ST 314 Data Analysis 01

Topics:
- Discrete Random Variables, Probability Mass Functions, Expectations.
- Lessons Covered: 1- 7.
- Textbook Chapters (Optional): 2 and 3.

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.

Deadlines:
- Final Submission: **Monday, October 5th**. All submissions must be PDF files.

Instructions:
- Clearly label and **type answers** to the questions on the proceeding pages, ***without*** question prompts, 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 Canvas link as a PDF. Verify the correct document has been uploaded. If not, resubmit. You can submit up to three times.

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.
- Outside resources need to be cited. Websites such as Chegg, CourseHero, Koofers, etc. are discouraged, but if used need to be cited and used within the boundaries of academic honesty.

ST 314 Data Analysis 01

Part 1. (8 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. A quality measurement for cabinet manufacturers is whether a drawer slides open and shut easily. Historically, 2% of drawers fail the easy slide test. A manufacturer samples 10 drawers from a batch. Assuming the chance of failure is independent between drawers, what type of distribution could be used to model the number of failed drawers from the sample of 10?

Binomial. This is because we are modeling something that has a binary outcome (success of drawer slide, failure of drawer slide), and we are just modeling it over a set of (in this case) 10 trials. Therefore, we use Binomial, as Poisson is for something that has a binary outcome measured over time.

Parameter Values

p(x) = 0.02
n = 10

Probability Mass Function

p(x) = (C10,x) * ((0.02)^x) * (0.98)^(10-x)

Random Variable 2. The warranty for a particular system on a new car is 2 years. During which there is no limit to the number of warranty claims per car. Historically, the average number of claims per car during the period is 0.8 claims. What type of distribution could be used to model the number of warranty claims per car?

Poisson, because we are measuring something with a binary outcome over a period of time or space (in this case time).

Parameter Values

λ = 0.8

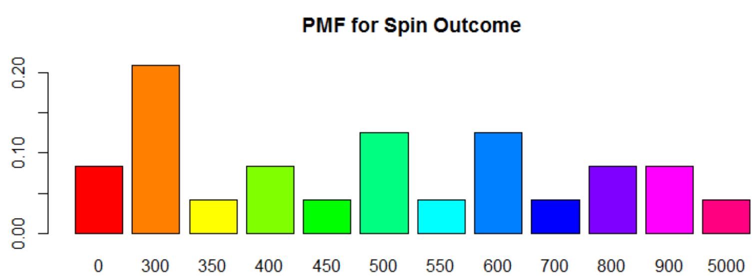
Probability Mass Function

p(x) = ((0.8^x) * (e^-0.8)) / (x!)

ST 314 Data Analysis 01

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, and Free Play = $500. There are 24 wedges on the wheel.



The following table provides the probability mass function for X. Round values to 3 decimal places.

x	$0	$300	$350	$400	$450	$500	$550	$600	$700	$800	$900	$5,000
Count	2	5	1	2	1	3	1	3	1	2	2	1
P(x)	0.083	0.208	0.042	0.083	0.042	0.125	0.042	0.125	0.042	0.083	0.083	0.042

a. (1 point) What is the most likely dollar amount?

$300

b. (1 point) How likely is it to spin Lose a turn or Bankrupt?

“where Bankrupt and Lose a Turn = $0”
0.083 = 8.3%

c. (3 points) What is the average dollar amount? Show work!

668.75

Mean of all runs = (2 × 0 + 5 × 300 + 1 × 350 + 2 × 400 + 1 × 450 + 3 × 500 + 1 × 550 + 3 × 600 + 1 × 700 + 2 × 800 + 2 × 900 + 1 × 5000) / (2+5+1+2+1+3+1+3+1+2+2+1)

d. (3 points) Suppose a contestant spins the wheel three times, how likely is it they spin $0 each time? Show work!

0.083 per spin is the likelihood that the player scores $0.

Therefore the odds that they spin $0 all three times out of their three spins is:

0.083 * 0.083 * 0.083 = 0.000571

When reading this question over again, I find I can also interpret the question as “what is the likelihood that each individual spin will result in $0”, in which case:

0.083 for all three

e. (3 points) Suppose a contestant spins the wheel three times, how likely is it they spin $0 at least one time? Show work!

The probability of spinning >=1 * $0 is the same as the complement (inverse) of not spinning a single zero over three spins.

The probability of not spinning a $0 over one spin is calculated as the odds that we spin anything but a $0:

P_1 = (.208+.042+.083+.042+.125+.042+.125+.042+.083+.083+.042)

We make three spins, then take the complement:

1-(P_1^3) = 1−((.208+.042+.083+.042+.125+.042+.125+.042+.083+.083+.042)^3)

= 0.229

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?

600

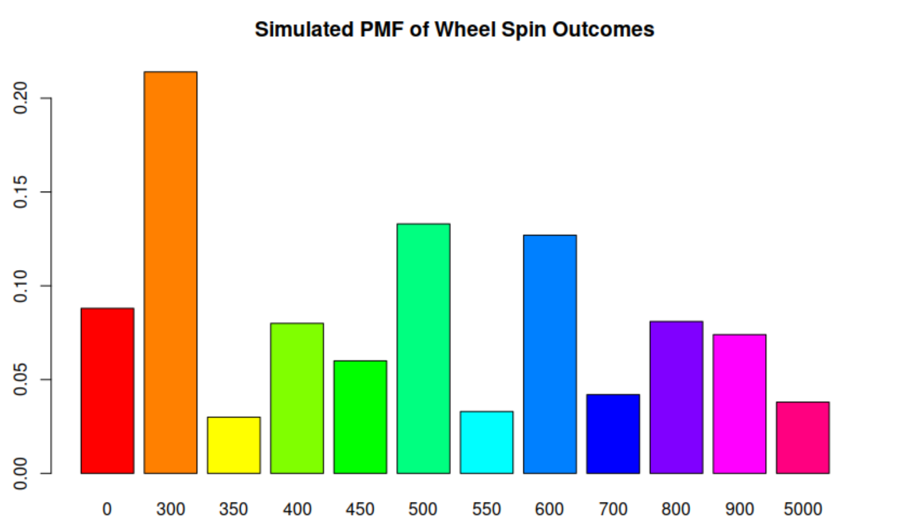
b. (1 point) What is the average of the 1000 simulated spins? How different is this from part 2c?

645.35

This was quite a bit lower than in part 2c. I re-ran with one billion spins, and came up with a much closer result.

c. (1 point) Paste the probability mass function and the plot of the probability mass function from R.

x	$0	$300	$350	$400	$450	$500	$550	$600	$700	$800	$900	$5,000
P_R(x)	0.088	0.214	0.03	0.08	0.06	0.133	0.033	0.127	0.042	0.081	0.074	0.038



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

x	$0	$300	$350	$400	$450	$500	$550	$600	$700	$800	$900	$5,000
Count	2	5	1	2	1	3	1	3	1	2	2	1
P_2a(x)	0.083	0.208	0.042	0.083	0.042	0.125	0.042	0.125	0.042	0.083	0.083	0.042
P_R(x)	0.088	0.214	0.03	0.08	0.06	0.133	0.033	0.127	0.042	0.081	0.074	0.038
P_R(x)-P_2a(x)	0.005	0.006	-0.012	-0.003	0.018	0.008	-0.009	0.002	0	-0.002	-0.009	-0.004

The probability mass function values were all within +/-0.02 of the theoretical probabilities.

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

Yes, the most likely outcome is the same on the plot, at $300.

This is also programmed into the R code, so it is highly unlikely that over 1000 runs, the data does not match roughly with the supplied data:

```
Wheel.Outcomes = c(
     rep(0,2),  rep(300,5),
     rep(350,1),rep(400,2),
     rep(450,1),rep(500,3),
     rep(550,1),rep(600,3),
     rep(700,1),rep(800,2),
     rep(900,2),rep(5000,1))
```

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

In the case where the theoretical values are programmed into R, the action that will make simulated results approach a perfect match of theoretical results is more runs.

In the case where the theoretical results are not programmed and are potentially incorrect, we cannot say what action will approach experimental and theoretical, as we do not know that experimental = theoretical at runs=infinity.