The Podcast *Quantitude*

with Greg Hancock & Patrick Curran

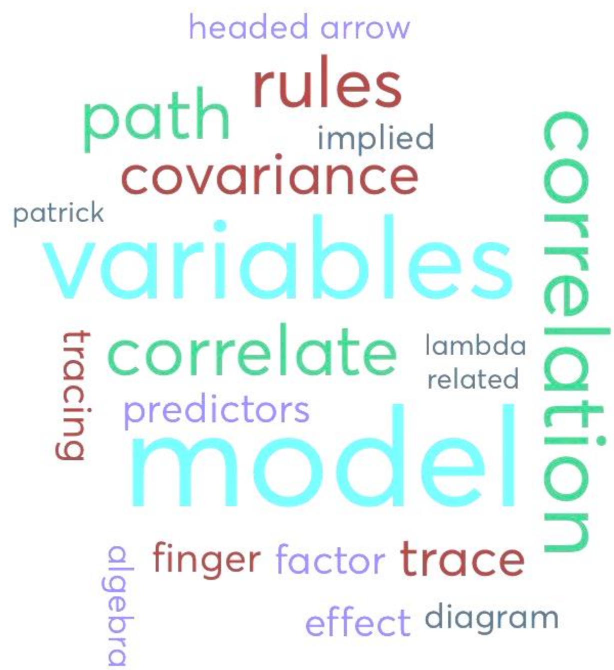
Season 3, Episode 24:

The Wright Stuff – The Beauty of Path Tracing Rules

Published Tuesday March 25th, 2022 • 1:00:10

SUMMARY KEYWORDS

model, variables, correlation, correlate, path, rules, covariance, trace, tracing, finger, predictors, effect, factor, algebra, headed arrow, diagram, implied, lambda,



Greg 00:01

Hi everybody, my name is Greg Hancock and along with my seasonal allergy affected friend Patrick Curran we make up quantitated we're a podcast dedicated to all things quantitative ranging from the relevant to the completely irrelevant. In today's episode we talk about civil rights path tracing rules as an alternative to covariance and matrix algebra, how the rules work and the tremendous insights they can provide toward understanding a model. Along the way, we also mentioned the Unabomber, Crate and Barrel, grocery Lane profiling, tedious as poop. Throwing dead cats. Senior animal husbandman, using your fingers, creepy guy in an alley, sweat pants versus suits of armor getaway car drivers, hold my guinea pig, chalkboard contests, western Kansas and getting tazed. We hope you enjoy today's episode.

Patrick 00:47

Sure, I gotta tell you a Spring has sprung in North Carolina and how you know that is my eyes are swollen shut and I'm on this bad trip. have flown ace Clarington and Benadryl ensconced in coffee. Uh huh. I am in a T shirt. I'm in shorts and bare feet. I have my feet up on my desk. I'm pouring new coffee. I would have to say looking at you as we speak. You appear to be wearing multiple hoodies and are shivering and have gloves. All right now you're just looking like the Unabomber at this point. The Unabomber with a lovely Hutch behind you with kitschy collectibles. Isn't that nice? So the Unabomber after going to Crate and Barrel? Right? Is there some story behind this?

Greg 01:37

Well, I am in Ohio where I come to help out with my elderly in laws from time to time, and I stay upstairs here. But unfortunately, the furnace has died. And outside it's currently 21 degrees Fahrenheit, which is like minus six degrees Celsius for our friends elsewhere in the world. In the bedroom. The top temperature is 45 degrees Fahrenheit, which is around six degrees Celsius. So right now I'm upstairs just shivering to the bone. I may have to go get a blanket here.

Patrick 02:11
Yes, he shivering needs a blanket and for our friends elsewhere. That means he's whining like a big baby. Whatever, just suck it up and get to work, dude.

Greg 02:22
But anyway, I'm here helping out and I will go to the grocery store in a little while. The grocery store here there's one lane, unlike at home where there's like eight lanes. And I don't know about you, I have this habit of looking to see which lane I should be in all the time. And then sometimes I'll zip over to another lane because I think it's moving faster. Try to figure out the average age of the customers in line to see which one might move faster.

Patrick 02:46
You're profiling. That's what you're doing your lane profiling in full

Greg 02:50
recognition of the fact that other people are doing that to me. But invariably, you know, what happens is I look back at the lane that I moved from, and I would have almost always already been through. If I just stick where,

Patrick 03:03
oh, it's an absolute rule. Yeah, you change lanes in heavy traffic, and the guy behind you is gonna pass. It's a law of life.

Greg 03:10
I used to drive my dad a lot of places and my dad was just the sweetest, quietest, most patient guy. But he got so mad when we would be out driving. I remember I was driving him someplace one time, he was in the passenger side, we were on the Baltimore beltway. And we were at a standstill. And this guy in a truck was just weaving in and out of lanes going up on the shoulder coming over my dad's in the passenger side. Just so angry at that guy who seemed to be taking advantage of the system and all the injustice is of 80 plus years come to that moment for my dad. And eventually when we make it up a little bit farther on the beltway, we're sitting right next to him. And my dad realizes that because the guy's truck is unmistakable. It's a hot day, you know, I've got the window down, and my dad from the passenger seat leans across me flips the guy off and goes, you can't you can't do that. Well, I am driving, again, really mild mannered peaceful guy. And I really don't want to have a throw down on the side of the bottom floor. But I just like the justice of the fact that the guy had been all over the place. And in the end, you know what, he's just right next to us,

Patrick 04:27
all of us at one point or another have thought, hey, I got a shortcut. Hey, there's an easier way of doing this. And then it turns out to be 10 times harder than if we had just sat down, rolled back our sleeves and done what needed to be done. And that's kind of what we're going to talk about today. It is

Greg 04:45
and for all these shortcuts, they're really turned out to be long cuts. Let me just say that there is one short cut that I think is infinitely better than the long cut or whatever. And that has to do with a little corner of Our world of structural equation modeling. So without giving it away yet, although I think we gave it away last week, in structural equation modeling the way you and I practice it, it is also historically referred to as covariant structure modeling, because so much of what we do is try to propose a model that explains why variables covary the way they do, or why variables vary the way they do as a whole system. So there are a lot of moving parts just in this covariance structure, not to say anything of an accompanying mean structure that we could get into. But even if we simplify that, and say rather than variance and covariance, even think about it, correlational Lee, you've got a whole system of I don't know, 20 variables, 30 variable, however many you might have. And we ask why do those variables correlate the way they do? I mean, you've got this monster of a model sitting there right in front of you. That's a lot of moving parts to try to think about and keep track of,

Patrick 05:52
you know, what's interesting, Greg is there are different ways that you can think about these models, even though they all are doing exactly the same thing. There are some black box kind of instruction out there where it's like luck, don't worry, your pretty little head. But here's your model, here's your chi square, do everything you can to get an RMSEA of point oh, four, nine, and you're off to the races. So kind of model a model B is a little bit like I was born and raised, which is starting with what are called path tracing rules. And that's what we're going to talk about in this episode. That's going to be our shortcut that has a lot of glorious advantages. And the third one that we can talk about briefly here, which is covariance algebra based, for those of you who haven't been exposed to this covariance algebra is a flavor of scalar algebra that we talked a little bit about in the first matrix episode. But it is with random variables. So certain new rules apply. And Dave Kenney in his 1979 book has a really wonderful chapter on this can Bolin in his 1989 book has three or four pages that are very, very clear. So what we're trying to do in SEM is parameterize, our model in a way that reproduces our variance covariance matrix as closely as possible with respect to what we observed in the sample now means also come into this but we'll focus on variances and covariances here, now picture a model where you have two predictors x one and x two. Now picture in your mind, you have COV(x1,x2) -- what's the covariance between x one and x two? Well, it is the covariance between x one and x two, it takes on the sample value, because it's not a dependent variable.

Greg 07:40
Yeah, the models not trying to explain why they covary just allow them to Now picture y one

Patrick 07:45
and y two. Right now, they're both dependent variables. And they can be dependent in any way that you want. Maybe they're indicators on a factor, or maybe they're in a path model where you have different predictors. But now picture COVID y one, y two, well, we can actually write an equation for y one, and we can write an equation for y two.

Greg 08:03
So the basic idea here is that if y one is a function of other things, and y two is a function of other things, then how y one and y two covary will be a function of the characteristics of those elements that went into making up y one and y two.

Patrick 08:16
And like any good mathematics, there's a set of rules. There are different ways of presenting it again, I really like how Dave does it in his correlation and causality book, he has a whole chapter on covariance algebra, but when you lay out these rules, then you can take this covariance between these two, what are now equations y one expressed as a function y two expressed as a function, and you can distribute things, you can factor out things, things go away, and then you're left with what is the model implied covariance between y one and y two, given how you structured your own model, right, that's what SEM is doing. It's doing this covariance by covariance variance by variance. It's freakin tedious as poop is the technical term is poop tedious can be very tedious.

Greg 09:10
That is a new level of disclosure.

Patrick 09:13
Now to be super clear, you get such a good understanding of what the model is doing. It is absolutely beautiful. What it didn't do for me was give you that 30,000 foot visual, though, totally, you're saying this variable that lives way over here in the model is actually correlated with this other variable that lives way over in this other part of the model. And you can do that through the covariance algebra, but Oh, I'd rather use my fingers.

Greg 09:46
Exactly right. This is an example of where we can get really good at doing things in the weeds, right? That's what these rules are these covariance algebra kinds of rules, but it's so easy to lose sight of what you're doing. So when you say you want to use your fingers I love that because to me, what you're saying is, you actually want to start touching what's going on in the model, so you can understand it a lot better. Exactly. And so we're going to go back to really the early 1900s, over 100 years ago, and over the summer, you and I talked about school, right? And school, right laid the foundations for all of this stuff. And I think some of the ideas that he had very early on, which we could look at as very elementary, I think they are as insightful as hell, and really worth resurrecting because I think people aren't exposed to some of his foundational ideas that really help you to touch and understand what's going on in any model.

Patrick 10:40
And to do that, I think we can go back to our friend, the path diagram, Oh, I love them. I know right? One of the beauties of SEM is it is so amenable to a visualization of what we're trying to do. FCM just lives within a path diagram.

Greg 11:00
Alright, so to make sure that everybody's on the same page here, let's just talk about the basic building blocks of these path diagrams. And honestly, if just about anybody out there goes to a journal in their field and throws a dart in an article and then opens it up, you're probably going to find some version of a path diagram. And the path diagram is going to be a way of communicating your beliefs about a system of variables. And some of those variables might be measured, some of them might be latent. When there is a box, a box represents a measured variable, something that you observe something that if you go to your data file, you actually have scores for people on there might be some missing data, but you have scores for them. So things in boxes, rectangles, squares are going to represent measured variables, and how we connect those will reflect our particular beliefs about how they're related in whatever mechanism we think is operating. For some of you out there, you might not refer to those boxes as variables, you might call them nodes, you might call them vertices. For our purposes, we're going to use the traditional path tracing language and refer to those as variables. And then how do you connect those variables? Well, there are two types of connectors that we tend to use in the system that we're talking about. One is a one headed arrow, a single headed arrow that communicates your belief about causality, that the variable at the beginning of the arrow is hypothesized to have a causal bearing a direct effect on the variable at the other end of the arrow, we call them arrows or paths. Sometimes they might be called edges or arcs, but we're going to refer to them here as single headed arrows. And the other type of connector then is a two headed arrow, a two headed arrow does not repeat does not mean that you think the two variables at either end have an influence on each other. A two headed arrow means that you have a theoretical reason to believe that two variables are related to each other, but not because one causes the other. It's usually because there's some mechanism outside the model that you believe is operating to give those covariation or correlation, when we start to get a little bit more daring, and believe that there are variables operating out there that we don't have direct access to, we introduce circles or ellipses. And those represent our latent variables, things that we can't see. But we really think that they're an integral part of the mechanism that's operating a part of the explanatory mechanism as to why the variables we are able to observe that we are able to hold in our hand actually correlate with each other, or covary with each other. So we have boxes, and we have circles, and we have single headed arrows and we have double headed arrows, there is another symbol that creeps into these models, in fact, plays a really, really important role in some of these types of models. And that would be a triangle, usually a representational convention for being able to bring the mean structure into our models. So when we're not just interested in understanding how variables vary in covary, or correlate, but are interested in the amounts of those variables, then a triangle which often has a one sitting inside, it allows us to be able to model intercepts and means associated with both measured variables and latent variables. We are not talking about those here in this particular episode. Although the path tracing that we're going to get into can have implications for and relate to those kinds of things. We're going to keep it at the level of the covariance or correlational structure that we've been talking about. So far.

Patrick 14:13
All of us to varying degrees are familiar with path diagrams, as Greg said, you would throw a dart for some reason I tend to throw dead cats is you throw a dead cat at a journal. And there's a path diagram. No cats were harmed in the making of this episode.

Patrick 14:31
Well, there are huge advantages to these. What is often used and I would say almost exclusively used in applied work is a path diagram is used to convey the subsequent model of interest, what are the predictors? What are the outcomes? What are indirect effects? What are latent factors, and this is a brilliant use of these. And indeed, I think that a path diagram is often what makes us seem so much more accessible in an applied setting. You can see what your model is implying from a substantive standpoint. But I feel like there are several generations of researchers who for various reasons, didn't get that same training, that these path diagrams, when properly constructed, are actually isomorphic with the equations that define the model. That is, if you have a properly defined path diagram, that implies the equations that govern the model, and the equations that govern the model, define the path diagram that goes with it, you can go back and forth. But maybe the most magical thing goes back to the 1920s. That using boxes and arrows and circles, and using our two index fingers, we can use a set of tracing rules. That gives us the covariance algebra. For every model implied relation in our set of observed variables, it is freakin magic, I had a really funny experience where Ken Bolan and I years ago, were working on a project together, and we were up on a big chalkboard, there wasn't even dry erase, it was just an old school chalkboard. And we were trying to figure out the identification of a model. And on the left hand side of the board, he was doing the covariance algebra. And on the right side of the board, I was frantically doing path tracing rules. And we were having this almost race. And it was just really funny as I've got my index fingers, as I'm tracing through this really complicated model. And I don't teach that to my own students. So I met with a student not long ago who had a complicated model, and it wasn't fitting and they have this big modification index that indicated misspecification. And I looked at the path diagram, and I said, well look at the tracing rules. The variable wants to go through here, but you're not allowing it. And they looked at me a little glassy eyed and said, I'm sorry, I don't know what tracing rules are. And I'm like, Oh, my God, who taught you SEM, and they looked at me glassy eyed and said, yeah, it was you.

Greg 17:07
Alright, well, this is your time to atone, then let's go,

Patrick 17:09
oh, no, I don't atone. No, no, I'm sorry. Okay. So why do we care, because if you have parental substance use disorder at the left hand side of your path model, and you have adolescent substance use way at the right hand to your model, and you have a field of mediators in between the two path tracing rules, like you understand all the ways that the parent predictor is influencing the child measure as it travels through this set of observed variables, right. So as every listener is fully aware, this entire arc of conversation inevitably leads us to guinea pigs. Now,

Greg 17:56
obviously, it's all about guinea pigs. So for those of you who didn't listen to our summer episodes, first of all, shame on you.

Patrick 18:03
Shame on you. Anyway, guinea pigs,

Greg 18:07
guinea pigs. Yeah. So path analysis, or as it was called the method of path coefficients, was developed by Sewell right. And we talked about school, right, a bit in the third of our summer episodes, the Summer of Love Episode. I know how you like that. So I'm not going to go back and rehash everything that we talked about there. But if you want to know why Patrick is talking about guinea pigs, and even more why he could have said guinea pig fetish, then you should go back and listen to that.

Patrick 18:35
Right? Not me, can we please clarify that?

Greg 18:40
Sewer right, developed this method of path coefficients in the 19. Teens actually, and codified it really, really nicely in a paper in 1921. And again, in 1934. His title when he was writing, which I love, was it he was the senior animal husbandman in animal genetics, at the Bureau of Animal industry in the United States Department of Agriculture, that was his title. And in the first paper that he wrote in 1921. Here's the setup that he gave, the present paper is an attempt to present a method measuring the direct influence along each separate path in a system. And that's a finding the degree to which variation of a given effect is determined by each particular cause. The method depends on the combination of knowledge about the degrees of correlation among the variables in a system with such knowledge as may be possessed of the causal relations. So he was coming up with a system for you to understand for you to keep track of for you to reckon why variables correlate the way they do, and in some cases don't correlate at all.

Patrick 19:41
And this was 100 years ago. This is the genesis of a set of heuristics of rules, that if you lay this out in this particular way, in terms of a path diagram, if you do the following, this is what your model is implying. It's really a rule markable work,

Greg 20:01
it is beautiful. And he came up with what have come to be called Rights rules or rights rules of path tracing. When I went through his work, I didn't find them so clearly articulated as they are now, but wasn't like they were buried in a footnote like Spearman did.

Patrick 20:14
The invention of factor analysis footnote to table

Greg 20:19
six. Yeah, if you go through and you read write stuff, you can pull out these particular rules, and they have been tightened up over the years.

Patrick 20:29
Again, going back to Dave, Kenny's book correlation and causality. He has a chapter on covariance algebra, and then he moves into path tracing rules. And Dave has a beautifully concise and elegant statement about these tracing rules. Dave focused on standardized effects. So correlations. Now all of

this stuff scales up to variances and covariances. But we're going to focus on the standardized like Dave did. So here's out of his book, The tracing rule is a simple non mathematical rule, the correlation between x sub i and X sub j equals the sum of the product of all the paths obtained from each of the possible tracings between i and j, the set of tracings include all the possible routes from exabyte X sub j, given that A, the same variable is not entered twice, and B, a variable is not entered through an arrowhead and left through an arrowhead. That's it. That is the distillation of writes tracing rules for standardized effects.

Greg 21:34
So you pick any two variables, and we ask why do those two things correlate? And what right said and what Kenny really nicely crystallized is that how two variables correlate are going to be a function of different ways in which those variables are connected in the context of your path model. And these rules layout, what will define those legal ways. So help us get into this, okay,

Patrick 21:54
what we're doing here is you're looking at your path model, whatever it might be, it might be a three mediator path analysis, it might be a bivariate, latent growth curve model. And what we're going to do is progress science through index fingers. So everybody hold up your left index finger, unless you're driving, everybody hold up your right index finger,

Greg 22:16
like you haven't driven with your knees. It's,

Patrick 22:19
yeah, never mind, what we're going to do was walk up to the whiteboard with our index fingers, and we're going to put one finger on one variable. And we're going to put another finger on a second variable and path tracing rules are going to allow us to rebuild the model implied correlation between those two variables. Now, we could do it with covariance algebra, we could do it with matrix algebra, but we're going to do it with our fingers. Now, use whatever analogy you would like. I love the Odyssey. Alright, so what we're going to do is we're going to hold one finger still, and the other finger is going to travel to meet it. It could be Odysseus, it could be my Lo and Otis, you know, like Disney.

MovieClip 23:06
Milo, I notice that traveling far away from home. And the only way back is together. Oh, not my low Notice again,

Patrick 23:15
it could be the Wizard of Oz, Dorothy is trying to get home is no please night. What do you got Hancock?

Greg 23:25
How about Patrick trying to get back to his dorm room during college on a Saturday at 3am.

Patrick 23:30
Okay, we're not going to get that complicated.

Greg 23:32
This is much simpler. This is a

Patrick 23:35
much simpler problem. Anyway, you've got one finger on one box and one finger on another box and you want to get from one to the other one finger is going to stay still the other finger is going to move, we can expand rights rules and Kenny's rules just to be a little bit more explicit in the things you can and can't do. Now, let's not forget our motivating goal, what we want to do is do a trace and a long that trace following these rules, we're going to pick up coefficients, maybe their regression coefficients, maybe it's a correlation. But as we go along a trace, we're going to gather those together and multiply them together. And the product of all of those coefficients is going to represent that first trace. Now, depending upon the model, that may be the only trace, but there might be another trace to get from one variable to the other. We're going to pick up the coefficients along the way and multiply them. And we're going to add that to the first trace. And we're going to do that for all the traces that are possible following the rules.

Greg 24:41
Yeah, two variables can correlate for more than one reason and that's what we're tracking here.

Patrick 24:45
Alright, so I'm going to do a couple of rules and then Greg can do a couple of rules. So we got to finger one place on the board. We got a finger on the other place on the board. We're going to hold one finger stable and the other one is going to be trying to find their way back home. Rule number one and we're going to Start by saying depending on the trace, because this depends on the characteristics of the model and what you're trying to achieve, you might choose to start going forward with your finger. Alright, so my left hand is on a box, my right hand is on a box. And I'm going to start moving forward down single headed arrows from one box to the next box moving forward, you can do that you know what, you can go forward as many times as you want. You know what the problem is, once you start going forwards, you can't go backwards, and you can't go across a double headed arrow. So once you're moving forward, you can only go forward, although you can go as long as you want. Yep. But you can't go backwards, and you can't go over a double headed arrow. Alright, So rule number one, once you start going forward, that's all you can do. Rule number two, you might, for various reasons, choose to start going backwards. So what backwards means is that we're leaving a box at the arrowhead, and we're just moving backwards from there. Alright, here's what's cool. You can go backwards all afternoon, as many backward arrows that you want. But here's the cool thing, you can actually do one of three things as you're moving backwards, you can go backwards, and if my lower notice, get back to your finger, you can just stop. Alright, so back, back, back, back, back back stop. You can go backwards, back, back, back, back back. And then you can start going forwards again. Oh, that's cool. So you can go backwards and then forward. But rule one applies to going forward once you start going forward. That's it, buddy. That's all you can do. Alright, so you can go backwards and stop, you can go backwards and forward. Or you can go backwards and span a double headed arrow, you

crossed that double headed arrow, you can stop, you can move forward again. But you can't go backward and you can't cross another double headed arrow.

Greg 26:53
What I like about these rules is that they completely map on to what makes sense when you're tracing forward, forward, forward forward. That's saying that those two variables correlate because of the influence that one has on the other direct influence. If it's one arrow, indirect, if it's multiple arrows, if you are going backwards, and then pivot to go forwards, what you're saying is that the two variables correlate because of the spurious influence of the thing that was up there where you pivoted. So these rules map on to the actual reasons that things should be correlating, I'm going to throw a third rule in here. And that is that once you cross a double headed arrow, or as you say, span a double headed arrow, you cannot span another double headed arrow. So if you cross the double headed arrow, as Patrick said, you can't go backwards. After that there is no backwards, you can only go forward or stop. One way to think about the idea that you can't go across to two headed arrows in the same trace is just because x and y correlate, which is what the two headed arrow is saying. And just because Y and Z correlate, which is what the two headed arrow saying doesn't mean that x has to correlate with z. If you think about it in terms of Venn diagrams, you could have X and Y overlapping and you could have y&z overlapping, but that doesn't mean that x and z are overlapping with each other. So in addition to Patrick's rules, a third rule is that you can only go across one two headed arrow. It's funny, we're taking what was a small set of rules that were making them a bigger set of rules. But I'm hoping that that makes it easier to follow as we're doing this on the podcast. So the next thing that I want to say is that as you're trying to get that one finger back home, you can only touch a variable once in that trace. So if there's anything about the trace, you say, but I followed the rules, I only did this, I only did that if you come back to the same variable as part of a trace, then that is not a legitimate trace. And that is sometimes referred to as the no loops rule. It keeps you from going round and around and around in systems where that would be allowed based on the structure of the path traces that you have.

Patrick 28:53
So you can go forward all you want, but nothing else, you can go backward or you want, you can stop span or go forward again. And you can't touch the same variable two times in a given trace, right? Those are the core rules. So what do we do now? Well, you do that for all the legal traces, you add them up and wait for it. That is the model implied correlation between those two variables. And that is the expression that you would get, if you did the covariance algebra that Bolan is doing to my left, they're the same thing.

Greg 29:30
They are the same thing. Except, and this is a really beautiful, except you can actually see the reasons in your hand, right? You go through all the covariance algebra and you get some stuff out and tada, you got the right answer. You go through the path tracing, and this trace might represent an indirect effect. And this trace might represent a spurious Association and this trace might represent something else. So you actually have this beautiful theoretical decomposition of the correlation into things that help you understand why those variables cause It's funny. So people who study mediation, really what they're doing is they're just isolating a specific type of trace between a set of variables. And if you go through it

at a covariance algebra level, or matrix algebra level that is buried inside there. So this is stuff people want. It's stuff that helps people understand the mechanisms that are operating.

Patrick 30:18
Okay, so in the spirit of trying to visualize these things, let's work through some examples. And we can start X predicting y, okay, there are only two boxes on the board, one is labeled X, and it has a single headed arrow pointing to why put one finger on X put one finger on why, what are all the possible ways we could do that? Well, we could start going backwards and it's y to x. Well, that's it, we're done. We start going forwards, right? That's an equivalent way of doing that trace, they're not separate traces. We could go x to y. Well, that's that relation. Think about when you teach this or if you've been a student in a class, and somebody may have told you at some point, well, in a one predictor regression, the standardized regression coefficient is the Bivariate Correlation. That's right. Well, this is highlighting that the only way we can get from x to y is down that single regression coefficient in a standardized metric, which is the Bivariate Correlation.

Greg 31:19
Exactly. Let's kick it up a notch. What do you say? Okay, Hancock,

Patrick 31:23
let's have X to Y and Y to Z.

Greg 31:27
All right. So the first thing that I'm going to say about this is that we're not just talking about one correlation here, we will be talking about what does the model imply about the correlation between X and Y, between y and z? And between x and z? I'll start with the correlation between x and y. What are all the ways I can trace one finger from x to get to another finger on why? And the answer is, there's only one way to do it, it's just slide and write down that single headed arrow. So the correlation between x and y that is implied by this model, just like in the first example we had is nothing more than that single path, that direct effect. Same thing holds for from y to z, if I put one finger on Y, one finger on Z, there's only one way to get between the two of those. And that would be whatever the path value is between y and z. Now the more interesting one is the path that goes from x to z, we asked ourselves, how many ways can my one finger on X get to my other finger on Z? And interestingly, the answer is there is only one way it is to go down from x to y, and then y to z. What that means is that the model implied correlation between x and z is going to be the product of the first path and the second path. And Patrick already talked about products as being the way that we aggregate these things within a single trace. And that totally makes sense, right? Because if it were the case, that x wound up not having any effect on y, then x would not have a reason to be related to Z either. So the idea of things being a product, even though it's rooted in algebra, and we could show why it's a product, it makes total sense, because any one of those pads breaks down, the entire chain winds up breaking down. So the correlation between x and z is just that thing that we might call a mediated effect, the path from x to y times the path from Y to Z, done and done,

Patrick 33:08
you just go forward from x to z, or you could equivalently start and go backwards from z to x. Remember, those are not separate traces. Those are equivalent ways of doing the trace. But there are a couple of really interesting things from this as well. What don't we have in that path model? Well, we don't have a direct effect from x to z that bypasses why, well, wait a minute, what does that imply? As Greg just said, The only way to get from x to z is through y. So it's a total mediation, that effect is completely mediated? Well, we have an over identified model there. Why? Because we imposed a restriction that x cannot directly predict Z bypassing y. That's right. There's no direct X to Z effect. Well, what are the implications of that? First, we have one degree of freedom. And so we have a test of model fit because we've imposed a restriction. And that restriction is we're not allowing that effect to be there. But what else does that mean? Well, as Greg implied, the correlation between x and z has to be completely reconstructed by the extra y and y to z effect. Well, we have a test of model fit. What if we get a test that indicates Ooh, wow, your hypothesized model does not fit your observed data? Well, well, that implies that the correlation that is imposed by the model does not closely correspond to that that we observed in the sample. Well, what does that mean? It's Whack a Mole. Right? It's we've talked about Whack a Mole before. That's right. If there is an x to z effect, and we remove that, that effect wants to be there to reproduce that correlation, but we're not allowing it

Greg 35:00
I love everything about this teeny tiny example, because it illustrates so many things already. You know, what we are saying with this model from X to Y to Z is that the three correlations that those variables have x, y, y, z and z are theoretically governed by only two moving parts, the path from x to y and the path from A to Z. So, we're saying three correlations are explained by two parameters. That's where one degree of freedom comes from, with this model, the idea that we have been done you call degrees of freedom measures of courage. Yeah. Isn't that how you refer to it, someone was courageous enough to say, any potential path from x to z is zero. That is guts saying that there's a path from x to y, or from y to z is saying, Well, I don't know. It could be something there. I don't know how big it is saying that there's no path from x to z is you going? It is zero, I have no need for that path. And so as Patrick said, that opens up the opportunity for misfit. But it also opens up the opportunity for our model to flex a little bit and say, Heck, yeah, I can do this, I can explain those three correlations with only two parameters. I got this. So assessing, fit and assessing misfit.

Patrick 36:06
And where do you pay the Reaper? Remember, folks, you always got to pay the Reaper. Let's say that X to Z wants to be there. Alright, so you bellied up to the bar. And you took two fingers of whiskey and threw it back and slammed the shot glass down and said, I believe the relation between x and z is wholly mediated by why? Well, what if it's not, but we force that onto the model? Well, the model is going to try to reproduce that x z correlation as closely as possible. Yeah, if there's a direct effect that you've pulled out, this is how Whack a Mole operates. That increases the potential that that x to y and y Dizzy relation is going to try to pick up the omitted X to Z effect, because it's frantically trying to get the model implied correlation as close as it can to the sample correlation. And you're going to get a biased mediated effect, because it's trying to pick up the effect of the omitted direct effect.

Greg 37:14
Yep, the two path values that you have to adjust are not just trying to serve the relations between the variables at the two ends of those paths, but between the two variables at both ends of that particular model.

Patrick 37:25
Alright, buddy, I'm going to shove you harder. Okay, when you got a picture on the whiteboard, we're going to add a second predictor. So we have x one and x two. And the two boxes are connected with a double headed arrow. Why? Because our exogenous variables are allowed to correlate with one another. But we're not saying one determines the other. Okay? They are just correlated. So we have two correlated predictors, x one and x two, both have a single headed arrow to y, and only y then predict Z. Okay? All right. So it's very similar to our prior one. But now y is jointly determined by x one and x two, walk us through the path tracing rules, because that just got more interesting.

Greg 38:09
Yeah, so it's kind of shaped like a hand mirror, or maybe you've got this thing up at the top that involves the two x's, it comes down to a point at y and then it goes from y down to the bottom of z, we got four variables and four variables are going to correlate a total of six ways x one with x 2x, one with y x one with z x two with y x two with Z, Y with Z six correlations that this model is trying to explain. And this model has four moving parts, the correlation between x one and x two, the paths that each of those have into y, and the path from Y to Z. That's all you got to build what you need here. So if I do my tracing, I can start with x one and x two, how many legal ways can I get from, let's say x one to x two? And the answer is one way, I just trace my little way across that two headed arrow that span and the two are joined together, there is no other legal way. If I take x one, and I go down to Y, I have gone forward, I cannot pivot and go back up to x two. So that's it. I'm done. x one and x two correlate because the model said they correlate Done and done as far as they're concerned, with x one and y, there are actually two ways that I can get between them. The direct effect that x one has on Y, and what sometimes called the unanalyzed Association, where you go from x one across the two headed arrow to x two, and then down from x to y. That's referred to as an unanalyzed. Association because it contains a two headed arrow. It's not unanalyzed in the sense of us not doing the math, it's unanalyzed in the sense that we're not literally modeling the mechanism that's responsible for that two headed arrow. So two sources of correlation, two legal traces from x one to y. And when we add those two traces together, that is the model implied correlation between x one and y. And you know what, everything I just said flips for x two also, there are two ways to get from X to Y. Now thinking about how things connect to Z, how would you get from x one to Z. Well, there's that indirect effect that goes from x one to Y to Z. That's a legal trace. And there's an unanalyzed association that goes from x one across the two headed arrow to x two, and then x two causally to y, and y, causally to z, I have two traces that get me from x one to Z, add those together, that is the model implied correlation between x one and z. And there's another one just like it from x two that mirrors that perfectly, so I won't go over it. The last piece of this puzzle is how Y and Z correlated according to this model. And I'm torn. Is that easy? Is that hard? Let me think about that for a second. Certainly, if I got one finger on why one finger on z, I can get those kids together pretty easily by just taking one finger from y and moving it over to Z. That's a direct effect. And that is an absolutely legitimate source of correlation between y and z. If I'm sitting on Y, though, it is completely legal for me to go back up to x one, that's a legal path. And it is legal for

me to go from x one across the two headed arrow over to x two, well, where do I go now? Well, I would come back down to why and then all of a sudden warning sirens would go off, because I would have touched the same variable twice in my trace. So if I was thinking, I was gonna take a little stroll around the park up by the x one and x two, come back to y and then go down to Z, and not allowed to do that. So that means that the model implied correlation between y and z is nothing more than the direct effect that Y has on Z. And that is a full reckoning of all six correlations among the variables as a function of the four parameters in this model.

Patrick 41:31
And think about the restrictions we've imposed. This model has two degrees of freedom. In the path tracing rules, it's very clear how we're getting those x one cannot directly impact z, x two cannot directly impact Z, they can only impact z through their influence on Y. So looking at the path diagram and thinking about the path tracing rules, you immediately see, that model has two degrees of freedom, because we removed those two effects Beautiful. Now when we talk about doing a priori tests of our hypothesized model, our chi square test statistic is evaluating what is the impact of removing those two parameters from the model in our ability to reproduce those six correlations, if air quotes it is non significant, the chi square, then we're tempted to conclude that we do a pretty good job reproducing those correlations without those two direct effects. If it is air, quote, significant, we're tempted to say, oof, you know what the model is implying that we're really not able to reproduce those x one to Z, or x two to Z correlations by omitting those direct effects. The modification index is saying, Hey, dude, right? The MMI is like this kind of creepy guy in an alleyway,

Greg 42:59
buddy, give me a kid, reduce your chi square by a

Patrick 43:03
few just put that single headed arrow, well, that's what it's literally doing is saying, if you were to put the direct effect from x one to z, then we would be better able to reproduce that correlation. But here's the poke in the eye is if you put both of those direct effects in the model is what we call saturated, we have imposed no restrictions. And you know what, we are going to perfectly replicate that correlation, because we've saturated every possible way we can get from x one to Z, and from x two to z. So how proud of yourself should you be by being able to reproduce that perfectly, none at all.

Greg 43:45
That's right, it's not a big achievement to be able to fit into sweatpants. Because they stretch all over the place, it's a big achievement to be able to fit into a suit of armor, right? The more restricted it is, the more of a test it places on you.

Patrick 43:58
The other thing to highlight is the importance of the correlations among your exogenous predictors. We forget about those a lot of times it's like, Oh, we got our predictors, their demographics, we're controlling and removing for them. But go back a couple of minutes to what Greg described. If we're trying to see how x one relates to Z, one way is the x one to y to z, but the other one is the X one's correlation to x two, that goes to y that goes to Z, right, as these two guys hang out together, they are

related to one another. And so what x two is doing impart relates to how x one is related to z itself. Now let's take a model that we wouldn't normally do, but it makes the point let's say that x one doesn't predict y or z, that x one is somehow just hanging out there and is correlated with x two, and x two predicts y to Z, the model still implies a correlation between x one and Z. How? Well x one is related to x two and x two predicts y that predict Z. Right? All right. This is like getting arrested for being the driver of the getaway car. It's like I didn't do it. I didn't hold up the bank. It's like, Dude, it's the friends you're hanging out with. And so one thing path tracing rules do is makes us really appreciate how all the variables relate to one another in our model, I like it. Okay, Wiseguy, things seemed easy enough when we have observed variables. Let's back the truck up a little bit and say, What about a confirmatory factor model. So we're going to go up to the whiteboard, and I'm just going to make things up, we're going to have five observed variables that load on factor one, and we're going to have another five observed variables that load on factor two, and we're going to log factor one to covary with factor two. So we have five indicators each on two latent variables, and the latent variables correlate with one another. Talk to me about how do we reproduce element by element or 10? By 10 correlation matrix,

Greg 46:15
you are giving me a 10 variable model to do in a podcast? Is that what you're saying?

Patrick 46:21
And what you're saying is, is you're gonna whine about it? Oh, no,

Greg 46:24
hold my guinea pig. Here we go. I got five indicators of factor one, five indicators of factor to what this model is saying. Overall, I'm going to ease into the path tracing in just a second. What this model is saying overall, is that the first five variables correlate because they're influenced by something in common a factor. The second five variables correlate because they have something in common, that second factor and the first set of variables Y one to y five, the second set of variables, y six to y 10, they correlate with each other, because the factors that govern them correlate with each other. That's the conceptual orientation. Now as a path tracer, I am not going to drag you through the what would it be like 45 correlations that we have operating here, 10 times nine divided by two, there we go. So I only need to take you through a couple because they stand for all of the rest. If I want to know why y one and y two correlate with each other, but one finger on y one, one figure and y two, they are both indicators, that first factor, I just move my finger from why one backwards up to the factor and forward down along the second path. And what do I have, I have a trace that includes the first loading times the second loading Done and done, there is no other legal way to get between variables one and two. In fact, the correlations among all variables Y one through five have that exact same pattern, and the correlations among all variables, y six through y 10. All follow that exact same pattern up to the factor and down, there's no more tracing that you can do every model implied correlation of variables within their respective sets are the products of two loadings. Now if I want to get from a variable that's in that first block y one through y five to a variable that's in the second block y six through y 10. Easy peasy. Let's say I got one figure on y one, which is the first indicator of that first factor, and I got one finger on y six, the first indicator of the second factor, how do I get between those two, there's only one way to do it, I take my y one, slide backwards up that loading path, pick up that loading, go across the two headed arrow, pick up the correlation between the two factors, and then slide down that loading path to y six,

that's it, there is no other way to get between the two. That means that the correlation between those two is a loading that we could call lambda if you wanted across the correlation between the two factors, which we could call psi, and then we could go down the other loading path, which is another example of a lambda. So in that case, the correlation between any variable in the first set and any variable in the second set is going to be a loading times the factor correlation times the loading in Greek lambda, psi, lambda,

Patrick 49:05
a couple of things that this then highlights first, Greg is using the rules, we can go backwards, we can go over a single span, and we can go forwards again. Now notice that Greg said, You go up in lambda, you go over psi, and you go down another lambda. So what is the correlation between two items on different factors? It's lambda psi lambda, one of the most famous matrix expressions in factor analysis is wait for it, lambda psi lambda prime, and then we have plus theta epsilon for reasons that we won't talk about here, but what is that lambda side lambda prime, when you do that matrix valued function of lambda psi lambda prime, lambda is going backward along that factor loading psi is going across the covariance in lambda prime is moving forward along that loading. To the other variable. So notice the lambda side lambda pattern in that it's exactly doing the path tracing rules out the level of the matrices. Yeah. Alright, smart guy, I'm going to change a factor correlation to our regression coefficient. Same deal, five indicators on the first factor five indicators on the second factor. But now factor one has a directed one headed arrow predicting factor two, it's a regression coefficient. What are the implications of that

Greg 50:32
a lot from a theoretical standpoint, where, you know, in the previous model, that was a confirmatory factor model where the two factors correlated with each other, what we were saying was that neither factor causes the other they likely are related to each other because of some forces that exist outside the model. This model where you say factor one influences factor two, that took some guts, right, that says that you have a theoretical reason for one of those things to be influencing the other. From a path tracing standpoint, it is not going to make any difference. If I want to know how the first five variables again, we'll call them y one through five, how those correlate with each other, it's still just going to involve two loadings tracing from one variable up down to the other variable, same thing for the variables that are in the other set. So then what happens when we want to get from variables one through five to six through 10, it's the same, imagine tracing your way from y one through five to y six through 10. You just go backwards up your loading, you go across the structural or causal path from the first factor, the second factor, and then downloading path to the other. So all we have really done is replaced what was a factor correlation in our trace previously that Patrick had called psi with a structural path that we could, for example, call gamma, everything tracing wise winds up amounting to the same thing.

Patrick 51:50
And once we're at this model, this is a structural equation model, we have multiple indicators on one factor, we have multiple indicators on another factor, we're regressing the second factor on the first now we're off to the races have a third factor and what we started with 15 minutes ago of x two Y to Z. Well, now you have factor one to factor two, factor three, everything scales up exactly the same picture, a

crazy complicated model. Because here's where Bolan ultimately won that contest, we had the chalkboard, because one of the challenges is in complicated models, you have to figure out all the legal traces. Yeah. And in a complex model, there might be six or seven or eight different traces that go into that component. But you know what, tough crap. Yeah, that's what the model is implying. What I mean is, is that's the point of path tracing is sometimes what you think is part of the model. That's way out in western Kansas that nobody's ever been to before. That actually relates to how two variables are correlated downstream, because Milo and Otis get thrown in a box and dropped in a raging river, and they go all the way out to Kansas, and then come all the way back. And that long trace is part of what your model is imposing about how the variables relate, not only from a statistical standpoint, but from a conceptual, substantive theoretical standpoint,

Greg 53:24
apologies to people in Western Kansas,

Patrick 53:26
you know, what is I have gotten two or three speeding tickets in western Kansas. One of which a state trooper asked me if I was able to read, I was taking exception to the ticket.

Greg 53:40
It's always a good strategy to take exception to the ticket.

Patrick 53:45
Especially if you like getting tased.

Greg 53:48
Officer, I have a couple of questions about this. But yeah, Patrick is exactly it doesn't matter how damn big the model gets, it just doesn't matter. Patrick, and I routinely work with people who might have 60 variables in a model. They're all over the whiteboard, the same rules apply. Why does each pair of variables correlate, you just go from one to the other end? That is what's under the hood, whether you express it in matrices, whether you express it in covariance algebra, whether you do it as path tracing, those are the mechanisms under the hood, those are the model implied correlations. Those are the reasons variables relate the way that they do or don't relate. Those are going to be the sources of your fit. Those are going to be the sources of your misfit. It's just a beautiful system.

Patrick 54:32
And it makes you so appreciate the fabric of how all of the variables relate to one another. So I met with a student who had a very complicated model and they were getting these residuals among their observed in model implied covariances at the end part of the model, the distal dependent part, and I made a comment about well, you may have misspecified some of your predictors and they kept insisting nobody Those aren't involved in this, those are over in this part of the model. And when you do the path tracing, and you see Oh, holy cow, I can go back back back over this exogenous covariance and then forward, forward forward. Oh my gosh, they're integrally related to how your dependent variables are related with one another. Yep. So why then, are we talking about this? Well, there are some advantages. Just to reiterate, as we head toward the exits here, the biggest one is just

developing this understanding of how every pair of variables in your model relates to one another. And in the covariance algebra, and in the matrix expressions, that is not always as salient as it is, when you take your two fingers and say, what are all the ways I can get from one finger to the other finger?

Greg 55:51
Yep, it's probably quite clear that I love path tracing, I love it. For some of the reasons that we've talked about that it helps us to really visualize all the sources of why variables are related, it also helps to tie together lots of ideas that exist elsewhere. So certainly that these ideas go on to the unstandardized model, where we talk about decomposing y variables covary. Specifically, rather than just correlate, we can decompose variances, all of those things. But all of those damn formulas that you had to remember, throughout all of your statistics classes, for example, you might have had to memorize the formula for a standardized regression coefficient in a two predictor model, where the formula for a partial correlation or a semi partial correlation, all of those things, and they existed as the separate things that you had to commit to memory. And what I will tell you is that all of those things are just path tracing every last one of them. If you want to understand how a partial correlation works, all you have to do is draw the picture for it. And the formula for partial correlation just falls out of the path tracing. Same for standardized regression coefficients as a function of other aspects of the model, the variance of a composite of a variety of variables that is just path tracing. So I love it when the world distills down to just a few very simple rules that govern everything else. And path tracing is just one of those elegant things that helps you understand so many other things.

Patrick 57:15
So when canon was doing the covariance algebra, and I was frantically trying to find all the possible traces through the path diagram, it didn't seem like a shortcut that turned into a lot more work. But oh, my gosh, we have this model implied correlation between these two variables. Here are all the components that are contributing to that. And here is what it implies for your theory. And your research hypotheses is I know, this thing is way over on the left, and I know this other thing is way over on the right, but this is how your model is saying these two relate to one another. If you

Greg 57:55
don't mind. Now, I'm going to go sit where it's a lot warmer, please. Sure,

Patrick 58:02
sure. Let's wrap up with it being about you.

Greg 58:05
Yeah. Before I remove my mitten, can you guess which finger I'm holding?

Patrick 58:11
As always, everyone, thanks for your time. We hope you've had half as much fun as Greg and I have had.

Greg 58:17
Alright everybody, take care. I'll talk to you later.

Patrick 58:20
Thank you so much for listening. You can subscribe to Quantitude on Apple podcasts, Spotify, or we'll come over to your house and do a live episode. Seriously. It has been a really long two years. And please leave us a review. You can follow us on Twitter we are at quantitated pod and check out our webpage at quantitated pod.org. For past episodes, playlists, show notes, transcripts and other cool stuff. Finally, you can get quantitated merch that is quite frankly ideal for all of your April Fool's Day gift giving needs at red bubble comm where All proceeds go to Donors Choose to support low income schools, You have been listening to Quantitude: the podcast equivalent to western Kansas. Quantitude has been brought to you by the Kansas State Board of tourism who would like to extend quantitate a warm invitation to visit the great State of Kansas and personally embraced the state motto and the Astra per Aspera which in the current context loosely translates to Quantitude can and then bend the barbed wire to in the wall placing the red hot horseshoe firmly on there. And finally they can with in both of their Kansas to the stars. By unstuck in time, a new plugin for Outlook that manually modifies the timestamp on any email sent during a zoom research talk. So you can at least maintain the appearance that you're not working while someone is generously sharing their time with you to discuss topics that are important to them, and by Patrick's turn to do the sponsors which are distinctly Jiffy free. This is most definitely not NPR.