18.05 Problem Set 2, Spring 2018

(due outside 4-174 Tuesday, Feb. 20 at 9:30 AM)

Problem 1. (10 pts.) **Employment benefits.**
In one state, 56% of workers have a workplace retirement plan, 68% have health insurance and 49% have both benefits. For a worker selected at random:

(a) What is the probability that they have neither benefit?

(b) What is the probability that they have health insurance if they have a retirement plan?

(c) Are having these two benefits mutually exclusive?

(d) Are having a retirement plan and having health insurance independent events?

Problem 2. (10 pts.) **Drug testing for athletes.**
Estimating the rate of "false positives" in drug tests for athletes is difficult or impossible, in part because a positive result is often not a random event but a consequence of chemicals present for "benign" reasons. For this problem we will use the often cited figure of 5%: that 5% of those who have not used banned substances will nevertheless test positive.

There are parallel problems with estimating the rate of "false negatives," but we will use another often cited figure of 10%: that 10% of users of banned substances will pass a given test.

At a certain school the percentage of athletes using banned substances is 4%. What is the probability of

(a) a person with a positive test being a user of a banned substance?

(b) a person with a positive test *not* being a user of a banned substance?

(c) a person with a negative test being a user of a banned substance?

(b) a person with a negative test *not* being a user of a banned substance?

Problem 3. (10 pts.) **Dice Games.**
I play a board game where on each turn a player moves some number of spaces along the board. The game has 'multiplier cards' and dice. To see how many spaces they will move a player first picks a 'multiplier card'. These cards are face down and each has a 1, 2 or 3 on the hidden side. Then, if they pick a card with 1 or a 2 they roll a 4-sided die. If they pick a 3 they roll a 6-sided die. Finally, the number of spaces they move is the product of the number on the card and the value of their die roll.

Suppose that the fraction of multiplier cards with a 1 is 1/5, with a 2 is 2/5 and with a 3 is 2/5. Let X be the number of spaces moved on my next turn. What is $E(X)$?

Problem 4. (10 pts.) **Bayesian Dice**.
I have a mixture of dice in a cup: there are 1 four-sided, 2 six-sided, 3 eight-sided,

and 4 twelve-sided. I take one die at random and, without showing it to you, I roll it and look at the result. Let S be the number of sides of the chosen die, and let R be the result of the roll.

(a) What are the probability mass functions (pmf) of S and R?

(b) Use Bayes' theorem to find $P(S = k \mid R = 3)$ for $k = 4$, 6, 8, and 12. Which die is most probable if $R = 3$?

Terminology: You are computing the pmf of 'S given $R = 3$'.

(c) Which die is most probable if $R = 6$? Hint: You can either repeat the computation in (b), or you can reason based on your result in (b).

(d) Which die is most probable if $R = 9$? No computations are needed!

Problem 5. (10 pts.) **Couples at dinner**
A total of n couples (more than one!) will attend a fancy dinner. The person in charge of seating did not know the couples, so they just randomly assigned people to places. Everyone will be seated at round tables (each with at least three chairs) with a person to either side of them. What is the expected number of couples who will be seated together?

Problem 6. (10 pts.) **R simulation of the length of runs.**
(a) Write down a sequence of fifty zeros and ones. Try to make one that looks like it was randomly generated by flipping a coin, with 0 for tails and 1 for heads.

(b) A run is a sequence of all 1s or all 0s. How long is the longest run in your answer to part (a)?

(c) Now we will use R to simulate 50 tosses of a fair coin and estimate the average length of the longest run. We give code below that simulates one trial. You will need to use a for loop to run 10000 trials, so we have included two small examples using 'for loops'. On the course web site you can find tutorials on both for loops and the rle() function used in the code. (This is on the "R in five easy pieces" subpage).

```
# R code to simulate 50 flips of a fair coin and find the longest run
# rle stands for 'run length encoding'.  rle(trial)$lengths is a vector
of the lengths of all the different runs in trial.
nflips = 50
trial = rbinom(nflips, 1, 0.5) # binomial(1, 0.5) = bernoulli(0.5)
maxRun = max(rle(trial)$lengths)
```

```
# R code demonstrating 'for' loops.
for (j in 1:5)
{
  print(j^2)
}
```

(Should produce: `1, 4, 9, 16, 25.`)

```
sum = 0
```

```
for (j in 1:5)
{
  sum = sum + j
}
print(sum)
```
(Should produce: `15.`)

Use R to simulate the average length of the longest run in 50 flips of a fair coin. Do this three times with 10000 trials each time and report the three results.

(d) Make small modifications of your code from part (c) to estimate the probability of a run of 9 or more in 100 flips of a fair coin. Do this three times with 10000 trials each time and report the three results.

Note: For parts (c) and (d) print out your R code and include it with your pset.