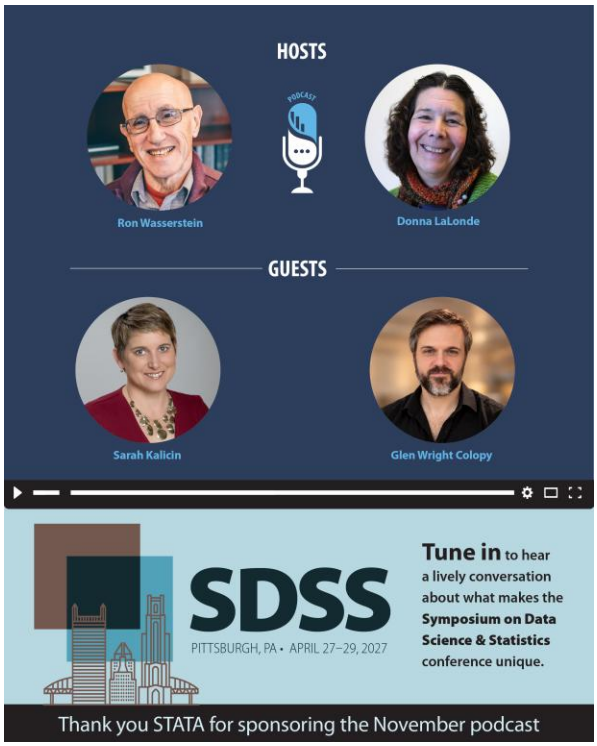
Practical Significance | Episode 71: Under the Hood—Statistics, Data Science, and AI at SDSS 2027



Donna LaLonde: All right, well welcome everyone to the November issue of Practical Significance. And we are super excited to be able to talk about all things SDSS. And so we have with us two members of the SDSS program committee, Sarah and Glenn. And we're going to get right into it, but we do always start by asking our guests to tell us a little bit about their day job.

So Sarah, I'll start with you. When you're not chairing the program committee, what are you doing?

Sarah Kalicin: Hi Donna. Thanks for hosting us. So, I retired a few years ago from Intel Corporation and part of my retirement was like, where can I give back?

So, I've been volunteering with the Red Rock Search and Rescue team here in Las Vegas. And one of the things that I then found out, search is very rich in analytics and data. So right now, since I discovered that I joined the team about a year ago, I've been building out a search analytic database to help the team use data and figure out how to execute and plan searches much more effectively and efficiently.

There is so much data, but nobody's organized it or analyzed it. So that's where I'm putting my passion right now. And with all the new technology with mapping, tracking, drones, I'm really eager to see how we can really improve search and rescue teams approach to searches. So that's what I'm doing for my day job. Plus a lot of traveling internationally, which by the way, I was in Poland and met some of my search and rescue counterparts there and we started talking about data and stuff. So I've been having a lot of fun.

Donna LaLonde: Always a data nerd. It doesn't matter where you are, right? That's great. That's wonderful.

Glenn, tell us about your day job.

Greg Wright Colopy: Yeah, so I work for two companies. The first is Polaris Data Labs. It's a bespoke data science and machine learning consultancy for life science companies. And then the other company I work for is called Igor. It's an electronic lab notebook and lab information management system. So basically the two different sides of helping life scientists deal with their data. One on the bespoke consulting, more traditional statistics, data science, analytics side, and then on the other side, it's the software, essentially building the products and tooling that they use to help record their data and get the most of it.

Donna LaLonde: That's fantastic. And it also sounds like both day jobs are great fits for being involved in SDSS. And so, Sarah, I'll start with you. You're chairing the SDSS 2027 program committee. Can you tell us a little bit about what that means and what drew you to the conference?

Sarah Kalicin: Yeah, I've been on the committee for the last three to four years. I've also organized some sessions a few years beforehand. What I love about the SDSS, it's where ASA gets to really showcase how statistics can really impact the broader field or really do collaborations across fields. The conference also bridges other fields of data science and AI along with statistics. The three fields are so closely related, but yet so different that we really need to have all these experts coming together and talking about how statisticians can positively contribute to data science and AI, while also showcasing how the new data science and AI field can impact statistics and allow us to be much more effective while also bringing really positive impact to the field of statistics.

So that's why I'm passionate about it, because there is so much going on, things are moving so fast that I see this conference as being the keystone of allowing stats to be relevant in this really fast-moving field.

Donna LaLonde: That's great. And Sarah, I'll give you a shout out for being extremely well organized. So you bring your data nerd hat to good use by keeping the committee on track, which can be a little bit like herding cats. So, yeah, so we appreciate it. Glenn, you're on the program committee and I believe you're heading up the social media outreach subgroup. So tell us a little bit about what drew you to SDSS and what your role is on the committee.

Greg Wright Colopy: So the track that probably would interest most people would be the AI and LLM application track, as in that's the track that I'm involved in that would interest most people.

What interests me about the conference is, you know, fundamentally, at my core, I'm a statistical data scientist.

I cannot unthink in Bayesian methods and things like that. So I can't not approach the applied data problems that I see in any way other than as a statistician and data scientist. However, there are these immense tools that are available out there with AI and machine learning that frankly are too exciting, too useful not to be using and making the most of. And also, they make beautiful, interesting problems.

And so I think what interests me about the conference is, and this is presumably what will interest a lot of people is that while these new tools are coming out and useful, I don't think it is useful to a lot of statisticians and data scientists to throw all their prior experience to the wind and jump entirely on a fancy new AI train. It does not make sense to abandon all that experience and just hop on what something that is very popular has a lot of hype on it right now.

And so, what SDSS is doing properly is that we are a good place where the traditional skill sets that have made data scientists and statisticians useful are still a core part of the value that people are talking about when they're coming and presenting a program. And that's one of the things that we're going to be looking at when there are submissions and things like that.

Those core values of statistics and statistical data science are still very appreciated in our conference while we're also not turning a blind eye to AI and machine learning, things like that. So sort of the snuggly feeling that I get when I think about the conference is that the core professional values that I've had, that my parents have had before me because they're from the same field again, that is still being represented and appreciated, or while we're also basically trying to feel our way to seeing how do these other tools work and help us out.

Ron Wasserstein: Thanks, Glenn. So, Sarah, thinking about someone who's never been to SDSS before, so for that person's benefit, how would you describe the meeting and how's it different from, say, JSM?

Sarah Kalicin: JSM is a great conference. Let me just start off that it's a conference to promote the broad advancement and application of statistics. In other words, it's where statisticians get to geek out on statistics. Where SDSS is different is it's really there to bridge the

fields of statistics, data science and AI. The focus is not just statistical methodology, it's how statistics is being applied to data science and AI and or how data science and AI is being applied to statistics so that we can highlight case studies and real-world problems that bring value to society. SDSS really is going to be focusing on four tracks.

There's actually four tracks. One of them is data science applications, which is basically the case studies and novel contributions involved in machine learning, computational statistics and statistical methodology. So really heavy duty on data cases and applications that demonstrate how these approaches are being used to address real-world problems. We're also looking at education and professional development as well.

So, this is like a novel approache in AI, data science and statistics education to basically promote the efforts in developing skills and workforce needs for the future. For anyone who's attending, we also have the systems and infrastructure. This is kind of really where things start getting a little bit more nuanced because in order to do a lot of statistical computation and AI and data science, the systems and infrastructure, the software and hardware, in order to make applied data science work, this is something else that we're also focusing in a track.

It enables the practitioners to really be able to use these methods in practice. And then finally, what everyone's talking about almost everywhere is a track around AI and large language models, LLM applications. So really looking at the applications of artificial intelligence, large language models and all the little nuances that we're seeing in data science research, practice and education, but also the nuance that studies stats can bring to these applications as well.

So, this is the kind of nuance with the symposium of Data Science and Statistics, or SDSS, that we're trying to focus on and really making sure that we are geeking out not just on statistics, but computer science, like the statistical computation and professional development, either from education into the real-world industry, government and pharmaceutical. So.

Ron Wasserstein: Well, thanks Sarah. It sounds like a great program. I'm already excited about it and I know that seems like a long way off now, but it'll get here quickly.

So Glenn, Sarah mentioned the idea of bridging that SDSS does, bridging statisticians, data scientists, computer scientists, and so on.

So from your experience, what does that mix from an on-the-ground standpoint? When you're in sessions, when you're in the hallways, how does that mix play out?

Greg Wright Colopy: Well, I don't think the mix on the ground deviates particularly strongly from the fact that these still are statistical data scientists who are showing up. You know, essentially if someone is purely interested in the application of LLMs, let's just say some non scientific field, for example, or something like that. I don't think that that's really going to be where the weight of the conversations are. The weight of the conversations are still going to be statisticians and data scientists trying to navigate their way through the world. Not too dissimilar. From how we were, you know, 10 years ago as well, you know, just with different types of tools and different types of challenges, different types of new interesting things to go through. So I think the people who are going to be coming are going to be people with that good solid statistical data science appreciation. They're self-selecting into the conference because they see this is a conference that appreciates the

statistical aspects, that appreciates the more traditional data science aspects. And we're trying to figure out what to do with this future that seems very uncertain right now. And so on the ground, that is what I am expecting, that's what I've seen before, that's what those conversations are about. And also just people continue to be people.

Their concerns are things like how do I secure my job, how do I continue to professionally develop, how do I continue to enjoy my job? You know, if you look at the Wall Street Journal back in 2010s where they called statistician the, you know, the sexiest job of the 21st century, I don't think that prediction lasted five years. I think then it became data scientist and then now there's the AI engineer. Somehow software engineers got totally skipped in that description,

which seems really strange since that's a profession that's powered a large number of these tools. But they're being paid well, so I guess they can go crying to their golden pillows. The point is that the fundamental cares that people have are still along, you know, essentially staying employed, still intellectually invested in what their jobs are. You know, essentially there is stuff like AI-based coding is cool, but I don't want to spend my entire life debugging AI code.

Essentially they like the speed up, but they want to make sure that it's now being spent on doing more interesting things.

So on the ground I think essentially it's still humans. On the ground it's still statisticians and data scientists. And so the people who self-selected into saying my craft as a statistical data scientist now what are the tools that I use to increase and continue to be good at my craft? I think that's what the conversations are going to be on the ground level.

Donna LaLonde: Thanks for that, Glenn. I'll just say that as a person who's attended a number of SDSS in the past, what stands out for me is it's a really nice place to find your people. If you're not quite sure who your people are, it's the right size in terms of being able to navigate and also being maybe able to be a little bit vulnerable about, well, I'm not sure I know that, but I'd love to learn it right

Greg Wright Colopy: on that point. The conference is an extremely personable group. Like, as far as personality goes, this is a good fit. There are people who I'm still talking to today. Where we met was it in Alexandria in 2018 when we had the conference grad students, Kristin Morgan at UConn. I met her via David Banks, who knew her mother. And then I met Kristen. And I was at a conference up in UConn Pharma DS a few years ago, and the first thing I did was I just popped Kristen email saying, hey, I'm in town. And she had not heard from me in a year. And we were having dinner with Junyan at UConn as well the next day.

And so, effectively, this is a very good conference to meet people. And one other thing is you'll start seeing them at one conference and you'll meet them at the next one, too, sometimes in a different country. I think the next time I saw Kristen after that was in Berlin. Again, this is a good place to meet people, go forth and make friends.

Donna LaLonde: So interesting getting both your perspectives on this. But, Sarah, I'll start with you. Since you've been involved with picking the tracks, and then also now as the program chair, I imagine that there are some folks that out there looking at that refereed abstract deadline of November 30 and thinking, oh, my work's just not right for SDSS. Where I suspect if they asked you, you'd say, yes, it is. It actually is. So help convince people that their work is actually right for SDSS.

Sarah Kalicin: Yeah, great question. Because, you know, the last few years going through and refereeing a lot of the submissions, I have seen that there are some really great fits, and then there's

some that haven't been necessarily a great fit. If your work is purely theoretical statistics in nature and you can't showcase how your work is bridging over to data science and AI applications, you're not a good fit for the conference. The other thing too is I've seen some really great abstracts out there. You know, I would sit back and go, how is this work really positively impacting society or the future of AI or data science work? And it wasn't really articulated very well. There's a lot of abstracts that get submitted and we have to do a cut. I think I just did the analysis where we only accepted 65% of the abstracts that were actually submitted. So if you have work that's either data science and AI applications and you can articulate the impact back to society or the future of either the field of statistics and or data science and AI. Then we really want you to submit your abstract and make sure that you have an impact statement like who is going to be impacted by this or how this is going to be impacted. Don't let the referee try to imply what you're trying to do. Be really articulate about it. You know, not just for

abstracts, but also for work and your resumes and stuff. Having impact statements of how you've made impact really sets you apart and it doesn't allow people to guess what that is, be explicit about it. Impact statements. Glenn, any thoughts on that you want to add?

Greg Wright Colopy: Yeah. Basically the way that you get your abstract or your proposal to stand out to reviewers, at least reviewers like me, is the same way that you try to stand out in life, which is just basically have a clearly articulated reasoning behind what you are doing and clearly articulate your critical thinking on that. This is something that pretty much the majority of like machine learning data science folks leave out when they're doing this.

You want to totally blend in, do this, make it a battle of the algorithms type paper where you basically say, you know, here's the application issue and we had old fancy model, which by the way was totally cool last year, and now here's new fancy model and oh, and I whack them together and we see which one wins. Like, you know, basically like when my son plays with Legos and he just whacks them together and then flings one and one's the victor and one isn't.

Oh, I don't think anyone who does journal review is surprised by this, where basically it's okay. Literally just battled the algorithms where people saying, I compared method one to method two. One of these is better, the end use method one.

But the point is, what they won't say is like, okay, what was the reasoning behind why one of these methods works better than the other? What is it fundamental in the nature of the data that one of these methods is more suited for the data, for the goal of using that data compared to the other?

And then how did I exploit that? You know, those are the types of things you will stand out if you can clearly articulate why you did what you did, why the algorithms and the data align, what the value is in the end is Sarah saying the impact of that, congratulations, you've won the game as far as making a submission.

And so I would just say put that in there and maybe just Quite selfishly, if you don't think that that fits for the traditional data science application, throw it in the AI one. If I see a good, scientific, well-reasoned paper, I don't care if it has to do with LLMs, I will grab that from my track because that's what I want in the track that I'm in charge of. So I'll just throw that one out there.

Sarah Kalicin: And if I can just add to something that Glenn said is if you do not know what the best track is, just do your best guess on that.

A lot of times when we go through and review abstracts, there's a good handful of them where the reviewers will say, this should be in this other track. It doesn't disqualify you. It's just like, hey, this is

a really great paper, but it's better in this track. So we will then decide to put it into another track and we'll let you know that congratulations, you got accepted, but we're putting you in this other track because we think it fits better.

Ron Wasserstein: So, Glenn and Sarah, thank you for spelling that out and for encouraging people to submit their work to the SDSS. Before I move on, I can't resist going back to a comment you made, Glenn, about people that you met at the conference in 2018 that you talked to then and you're still talking to now. I have sort of the reverse of that. There were people then that weren't talking to me, that still aren't talking to me.

So, Sarah, one of the cool things that presenters can do is to submit their work to the Journal of Data Science. And so how does that fit into this picture that you've been weaving about the conference and about the things that you want to see in there, especially for those people for whom a publication after the conference would be valuable.

Sarah Kalicin: Yeah, so this is really exciting. So back in 2022, the Journal of Data Science has partnered with SDSS, where the first special publication was published in April of 2023. So basically the April issue of the Journal of Data Science is just dedicated to presentations from the previous year's SDSS. And this, it's a really great way to be able to showcase your work and be recognized by your peers. And not just at a conference, but then more broadly in a journal.

So this is another incentive of why to submit a refereed abstract, because you have a chance to also appear in a dedicated journal.

With that, on our committee, we have a Journal of Data Science editor and they do want me to say that we are looking for guest editors, especially those that can review computer science and data science research. So if you're interested in participating in this year and have experience in computer science and data science research, please reach out to me and I'll get you contacted with Mani and June so that you can participate. Because we are looking for folks that bridge that gap and that can do reviews of the journal articles.

Donna LaLonde: And Sarah, I would just say come to SDSS in April and meet Moni and June in person and Sarah and Glenn and all the other wonderful people and we'll get you connected in all sorts of ways. Right?

Sarah Kalicin: Yeah, definitely come to SDSS because then you can also see the papers in person and be able to talk to folks as well.

Donna LaLonde: So I know both of you are very generous with your time and mentoring supporting early career folks. So I wanted to ask you a question about what would you tell someone who is an early career professional who's wondering, is my work good enough?

So, Glenn, I'll start with you for this one. What would you say?

Greg Wright Colopy: Typically, I find early career people's work often better and better explained and better articulated than many later career people. And so I would strongly encourage that. In fact, one of the really nice things we had with the continuing education courses was that a grad student had the guts to make a submission.

And I thought it was really good. And just for what it's worth, I put a comment there saying I want more grad students, for example, offering continuing education courses. And the reason for that is because one of the best teachers that I had when I was in school was a grad student and I took the exact same course with an extremely tenured professor and I didn't learn anything. And the reason

for that, I believe, is that frequently the earlier people are in their career, the closer they are to the student, they're still learning. They remembered how they got on board in the first place and they remember the aha moments and the key intellectual dots that they had to connect in order to understand their craft.

And what's nice about that is that transfers really nicely into presentations because those are the key points that they will walk you along when they present. And so that's quite useful. You know, back to what was said before, you know, clearly articulated reasoning. And so if you are early career, early professional, even a student, I don't care undergrad, there's a track for you. Submit to it, clearly articulate what you're doing. If you follow those rules, you will be ahead.

And I'll just say this from my perspective; I'm not looking for the word doctor in front of the name of a person who is submitting. If I name the top three data scientists I know professionally, like actual professional working machine data scientists, of which I'd like to of course include myself as number one. But the other two behind me do not have doctorates. They are extremely good and I will guard their names because I want to always have access to them in their work.

But the point is, at least as far as my track is concerned, we are not looking for doctors in front of the names. We're not looking for professor or something or other. A student who clearly articulates what they're up to, the value of it, and again, the connection between their quantitative work and the domain will have a special place in my heart. In fact, I would give them the edge. I would emphatically impress upon students. Come here. People are very eager to meet you. People love talking to students because they are the most exciting part of our field.

Sarah Kalicin: Awesome. Thanks, Glenn. As a past student, many, many, many, many years ago, conferences were a great way for me to do a lot of networking and mentoring and so forth. And one of the things about SDSS is we do have a mentoring program that you can get connected, could be a student, you could be early career, you could be also in an advanced career and we'll have mentoring for you because I believe anytime, anywhere in your career, you can benefit from mentoring, sponsorship and so forth.

And it's really all about networking and meeting new people. And as for is my work good enough? I'm going to say yes. If your work bridges the field of statistics and data science and AI and you show impact, you are invited to submit your work. And just because you may be an undergrad or graduate student, we have three levels of abstracts that you can specify what level of abstract it is. Beginner is for undergrads and graduate students, Intermediate is guideline for early career, more advanced grad students. And then we also have advanced abstracts for more cutting edge applications.

And we make sure that we have a balance of these presentations throughout so that we're hitting all these different levels and really showcasing the breadth of what people are doing in those areas.

And even if your abstract isn't accepted, submit it to a lightning session. You still get to present for like, I believe, five minutes and then get to go into this large conference room and talk about your research much more in depth and get to meet people. And in fact, lightning sessions are actually really fun because you don't know who's actually coming up and talking to you. And in fact, one of the most memorable lightning sessions that I saw was this undergrad talking about, you know, rock climbing training.

I'm a rock climber and I was just geeking out over. You're like, this is cool. I've thought about doing research like this, but I didn't even know how to even start this. And this undergrad person had a great presentation of showcasing the basis of how to even start going around and training for rock climbing. So it was a great application and something that I actually really enjoyed. So submit it. If

you don't get accepted, then go for a lightning round, or if you don't actually know, just show up and go to the conference and then seek what other people in your level are presenting and maybe get inspired to submit something. Some of your work.

Greg Wright Colopy: And just to double down on, Sarah said both the lightning rounds and just the act of submitting your abstract are both stepping stones to progress. So basically people think, oh, I submitted something. It didn't get in, it didn't get accepted. That means three people who must of course be genius experts voted against me and did not want to see me there.

That's not quite right. You might have had two people who freaking loved what you did or one person who, like, was Sarah with the rock climbers? Like, oh, that's really cool. So essentially you're putting out your feelers there. You might fascinate one of the reviewers who now knows your name and keeps an eye on you. Quite similarly to, for example, the continuing education courses and the student who submitted his abstract. I know that name now. It's the grad student who had the guts to submit what I thought was a really good proposal. And I bet he would have done a great job.

Lightning rounds are the way for students to get noticed because basically you go in, you present a clean elevator pitch of what you're doing. You work at it, you don't take it lightly. You go and be like, look, I'm going to give a lightning session that's going to be better than the keynote. That's the way you approach it. And people will notice you. And if you do a good job, they will want you back next year. Take a shot on yourself.

Donna LaLonde: And if you come to Pittsburgh, Sarah and I will help find a climbing gym and find a rock wall, right?

Sarah Kalicin: Oh, yeah, we will definitely do that.

Ron Wasserstein: So thanks for those very helpful and encouraging responses. You've given a lot of clarity about what you're looking for in the conference. You've made it really clear that there's a tremendous depth and variety of opportunities for people to participate. So now I'm going to ask you for a sentence. Okay, so a sentence of encouragement for someone on the fence about submitting or attending. What would that sentence be? Glenn, let me call for your sentence first.

Greg Wright Colopy: That's tough because as you know, I'm very verbose and confining me to a single sentence is not my style. The thing that came to my mind was, first of all, why not try put yourself forward? You have nothing to lose. You have only things to gain from that.

Ron Wasserstein: And why not Is a perfectly good sentence for this, I would say. [00:28:52] Donna LaLonde: I think that's an exceptional sentence.

Ron Wasserstein: Yeah. And I don't even know how easy it will be for you to add to that, Sarah, but what do you have in mind?

Sarah Kalicin: Actually, I do. Though I don't know if I can get it into one sentence. So this year, this is the reason why you need to come to SDSS.

And we all know that the field of statistics, data science and AI is moving so fast and that education and building our skills are so important.

My sentence is: The continuing education sessions are free and included in registration.

So that is my one sentence. Come for at least attend one continuing education session because. And that's usually $300, so. But it's included in your registration.

Ron Wasserstein: Free is indeed my favorite four-letter word in the English language. So thanks for that.

Donna LaLonde: Well, Sarah and Glenn, this has just been a fantastic conversation and I can't imagine that anyone is actually still on the fence about attending SDSS or submitting an abstract, which is great.

So we have some traditions on Practical Significance, and two of those traditions are how we like to end the podcast. And the first is we love to ask our guests about what they're reading, listening to, watching all of the above. So what's on your TBR, your playlist, your YouTube video to be watched? And Sarah, I'll start with you.

Sarah Kalicin: So I am like over my head on search theory and lost person behavior, along with journal articles from my fellow search and rescue colleagues. So one of the big challenges in search and rescue is making data actionable in real time.

So I'm looking at how data can be used, how do we organize it, how we can analyze really complex data and making sure that we have the infrastructure to make it user friendly in real time. So those are the things that I am reading. So like urban searches, intelligent searches, lost person behavior, and all the statistics behind it. I'm just geeking out on all that right now.

Donna LaLonde: That's very exciting. So Glenn, what about you? TBR

Greg Wright Colopy: yeah, absolutely nothing to do with data science or any of that. I'm actually right now listening to Njáls saga. It's one of the Icelandic sagas. It's long. People are scrapping, they're getting to feuds and sometimes they work it out.

Donna LaLonde: That's great. Again, thank you both very much for joining us. And our other tradition for ending is to have Ron's top 10. So I will turn it over to my colleague Ron for the top 10.

Ron Wasserstein: Thank you, Donna. Well, I can't seem to log into anything without being asked to prove I'm not a robot. Is that a stop sign way in the back of the picture? Does that sliver of a traffic light in the upper left corner count? I guess we just have to deal with the amusing, annoying reality that machines are asking us to prove we are not machines. As always, the Practical Significance podcast wants to help. Here are the top 10 ways for statisticians to prove they are not robots.

10. Select all squares containing outliers.

9. Guess the correlation for this scatter plot.

8. Select the sequences that converge almost surely.

7. Which of these graphs have been overfit?

6. Which of these characters are parameters and which are estimate

5. Impute the missing value and justify the mechanism used.

4. Select all the pie charts that should have been bar charts.

3. Explain the implications of failing to reject the claim I am not a robot.

#2. Spot the p-hacked result in these six abstracts.

And the #1 way for statisticians to prove they are not robots. Check a box saying I actually listened to Wasserstein's ridiculous Top 10 Lists.

Well, that's it for this episode of Practical Significance. We look forward to continuing the conversation next month.

Speaker B: thank you for listening to this edition of Practical Significance, the podcast of the American Statistical Association.

A new episode will be coming your way next month from Amstat News, the ASA's monthly membership magazine.