Codes: NPC: no partial credit given

Problem 1 This short-answer section will ask basic questions about probability.

(a) [3 pt] What is the definition of probability? (you may use any of those that we discussed in class, but please explain to the best of your ability)

Prob is long term frequency

$P(A) = \lim_{n \to \infty} \frac{\#A}{n}$

1pt for something reasonable that was not one of the P def's

(b) [2 pt] Below is Boole's inequality (AKA Bonferroni's inequality):

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$$

Express this inequality for $n = 2$ *i.e.* for the events A_1 and A_2.

$P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$

(c) [2 pt] Under what condition(s) is Boole's inequality an equality (*i.e.* = and not <)?

If A_1, A_2 disjoint

(d) [2 pt] Under what condition(s) is Boole's inequality a strict inequality (*i.e.* < and not =)?

If A_1, A_2 not disjoint

(e) [2 pt] Draw the sample space for a coin flip and spin of a spinner with sections for "A", "B", and "C".

Ω

HA	HB	HC
TA	TB	TC

NPC

Problem 2 You want to survey the residents of Philadelphia in order to find out about their dieting habits. You get a list of all apartment buildings nearby and randomly pick 40 buildings. You then make a list of every apartment in each of the 40 buildings and you randomly pick 250 from this list using a sorted random number column in JMP. You then survey those willing to talk with you.

(a) [2 pt] What is the maximum size of your sample?

$\boxed{n \leq 250}$

(b) [5 pt] If your intention is to make inference about the diets of all residents of Philadelphia, is this a properly designed survey? If not, explain the problem(s) with the survey.

(1) Not representative since only apt bldgs (2)
(2) Convenience - nearby (2)
(3) non-response - only those willing to talk to you (1)

(c) [2 pt] If they exist, can the problem(s) be fixed in order to make non-biased inference about dieting habits using this data?

NO

(d) [3 pt] If instead you sampled 10 apartments from each type of building (luxury, storefront, federal housing, etc), what kind of sampling would this be called?

Stratified / cluster sampling

Problem 3 Adam buys 10 shirts: 2 are red, 3 are blue, 2 are black, and 3 are white.

(a) [3 pt] When Adam dresses for a lecture he picks a random shirt out of his closet. After the lecture, he puts it back in his closet. Every morning he forgets which shirt he wore the previous day. Consider each shirt unique even though they may be of the same color. If there are 23 lectures, how many different ways can he wear shirts for the entire summer session?

$\boxed{10^{23}}$

NPC

(b) [3 pt] Consider the unlikely situation that after a lecture, he throws the shirt in a hamper and does not wear it the next day. How many different day-shirt combinations are there in one week of lectures (*i.e.* four lectures)?

NPC

$$\frac{10!}{6!}$$

(c) [4 pt] What is the probability he wears two red shirts in a week of lectures?

$$P(2 \text{ red shirts}) = \frac{\binom{2}{2}\binom{8}{2}}{\binom{10}{4}} = .133$$

(1) making a prob statement
(1) correct denominator
(2) correct numerator

(d) [4 pt] What is the probability he wears one red shirt in a week of lectures?

$$P(1 \text{ red shirt}) = \frac{\binom{2}{1}\binom{8}{3}}{\binom{10}{4}} = .533$$

(ditto)

(e) [3 pt] Red shirts are special because they have to be dry-cleaned in a special way. Other shirts can be washed. Consider the cost of washing negligible since he has to do a wash anyway. Each red shirt worn the week prior will become dirty and costs \$3 to be dry-cleaned.

Create a r.v. X that models the cost of Adam's dry-cleaning bill. Use the $\sim$ and brace notation used in class and indicate units. If you do not know how to do this, make up a r.v. X that is *reasonable* for this model so you can continue with the rest of the problem after.

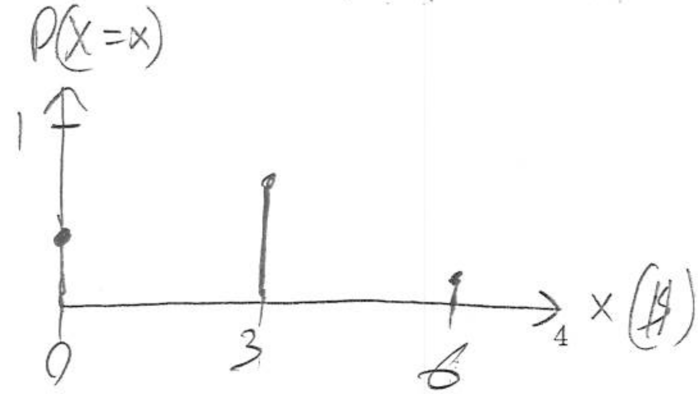
(1) Calculating P(0 shirts)
(2) Creating r.v.

$$\Rightarrow P(0 \text{ red shirts}) = 1 - .133 - .533 = .333$$

$$X \sim \begin{cases} \$6 & \text{w.p. } .133 \text{ or } \frac{2}{15} \\ \$3 & \text{w.p. } .533 \text{ or } \frac{8}{15} \\ \$0 & \text{w.p. } .333 \text{ or } \frac{1}{3} \end{cases}$$

(f) [2 pt] Draw the PMF for X. Indicate axes and units clearly.

NPC



(g) [2 pt] Adam teaches for five weeks. Each week Adam drops off his dry cleaning on Thursday and picks it up on Sunday. Denote the first week's bill by X_1, the second week's bill by X_2, ..., and the fifth week's bill by X_5. Are $X_1, X_2, ..., X_5 \overset{iid}{\sim}$ with the PMF you just drew? Justify each assumption or explain why it does not hold.

Yes as long as Adam doesn't remember what he wore last week and if he continues drawing his shirts at random

(h) [5 pt] Regardless of what you wrote in the previous part, assume the weekly dry cleaning bills are $\overset{iid}{\sim}$ r.v.'s with the PMF from part (f) from now on. Find the expected total dry cleaning expenses and the standard deviation of total dry cleaning expenses.

(2) exp

(3) SD

$E[S_5] = 5\mu = 5 \cdot \$2.40 = \boxed{\$11.99}$

$SD[S_5] = \sqrt{5}\,\sigma = \boxed{\$4.38}$

$\mu = E[X] = 3 \cdot .533 + 6 \cdot .133 = \$2.40$

$\sigma^2 = Var[X] = (0-2.40)^2 \cdot .333 + (3-2.40)^2 \cdot .533 + (6-2.40)^2 \cdot .133 = 3.833 \$^2$

$\sigma = SD[X] = \sqrt{3.833 \$^2} = \$1.96$

(i) [3 pt] For those five weeks also find the expected average of his weekly dry cleaning bill and the standard deviation of the average weekly dry cleaning bill.

(1) exp

(2) SD

$E[\bar{X}_5] = \mu = \boxed{\$2.40}$, $SD[\bar{X}_5] = \frac{\sigma}{\sqrt{5}} = \frac{\$1.96}{\sqrt{5}} = \boxed{\$0.88}$

(j) [3 pt] Adam manages to get a 10% discount and also tips \$8 at the end of the summer. Find the expected total bill and the standard deviation of the total bill.

1) for exp

1) SD

1) correctness

$T = 0.9\, S_5 + 8$

$E[T] = 0.9 \cdot E[S_5] + 8 = 0.9 \cdot \$11.99 + 8 = \boxed{\$18.80}$

$SD[T] = |0.9|\, SD[S_5] = 0.9 \cdot \$4.38 = \boxed{\$3.94}$

Problem 4 Dell, Inc. was founded by Michael Dell in 1984 while he was a student at UT. By 1988, it was a publically traded company. In this question, we will be investigating a model for technical support by phone.

(a) [2 pt] In 1984, Dell got his first three customers — three unrelated business managers in the Austin, TX metro area. However, his computers weren't polished products and had many defects. During any business day, his three customers could call him personally asking for help. There was a 10% chance each would call in. How would you model these three customers' possible calls? Use the notation X_1, X_2, X_3 for the three customers and use the r.v.'s we learned about in class.

(1) saying Bernoulli
(1) correct notation

$X_1, X_2, X_3 \overset{iid}{\sim} \text{Bernoulli}(10\%)$ where X_i is $\begin{cases} 1 & \text{if called in} \\ 0 & \text{if not} \end{cases}$

(b) [3 pt] What is the probability that two customers would call in in a single day?

NPC

$P(S_3 = 2) = \binom{3}{2} \cdot .10^2 \cdot .90^1 = 3 \cdot .01 \cdot .9 = \boxed{.027}$

(c) [4 pt] By the beginning of 1985, he had 100 customers and his computer-assembly process was greatly improved. Now there was only a 3% chance each would call in. What is the expected number of calls per day and the standard deviation of the number of calls?

(2) exp
(2) SD

$S_{100} \sim \text{Binomial}(100, 3\%)$

$E[S_{100}] = np = 100 \cdot 3\% = \boxed{3}$

$SD[S_{100}] = \sqrt{np(1-p)} = \sqrt{100 \cdot 3\% \cdot 97\%} = \boxed{1.706}$

(d) [4 pt] By the middle of 1985, he had 577 customers and improved his computers to the point where only his customers called in daily with 1.5% chance. What is the probability he got more than two calls per day?

-2 for only $P(S=2)$
-1 not incl $P(S=2)$
-1 for some calculation w/ binomial

$S_{577} \sim \text{Binomial}(577, 1.5\%)$

$1.632E-4 \quad .001434 \quad .006289$

$P(S_{577} > 2) = 1 - P(S_{577}=0) - P(S_{577}=1) - P(S_{577}=2)$
$= 1 - \binom{577}{0}(0.015)^0(.985)^{577} - \binom{577}{1}(0.015)^1(.985)^{576} - \binom{577}{2}(0.015)^2(.985)^{575} = \boxed{.99211}$

(e) [2 pt] Business kept growing and growing for Dell and his ability to make better computers also improved. By 1988, Dell had 400,000 customers and the probability of each of them calling in was 0.7%. Model a r.v. N for the daily number of calls in 1988 using an approximation we learned about in class. Indicate the parameter(s) clearly.

(1) saying Poisson
(1) correct notation

let $\lambda = np = 400{,}000 \cdot 0.7\% = 2800$

$N \sim \text{Poisson}(2800)$

(f) [4 pt] Using the approximation from part (e), find the probability exactly 100 people call in during one day in 1988.

(2) Poisson PMF
(2) For correctness

$$P(N=100) = \frac{e^{-2800}\, 2800^{100}}{100!} \approx 0$$

+1 for correct binomial

Problem 5 A drivers education website requires students to read an essay on drunk driving as part of its curriculum. Below is an excerpt:

> As more alcohol is consumed the risk of getting into a vehicular accident if the person drives grows. For example, a man that weighs about 160 pounds would have a BAC of 0.04 an hour after drinking two beers. Its still way below the limit of driving under the influence but the likelihood of getting into an accident is 1.4 times more probable than [the national average]. Add two more beers then the probability goes up tenfold. Make it a six pack with two more beers, the drinker reaches the limit of 0.10 BAC and the risk is now 48 times more than [the national average]. Add two more for the road and you reach 0.15 BAC well above the legal limit and the risk is now 380 times than the [the national average]. Drunk driving is never an option...

During another part of the curriculum, they read excerpts of the National Safety Council's (NSC) report on traffic fatalities countrywide: $P(\text{accident}/\text{mi}) = 1.54E-8$

> The motor-vehicle death rate per 100,000,000 vehicle-miles was 1.54 in 2005...

Consider the NSC's statement to be accurate as an average across all accidents, including those under-the-influence of drugs and alcohol.

Another part of the curriculum is learning about the prevalence of drunk driving:

> WASHINGTON (AP) An estimated 17 million people have driven while drunk at least once on U.S. streets and highways in the course of a year, according to a government study released Wednesday... $P(D) = \frac{17m}{300m} \cdot 1\% = 5.67E-4$

We estimate that 1% of the people that "have driven drunk" drive with 0.15 BAC or over and there are 300 million people in United States.

Consider the drunk driving risk literature as well as the newswire to be ballpark-accurate for the year 2005. Consider the distance between your friend's apartment and your apartment to be 10 miles. Denote the event of getting into an accident during these 10 miles as "A". Denote the event of driving while completely drunk (*i.e.* 0.15 BAC) as "D". Please use this notation.

(a) [3 pt] Use the NSC report excerpt to find the probability of getting into a fatal accident while driving from your friend's apartment to your apartment.

+1 correct prob statement

-2 for not mult by 10

$$P(A) = 1.54E{-}8 \cdot 10 = \boxed{1.54E{-}7}$$

(b) [3 pt] What is the probability of getting into a fatal accident while driving home from your friend's party on Saturday night after you drank heavily?

1) correct prob statement

1) follow through

$$P(A|D) = 300 \cdot P(A) = \boxed{5.85E{-}5}$$

(see first excerpt)

(c) [4 pt] If you got into an accident, what was the probability you were drinking heavily?

1) correct prob statement

1) correct $P(D)$

2) follow through

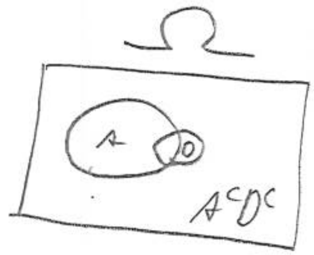
$$P(D|A) = \frac{P(AD)}{P(A)} = \underbrace{\frac{P(A|D)}{P(A)}}_{300} P(D) = 300 \cdot 5.67E{-}4 = \boxed{.215}$$

(see first excerpt)

(d) [4 pt] What is the probabilty of *not* getting into a fatal accident while driving from your friend's apartment to your apartment after you did the Stat 101 homework (*i.e.* *no* drinking)? Hint: you may want to leave the values in the calculator and keep values to many significant digits.

+1 correct prob statement

$$P(A^c|D^c) = \frac{P(A^cD^c)}{\underbrace{P(D^c)}_{1-5.67E-4}} = \frac{.999}{.999} \approx \boxed{1}$$



$$P(A^cD^c) = 1 - P(A) - P(D) + P(A,D)$$
$$= 1 - P(A) - P(D) + P(A|D)P(D)$$
$$= 1 - 1.54E{-}7 - 5.67E{-}4 + (5.85E{-}5)(5.67E{-}4)$$
$$\approx .999$$

Problem 6 Consider a the following data of heights in inches of an NBA team's roster. The data is already sorted from smallest to largest:

$$63, 72, 74, 75, 76, 76, 77, 78, 78, 79, 80, 80, 80, 82, 82, 83, 84$$

Here is an excerpt of `JMP` output for the above data.

Quantiles			Moments	
100.0%	maximum	84	Mean	77.588235
99.5%		84	Std Dev	4.9882214
97.5%		84	Std Err Mean	1.2098214
90.0%		83.2	Upper 95% Mean	80.152942
75.0%	quartile	81	Lower 95% Mean	75.023528
50.0%	median	78	N	17
25.0%	quartile	75.5		
10.0%		70.2		
2.5%		63		
0.5%		63		
0.0%	minimum	63		

(a) [1 pt] What is the modal value?

NPC

80

(b) [1 pt] What type of data is our univariate dataset?

NPC

interval (numeric)

(c) [1 pt] Begin to calculate the sample standard deviation, s. Write out the first few terms but do not compute.

NPC

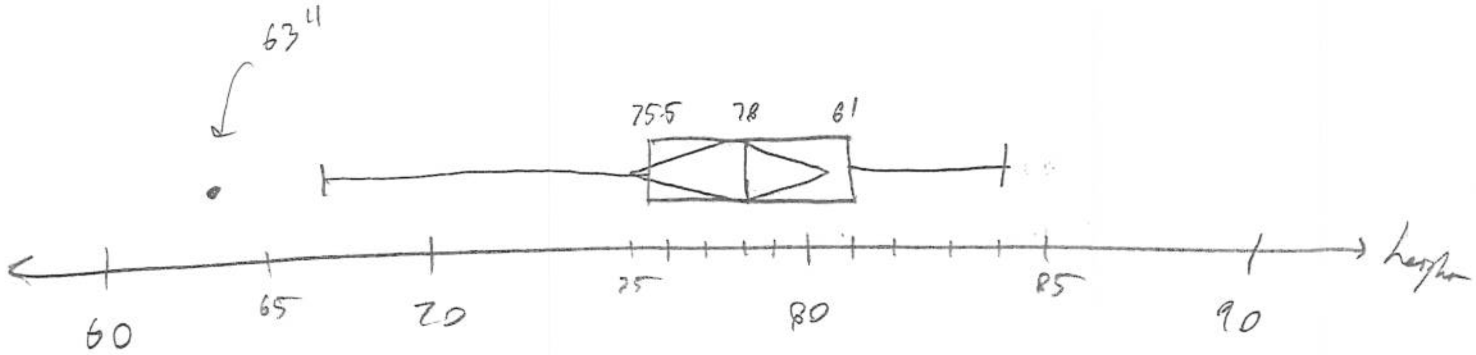
$$s = \sqrt{\frac{1}{16}\left((63 - 77.588)^2 + (72 - 77.588)^2 + \ldots + (84 - 77.588)^2\right)}$$

(d) [1 pt] Is there a skew? If so, what type of skew?

NPE

left skew (since mean < median)

(e) [8 pt] Draw a box and whisker plot for this dataset to scale. Use 2.4 as the half-width of the diamond. Denote the outlier(s) and the salient points of the box (do not denote the values at the ends of the whiskers). Use the JMP output; do not compute the percentiles by hand.



IQR = 5.5

whisker length = 8.25

-2 for outlier not present

-2 for right whisker

-2 for no denotation

-1 for incorrect diamond / scaling