ST 314D Data Analysis 01

Topics:
- Discrete Random Variables, Probability Mass Functions, Expectations.
- Lessons Covered: 1- 7.

Grading:
- Points are listed next to each question and should total 25 points overall.
- Grading will be based on the content of the data analysis as well as the overall appearance of the document.
- Late assignments will not be graded.

Instructions:
- Clearly label and **type answers** to the questions on the proceeding pages in Word, Google Docs, or other word processing software.
- Insert **diagrams or plots as a picture** in an appropriate location.
- Math Formulas need to be typed with Math Type, LaTeX, or clearly using key board symbols such as +, -. *, /, sqrt() and ^
- Submit assignment to the Gradescope link as a PDF. Indicate the pages to the individual questions and also verify the correct document has been uploaded. Failing to follow this direction may result in point deductions.
    - If you are unable to answer any questions before the submission deadline, please still match the question location to the correct page when submitting in Gradescope.

Allowances:
- You may use any resources listed or posted on the Canvas page for the course.
- You are encouraged to discuss the problems with other students, the instructor and TAs, however, all work must be your own words. Duplicate wording will be considered plagiarism.

This assignment will require you to do a bit of coding in RStudio. If you have not already worked through the R Tutorial page in the Start Here module on Canvas, take a break from this assignment and go work through that tutorial.

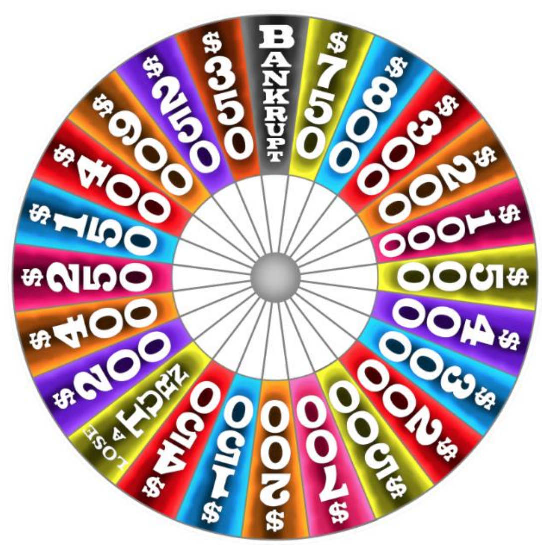
Part 1 (6 points)
For each random variable, state whether the random variable should be modeled with a Binomial distribution or a Poisson distribution. *Explain* your reasoning. State the parameter values that describe the distribution and give the probability mass function.

Random Variable 1 (3 points)
A software development team is concerned with the performance of their code. Historically, 3% of code modules exhibit performance issues. A development team samples 25 code modules from a project. Assuming the chance of performance issues is independent between modules, what type of distribution could be used to model the number of successful code modules from the sample of 25? Hint: the event of interest here is a code module performing well (i.e., not exhibiting performance issues).

This random variable should be modeled with binomial distribution. This is because we have a fixed number of trials (25), and the probability of success is the same throughout the trials.

Random Variable 2 (3 points) A brand of smartphone has a warranty period of 2 years. During this time frame, there is no limit on the number of warranty claims that can be made. Historically, the average number of claims per smartphone during the 2-year period is 1.2 claims. What type of distribution could be used to model the number of warranty claims per smartphone? Hint: the event of interest here is the number of warranty claims made in a 2-year period.

This random variable should be modeled by Poisson distribution. This is because we are modeling the number of events that happen in a fixed interval of time (2 years).



Part 2: (11 points)
Wheel of Fortune is a popular game show on Television. Contestants spin a wheel and try to guess a correct letter from a word puzzle. If they guess correctly, they earn the dollar amount from the wheel. If they spin "bankrupt" or "lose a turn" they get nothing and can't play. To the right is an example of the wheel. Watch this video to see an example of someone spinning the wheel.
https://www.youtube.com/watch?v=_Pv33JWBdY8

The outcome of a spin on the wheel is a discrete random variable. Consider X the dollar amount spun on the wheel, where Bankrupt and Lose a Turn = $0. There are 24 wedges on the wheel.

a. (2 points) The table below displays the values the random variable *X* can take on, along with the number of times each value appears on the wheel (labeled Count). Fill in the last row of the table with the correct probability mass function values. That is, for each value *X* can take on, determine the probability the wheel is spun and lands on the given value, *p(x)*. Round values to three decimal places.

x	$0	$150	$200	$250	$300	$350	$400	$450	$500	$700	$750	$800	$900	$1000
Count	2	2	4	2	2	1	3	1	2	1	1	1	1	1
p(*x*)	.083	.083	.167	.083	.083	.042	.125	.042	.083	.042	.042	.042	.042	.042

b. (1 point) What is the most likely dollar amount the spin will land on?

The most likely dollar amount to land on is $200 with a probability of 0.167.

c. (2 points) What is the average dollar amount? Show work!

The average dollar amount can be calculated with the equation ∑(x * p(x)). Using this equation, we get an average dollar amount of $348.9.

d. (3 points) Suppose a contestant spins the wheel three times, how likely is it they spin $500 each time? Show work!

To calculate the probability of spinning $500 3 times, we need to cube the probability that you will get 500 which is .042 . If we cube that, we get .000074.

e. (3 points) Suppose a contestant spins the wheel three times, how likely is it they spin $500 at least one time? Show work!

To calculate how likely it is they $500 at least one time we can we can take the probability of not spinning 3 times and subtract that from 1. To calculate that probability. We do $(1 – 0.042)^3 = 0.879$. We then subtract that from 1 which gives us our answer of 0.120 .

Part 3: (6 points)
The PMF in Part 2 is based on probability theory. Do these probabilities stand up when a contestant actually spins the wheel? Go back to the Data Analysis #1 instructions page on Canvas, download the R script titled: Wheel_of_Fortune_Spin_Script.R , open the file it will automatically open in R. You need R software on the computer to open the script window. Follow the instructions in the code then answer the following:

a. (1 point) What value did you spin?

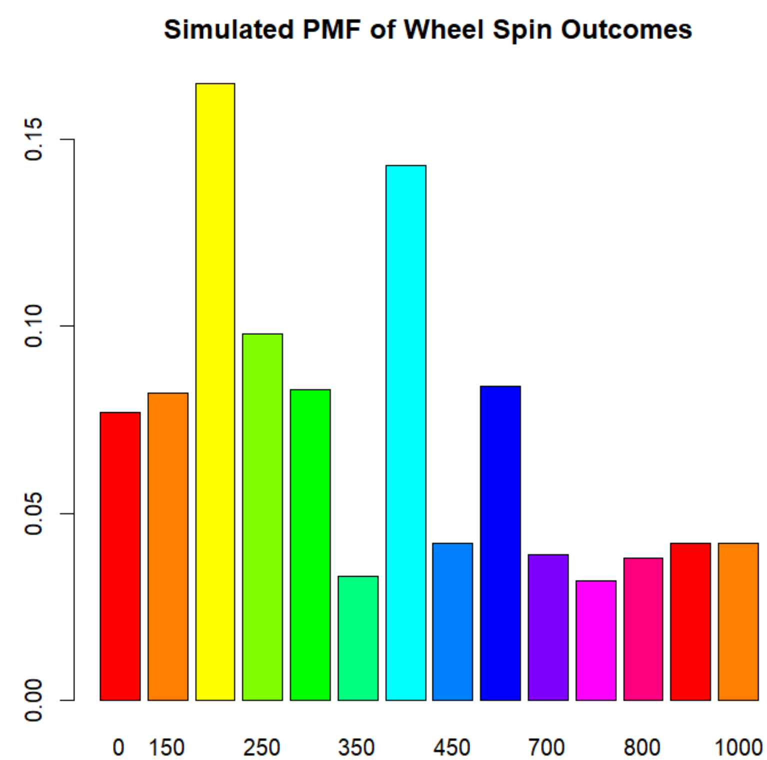
I spun 500.

b. (1 point) What is the average of the 1000 simulated spins? How does this value compare to your answer in part 2c?

I got 385.85 as my average when I spun, which is $40 off from my answer in part 2c.

c. (1 point) Include the simulated probability mass function AND the plot of the probability mass function from R. For the probability mass function, you should include a table with the all possible outcomes and the proportion of times each outcome was spun (output from line 45 of the code).

x	$0	$150	$200	$250	$300	$350	$400	$450	$500	$700	$750	$800	$900	$1000
Prob	.077	.082	.165	.098	.083	.033	/143	.042	.084	.039	.032	.038	.042	.042



d. (1 point) How do the simulated probabilities compare to the theoretical probabilities in part 2?

The simulated probabilities are very close to the theoretical probabilities as expected. We saw that the probabilities of getting 200 and 400 were the highest in the graph and that was as we had calculated.

e. (1 point) Based on the plot is the most likely outcome the same as it is in part 2a?

Yes, $200 was the most likely outcome in the simulation which is the same as the theoretical in part 2a.

f. (1 point) In general, what action will make the simulated values more like the theoretical ones?

The more spins we do, the more accurate the simulate values will be to the theoretical ones.

Gradescope Page Matching (2 points)
When you upload your PDF file to Gradescope, you will need to match each question on this assignment to the correct pages. Video instructions for doing this are available in the Start Here module on Canvas on the page "Submitting Assignments in Gradescope". Failure to follow these instructions will result in a 2-point deduction on your assignment grade. Match this page to outline item "Gradescope Page Matching".