Episode 17: Dr. Jason Osborne

# KL: Katie LinderJO: Jason Osborne KL: You’re listening to *Research in Action*: episode seventeen.

# [intro music]

# Segment 1:

# KL: Welcome to *Research in Action*, a weekly podcast where you can hear about topics and issues related to research in higher education from experts across a range of disciplines. I’m your host, Dr. Katie Linder, director of research at Oregon State University Ecampus.

On today’s episode, I’m joined by Dr. Jason Osborne, Associate Provost and Dean of the Graduate School at Clemson University in Clemson, SC, where he is also Professor of Applied Statistics in the Department of Mathematical Sciences and in Public Health Sciences.  He is author of over 70 peer-reviewed articles and seven books, many of which focus on best practices in statistical methods. He has also been active in research related to social justice, educational psychology, and evaluation. His work has been cited in scholarly publications over 10,000 times according to Google Scholar, and he is also an Accredited Professional Statisticiantm (awarded by the American Statistical Association).  Jason is a 3rd degree black belt in Songahm Tae Kwon Do, and the proud father of three, each of which he considers an outlier in the positive tail of the distribution of awesomeness.

Thanks for joining me, Jason.

**JO**: It’s my pleasure to be here.

**KL**: So Jason I had actually reached out to you because I read your book, Best Practices in Data Cleaning and just found it to be helpful, and credibly practical, and I really enjoyed it. I knew I had to have you come on the podcasts. I’m so glad that you can join me. One of the things that really drew me to this book, is you talk about very early on in the book that one of the reasons that researchers might not be using best practices in data cleaning, is a lack of clarity in exactly how to implement those practices. And that this is something many of us are not directly trained in as researchers. Why don’t we start out just by defining data cleaning, what does that mean?

**JO**: Sure, so data cleaning in my mind, in some fields there are some very specific things that constitute data cleaning. What I tried to do is be a little more inclusive in talking about data cleaning to include everything that you need to do, to make sure you’re ready for, what most of us really want to do. Which is the analysis and to find out, did my hypothesis get support or not. So, just as a very kind of broad overview, I view data cleaning as whatever process you go through to understand your data, to become more intimately aware of what your data is telling you, and to make sure it’s got reasonably good quality to it.

**KL**: One of the things that I think can be kind of attractive to researchers, especially those who might have graduate students, is to delegate data cleaning to someone else. But one of the things that you’ve just mentioned and you talk about in the book is how data cleaning really helps you to get to know your data on a really different level. Is that something that you think that all researchers really need to kind of dig in and get into the data through the data cleaning? Is that something that can benefit their analysis in the long run?

**JO**: Well obviously I think so, I wrote a whole book about it. But let me speak a little bit more to that. As I was coming up through graduate school and beginning to embrace quantitative research, the reason why some of us get into this is because some of that moment of epiphany that we have where, we’re talking in the computer age now where we do analysis with statistical packages. We push run and then we almost immediately see the result. We have that moment where, we’re the first person to see new knowledge and that’s a really exciting and almost addictive thing; at least it was for me. Many people try to get to that as quickly as possible and what I’m trying to do is to communicate the fact that, yes we need to celebrate that, and that’s why we’re in this business, and why we do what we do, but you can’t jump right to that. Or at least you can’t jump right to that and then say that’s the final word. It does really help to understand your data in a deep way, to play with it. I don’t think we do enough as we’re training researchers or has practicing researchers to just play. It doesn’t hurt anything and it really doesn’t take a whole lot of time given modern computers to just throw some plots together, look at graphs, scan the data. Just play, we have this really fun profession and I think we should encourage that more. Getting back to your question, all of that helps you understand your data in a deeper way that them helps you understand the nuances and the trustworthiness of that final epiphany moment.

**KL**: I think that’s such a great point, that we don’t want to rush that, we don’t want to really be looking and also just clarifying, what is the direction the data is pointing once we’ve gone through and have a really deep understanding of it. One of the things that you do in this book and you give some really excellent examples of this, is just the different ways that data can be interpreted before you clean it and then after you clean it. And you give several examples where there are drastically different outcomes with the data and sometimes opposite outcomes.

**JO**: Yeah and this comes out of, this is probably just my therapeutic attempts to recover from several decades of quantitative research, and battles with peers, and journal reviewers, and things like that; but what I’ve come to believe and I think I can demonstrate, is that in the rush to want to know, we may overlook a few things that actually impede our abilities to know or at least our ability to be confident in what we know. And so none of this is to say that we shouldn’t embrace that but what I wanted to do is kind of bring us as a field. And by field I mean a broad group of quantitative researchers, back to our roots a century ago; where we did think about, what does it take to get a good trustworthy result from this particular analysis? And I do try to be very pragmatic and motivate people who somehow happen to read my book to want to do this. I’ve tried to come up with authentic examples from my own data or from real data. I don’t just make up data for the most part, to motivate people to say, “Oh yeah that could be valuable to me.” And what’s been fascinating to me is I really didn’t purposely start down this path. It just sort of happened that I discovered the importance of this. By discover I mean it was reinforced to me; it’s been out there for a century in the literature. But back to your point, I don’t think that, at least in my experience, we’ve done a good job of acculturating researchers to the importance of just looking at very basic things. And as I’ve gone through, the data cleaning book was the second book I wrote, and then as I’ve gone through research and other books, I just keep finding more examples of little things you can do that have sometimes profoundly impactful facts on your data.

**KL**: One of the things that you mentioned a little bit earlier but also talk about in the book is that the kind of the rise of these software packages that allow you to run statistical analysis on data. One of the challenges is that sometimes people get trained in these software packages but that means they don’t really know how to do hands on work with the data aside from those software packages. What are some of the things that you might offer to beginning researchers in terms of just thinking about what they need to know beyond how to use these software packages, kind of the fundamentals of working with the data.

**JO**: Right, so I’ve had a lot of conversations over the years with colleagues, and mentors, and things about this because we’ve, at least I lived through, I see change in terms of my career where, when I started doing research as an undergraduate, it was very common for us to do some basic analysis by hand. And I learned matrix algebra and all of this, and then pcs and Macintoshes started coming on to our desks and all this stuff happened. And so people from previous generations have this discomfort with people jumping right into statistical computing without knowing the formula and the matrix algebra, and being able to do things by hand; so I get that. The way I tried to think about it and talk with my colleagues largely about it, is driving an automobile. Back in the old days, you had to really know a lot about a car because at any given moment it could break down. And you had to understand how things worked because otherwise you’d be stranded in the middle of nowhere. Today we don’t really have to understand internal combustion engines, and calipers, and fuel injection systems. We have to know how to be responsible users of that system and safe users. And so, the simile or the analogy I try to draw is, we don’t need to, I don’t think most people need to understand how to do these things by hand. I think we need to help them transform into responsible users of the system and part of that is understanding what goes into a good result and all the steps. And I also think there’s some real benefits to the era we live in now. If you wanted to do plots forty years ago by hand, that was a very time consuming thing. Today, in several seconds, you can create a scatter plot and this kind of plot and that kind of plot. So, you can really play with your data in a way that you couldn’t a generation or two ago. I think people may have lost some deep knowledge of the actual mechanics and we have mathematical statisticians for that. The vast majority of researchers I’m aware of don’t need to understand that.

**KL**: That makes total sense to me. I love that analogy of thinking about cars, and technology, and what do we need to know, and what do we not need to know to be responsible users. We are going to take a brief break. When we come back, we are going to talk a little bit about some myths and practicalities of data cleaning, back in a moment.

# Segment 2:

**KL**: Jason, one of the things that I though was really interesting about your book Best Practices in Data Cleaning was that you framed it around several myths that you say have undermined data cleaning as a standard research tasks. Of the myths that you discussed in the book, what do you think are the most pervasive?

**JO**: I think there’s a few that are probably widely held at least within the disciplines that I’ve been associated with, which involves health sciences, social sciences, education, that constellation. Some of the myths are things like; if we clean our data we make the results less reflective of the population. That’s one I hear all the time from journal reviewers, and colleagues, and so one of those, that was one of the first myths that I thought about and has led to several articles, and a real perseveration on data cleaning. I think another one that’s particularly challenging is the myth of robustness. That in 2016, we don’t really need to worry about testing all these lists of assumptions because, well they are largely robust to violations of the assumptions. And part of, in the previous segment you asked about software; I think something that we have lost is the procedural comfort of going through typical steps where you get your data, you look at the data, you test assumptions, you calculate the results, and so on. With software we could do a much better job of scuffling these types of things, where the software could actually que you to do certain things. What we have is, we have a generation or two of researchers who will say, “I’m suspicious of what you are doing because I don’t believe it’s important.” If I publish an article where I say, “Well you know there were several influential cases and I removed them, they are suspicious that I’m cooking the data.” And so one of the things that I tried to set out, to do in this book, was to actually show using evidence, that doing this may make your results much more reflective of the population for the group you want to generalize to. I also spend some time focusing on the importance of testing assumptions because there really is no way to assume results are robust to violations. Particularly if you have examine whether there are violations, which is another thing that I tend to focus on.

**KL**: One of the things that you talked about in this book, which I just found really fascinating, is that we are, we’ve kind of gotten to a point where we assume that people are doing some degree of data cleaning but we don’t describe it in our publications. And you talk about, kind of reviewers not really asking these kinds of questions when they get articles that don’t talk about it. And you’ve actually done some meta-analysis of articles in different fields and really noted that you know the majority of them, the vast majority of them are not talking about testing assumptions or data cleaning. And one of the goals actually of your book is really to get people, to get back into that practice. Why do you think that we’ve kind of fallen away from that or were we ever in a place where we were describing those things in our publications?

**JO**: I think there was a time, a long time ago, when a lot of these statistical techniques were being invented and they were new that I think people did take pains to at least communicate some of their assumptions and whether the data met their assumptions. I think that we’ve had this tension in our publication, particularly in paper publications where journals don’t want to spend pages and pages in each article having authors describe what they’ve done to clean the data. They want to get through, why people should care about this thing and then what is the result, and your conclusions? And so I think that there’s been several things, including that tension for the space on the printed page that helps squeeze that out but I also think that, I think there’s a couple things going on. I think there are a lot of really good researchers that are doing this kind of thing and mentoring their students, and they are not reporting it. So that’s one possibility. There’s a minority of researchers who do report it and that varies by field and discipline in general. And then there’s probably a group who never really thought about it, or who weren’t trained to, or who aren’t convinced that it’s valuable, so they don’t do it. I kind of take two tacks in my writing; one is to encourage those people who are doing it, to at least very briefly report it. And this can be as simple as a sentence or two where you say, “I looked at my data and there don’t seem to be any influential cases, and I’ve also tested the assumptions, and I assert that I’ve met the assumptions.” Something as simple as that, can then give me as a reader a lot more confidence in your data. Whereas if I don’t see that, I’m wondering and I’m not sure whether your results are trustworthy.

**KL**: Absolutely. In your book you talk about how text book seem to skim over the important details of data cleaning and this means that researchers, especially kind of junior researchers, leave either avoiding doing those things all together or they spend a lot of time trying to figure out how to do them on their own. I’m wondering for people who have little experience in this, and maybe they are listening to this and going, “I had no idea. No one told me; I didn’t know I was supposed to be doing this in the first place.” What are some practical first steps that researchers can take if this is an area where they have little experience.

**JO**: I think the first thing, that we need to get that group of folks onboard with, is buying into the notion that’s important. Once you get someone to think about the concept of testing assumptions and looking at the quality of your data as an important thing. Then there are lots of things that they can do and I really do describe it as both a science and an art. Because I don’t think there’s any real right ways to do it. I think that the most important thing to do is to just to do something. What I would suggest, if you don’t have a clear direction, is to simply start looking at scatterplots or histograms of your data. To look through your actual data spreadsheet and see if there’s anything surprising or comforting; to look at simple descriptive. We all want to jump into that latent variable model or you know the sexy stuff but sometimes just looking at simple correlations can help; you know just playing. I think we have to bring, I’m sorry I keep harping on this, but I think we have to bring the notion of fun and play into quantitative methods because there really is no wrong way to approach it and it can be kind of fun, at least kind for me. That might be revealing some kind of neuroses that most people don’t have. But I think playing around is a good first step, but playing with a goal. Your goal is to see if the data are what you expect and if the assumptions are reasonable to say that they’ve been met. If you look at a variable like, how many friends does a typical 16 year old report having? And most will say, I don’t know, six to eight, to ten, to twelve and then there’s a ninety-nine. Well you might want to think about whether that case is something that you expect and is a legitimate member of the class that you want to look at or whether that’s something that you want to separate our for closer examination.

**KL**: one of the things that you mentioned and I think that many people don’t associate quantitative methods with play or fun; particularly people who may feel a little fearful of statistics or of their abilities in that area. And I think one of the things that might be helpful for listeners, is I want to refer them to episode eleven where we talk a little bit with Steve and Tuyl about data management. And one of the things we talk about is backing up your data and having your raw data backed up in a way that you feel like you should be able to play with it and not be afraid that you’re somehow wreck it or ruin your data. If folks are interested in learning a little bit more about that, I recommend listening to episode number eleven. But I think that that’s such an excellent point Jason about really getting to know your data. There’s nothing wrong with getting your hands in there and really thinking about, what does this mean? Before you run the analysis and that allows you to then look at the analysis and say, does this backup what I thought was there; which is also a very important component of research.

**JO**: Well you make a very important point and I want to really emphasize this. I’ve seen many things go awry when people play with their only copy of their data. Absolutely back up your data and put it in your software or wherever you put it in, in the cloud and then play. The nice thing about this modern age of statistical computing, is that you can just play and it’s not like you can break the software, and you can’t break your data because you always have the backup copy. So there’s really nothing at stake and people get very fearful around mathematics, especially in the United States. But this isn’t really math, this is science and there’s numbers but I always argue that statistics is more about being able to explain what those numbers mean rather than being able to complete equations and get the right answer. I think if people bring that notion, then it’s okay to just play, you can’t break anything. I think that relieves some of that anxiety.

**KL**: That’s such a great point. We’re going to take another brief break. When we come back we’ll talk a little bit about Jason’s most recent book and some of the things that are coming up for him, back at a moment.

# Segment 3:

**KL**: Jason as we record this episode you have a new book that just came out within the last few weeks; Regression and Linear Modeling Best Practices in Modern Methods, this is from Sage. I’m wondering if you can tell us a little bit about this book.

**JO**: Yeah, so this book was really an evolution from the previous books. I feel like I’ve been on a journey from Best Practices in Data Cleaning which was my first book through this one. And my goal here was to create a book. I often teach regression courses and things like that that are very applied and there are great books out there, classic books, I love them. But I never felt like I had the right book for me and so this book has been kind of the summation of my journey over twenty six years now. I created this book to try to mentor the reader into a larger world. Many people are familiar with simple regression and what we call ordinary least squares regression which is the most common in any sciences. And then that’s where they stop. But there’s a whole universe of regression types that are really useful when you have different types of data. To be very brief, what I tried to do was to develop a best practices and regression theme and then apply that to a variety of different linear models. And all the time, by the way, I keep this focus on data cleaning and testing assumptions of each of these different models. So trying to really gently open the reader to what’s a very large universe of tools.

**KL**: I love books like this that are so practical and the idea of you know, as an author, also being a mentor, I think is super important especially with research methods that can seem really abstract to people. I realize we may have some listeners who are hearing regression and maybe tuning out a little bit because they are not quite sure what it means. Can you briefly describe; when is something like regression used in terms of data analysis?

**JO**: Regression is a very flexible tool that we use to look at relationships between multiple variables. The easiest example is your height and your weight, are probably related. People who are taller tend to be heavier; people who are shorter tend to be lighter; although there’s a lot of variance around that, right. That’s a general relationship and those are two continuous variables; things that can vary across a broad range. So we would use ordinarily squares regression to look at that simple thing.

**KL**: And by using that regression to look at it, what are the kinds of things that it can tell you?

**JO**: It can tell you several different things. First of all, is there a relationship? And what is the nature of the relationship? In this case I described a positive relationship where, as height goes up, weight also goes up. And there are, what we call inverse or negative relationships, whereas one goes up, another thing goes down; so, you get that. And then you also get strength of the relationship. As we as a society get a lot more variance in terms of our body shapes and types, someone at five foot four for example, can be a hundred pounds or two hundred pounds. So there’s variance around that relationship and it can quantify how strong or weak of a relationship that is.

**KL**: One of the things that I think might be of interest, I mean it’s certainly of interest to me, maybe of interests to our listeners as well; it’s just this idea of writing about statistics. I think that it’s something that most people don’t necessarily think about, writing about research methods as a thing, you know as something that you can do as a researcher. So when…..

**JO**: Including my wife she doesn’t understand why anyone would want to read or write about this stuff.

**KL**: Well I think that its, there is such a need for these practical guides and there are new techniques and things that are being developed. And I’m just curious if you can talk a little bit about, how you decide to move into a new project? Or how you decide, you know clearly these books and articles that you are writing are linked and as you pointed out, there is a kind of trajectory or journey here for you. But what kind of leads you into the next thing you decide to work on?

**JO**: You know that’s a great question that many scholars wonder about. When you are on the journey, sometimes it becomes self-evident what the next step is. And I’m not sure I can really articulate it but what I will tell you is that I’ve been very lucky and blessed to have been at institutions where I can teach amazingly intelligent graduate students and I have wonderful colleagues. A lot of my articles, especially early on, were the result of a student or a colleague just asking a simple question. And either me not having an obvious and clear answer for it, or seeing that there really isn’t an obvious and clear answer in the literature or in text books as to either why something would be the way it is or how to do something. And so, as I taught courses and interacted with people, you just get all these really interesting questions that seem very basic until you dig into them. That’s really been a lot of my journey and I’ve been fortunate, I will also say, to co-author with some of my students who get interested in this and have interesting questions and so we take a journey together for a while.

**KL**: One of the things that, I know your pipeline is so full. Clearly you have some ideas for future directions of where you’re going to go. Can you tell us a little bit about that?

**JO**: I do and part of the theme of recent books, and some articles, and conference presentations has been curvilinear effects in a linear world; I’ll just throw that out there as a phrase. And what I mean is that most of our analysis, like we call regression linear regression, we assume that there is a linear relationship between something. And I think most people intuitively understand that things are not linear in life and we can talk about many different examples such as the learning curve. You know as people are learning to read, they learn one word, two word and then all of a sudden they take off on this exponential rise and the number of words they can process. You know when we are learning a sport, there’s a growth curve. We intuitively understand that things are curvilinear yet we often don’t model them. And so one of the things I have spent some time in recent books is really helping the reader understand that it’s not very hard to do that and it can be really fun, when you see a curvilinear affect that just makes you stop and think or that can be very useful. For example, in my new book there are a few themes data sets that I look through and one of them is the relationship between smoking marijuana and then achievement test scores; something like that and there’s a really interesting curve in there. And it’s not just because the curve is shaped an interesting way, but if I actually knew anything about drug use or achievement, I could probably use that to really focus in on a particular area where an intervention can be most effective. Whereas if you just have a line, that doesn’t give you any indication as to where you might get more bang for your buck.

**KL**: Well I think that the passion that you bring to data cleaning and to statistical methods in general is incredibly inspiring to me. So I want to thank you so much Jason for coming on the show and I highly recommend to our listeners Jason’s book, especially one of my new favorites, this Best Practices in Data Cleaning. So, thanks so much Jason.

**JO**: Thanks for having me, it’s been a pleasure.

**KL**: …And thanks to the listeners joining us for this week’s episode of *Research in Action.* I’m Katie Linder and we’ll be back next week with a new episode.

Show notes with information regarding topics discussed in each episode, as well as the transcript for each episode, can be found at the *Research in Action* website at [ecampus.oregonstate.edu/podcast](http://www.ecampus.oregonstate.edu/podcast).

There are several ways to connect with the *Research in Action* podcast. Visit the website to post a comment about a specific episode, suggest a future guest, or ask a question that could be featured in a future episode. Email us at riapodcast@oregonstate.edu. You can also offer feedback about *Research in Action* episodes or share research-related resources by contacting the *Research in Action* podcast via Twitter @RIA\_podcast or by using the hashtag #RIA\_podcast.  Finally, you can call the *Research in Action* voicemail line at 541-737-1111 to ask a question or leave a comment. If you listen to the podcast via iTunes, please consider leaving us a review.

The *Research in Action* podcast is a resource funded by Oregon State University Ecampus – ranked one of the nation’s best providers of online education with more than 40 degree programs and over 1,000 classes online. Learn more about Ecampus by visiting ecampus.oregonstate.edu. This podcast is produced by the phenomenal Ecampus Multimedia team.

# Bonus Clip #1

**KL**: In this first bonus clip for episode 17 of the research in action podcast, Doctor Jason Osborne describes some assumptions of measurement, take a listen.

Jason in quantitative analysis, measurement is such a large component that many people are relying on. I’m wondering if you can talk a little bit about some assumptions that might be being made about measurements that could be harmful to the overall analysis of quantitative data.

**JO**: Absolutely, so in science we assume we’re measuring what we think we are measuring and that we are measuring it well. And in almost every quantitative analysis, that’s just an assumption and most people don’t actually test it but if you think carefully about what we are doing here, everything that we do is predicated on the notion that we’re measuring things well and that we are measuring what we think we are measuring. Now in some fields like bench sciences, that’s sometimes an okay assumption; you stick a thermometer in something and get a reading, if you’re well calibrated you can trust that. When you’re in health sciences, social sciences, and education, when you’re dealing with human beings, sometimes that assumption falls apart. And I don’t think we take enough care in our research to make sure that we have good measurement and valid measurement. And so, one of the things I talk about in some of my books in writing, is the need to really focus on the quality of measurement because if you have lousy measurement, nothing else matters after that point.

**KL**: that’s such an important point and I would actually also point our listeners to episode seven where we have Doctor Josh Weller talking about psychometrics. Which is also very much about making sure that your measurements are measuring exactly what it is you are hoping to measure and that your measurements are valid; I think that’s such an important point.

You’ve just heard a bonus clip from episode 17 of the research in action podcast with Doctor Jason Osborne describing assumptions of measurement. Thanks for listening.

# Bonus Clip #2

**KL**: In this second bonus clip for episode 17 of the research in action podcasts, Doctor Jason Osborne discusses the components of testing assumptions, take a listen.

Jason one of the fundamental components of data cleaning is this idea of testing assumptions. I’m wondering if we can talk a little bit about, what do you mean by assumptions. What are the kinds of assumptions that people might make about their data?

**JO**: Right, so in order to understand assumptions, we need to understand how these statistical tests came to be. And basically at the end of the day, a hundred and twenty years ago, there were a bunch of really smart people who were trying to figure things out. And they were mathematicians who were trying to summarize patterns, or to test hypothesis, and producing a number to do that. Well in order to have those formulas be valid, they had to make certain assumptions about the data or else none of the math worked. At the very basic level, the assumptions are cooked into the test we’re using. And what we are not doing is actually evaluating whether our data meet those assumptions and then whether it’s legitimate for us to use this particular test, whatever test we are talking about.

**KL**: That’s really helpful Jason. Can you offer maybe a specific example of a certain assumption and maybe why you might want to test that assumption with your data?

**JO**: So keeping in mind that every test has different assumptions, there are a few common ones that seem to pop up again and again. And there’s a whole class of tests called parametric statistics, which assume certain distributions. So for example, if you’ve ever taken a statistics class, you probably heard about a normal distribution. It’s also called a Gaussian or a bell shaped distribution. And what that really means is that, the data can form to a particular shape that the majority of cases are around the mean and then they feathered down to relatively rare cases being far from the mean, and it’s symmetrical and has certain mathematical properties. Well we perseverate about that in this particular case because if your data don’t conform to that, all the tests that flow from those data are miss-estimated. So all of our calculations are, we haven’t talked about significant tests or P values but those are built upon a normal distribution. So to the extent that your data deviate from a normal distribution, then all of your tests are either underestimated or overestimated and unreliable.

**KL**: That’s such a crazy thought that people aren’t necessarily testing this assumption and then all of the things that they are running after that are maybe, incorrect.

**JO**: They might be and if you’ve been paying attention to the scientific literature, there has in recent years been a huge controversy over replicability, reproducibility of effect; which is a basic hallmark of science. And part of that could be just a simple that we’re not starting with clean data that represents the population we want to talk about. And if two researchers have data that are nonconforming in different ways, you can get vastly different results and a huge controversy over who’s right, when in reality they both could have found the same thing.

**KL**: So one of the ways to maybe test this particular assumption of normal distribution and it maybe be a very simple way; could be just to plot the data and look at it and see does it fall visually in what looks like a normal distribution. Are there other ways to test this assumption?

**JO**: There absolutely are and there are a lot nuances to this but there are some very simple mathematical tests you can do. Such as, well in statistical packages there are numbers that will tell you how close you are to conforming to the standard normal distribution. You can look at standardized ways of knowing whether data points are farther from the mean than you would expect. And if you are doing something like regression, a really fun thing that a lot of people don’t do is to look at residuals and what we assume in regression is that you have normally distributed residuals. And so you can look those and see whether they are normally distributed and wildly out of shape and that can tell you other interesting things about your data like, whether you might have a curve in your data or whether you might have outliers. My focus in preservation on outliers can partly flow from wanting to make sure we have data that’s appropriately distributed.

**KL**: Well Jason thank you so much for sharing a little bit more about what it means to test assumptions.

**JO**: Happy to talk to you, thanks.

**KL**: You’ve just heard a bonus clip from episode seventeen of the Research in Action podcasts with Doctor Jason Osborne discussing components of testing assumptions. Thanks for listening.

# “Research in Action” transcripts are sometimes created on a rush deadline and accuracy may vary. Please be aware that the authoritative record of the “Research in Action” podcast is the audio.