2025 Statistics Club Case Competition

Introduction

Thanks for your interest in joining the 2025 Statistics Club Case Competition! Fill out this survey to sign up!

Important Dates

- **March 6th:** The Case Competition Kick-off will be at 5pm in WVB room 1159. This is where questions can be asked, tips will be shared, and team members can be found. Additionally, the case prompts and rules will be introduced to teams. Treats will be provided!
- **March 14th:** Sign ups for new teams closes. All teams will be finished being assigned by this time.
- **March 27th:** Presentations to judges for the case competition will be held in the WVB at a time TBD. Food provided! The winning teams will also be announced at the end of the activity.

Registration Rules

Groups are 3–4 people (hard cap at 4). You can make your own group and report it in this survey or be assigned to a group. If you only have 2 people in your group, sign up anyway and you'll both be added to part of another group. Only one person per group needs to sign up.

There are 2 competition tiers, one beginning/intermediate and one intermediate/advanced. To preserve fairness and competitiveness in the first competition tier, any group with a member who is currently enrolled in or has taken an advanced Stats Class (Stat 348, 386, 486, or 469, CS 472, or MATH 404/580) is disqualified from winning Tier I (the lower tier). This is not an exhaustive list of disqualifying classes outside of the stats program, and some high level coding or modeling classes in other majors might also disqualify you. If you have any concerns about other such classes, contact Dr. Richardson at `richardson@stat.byu.edu`. Anyone is welcome to join Tier II, no matter their experience level.

Graduate students are disqualified from participating since they are too cool and know too much.

Expectations

You will be given a scenario and, with your team, create the best solution to the problem at hand. You will then prepare an 8-minute presentation with your findings for the judges. More details on judging criteria and presentation details will be announced when we release the scenario.

Data and Methods

There will be data provided for you that will be sufficient to answer the question, but there are no limitations as to what data you can use. If you find other public data relevant to the problem you are trying to answer, you can include it in your analysis and your conclusions. You are not required to use any specific method. An advanced machine learning method will be weighted as equally as a useful figure as long as it is used correctly, well justified, and effectively answers the problem you are given. However, it is mandatory to include modeling in your results.

Presentation

Each individual in the team must contribute to the presentation. Presentation skills will be evaluated as part of the competition. A presentation will consist of a maximum of 5 slides and will take a maximum of 8 minutes with a 2-minute question and answer session. One possible (but not required) way to write 5 slides is:

Slide 1: States the problem you have been given.

Slide 2: Describes the data you use and its appropriateness for answering the questions.

Slide 3: Describes the modeling and/or summarizing efforts that were taken, along with useful figures or relevant model results.

Slide 4: Continues describing the modeling and/or summarizing efforts, along with useful figures or relevant model results.

Slide 5: Includes a summary and conclusions that directly relate to the problem you have been given.

Presentations will be judged on slide quality, presentation quality, data handling, methodology, and conclusions made.

Prizes

The winning teams will win $150 per person!

Contact Information

If you have any questions about registration or concerns regarding the competition, please contact:

- Chloe Walcott (`walcottc@student.byu.edu`)
- Dr. Richardson (`richardson@stat.byu.edu`)

Good luck!

Prior Judge's Comments

Last year, I asked the judges to write down any comments they had about the presentations they heard. For context, the prompt last year was how weather and phases of the moon affect crime. This is what they shared:

> "For presentations, ensure that the goal of the project is clearly stated in the slides and that the conclusion includes suggestions of actions to take. Make sure to use plots to display data and not code snippets. Ensure that the images are easily visible. Explore the data well and don't be afraid to try different approaches. There was a clear difference in the students who spent more time and effort on their projects."

> "A good presentation should start with the conclusion and then justify it. To really grab the audience, slide 1 should be: 'In order to reduce crime, we found the most promising ways are A, B and C and as I go on I'm going to show you why.' Then unpack A, B, and C, touching

upon the data and modeling evidence that support it. But the police chief doesn't need a full recounting of every modeling step.

At least in my room, all the presentations were more of a story about what they did in the order they did it. E.g., first we got the data, here's the ways we cleaned it up and prepped it, then we fit some models, here's the models we fit and what their diagnostics and p-values were, then most just stopped there.

It seemed the general idea in the students' minds of what counts as a 'result' is a p-value, or an MSE score, or some other model diagnostic. That might be true for a homework assignment, but if you're trying to present to a stakeholder, you need to speak to things they can do (e.g., put more officers in ABC location, or have more during XYZ holiday/event/weather) vs. the unactionable p-value for weather was 0.04."

"I know this was a stats event but I think the focus should be more on the recommendations/conclusions than the methodology. The two things I'd wish I'd seen more of were: (1) I wish they dug deeper and looked at things from different angles (broken down by location, trends over time, demographics, etc.). And (2) I wish they had brought in more 'industry knowledge' like sources that backed up their conclusions or at least acknowledged general thoughts within criminal justice. For example, something to the effect of 'we found this trend x was different than what some experts believe, but this is why we disagree.' "

"I generally found their analysis to be good, but it feels like they would benefit from a training on how to Storytell with Data. For example, just how to structure a slide and a graph would have gone a long way. I'm happy to do this (I already teach an MBA class on this). I saw lots of screenshots from different tools, or graphs that were created in Excel that were extremely difficult to read."

"All teams talked about the model, without looking or showing the data and doing some basic thinking about the data. Additionally, their conclusions were not stated like a statistician should state; i.e., there was not significant evidence to show that this was a factor (based on the model that we used). But overall, the presentations were very professional."

"For one, I feel like most of their figures (and/or labels on their figures) were generally too small, sometimes from trying to fit too much into one page. Figures are supposed to be simple and useful, not flashy, complicated, and just for show. More thought and care should be put into the slides themselves. There were multiple times where there was just a screenshot of R console output—which is pretty unacceptable in a professional setting. Labels such as 'sin_cos_trans_day' should be put in human-readable form for the presentation. Remember the audience and purpose.

Another thing is I feel like more should be focused on what to do with the data they are given, prior to fitting any models. All the teams I judged focused far too little on what the variables actually were, and what was the consequence of fitting them at face value. I got a lot of 'we fit a linear model'... on... everything? And they ended up with these giant overfit models, which hardly got any cross validation. The best teams dove into things like which holidays were important. Explained which trends were found in the data, instead of talking about 'my model predicted this well.' Full moon doesn't matter."

"I'd say one area that all students would benefit from is telling better stories with their data. This includes how they build their charts, their color selection to show where the story is, etc. Here's a short article I wrote on it you could share with students: `https://unwrittenbusinessguide.substack.com/p/5-ways-to-tell-better-stories-with?utm_source=publication-search`

I fully realize these are undergrad students doing their first case competition, so I'm just sharing this in an effort to help them really wow the judges, do well in their first internships, etc."