MATH 201: Lecture 4a Handout

Section 3.5 Continuous Distributions

Name:_____ Date:_____

Learning goals for today

By the end of this activity, you should be able to:

- Calculate numerical summary statistics (mean, median, standard deviation) and plots using technology.

Biased AI detection tools data

In this activity, we will analyze a real dataset from a research project studying AI writing detection tools. As AI-generated text becomes more common, several companies and researchers have developed "detectors" that attempt to classify whether a piece of writing was produced by a human or by an AI model. These detectors output a probability score indicating how likely the text is to have been written by AI. The study can be found at: https://arxiv.org/abs/2304.02819 and a description of the dataset can be found at: https://github.com/rfordatascience/tidytuesday/blob/main/data/2023/2023-07-18/readme.md.

Each row in our dataset represents one essay that was evaluated by a detector. The variables we will use today are:

- kind – Whether the essay was actually written by a Human or by AI.
- .pred_AI – The detector's predicted probability (between 0 and 1) that the essay was written by AI.
- .pred_class — the detector's class decision (AI or Human) using a 0.5 cutoff.
- detector – The name of the AI detection tool used to generate the prediction.

For example:

- A value of .pred_AI = 0.95 means the detector believes there is a 95% chance the essay was written by AI.
- A value of .pred_AI = 0.10 means the detector believes there is only a 10% chance the essay was written by AI.

Step 0: Load the Data into CODAP

1. Go to: https://codap.concord.org/
2. Click **Launch CODAP** and **Create new document**.
3. In the top right click the three horizontal bars (the menu), **import**, and **url**.
4. Paste the dataset URL: https://raw.githubusercontent.com/rfordatascience/tidytuesday/main/data/2023/2023-07-18/detectors.csv
5. Click **Import** and tell it to import all rows.

What do you notice in the case table (number of rows? columns?)?

Part 1: Numerical Summary of AI predictions

1. CLick `Graph` and either drag the `.pred AI` column onto the empty plot or click the x-axis and choose the variable.
2. In the right panel of the plot click the fourth option (configuration) and select both "Group into Bins" and check "Fuse Dots into Bars". You now have a histrogram!
3. You can play around with different bin widths (try a few to see the difference). What appears to be a good bin with to visualize the data?

4. Describe the distribution of prediction probabilities.

5. In the right panel of the plot click the first option (measure) and select "show measure labels" and both "mean" and "median" under "measures of center". Record the values here.

6. Click "standard deviation" under "measures of spread". Record the value here:

7. Click "Box plot and normal curve" and select "box plot" to overlay a boxplot. (You can uncheck the measure boxes now to unclutter your plot). Describe some features of your boxplot.

8. Now drag the `kind` column header from the table onto the plot. Drop it on the right vertical axis (opposite the y-axis) and you should end up with two separate histograms/box plots, one for human created documents, and one for AI created documents. Describe differences between these plots.

9. Now drag the "native" header onto the top of the plot and drop it as the side-by-side layout by native. Describe each of these plots (You may need to readjust the bin width). Do you think the data suggest bias for non native speakers? Explain your answer either way

10. Keep exploring the data with CODAP!