicized or bold, that's going to be a scalar. So lowercase x, just like if you type it into your computer, that equals five, it's just a number, we're all going to agree that a lowercase letter that is bolded is going to represent a vector and an uppercase letter that is bolded, then that means it's a full matrix with multiple rows and columns. So lowercase, not bold as a number lowercase bold as a vector uppercase bold as a matrix, a lot of times, we don't want that column vector, we want to make it a row vector. So we're going to do a thing called a transpose. And it's super simple, just walk up to it and shove it on its side. So where we have a little bold X, now we're going to put a prime sign next to it x prime, when you see that no matter where you are in the world, is if you see a bold X with a prime, that's now a row vector, it was an N by one vector. Now it is a one by n vector. But we can do that with a matrix as well. So if we have an n by p matrix, x, x prime is a p by n matrix, all we do is we make the first column of the matrix, the first row of the transposed matrix, the second column is the second row, and so on. And then a little bit when we wander in this way, we're going to make huge use of transposes. Because that's going to allow us to do all sorts of really cool things when we start to manipulate our data.

Greg 22:48
So there are two roles that matrices play in our world. One is, it's going to serve as an agent of transformation. For other things. I'm going to hold off on that for just a minute. But I do want to think about it in the context of our data matrix. When we talked about a data vector, we had one column, right where everybody had a score on x. Well, now we have an x one and x 2x 3x 4x 5x p, right, as many as

we have, I'm going to go back to that space that I defined, where I did it for three dimensions for owl, and Beth and Carl, but we could have it as many dimensions as there are cases right and dimensions. But I'm going to do the same thing for each of the columns, I am going to trace a vector from the origin out to wherever that point is, that represents everybody score on x one. And then I'm going to trace a vector out to the point that represents the score that everybody has on x two, and you know what x two might be in completely different units than x one. So it might go on for miles, whereas the one for x one might be this stubby little thing. We'll talk about that in just a second. And I do it for x three, and x four and x five, and I get this whole family of vectors. And they might be shooting off in very similar directions, or they might be shooting off in very, very different directions. And those directions that those are heading relative to each other are going to tell us something about the relation that those variables have with each other. Sometimes we talk about data rather than being in their natural metric as being in a standardized metric, right, our old friend the z score, what the z score does in this particular context is it gives all the variables the same units. So when I am plotting owl's Z score, and Beth Z score and Carl Z score, I will have the same units for every person, every variable. And in the end, the vectors that I draw out won't be of different lengths, the vectors will all wind up being of the same length. So one of the benefits of standardization just visually, is it makes all of the data vectors that are heading out, they're going to still be heading in the same direction as when the variables were in whatever their natural metric was, but it brings them in so that they're the same length. So standardization is really helpful. thing just in terms of imagining all of these vectors heading out the same distance,

Patrick 25:04
take those two variables, and you do your clusters in the way you described. And you draw those vectors, right? It's almost like it's a lightsaber. Have you ever go out in the backyard on a misty night, and you have one of those tack flashlights, and it looks like, okay, so maybe I've done that, wow, that's pretty cool. Think about two variables them that are uncorrelated, that they have no relation to one another, well, those vectors are at 90 degrees, they go off forever, never to me,

Greg 25:32
that's right, 90 degrees, they're perpendicular, or as we say orthogonal, as they get

Patrick 25:36
a little bit more related, well become a little closer to each other. And a little closer, a little closer, at the extreme. If those two variables are correlated 1.0, there's a single vector that's needed to run through those because one is a carbon copy of the other, it's really neat to think about the correlations in terms of the geometry of the angle between the two vectors that are associated with each of the observed measures.

Greg 26:07
And in fact, you could see it in your data matrix. If you converted everything to Z scores, because of two things correlated 1.0, the columns of z scores would be identical, they would map one to one. So that's why when you draw that vector out there, when it's standardized, you're drawing the exact same vector, whether it's standardized x one or standardized x two. And when they fall at right angles to each other. Any two vectors that fall at right angles to each other are really just like the axis that defines owl and

Beth and Carl, those are at right angles to each other. Also, because those people are independent. When our data vectors are at right angles, it's the same idea that they are also assumed to be independent. Now the cool thing that you should all go let your high school trigonometry teacher know is that you are using trigonometry right here. As Patrick said, the angle between two data vectors is extremely telling specifically breezed through it.

Patrick 27:05
This is my favorite thing. The only time in 30 years I use cosine.

Greg 27:12
Yes, exactly. It's the cosine. You told yourself, you would never have any need for this in real life. This is it. Today's the day cosine, the cosine of the angle between two vectors is I don't mean it's like it is the correlation between two variables. And when those variables are right smack on top of each other, I have zero angle between them, the correlation is 1.0. And when the angle between two vectors is let's say 45 degrees, quick punch it into your calculator, the cosine of 45 degrees, I think it's point 707. And when they're right angles, the cosine of 90 degrees is zero. And I'm getting so happy talking about angles between these things. What do you suppose happens if I have two vectors that are heading off in opposite directions? What would your gut tell you that is negative? Exactly right? And if they're headed in precisely opposite directions, so that they actually fall in the same line? You would guess?

Patrick 28:11
I'm negative 1.00. Look

Greg 28:14
at you. Yes, exactly right. If there's 190 degree angle between these two data vectors, then the cosine of 180 degrees is going to be a negative one. And if you went to your standardized data, just for simplicity, and you looked at the columns of numbers there on x one, when someone had a 2.5, on x two, they'd have a negative 2.5. If someone has a minus 1.2 on x one, then they have a positive one purchase everything there just flips. It is so beautiful.

Patrick 28:42
This is the matrix, right? When he sees it is the one.

Patrick 28:55
When he sees all the things coming down, he's like, whoa, right is no wait, maybe that was Bill and Ted's Excellent Adventure, I get those mixed up.

Greg 29:05
I like that you went with matrix, you put an OOM loud over it to give it more of a European flair.

Patrick 29:10
Whatever you agreed on,

Greg 29:12
I forgot the matrix. Let's go with Ma Trix matrix. I like it. All of

Patrick 29:17
these are puzzle pieces that click in together,

Greg 29:21
there are two other things I want to put in place. And then we can talk more about operations. One is that if you had a whole bunch of variables whose vectors were headed off in very similar directions, we know now that that means that they're very highly correlated, but start thinking about the analyses that we do, right? What does it mean when our x's are correlated, or our predictors are correlated, that's going to be relevant for some of the statistical things that we're going to talk about in a little bit. The other thing that I will mention is that when you learned about cosine, you learned about it in the context of a right triangle. And you might have even had the what's it called? What's it The mnemonic, yeah, you might have even had the mnemonic,

Patrick 30:03
you need a way of remembering that.

Greg 30:07
That was good. So you might have even had the mnemonic device of SOHCAHTOA, sine is opposite over hypotenuse, cosine is adjacent over hypotenuse, and tangent is opposite over adjacent. So the Cosine being the adjacent over the hypotenuse, what that really means in the context of two variable vectors is that the cosine of the angle is also the length of the projection that you would get if you projected one of those vectors on to the other projected it on there so that it formed a right angle. So the length of that projection represents the correlation. If two vectors are right angles, there is no projection of one vector on the other. And if they're right on top of each other, the projection is the full length of that vector, the full one unit of that. So just keeping that in mind, too, that this correlation also represents the length of a projection, let's drag

Patrick 30:58
us kicking and screaming into say, third grade math. We're gonna start with our data matrix. Typically, we have x all the model building that we do all the model fitting all the testing is all based on this n by P data matrix. But to do that, we need to start working on ways of combining certain things may be vectors may be matrices. So addition, subtraction, multiplication, and division before you're the high school math teacher, why don't you walk us through that? Well, back in my day,

Greg 31:33
we didn't have all four of those. I love it when people say didn't use to teach chemistry, I'm like, Yeah, but there was only 18 elements. All right, so addition, addition, whether we're talking about adding scalars are adding vectors or adding matrices, the main idea is that they have to be of the same dimensions, right? If you're adding scalars, well, that's what you were doing back in second grade third grade. Adding vectors just means that you take all the corresponding elements, and you add them together. So if you have one column vector, and another column vector that is of the same length, the

sum of those just means adding the corresponding elements and getting out a vector that is exactly the same size. So if you had a data vector x one, right next to it x two, and you want to create another thing, that is the sum of those great, that would be the sum of two vectors. And then for matrices, it's just the same idea that you've got a matrix, let's say it's an n by p matrix, it doesn't have to be it could be any dimensions. Adding that to another matrix requires that the other matrix be of the same dimensions and by P, and then you just do it element by element. And that's it. That's matrix algebra, addition Done and done.

Patrick 32:40
Another term, just to confuse you is we often talk about the dimensions of the matrix of the rows and the columns, we talked about that as the order of the matrix will be order is n by P, or it's P by P, or it's k by k, or whatever you might have. And to add two matrices, they have to be the same order, how I teach it is in your mind's eye, lay one matrix on top of the other, and just add each of the elements that overlay, if they're not the same order, they don't cleanly lay on top of each other. So you can't add them subtraction, knock yourself out a minus b, same thing, just subtract b from anything to add to that, not at all. That's addition, that subtraction, multiplication. Now we start to mess with you as multiplication of two matrices is actually very different.

Greg 33:31
multiplication of matrices with other matrices or matrices with vectors starts to get a little bit into the weeds. And it doesn't follow rules that are intuitive. And I will say the words that go with it, but I'm not planning on doing much more than that, instead, I'm going to transition into a geometric explanation. So let me start with the words that correspond to the operations. But let's say I have a vector, and I would like to multiply that by a matrix. And what I do is if I'm multiplying, let's say a matrix, big a capital A, I think Patrick said it has to be bold in capital by a vector x, a lowercase x that's bold. In order to do that, I have to make sure that the number of columns in the matrix match the number of rows in the vector. So it is a one to one matching, where you multiply each corresponding element going across the first row of the matrix by each element going down the column of your vector, and then you add all of those together, who all right, there's a lot of stuff going on there. And that's an absolutely necessary thing to do for a lot of the manipulations that are going on under the hood in statistics. Geometrically, I would like to come back to just the idea of what a matrix is. And I talked about it as a collection of vectors, but it also serves a role when we multiply by a matrix. So imagine I have a vector in space could be a data vector could be any other vector. When I multiply that vector by a matrix, what that matrix is actually doing is transforming that vector, it might be transforming that vector in terms of location, so it might rotate it to a different location, it might stretch that vector, it might shrink that vector, it might flip that vector. So the idea of multiplying a vector by a matrix is transforming it in some way. When I am talking about multiplying multiple matrices, we can think about it in a couple of ways. One is, we can think about doing the transformation of one matrix on all of the vectors contained in the other. That's not a bad way to think about it geometrically. The other though, is that if one matrix represents a transformation when we apply it to some vector, and then we have another matrix that we also want to use, when you are multiplying two matrices together, all you are doing is compounding their transformations. So if one matrix rotated everything clockwise 90 degrees, and another matrix rotated everything back 45 degrees, the matrix that you would get when you multiply those together just does the compound operation, which would be just rotating everything 45 degrees. So the idea of matrix

multiplication is really just taking the transformation that each one represents smooshing it all together into what the final transformation would be.

Patrick 36:19
The weird thing about multiplying is, as we just described with addition and subtraction, you take an m by p matrix added to an m by p matrix, and you get an n by p matrix or whatever those dimensions might be. Well, often, you'll multiply two matrices of a particular order and get a matrix of a different order. And it goes toward what you're saying and transforming one as a function of the other. If our first matrix is n by p, our second multiplier has to start with P in its order. So the inside numbers on those orders have to be the same. So you could multiply an n by P by A p by K matrix. So those peas are the same on the inside. And what that means is the columns in the first matrix have to match the number of rows in the second matrix. But what's really cool is the outside numbers are going to give you the dimensions of your new matrix. So if you have an n by p times p by k, you're going to get an n by k matrix. We will commonly in this happens, oh my gosh, all the time x prime x? Well, what we're going to do is combine a transpose when we multiply A by itself, well, x is n by p, so x prime is going to be p by n. And then we're going to multiply it by x, which is n by p. So the inside ends match, but our resulting matrix is going to be P by P.

Greg 37:47
Well, so let's transition when we go from addition to subtraction, we go ah, it's just, you know, sort of adding the opposite. When we go from multiplication to division. How do you explain that one?

Patrick 37:59
This is a new ballgame is addition of subtraction or easy multiplication, you're like, oh, yeah, I see this. In Division, we're kind of off to the races. The short description is there's not a direct corollary of dividing two matrices. So we're going to trick it by multiplying by the inverse of a matrix, which means we got to get an inverse of a matrix.

Greg 38:26
And the task is kind of a formidable one and requires a lot of math, two things that I will say, one is that we're talking about inverses of square matrices here, the concept of an inverse of rectangular matrices exists. We're not talking about that. The thing that I will say geometrically about it, though, is like I mentioned before, a matrix can be thought of as this transformer, that when you multiply it by a vector or some other matrix, what it is doing as it is transforming it, it's rotating things, it's stretching things, the idea of an inverse of a matrix. So if a transforms something when you multiply it by A, A inverse undoes that. So whatever a did a inverse takes it back to the original starting point. For that reason, when we have a times its own inverse using matrix multiplication, that nullifies whatever you did, it's as if you are multiplying by i don't want to say nothing, but you're essentially leaving everything in identically the same place. And so when you multiply A times A inverse, or inverse times A, you're left with a matrix that just has a bunch of ones down the diagonal, and it is called the identity matrix, because it leaves anything, you multiply by it in identically the same place as it started. But now if we talk about the matrix A divided by the matrix B, we don't literally do that. Instead, we do a times the inverse of b. And we're familiar with doing that in regular old math. If someone says I want you to take two and divide it by three You know that that's like taking two and multiplying it by the inverse of three? Well, it's the same idea

here. A divided by B in terms of matrices is just a times the inverse of b, where the inverse of B is that matrix that undoes whatever b itself could

Patrick 40:16
do. Imagine you're in middle school, and you have 5x equals y, and you want to isolate x, well, all of us look at it and say, Oh, we divide five, and so x equals y divided by five, so you isolate it x, y, we use these inverses. Often, not always, but often is to solve unknowns in terms of knowns. So you have 5x equals y, we divide both sides. Well, the general rule is we can't literally divide two matrices, so that 5x equals y, we can multiply both sides by the inverse of five, which, of course, is 1/5. What is five times five inverse equals? Well, that equals one, right? And that's what Greg is saying is, if you have A inverse, that equals the identity matrix, it's a square matrix, it has ones on the diagonal zeros everywhere else, it's the matrix equivalent to the number one. But on the right hand side of the equal sign, now we have y times five inverse, which in our scalar algebra, we know is the same as y divided by five. So that's what we're doing. It's a little bit of a Penn and Teller Sleight of hands is saying, Okay, if you're going to be a winner, and not let us divide by a matrix, well screw you, I'm going to multiply by the inverse of a matrix. And let's see how you like me now.

Greg 41:46
But those are the operations right? We have addition, subtraction, multiplication, and division. And those get us pretty far. There are other things we do with matrices, though, also that are kind of specialized. And it seems worth mentioning them, they occur occasionally something called a trace, which just means you take a square matrix, and you add up all the elements on the diagonal that comes up in some of the derivations that we do. Something called a determinant is a little bit messier. How do you think of the determinant,

Patrick 42:14
one thing that's confusing about the determinant of a matrix is no matter the dimensions of the matrix, so we're still talking square matrices, the determinant is a single numerical value, it's not a matrix, it's not a vector, it's a single numerical value that captures the general variance that is free in that matrix independent variants in that matrix. The numerical value itself is almost always uninterpretable. It's not like I can compare my determinant to your determinant, because this is scaled the in whatever the scales of your measured variables are. So we often have a determinant of a covariance matrix, or a correlation matrix. And what we do know is, the larger the value of the determinant, the greater the general variance, the independence is among the columns. In that matrix, the smaller that determinant is those vectors of yours that you so nicely described, they start collapsing on one another, and they start getting closer and closer and closer. So when I teach it, I have a quick way of thinking about it. Imagine we had a two by two matrix, right? So super simple, two variances in a covariance. All right. Now, again, I'm asking a lot here. But as you're driving, just picture in your mind's eye, not your third eye, because not all of us have third eyes. in your mind's eye. The first row, we're gonna denote a and b. And the second row, we're gonna do note C, and D, A, B in the first row, CD in the second row. So the diagonal is a and d, alright, those in a covariance matrix represent our variances, and then those off diagonal is going to represent a covariance for that simple two by two matrix, the determinant is the product of the diagonal minus the product of the opposite.so A times D minus B times C is the determinant of that matrix. Now as you go to the three by three, four by four, who boy gets much more

complicated, yeah, let's put some numbers in there. So the first row is going to be two and three, and the second row is going to be three and eight. So the determinant is going to be two times eight. That's the diagonal, two times eight is 16. Minus three times three, which is nine. Alright, so we get a determinant there of seven. So there's some independence between the two. Alright, let's go to the extreme. What if it is two and zero and zero and eight? Alright, so two and eight are the diagonal elements, zeros, the off diagonal, those are Greg's 90 degree vectors. Yep. Well, two times eight is 16. That's the diagonal minus zero times zero. Well, all of that variance is independent. Two times eight is 16. We're not subtracting in anything, it's minus zero, because those are traveling off infinitely in 90 degree angles. So the third one, picture two, and four, and four, and eight. So the first column is two and four, the second column is just two times the first column for an A, well, now the product of the diagnose two times eight is 16, we're going to subtract the product of the off diagonal, which is four times 416 16 minus 16 is zero, the determinant is zero, there is no general variance, because one column is simply two times the other column. So I kind of like that way of thinking about a determinant is representing the generalized variants that exists in a p by p matrix.

Greg 45:50
And that already has some geometric spin to it. So yeah, you, I'm going to add a couple of more geometric pieces for those people who find it helpful when you were describing your three examples, which I love. For the one where you had the orthogonal vectors, I think one was

Patrick 46:06
two and zero and zero and eight.

Greg 46:09
So that actually makes a rectangle that's two on one side, and eight on the other side, and the area of that rectangle is 16. That would be the determinant. When Patrick did some vectors that were not orthogonal to each other and computed the determinant, those could form a parallelogram rather than a rectangle, the area of that parallelogram is what the determinant is telling you. Now, as the vectors move closer and closer and closer to each other until they're on top of each other, what is the area of a parallelogram that's defined by two lines that are right on top of each other? The answer is there's no area, the determinant goes to zero. So the determinant can be thought of as the area of the rectangular parallelogram in the two dimensional space that Patrick described, or the volume of whatever the shape is, when we move out to more dimensions, it becomes instead of a parallelogram, it becomes what's called a parallel a pipette. But then when I asked you what is the volume of something that is flat, there's no volume. So when one of the dimensions collapses in, boom, the determinant goes to zero. The other thing that I just want to reiterate that Patrick mentioned, that is so hugely important, is the idea of the determined is representing that thing we call generalized variance. And forgetting about geometry in the way that I've been talking about in terms of vectors, just think of a scatterplot. And if we wanted to talk about just in general, how poofy the data are in that scatterplot, the determinant of the correlation matrix, which would be in a standardized metric gives us some sense of how big that cloud of points is. And as our scatterplot gets narrower and narrower and narrower, as our two variables get more and more correlated, things vary a whole lot less, the determinant gives us that sense, the determinant goes closer and closer to zero as those variables relate more strongly. When we have a point cloud that is in three dimensions, were hacked, just go all the way to p dimension, this p

dimensional point cloud, the determinant tells you how much variability there is overall. So it's a nice way to take the idea of variance and extend it to multiple dimensions.

Patrick 48:16
This isn't like a pedagogical mechanism of going out to an up eight and multiplying it to get 16. That's it. That is what we're doing. Now. Let's then turn briefly to So what So what do we do with the determinant? It turns out to invert a matrix, you must have a nonzero determinant. You cannot invert a matrix, if you have a zero determined,

Greg 48:46
do not try and bend the spoon. That's impossible.

Patrick 48:50
Not often do we get exactly zero determinants? We certainly can in our day to day life, what hassles us is when we have near zero determinants, and this often comes up in multicollinearity. variance inflation factors like this stuff has its fingerprints everywhere. For those of you who use sem programs, when you get that error message that says Cy is not positive definite cannot be inverted.

MovieClip 49:19
How about I give you the finger and you give me my phone call?

Patrick 49:25
This is what's going on. If you have a determinant at or near zero, and you cannot invert that matrix, the term you have is that matrix is non positive definite in P D. What that means is there some linear dependency among the columns of that matrix, and you actually have problems that's like a deal breaker. Again, that OLS regression coefficient vector is x prime x inverse x prime y. Well, we have that x prime x with is now a p by p matrix that we have to invert. If there's a linear dependency among those columns, right, let's say you have two predictors in your data matrix that are highly correlated, that determinant is near zero. You can't invert it, and you can't do your OLS regression.

Greg 50:20
Oh, yeah, everything goes to hell, right? Because what you're determined it is telling you is that even though you've got p dimensions worth of variables, you don't have p dimensions worth of behavior in those variables, they're starting to collapse on you. And it might be as simple as two things that are nearly perfectly correlated. But more likely, it's some kind of multicollinearity. That's really hard to observe just by inspecting things,

Patrick 50:43
that's what's really hard is, I will have a student come and say, I can't invert this matrix. But I look at the correlations and nothing is bigger than point five, eight. And so I know it's not multicollinearity. And what I have to remind them of is the multi part. Yeah, because Greg has been giving these wonderful descriptions of the vectors. But notice listeners, he's never moved beyond three dimensions. Why? Because he and I are not Stephen Hawking. If you

Greg 51:14
are looking for trouble. You found it. Yeah, just trying

Patrick 51:17
to use Oh, what about four dimensions, or five dimensions, or six, that's the point of the multicollinearity. You don't have an Agatha Christie smoking gun, oh, x one and x three are correlated, point nine, seven, the drunken punch in the faces. When you do this across the entire matrix, they combine in a way to say dude, you have a p by p matrix, you do not deserve p by p independent dimensions, that there's an overlap in a way that is just going to jerk our chain. I like that. Okay, we got addition, we got subtraction, we have multiplication, we have division, we have inverses. And we have determinants. Do you know what that brings me to what we are at minute 85 in recording? And they haven't started talking yet about what the point of the episode is?

Greg 52:14
Oh, well, what did they do in the movies when that happens?

Patrick 52:18
Oh, dude, there you go, the Metrix Part Two

MovieClip 52:23
or part two missions. You got to come walking into this

Greg 52:34
absolutely. I smell

Patrick 52:35
a sequel to this for all it's worth. If we didn't hemorrhage money on this out of our own personal bank accounts. This would be like a pay day. It's gonna be part two. Uh huh. Did I tell you I went to the movie with my wife of the first of The Hobbit. Oh, yeah. Right. So this was years ago. There's all the action and everything and then like the little hobbit friends are all together. And they walk off in the field and it's like, oh, Bilbo, you're my friend and no, whatever. You're my friend too. And they walk into the field and they start running the credits. And I was like, son of a I had no idea was a freakin trilogy. I had to sit through like six hours to figure out if he was gonna throw the ring into the volcano. So yes, this is like little Bilbo Baggins is gonna walk off into the field. I'm glad you're with me. When looking for that clip and post processing, I realized that confused the movies of The Hobbit with Lord of the Rings and Bilbo Baggins with Frodo Baggins on behalf of quantitated. I apologize. Please direct all complaints to gr Hancock at quietude pod.org. And we are going to come back next week and we're going to say, Okay, we've got this whole architecture now that we built. What are we going to do with it? How do we use this to our advantage? The sequel UI I told you to stop doing that that weirds me out.

Greg 54:16
You still don't like my my sexy, sultry voice. Next week on quantity,

Patrick 54:21

buddy. We're gonna call it here. My trick I shut things down. We hope this has been of some use. I know it's like weird Oh town to do matrix algebra while you're driving from the 110 to the 10 to the 405 Dad's gonna be John. As always, thanks for your time, and we will start this conversation next week.

Greg 54:48
Thanks everybody.

Patrick 54:49
Take care God help me

Greg 54:59
nope, nope, nope. We can't do that can't do that.

Patrick 55:03
Thank you so much for listening. You can subscribe to quietude on Apple podcasts, Spotify, or wherever you download playlists for your upcoming Spring Break road trip and please leave us a review. You can also follow us on Twitter we are at corner tude pod and visit us on the inter to back Conner tude pod.org for past episodes searchable playlists, show notes and newly available transcripts and other cool stuff. Finally, you can get ridiculously fun quantitated merch at red bubble comm where All proceeds go to Donors Choose to support low income schools You have been listening to Quantitude. Even if you intentionally try to mispronounce us we still come out as Quantitude. Today has been brought to you by Quantitude Brothers Studios, created after the surprising success of the Mattrix, QBS is proud to bring you upcoming releases including Sheepless in Seattle, Citizen Kanye, Star Warts, Apocalisp Now, the God Feather, and Shaving Ryan's Privates. By Third Eye Prosthetics: Do you have a loved one who suffers from the complete and total inability to introspect or take viewpoints beyond whatever is sitting directly in front of them? founded out of sheer frustration by Greg and Patrick's wives, Third Eye Prosthetics can help. And by eigenvalues and eigenvectors who would like to let you know that in next week's episode, they are going to [delete] you up. This is most definitely not NPR