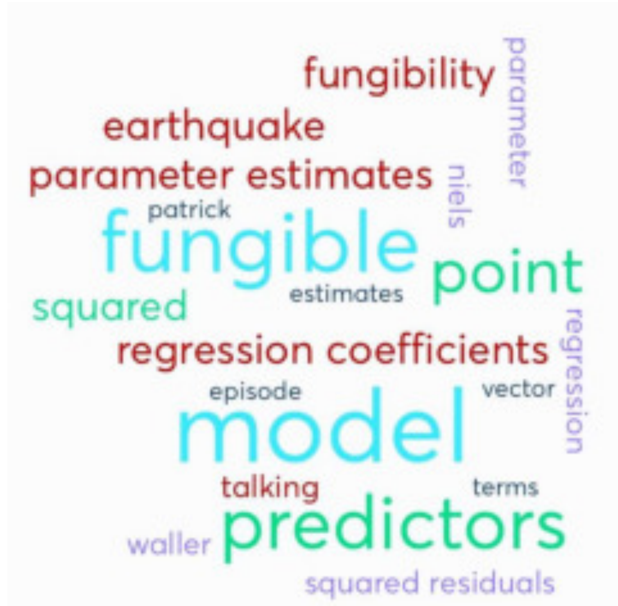
The Podcast *Quantitude*

with Greg Hancock & Patrick Curran

Season 3, Episode 25:

Fungible Parameter Estimates & Earthquake Waller



Published Tuesday April 5, 2022 • 55:50

SUMMARY KEYWORDS
model, fungible, predictors, point, squared, fungibility, parameter estimates, regression coefficients, earthquake, talking, squared residuals, regression, Niels Waller, parameter, terms, episode, estimates, vector

Patrick 00:01
Welcome, my name is Patrick Curran and along with my non fungible friend Greg Hancock we make up quantitative food we are a podcast dedicated to all things quantitative ranging from the relevant to the completely irrelevant. In this week's episode we discuss the sometimes terrifying issue of fungible weights and multiple regression and structural equation modeling in which selecting a trivially worse criterion of fit can often lead to radical changes in the corresponding parameter estimates. Along the way. We also discuss competitive family Wordle disambiguation in flammability, perpetuity, being nonplussed, running laps after practice schmungible, audio eye rolls, haystacks at sunset, hyper eggs, the spider verse, mountain moon rises, tin cans and strings, and earthquake Waller. We hope you enjoy the episode. You're a guy who loves words, right?

Greg 00:59
I love them.

Patrick 01:00
And are you still doing the Wordle thing?

Greg 01:03
Yes, our whole family does it. We compete. This morning I lost

Patrick 01:06
and how does that differentiate that from any other morning? So what are some of your favorite words?

Greg 01:12
There are words that I have used previously, only to find out later that I'm an idiot and was misusing them in public. So I don't know if those are my favorite words. But I certainly have some memorable words.

Patrick 01:24
I couldn't care less what some of your favorite words are. I want to hear a story about where you've misused a word.

Greg 01:30
I'll give you a little one and a bigger one. The little one is, you know the word in firm like when someone gets older, they're in firm. I always said in firmed I would say yes, that person is infirmed. And Goldie had never heard me say it before. And then one time I mentioned she looked at me like did you say in firm? said Yeah, yeah. Why? Because that's not a word. That's not a thing? I'm like, Yes, it is. Of course, it's so you know, we look it up. And I was mortified because I played through all the people I had said that in front of and then I felt like a complete idiot. So that's a small one. A bigger one. I was looking up something statistical. And it was a pretty basic idea was a number of years ago. And I would just wanted to get a good description of it. So I got on Wikipedia, I got on early Wikipedia, and I typed it I don't remember what the statistical concept was. And the page popped up. And it had the title of what I was looking for. But there wasn't anything on the page, except one word disambiguation. And I looked at that, and I thought, hell yes, that is exactly what it is. It is disambiguation. That is the perfect description of what we do. Oh my god, Wikipedia is amazing. And I was describing some phenomenon to a class to colleagues talking about disambiguation. It's like a philosophy. It's the way we perceive the world and what our goal is within. And I came later to realize that Wikipedia just uses that word to mean that yeah, there's some stuff on this page. He's. But by God, it's still a good word out. It's

Patrick 03:02
a great word that reminds me of I don't even remember it was decades ago, but there was some sitcom and I watched an episode of it on an airplane going somewhere. Two of the workers bought their boss. So word of the day calendar that they custom made. That was all made up words. And then every day, she would use one in a sentence. Wow. Aren't you the perpetuity one? Two funny things with me and words is one, I will learn a word that I think I know the meaning and that I like it. And oh my gosh, I am not shy in using it. When Andrew was pregnant, we have twins, we were out shopping for cribs. And for any of you out there who are expecting or have a newborn, as you know that you obviously don't want your child if you buy the cheaper version of the product. And we're looking at cribs and you can get this awful death mattress for some expense. But if you love your child, there's this improved mattress, and I said well, I assume it's inflammable and looked at me and there was a long pause and she said, No, it's not it's not inflammable and I said really, you would think by now that the technology and you know, Andrew is like that's not what inflammable means. There's flammable and there's inflammable and flammable turns out to easily bursts into flames. So this nicer mattress so it easily burst into flames, I assume. No. No, it doesn't. Another one that I have misused repeatedly is nonplussed. Okay, so I've used this for years. Yeah, in a stressful situation. I am totally nonplussed. wrong. Oh my God tell me easily confused, surprised or stunned.

Greg 05:05
I thought it just meant unfazed, bewildered, confused,

Patrick 05:10
easily surprised.

Greg 05:14
Oh, do you remember this teachable moment?

Patrick 05:19
Myself as nonplussed? You can hear us in the horn. I don't think I've missed what your thinking is.

Greg 05:25
Wow.

Patrick 05:26
But here's where it gets worse. How many times have you been in a situation where someone else has used a word that you don't know what it means that you're just gonna go along for the ride?

Greg 05:39
Oh, there's that inner dialogue where you're like, oh my god, do I just fess up? Or do I play it cool. And once you play a cool, you pray you don't get called on this?

Patrick 05:47
Well, so two things First, is you never Fess up. Dude, I'm surprised you even gave that a person the number. Never Fess up. All right, in second, I gotta tell you from personal experiences by not fessing up. It just gets worse by the minute as the conversation proceeds. Oh, yeah. So that's what brings us to our topic for today. Number years ago, I don't know how long I was at a conference and one of my favorite people came up to me in the lobby, there was some Danishes and green hard bananas and bad coffee. And it was Nils Waller from the University of Minnesota and Niels His name has come up before we were talking about his work with Allie on Hayward cases in FFA. We had a whole episode on that. That's right. And a great guy and a great guy. Niels is funny. He is smart. He is creative. I really, really liked Neil's we hadn't seen each other in a year or two, and we do the mandatory Hey, man, how you doing? What are you up to? And I said, are you presenting it later? And he said, Yeah, I'm on some recent work I'm doing I'm Wow, what are you up to? I'm looking at fungible weights. And I was like, Oh, really? Alright. Fungible? Mm hmm. Looking back over the tapes, if we were to review them on Monday after the game, and the coach slows it down in slow motion, and says, Alright, you had an opportunity here. And here's what you did.

Patrick 07:16
Here's the choice you made in this situation, rest of the team. What else could he have done here? And how would it have changed how the play unfolded? Let's just say that I'm going to be doing laps after practice on this one, I have the opportunity to say what is fungible, but I'm a male. I'm a faculty member, I'm occurring. I'm never going to say, I don't know what that means. And I'm like, Oh, really?

What are you doing with them? Having no idea what the word fungible means. And he goes into this really wonderful description about models and how they're best linear, unbiased estimators, but that we need to factor in fungibility and to think about what the implications are. I'm like, Well, you know, what, I always say, fungible. munchable. Yet, obviously, you know, who am I right? Or am I right? Miss point, I'm like, I can't even climb out of the hole that I'm right. And it's kind of like, where you're on the third date, and you don't remember how to pronounce her name, but you're way too far in to clarify that. You know, and so you try something like how does your mother announce your name? And she looks and says, Jane,

MovieClip 08:37
You don't know my name do you? Is it Mulva?

Patrick 08:47
So I'm like way in the third day, I've forgotten her name. And so we go back in and he gave a brilliant, wonderful talk, where I did figure out what the word fungible actually means. And that's what I want to talk about today is fungibility. Okay,

Greg 09:03
how do we get into this topic? Do we start with the Merriam Webster dictionary definition?

Patrick 09:08
That would imply I looked it up at some point over the last 15 years and would know what fungible means. Okay, this is really awkward. I probably should have looked that up before we started recording.

MovieClip 09:19
This is so David Attenborough that what the data is fungibility, the definition of which seems rather important given that today's constitute episode is built around it. fungibility is the characteristic of being able to be replaced by another identical mutually interchangeable item.

Patrick 09:37
Well, the irony is, it's more widely understood now because of NF T's. Everybody has heard about non fungible tokens. Yeah. Nf T. Wow. Okay. That was an eye roll, folks. I don't know if you felt the eye roll in the audio, but visually, I'm kind of impressed. Your eyes came back down out of the back of your skull.

Greg 09:58
Yeah, I don't even know what to do with that hole. Topic. It's just weird to me.

Patrick 10:02
I don't have Merriam Webster's, but it means interchangeable. Mutually interchangeable one is as good as the other. If two things are fungible, it doesn't matter which one you have they're exchangeable. They're interchangeable. That's why the source of the definition of non fungible token is if it's non fungible, there's only one and it's not equivalent, if it's exchange non fungible is unique. How I had someone described the non fungible token to me is it's as if you had a Monet print, where the prints are

fungible, you go into the museum store, and you buy haystacks at sunset, and the whole stack of the prints are fungible because any one of the prints is exchangeable with the other one, but the original is non fungible. And I got to tell you, we had an episode. I don't know how long ago, but it was on equivalent models. Remember, these are not alternative models, they're equivalent models. They're models that you rearrange observed variables, you change single headed arrows to double headed arrows, you change one way effects to two way effects, and they give you the same log likelihood, we figured if that didn't scare the crap out of you. As a data analyst, we're gonna up the ante Oh, yeah, we're not gonna move boxes, we're not gonna move arrows. We're gonna say okay, pretty boy. You want your precious little model go up in Sharpie it in on the whiteboard for me. Yeah, we're gonna give it to you don't change a thing. Good for you, you got it right for you What could possibly go wrong. And your precious little model had an R squared of point 256. We're gonna change that 2.253. And we're going to show you there are an infinite number of regression coefficients that will give you that r squared. And let's see how you sleep at night now. Thank you, Niels. Like me and Hershberger didn't keep us awake at night. Niels now is like holy crap, your precious little regression coefficients. If we move the R squared point, oh one, we now have an infinite number of regression coefficients, that would give you exactly that r squared. How are you like me now?

Greg 12:37
Okay, so that's a lot to take in. How about if we let's wait for the ground to stop rolling here. Because this feels like there's a worldview about to change.

Patrick 12:50
And that's a really good analogy. People talk about where you have the salient events, and it forever changes how you view a prior belief. One is earthquakes and people will talk about seeing the ground wave like waves on a shore at a beach, and that for the rest of your life, the ground is no longer a stable solid thing. Because you saw surge and wave during an earthquake. I think this is very similar. Those regression coefficients that you get in the standard errors and the confidence intervals are maybe not what you think they are who so Niels is that earthquake. Thanks, buddy. All right, so what's orient before the earthquake, they don't name earthquakes like Hurricane, they should. If they named earthquake, this would be the Nils earthquake. Alright, so we're gonna name we're gonna name is the Niels earthquake, okay, as Niels points out in his own writing, oh, my gosh, he is such a wonderful writer. And it is so engaging. And even with his most technical stuff, he kind of holds you by the hand walking through it, and you feel like you actually understand it. So well, well done. But we're still going to name the earthquake after you let's pre earthquake think about a lecture that all of us have either given or have been part of. So we're going to start very simple with a one predictor regression. And we know that ordinary least squares is we want to select the value of a regression coefficient that minimizes the sum of the squared residuals. And the residuals is the difference between what we observed on y and what is predicted in Y hat. So Y minus Y hat is the residual IE, we squared the residuals z squared, we add them up as the sum of the squared residuals. And we want to pick in a one predictor case, that coefficient that gives us the smallest sum of the squared residuals possible, where you may have either given this as a lecture or been part of it, is you go up to the board and you draw an X Y axis and on the x axis, you put B sub one as your regression coefficient on the Y axis, you put sigma e squared, which is the sum of the squared residuals, we can do what's called brute forcing it. And we say, well, what if the regression coefficient were to? And I walk up to the board and I put a

little.on, the board? And I say, Well, here's the sum of the squared residuals, that a coefficient of two would be, well, what about negative one? And I put another dot, and I said, Well, what about point five and I put another dot, and you start to fill in the dots. And if you do it more or less, right, it starts to outline a parabola, actually a really beautiful parabola, and you say, Holy cow, instead of brute forcing it, I wonder if I can write an equation for this. And then you put in this really pretty parabola, it goes down, down, down, there's some bottom and then goes back up, up, up, up, up, and you say, Wow, I could brute force this, but I remember from high school calc that I can write a derivative for this, ooh, the derivative is the slope of the tangent line at a given point on the curve. And I can say, well, for trying to find the bottom of that curve, that is minimize the sum of the squared residuals, we write out the equations for that derivative, and we fix it to zero, because that's where the slope is zero, it's horizontal. In regression land, we call that a normal equation, we solve it. And it gives us that point for B one in which no value gives a smaller sum of squared residuals than that value will be one. And all of us wax poetic about it. Nowhere in this universe or any spider multiverse. Is there a value of be one that gives you a smaller sums of squares, that's one predictor. Two predictors, you have x one corner of the room next to is the other corner, why you go up, remember my plywood and tennis balls from a prior episode? It turns out, I know nothing about statistics, because I'm gonna generalize plywood in tennis balls into an egg. Okay, that parabola for one predictor now becomes a three dimensional egg, you're trying to feel along the sides of the egg to find the bottom of the egg for what is the joint value of b one and B two that together, give the smallest sum of the squared residuals, you have three predictors? Well, now it's a hyper egg,

MovieClip 17:27
That is not actually a thing. Patrick has made that up. But it is a lovely notion.

Patrick 17:32
And what all of us have learned is under assumptions, there are no combinations of values for our regression coefficient that will jointly provide a smaller sum of the squared residuals than those OLS estimates.

Greg 17:50
I love the way that you're describing it, you use the term brute force at the beginning, almost as though you are wondering this egg trying to find that lowest point. The beautiful thing about this in the regression system is that there's actually a closed form solution for that, that I think you beat people over the head with in the matrix episode, right? That's

Patrick 18:08
right, we have the function for the minimizing the sum of squares, they're partial derivatives. You take a derivative of that function with respect to each predictor, you get the derivative, you set those to zero and you solve and that is the bottom of the hyper egg.

Greg 18:26
X prime x inverse x prime y.

Patrick 18:31

There you go. That saves you the trouble of doing the derivatives. One thing about moving to the fungibility is there is a breathtakingly beautiful geometry to it. So I did tennis balls, plywood and a hyper egg in the spider verse. Can you maybe formalize that a little bit with the geometry before we start talking about earthquake Neil's

Greg 18:56
I'm going to harken back to something that I talked about in the last episode that we had the matrix part two, as we described in that episode, every variable that we have can be thought of as a vector in space, a space that's defined by the dimensions of all in people or in cases that we have. So there's a vector that goes out for x one, there's a vector that goes out for x two, and we're gonna stick with those for right now. And the goal is to try to understand the relation that they have with y. And as Patrick just talked you through, what he was looking for was combinations, this cocktail of x one and x two, a linear cocktail, where we have some weight that goes with x one some weight that goes with x two, and we're trying to build something out of those that gets as close to y as possible. And we call that thing y hat in the geometric world that I described previously. We have this vector x one and this vector x two, and they might be very close to each other. If they correlate highly, they might be closer to right angles if they're more orthogonal or uncorrelated with each other. And then we have this Y vector. The way that we described it in that episode was think about x one and x two To those two vectors is defining a plane, or a tabletop. And that plane or tabletop represents all possible linear combinations of x one and x two. And the y vector as being a vector that sticks out of that particular table top y has an infinite number of shadows on that plane that's made by x one and x two, if you were holding a flashlight above Y, and moving your hand all over the place, you would get these shadows in that plane made by x one and x two, there's only one shadow that is as close to that y as possible. And that's the shadow directly beneath it. So when Patrick wanders the space sort of metaphorically saying how much x one do I need, how much x two do I need, what he is doing is arriving at that place. That is the perfect straight down perpendicular projected shadow of y into the plane defined by x one and x two. And that is a linear combination of x one and x two with weights that are the exactly the same as Patrick described. One last reminder about that is that once we have identified where that perfect shadow of y is that y hat, we can measure the angle between the actual y and the y hat and the angle between those if we take the cosine of that becomes the multiple correlation, big R, that big R is as good as it's going to get right of all the linear combinations of x one and x two you could build that is the one that has the highest correlation with the actual y. But of course, there are other shadows that might have slightly lower values of are slightly bigger angles with that actual y. And we're going to talk about those.

Patrick 21:44
It truly is a remarkable characteristic of ordinary least squares regression that there exist no set of regression coefficients that lead to a smaller sum of the squared residuals. If you go a billionth of a point to the right or the left on that hyper egg, the R square will get smaller. There is no combination of coefficients that give a smaller sum of the squared residuals. But holy crap, there are a boatload of them that give almost the smallest sum of the squared residuals, but may lead us to write a different discussion section oof, that's where the word fungible comes in. And this is where Niels trots out on the field with his 2008 psychometric paper, a fungible weight is one that can be exchanged with another, but has no difference in the impact on some outcome that we describe. That's the fungibility. So what does that mean? Well, let's say that we take our multiple R squared that we have, and I'll just make up

a number and it's point two, five. Now it is true that there is no combination of our predictors that will give more than an r squared 2.25. But what if we make it trivially lower? Let's say what we obtained was point two, five, and we say all right, well, what about point 245? That's close what's 5/1000 of a point between State University employees the difference between point two five and point 245 trivial difference, Neil's prove that for a model with three or more predictors, there are now an infinite number of regression weights that will result in an R squared of point 245. The pictures on your wall are starting to jiggle. An infinite number of regression coefficients might I say fungible regression coefficients, that all lead to point 245. And here's the drunken punch in the face that's 5/1000 of a point in multiple r squared. But some of these may combine where you would write a different discussion section for your model results.

Greg 24:12
Even though the R squared is only five thousandths off the ground is so unstable when I talked about there being this one perfect shadow underneath why and that its angle conveys information about the multiple correlation. I also said there were a number of other shadows. If we moved our flashlight if we didn't hold our flashlight directly above Y but moved it a little bit to the side, we would get different shadows that have different angles with the Y vector. Every one of those shadows represents a different linear combination of x one and x two. If Patrick says that he wants something with a squared multiple correlation of point 245 Rather than point two five, it would come up with a correlation that is lower what that means in turn terms of projections Is that why would make an angle with such a thing that is actually larger? Right? That is slightly sub optimum in this space that we're talking about right now, where we have two variables x one and x two defining the tabletop defining the plane. If Patrick says, Yeah, I don't want that perfect angle, I want something that's just a little bit bigger. The answer is, well, there's something that's just a little bit bigger off to the left, and there's something a little bit bigger off to the right, I can move my flashlight left or right to create that perfect shadow that makes the angle with the correlation that Patrick describes, and there are two of them. And that's it in the world with only two predictors. That to me, is already ground shaking that you tell me Wait, wait, wait a minute, wait a minute, the correlation is off by a tiny bit. And now there are two solutions, not only are there two solutions, they can be very different from each other, right? Because one of them veers off to the left of the perfect shadow of why one of them veers off to the right. And those themselves can be quite different from each other, meaning that they're very different in terms of the slopes. And so they're potentially, as Patrick said, very different in terms of shaping your discussion. Now, if you move that up to three or more predictors, there isn't just a shadow to the left, or a shadow to the right. There are shadows in multiple dimensions. And there are lovely explanations that involve ellipsoids. And other completely made up words. But the idea then is, as Patrick said, you get an infinite number of regression weights, all of which yield the exact same r squared. And even if it is a tiny bit sub optimal, they can be very different.

Patrick 26:36
And they're not always very different. That's right. This is actually a measure of model sensitivity. It's almost like a diagnostic. Okay, you have your magical vector of OLS regression coefficients under a glass dome on your desk. Well, how sensitive are those 251 1,000th of a point in our squared terms,

Greg 27:00
and I think there's something we need to be very clear here in distinguishing This is not sampling variability. All

Patrick 27:07
right, so hard to get your head around that. That's right.

Greg 27:11
When you find that optimum spot when you're doing OLS regression or other things as well, when you find that optimum spot that's optimum for the sample that you have. And if you get another sample, then you will find optimum for that sample and then optimum for the next. So upon repeated sampling, there is natural variability that you would expect in the slopes that we're talking about. And as you got bigger and bigger and bigger samples, you would expect that natural variability get tighter and tighter and tighter, so that you don't expect there to be a whole lot of uncertainty in the slope estimates based on random sampling variability. This does not have to do with that. Go ahead and imagine that you have got giant samples. So you've got these tight little confidence intervals around your slope estimates. Go ahead. Let's get that out of the way. What Patrick is talking about is uncertainty on top of that, that is due to this fungibility issue.

Patrick 28:08
There are a lot of really nice recent papers on this. And some will comment on here. If we don't comment on one, that doesn't mean anything other than we just didn't comment on one. But we'll put a long list in the show notes. But in the last three, four or five years, there's been a burst of really interesting work. And related to what Greg just described, Joe LINPACK has a really nice paper where she talks about the very issue that you described, and she differentiates fungible parameter estimates. And she distinguishes that from what she calls confidence sets. And those are confidence intervals and confidence regions on the parameter estimates themselves. So what that means is you can have a large sample size, you can have a small standard error, you can have tight confidence intervals, you can have a well fitting model, do you know what the statistical phrases of how that links to what we're talking about with fungibility don't matter, you can get these differences in interpretation at 5/1000 of a difference in r squared, because this is operating in a different part of the model. I don't care if you have a large r squared and a large sample size, that is not the governing force at work here.

Greg 29:31
There's this term that you have mentioned explicitly. And it really is a theme of where we're going with all of this and that is sensitivity analysis. We talk about sensitivity analysis. We use that expression to describe a lot of things, but it usually takes a different form. You know, we might say something like well, would your results have been different if the distributions of your variables had been different or if you had left out one of your predictors or if you hadn't categorized that variable or as etc, etc, etc. So we talk about sensitivity as this robustness to different changes that we might make in the circumstances and see whether or not the ultimate inference that we make stays the same or doesn't this is in that family, but it is just so potentially not necessarily dire in terms of its consequences. And I think we have no awareness of that. So can you talk about this in terms of sensitivity analysis and how it can be helpful for us?

Patrick 30:27
Well, I mean, I didn't want to interrupt and correct you is I think it's less sensitivity analysis and more pokin stick. Right, because we've talked about that a lot. And, and we've gotten yelled at on Twitter a lot about this pokin stick thing, you're exactly right, we do our sampling, we do our measuring, we do our modeling, and we get a final model that we want to believe in. And then we pull out our pokin stick, and maybe we add an interaction or a curvilinear term, or maybe we do a nonlinear power transformation. Maybe we add a couple of nuisance parameters that moderately improved fit and see what the effects are. Maybe we omit a variable. So we have regression diagnostics, and you do studentized residuals, and you look at outliers, and you look at DF fits and df, betas and all of these things. And the point is, you kind of slap the model around and see well, how stable are my results in how I would write my discussion section. That's how I often think about it, it's sure things are gonna change here, there. But how would I change my discussion section, and if you can jab at it with a pokin sticking, you don't get bit or stung, then you have some confidence in it. This is a similar kind of thing. But we're approaching it from a different angle. The logical syllogism is pretty straightforward. You fit in OLS regression, and you get your r squared, you subjectively pick a trivially different r squared. And then you find all the regression coefficients that would give you that trivially different R squared now with three or more predictors, or an infinite number of those, and so they're different strategies for picking candidate values. Niels has another paper where they derive the maximal difference between regression coefficients, which is kind of cool, right is how far apart can those be? Oh, that's just crazy. Scary. Yeah, well, that is where we move from the picture frames raveling to the ground is now weaving and liquefying, thank you Niels. But you can derive those. And then that becomes a measure of sensitivity, not sensitivity to outliers, not sensitivity to an omitted parameter. But how sensitive are your parameter estimates to where this global minima resides, versus being really, really close to that global minima to the point that nobody cares, right? There's the famous Howard Wainer line of it don't make no never mind. That's right, right is okay move point oh five, nobody gives a crap. But oh, my god, earthquake O'Neill shows us that one of your precious predictors, is positive and significant with an R squared 2.25 and is negative and significant at an R squared of point 245. Go ahead and hang that big heavy picture over your bed and have a nice night's sleep.

Greg 33:39
So there's an old expression a difference isn't a difference unless it makes a difference. We can talk about a difference in terms of r squared here of being whether it's five thousandths or point oh one or whatever heck difference in R squared of point oh, one is nothing. It is nothing in the grand scheme of our lives. But as Patrick is saying, it can make a vast difference in terms of the weights that we have right now. That is where the earth rattles because ultimately, our goal is to understand how things work. And it's one thing if our only goal is to know whether or not a variable is a statistically significant predictor. But that really shouldn't be our end goal, right? Our end goal should be to understand the magnitude of the relations here. In the end, we might find that oh, yeah, the same things are statistically significant. That's great. That just means you probably had enough power still, but the values that are associated with them can lead to fundamentally different assessments, it can lead you to conclude that one variable is much more important and understanding why than the other, whereas the opposite might be completely true. This is really getting at the core of generalizability. And understanding how systems function and replicability and all of this still remain scary as hell to me.

Patrick 34:48
So everything we've been talking about just to get our head around the problem and also to stay consistent with Niels his initial work was in linear regression, right? We have a single dependent variable we on multiple r squared, there's a follow up paper that Niels did with a colleague of his, and it's Jones and Waller that was in Psych methods in 2016. And that generalizes that to logistic regression, and logistic regression, of course, is a regression model for a discrete dependent variable, but we still only have one dependent variable. And they address some complexities in logistic that we won't belabor here, a whole lot of us say, Yeah, I mean, regression is nice, but I've got an SEM, I've got a path model, I've got multiple indicator latent factors, you've got an r squared. But now if I have four dependent variables, while I have four r squared, there's one for each dependent variable. What do you do with that,

Greg 35:45
right? First of all, you pay less attention to the r squared, right? If we think about a larger measured variable path model, or confirmatory factor model, or latent variable path model, or general structural equation model, however you want to call it and all of the other extensions as well, the growth modeling things that you and I tend to like a lot, we don't tend to gauge the worth of those models by the R squared associated with every endogenous variable, every dependent variable that we have in the system, maybe we should pay more attention to that, right. If we listen to some of the things that Doug Stanley had to say, a couple of episodes ago,

Doug Steinley 36:18
I feel like we just want to boil it down to regular regression, if we get significant predictors and regression is going to be related to like a significant test for the model. And that means that we're minimizing the mean squared error to something, which is also exactly what a good prediction is, right? A good prediction is going to be minimizing the distance from the predictive point to the observed points, they're going to be minimizing the same residual. And I feel like tell me a model that you think is a really good model building on theory, but it can't do a prediction. Why is it a good model? That's really hard for me, just from a practical point of view is like somebody presented say, Hey, I got this new model. I'm like, What can I predict? I'm like, Man, nothing really. But these things are significant. Is it a good

Greg 36:55
model. But in the structural equation modeling world, there was a decision made a long time ago to move from an r squared way of thinking about the value of a model to something that was more global. There are a variety of indices as we've talked about before, we had a whole episode first season, but one of the ones that has been around for a while and doesn't seem to have any signs of going anywhere is the root mean square error of approximation Steiger and lens RMSEA. It is a measure that combines an assessment of the fit of the model overall, the absolute fit in the form of the model chi square, if you want to think about that, or really the model fit function, it has sample size involved in it, it has degrees of freedom involved in it. And by virtue of its inclusion of degrees of freedom, there's an element of parsimonious correction that is tied to it as well. Some of the wonderful work that followed from Waller by Tae-Hun Lee, Bud McCallum and Michael Brown extended these ideas of Wallers into

the structural equation modeling world where it's not just a family of regression coefficients that you have, it's actually all the parameters in your model, we have this fit function, and we're trying to minimize this fit function, which simultaneously maps on to maximizing the likelihood associated with our data. Well, one way to measure how well we've done that is this root mean square error of approximation. And let's imagine it came out to be point o four, and we say, that's a pretty good RMSEA. In fact, that's the best RMSEA we could possibly get with these data. Well, what Lee McCallum and Brown said, what if it's not point oh, four? What if it's point oh, four, three? Would you lose sleep over that? And the answer is, heck, no, I wouldn't lose any sleep over that, I'd be pretty much just as happy with an RMSEA of point oh, four, three, as I would with point o four. And what they showed, which is the extension of Wallers work is that it can not necessarily but it can make a world of difference to the parameters that are in your model, the parameters that you are using to try to understand a complex system and earthquake, Waller just had repercussions on the other side of the globe,

Patrick 39:00
I so strongly recommend looking at this paper, I've actually got a copy right here in front of me, it's 2018 psych methods. Again, if you're somebody looking for an area of research that is moving quickly, and there's lots of good work yet to be done, this issue of fungibility across a whole array of types of things that we do. There's a lot of really interesting things to be done right. So this is 2018. So we're not talking about stuff from the 90s. This is pretty rapidly moving, if you're looking for new gig is there's some pretty cool stuff to be done here. But it's wonderfully written it is very, very clear. Tae-Hun is a wonderful guy. He was here at Carolina, he got his PhD under bud McCallum. He and I work together quite a bit and he's just a lovely, lovely guy. He's at Chung ang University in Korea. They have some figures and plots in there. They do a really nice job of laying out what we tried to describe earlier about how you have a parameter on one axis or parameter on the other axis. Now, the maximum likelihood estimates are a point on that plot, right? But they're not closed form anymore. So you alluded to this earlier in the conversation, is that one of the beauties of oh, well, last, as long as you have one observation more than the number of predictors, you will always get a solution in a while less now might be a horrible bad solution. Well, in in SEM, we're very often in a situation where there aren't closed form. And so we get maximum likelihood estimates. So I'm looking at their figure one, and in the maximum likelihood estimate, you get that joint pairing of what they refer to as theta one and theta two as the two parameters. And there's a single dot that represents that vector. Well, then in the next panel, as you describe the move, vector out some degree of misfit that we're willing to tolerate. And so an RMSEA, that differs by point oh, five, who cares? Nobody cares, then you can paint the boundary of an ellipse where anywhere along there, that pairing of coefficients gives you exactly the same RMSEA. And then their final plot is really cool is drawing on Niels, His work of deriving the maximal distance is you can actually get the endpoints of that primary axis on the ellipse and say, Holy crap, we can be way over here on the ellipse, we can be way over on this other side of the ellipse, and it gives you exactly the same RMSEA. Now, what is the complication in the SEM, because you don't have closed form, these ellipses and hyper ellipses start getting really complicated. They have an example that they use that has 67 free parameters. They do two things. One is they focus on a small number of ones that are of primary theoretical importance and consider the other nuisance. And second, they brute forcing. And what that means is they write code to plow through 10s and hundreds of 1000s of different values that you can get. And then they have a way of graphically demonstrating that and trying to identify this in a

real data analysis. So it's really, really cool is everything you think about that SEM, we can just scale that up as another measure of sensitivity. There is a really cool paper that's co authored by colleague of yours, Jeff herring. Yeah. And again, 2019. Come on, folks, if you're looking for a dissertation, if you're looking to contribute, oh, my gosh, there's so much interesting work to be done here. Yeah, but Prendez and Harring in SEM, the journal measuring parameter uncertainty by identifying fungible estimates in SEM, they have a wrapper package for Lavonne. In our that tries to quantify this,

Greg 43:14
I think it's very clever with it did because the work by Lee McCallum and brown admitted that the mathematics of identifying this whole collection of fungible parameter estimates is really, really formidable. And they even said that what they have is important, but kind of clunky for the user. So along come print Dez, and herring. And what they said essentially, is, you have already wandered that space, you have already wandered that mountain in the iteration process that has already occurred on your way to that optimum RMSEA for your sample, you have hit sub optimum RMSEA. So what happens at those sub optimum RMSEA is, we don't usually see any record of that right, all we get is where we are at the last iteration, the only time we see parameter estimates at some sub optimum RMSEA is when our model fails to converge. And the output says, well, here's where you left off, you might want to pick up from here, if you're going to iterate some more, that's about the only time we see that. And what Prendez and Harring said is, well, maybe we can find a way to collect that information throughout the iterative process. And then once we have that, we can know what the Parameter Sets are at these different altitudes of our Fit function. If you go stand on the top of the mountain, you are in one location. Let's say that mountain is 14,000 feet. There you are. If someone says go down to 13,900 feet, will you go What do you mean, where I could go down this way, this way, this way this would be and be on very different sides of the mountain. What Prendez and Harring did is they said, Well, yeah, and he just walked that whole mountain all over it. Let's take that information and give you a sense of what those parameter estimates would be. And I thought it was terrific that they created that package. It's called PSINDEX. And as you said, it goes right along with lavaan. And it can provide that information for you, along with your optimum parameter estimates. And I think that's exactly what we need to be doing, not just sounding the alarm, right, not just cursing the darkness out there, but actually putting things in the hands of users so they can understand in the end, is this going to substantively alter the inferences I'm going to make? Or do I have more confidence in the claims that I'm going to make about the relations that I'm studying in this larger model, I love

Patrick 45:31
your description of the mountain. And it reminds me of something that happened with my wife and may 26 years ago, we went to Steamboat Springs in Colorado for our honeymoon, we were out on the first night after our wedding, it was a cold night high in the Rockies. And we were out on a walk and the full moon was just peeking up behind the mountain. And we huddled up it was cold, we sat there and we watched the full moon rise and it was glorious. We drove back to our cabin, got out of the car, and the full moon was just peeking up behind the mountain. And we were both like, Ah, screw this we're going inside. It's like it's relative, you go down a little bit on one side, you go down a little bit on the other side. This really is a form of pokin stick. We have really well developed diagnostics and outliers in DF fits in DF betas and modification indices. And we have cross validation indexes. And we have a whole focus now on reproducibility and replicability. This is not histrionic, hyperbolic, your model is crap. It doesn't

mean anything. Because a number of these published applications, reanalyze prior models and show that you can mess around with these criterion. And they're really stable in the conclusions that you would draw? Well, that's great. That's a really good thing to find about your model. But they have other examples where you change these criterion by just a little bit. Yeah. And you would change your discussion section in that sense of you need to be aware, right, Doug Stein Lee, when he was on the microphone, that was an old tomato soup cans, and a string in the bottom of a dumpster in Columbia, Missouri. He talked about two things that are related here, we should start with eigenvalues and eigenvectors of our data matrix, no matter what we're doing, just to orient to the data, and a couple of pages into Niels his first 2008 paper, they're set Eigen values and Eigen vectors, these play a critical role in this. The other thing Doug groused a lot about was the difference between prediction and explanation that kind of relates here as well as we reify these parameters that we get right they are the truth is God sees it. And it's our own fault. Because we say there are no other collection of values that give you a smaller sum of the squared residuals. That's right. But the second half of the sentence we leave off is Oh, but there are an infinite number that give us a sum of the squared residuals that are trivially different from this one and nobody would care about C on Tuesday. Be sure to get the problem set in to me by nine o'clock. So what are some take home points here? First, I've been saying earthquake Neilson you said earthquake Waller. Oh, did I you did but it rolls off the tongue better. Earthquake in Neil's has a harder transition. So we're gonna go with earthquake Waller. Okay, that rolls off the tongue. So what are the implications of earthquake Waller? Well, one is let's not quite brag so much about those final estimates we have. Because yes, those are final estimates that meet some criterion that we have defined. But holy cow, there are a whole bunch out there. And by whole bunch, I mean, an infinite number that are damn near exchangeable with those Yeah, and we need to be aware they exist,

Greg 49:24
you can't unsee this, you can't unhear this, once you're aware of this particular problem, I think it will start to become your responsibility to understand the extent to which this is going to influence the inferences that you have. Ultimately, I'd like to see this as a standard kind of output associated with any modeling package, quite frankly. But until that time, I would really encourage researchers out there to hold their breath, go in and request that kind of information. So that in the end, on the one hand, it might lead to tremendous instability and uncertainty regarding what you can conclude regarding particular predictors. The role that variables play in some larger system, on the other hand, it might actually bolster your ability to make claims, and how nice that would be as well. So I think this is moving from this curious thing that some nerds sort of figured out to something that is going to be shifted as a burden of responsibility on the researcher.

Patrick 50:19
And it needs to be supported, right? I don't want to get into the rabbit warren about how we're anti preregistration. We're not, we're not anti preregistration. What we were talking was paying attention to your data and the sensitivity of your models to your results and feeling comfortable navigating those waters, in trying to understand the extent to which your model is sensitive to characteristics of the data characteristics of the parameterization. And now characteristics to the sensitivity of your parameter estimates. This is information to have now what to do with that? That's a harder question. Many of us have taught regression diagnostics. And maybe you have an example where there's an effect that

significant in the presence of the full sample, but it goes to nonsignificant. If you drop two observations, and then people say, but what do you do? What do you do? You shrugged, insider know, right, you've got to make an informed decision. You've got to convey that to the reader, we have to turn toward replication. What do you do if you do these kinds of things? And you find that your parameters are not sensitive in this way? Well, that's good. As Greg said, that's really important to know, you know, what if you do this and you find out holy cow, if I give up point O two on my RMSEA, but my core predictor goes from significant to non significant in a large proportion of the fungible estimates. That's something you need to communicate to the reader, it doesn't mean that your predictor is not sailing into an important, but it means it's unstable. We need to know this when we're interpreting the results. Man's got to know his limitations, especially as we continue to build a replicable science.

Greg 52:13
And on that last point, thinking about each study being this pebble in the pile of science that each study is a data point. And when we tend to approach these things, meta analytically, and aggregating stuff across all of these different studies, it would seem to be so informative if we had this kind of information along the way, also, to get a sense of how much weight to attribute to each of these estimates that comes out. So maybe

Patrick 52:39
as a final point is reiterating what's come up a couple of times, is oh my gosh, there is so much interesting work to be done here. If you go back to Niels original work, he has a couple of follow up some collaborative work with Jeff Jones. If you read Tim Hortons paper, read some by Joe Lynn pack by Jordan Prendez. And Jeff Harring, go orient to these papers. And there's just half a dozen things that fall out of work that has yet to be done. That would have a very high impact. As Greg just alluded to, this should be part of standard output from an SEM package. What would that look like? What are criteria we would use? What are measures that might indicate a problem versus not a problem? How could we make this part of reporting standards? There's a lot of work yet to be done on this.

Greg 53:33
You sounded really excited talking about that topic. This is cool. It is very cool. I would say that you were grunt old.

Patrick 53:40
You're making up words again.

Greg 53:42
No, if there exists disgruntled, and today you are completely grunt old and I appreciated that

Patrick 53:48
then I am Gruntal than nonplussed. If you can learn a word from Wikipedia, I can learn a word from Are you are officially deprecated

Greg 53:58
Did you say deprecated? Thanks, everybody. All right. Bye. Bye. Bye. Thanks so much for joining us. Don't forget to tell your friends to subscribe to us on Apple podcasts, Spotify, or wherever they go to learn new words or new definitions for words they already had perfectly good definitions for. You can also follow us on Twitter where we are at quantitated pod and visit our totally redone website quantitative pod.org where you can leave us a message find organized playlists and show notes. Listen to past episodes, get transcripts of recent episodes and other fun stuff. And finally, you can get amazing quantitative merch, like shirts mugs and notepads from red bubble comm where All proceeds go to donorschoose.org to help support low income schools, you've been listening to Quantitude: The podcast that listener Caleb Chang of Cincinnati, Ohio said makes a seven mile run feel like a 10 mile run. Today's episode is sponsored by statistical terms that don't exist, but really should. A model may be recursive or non recursive. But how about a model that is just playing cursive? It looks quite lovely but it's estimation leaves you cursing, ever even starting the research project in the first place? A cursive model and correlation matrices might be positive definite or non positive definite. How about a positive non definite correlation matrix? We have correlations above one below negative one, or just that make no damn sense at all, a positive non definite correlation matrix. And finally, for missing data, there's MCAR, MAR and MNAR. There should also be MWTF missing what the when you really have no freaking idea why data are missing? MWTF missing data. We highly encourage you to integrate these terms into your daily statistical lexicon. This is most positive non definitely not NPR.

MovieClip 55:46
Regularly you are inflammable and a bit perpetuity