Homework 1 KEY

Due: Sunday, Sept 7th 11:59 PM

Using the provided Qmd template, complete the following exercises and submit the document with your answers on Gradescope, which you can access through Canvas. You must show your work for all problems.

1. In a Japanese cohort study, 5,322 male non-smokers and 7,019 male smokers were followed for four years. Of these men, 16 non-smokers and 77 smokers developed lung cancer.

    a. Fill in the 2x2 table for this problem:

	Developed lung cancer	Did not develop lung cancer	Total
Smoker	77	6942	7019
Non-smoker	16	5306	5322
Total	93	12248	12341

    b. What is the probability that a randomly-selected non-smoker from this group developed lung cancer?

```
16/5322
```

```
[1] 0.003006389
```

    c. What is the marginal probability of developing lung cancer in this group?

```
93/12341
```

```
[1] 0.007535856
```

    d. Are the events "smoker" and "developed lung cancer" independent?

    P(developed lung cancer) = 0.008 (from part b)

    P(developed lung cancer | smoker) = 77/7019 = 0.011

    These probabilities are not equal, so the two events are not independent.

2. Suppose a particular disease is prevalent in a population with 60%. The sensitivity and specificity of the medical test for this disease are both 0.9. A patient from this population is visiting the physician and is tested for the disease.

   a. What is the probability that a person has the disease given that they had a positive test result?

   $P(D+|T+) = \frac{P(T+|D+)P(D+)}{P(T+)}$

   $P(T+|D+) = 0.9$ (given)

   $P(D+) = 0.6$ (given)

   $P(T+) = P(T+|D+)P(D+)+P(T+|D-)P(D-) = 0.9*0.6+(1-0.9)*(1-0.6) = 0.58$

   Therefore, $P(D+|T+) = \frac{0.9*0.6}{0.58} = 0.93$

   b. Drop the given statement that the specificity is 0.9. If the sensitivity is 0.9, what is the minimum required specificity of the test to know with at least 95% certainty that the patient has the disease given that they have a positive test?

   Want $P(D+|T+) \geq 0.95$

   Using part a, we drop the (1-0.9) in the denominator calculation and replace with $1-s$ and solve for s

   $0.95 \leq \frac{0.9*0.6}{(0.9*0.6)+(1-s)(1-0.6)}$

   $0.95 \leq \frac{0.54}{0.94-0.4s}$

   $0.95(0.94-0.4s) \leq 0.54$

   $-0.38s \leq -0.353$

   $s \geq 0.93$

3. A manager at a restaurant is training her new hostess and during the discussion the new employee asks about the typical size of a group that needs to be seated. The manager says company research indicates that one of the following three probability models is appropriate.

Group size	1	2	3	4	5
Model 1	0.1	0.25	0.2	0.2	0.1
Model 2	0.15	0.35	0.25	0.25	0.1
Model 3	0.1	0.35	0.15	0.3	0.1

a. Identify the correct probability model and explain the problems with the other two models.

The third one is the correct probability model. The probabilities of the other two models do not sum to 1.

b. What is the expected size of a group that needs to be seated? Is this different from the most likely size?

$\mathbf{E}[X] = \sum_x xPr(X = x) = 1 * 0.1 + 2 * 0.35 + 3 * 0.15 + 4 * 0.3 + 5 * 0.1 = 2.95$

```
1*0.1+2*0.35+3*0.15+4*0.3+5*0.1
```

```
[1] 2.95
```

The expected size is the mean size, which is different from the most likely size. The most likely size is given by the group size with the highest probability, which is 2.

4. A potato chip manufacturer decides whether to purchase a truckload of potatoes by selecting samples and inspecting them to see if they meet the company standards. Suppose that 20 potatoes are randomly selected and the company policy is to purchase the load of potatoes if fewer than 2 of the potatoes are deemed unsatisfactory. Find the probability of purchasing a load of potatoes for which

a. 5% are rotten

```
pbinom(1,20,0.05)
```

```
[1] 0.7358395
```

b. 10% are rotten

```
pbinom(1,20,0.1)
```

```
[1] 0.391747
```

5. A soup company claims that one of its products contains 480 mg of sodium per serving with a standard deviation of 6 mg. Assuming that the amount of sodium in a particular serving varies according to the normal distribution, what is the chance that a particular serving contains:

a. fewer than 475 mg of sodium

```
pnorm(475,480,6)
```

```
[1] 0.2023284
```

b. at least 490 mg of sodium

```
1-pnorm(490,480,6)
```

```
[1] 0.04779035
```

c. Find the interval of sodium levels that contains 80% of the values for the amount of sodium in a particular serving.

```
qnorm(0.1,480,6)
```

```
[1] 472.3107
```

```
qnorm(0.9,480,6)
```

```
[1] 487.6893
```