Complete the following tasks. Show your work where applicable. Some answers have been provided. Assemble your work into one PDF document and upload the PDF back into our CatCourses page.

1. (*Bayes Rules!* exercise 2.4) Edward is trying to prove to Bella that vampires exist. Bella thinks there is a 0.05 probability that vampires exist. She also believes that the probability that someone can sparkle like a diamond if vampires exist is 0.7, and the probability that someone can sparkle like a diamond if vampires don't exist is 0.03. Edward then goes into a meadow and shows Bella that he can sparkle like a diamond. Given that Edward sparkled like a diamond, what is the probability that vampires exist?

   If we let S be the event of sparkling and let V be the event of a vampire, then we can employ Bayes' Rule:

$$\begin{aligned} P(V|S) &= \frac{P(S|V) \cdot P(V)}{P(S|V) \cdot P(V) + P(S|V^c) \cdot P(V^c)} \\ &= \frac{(0.7)(0.05)}{(0.7)(0.05) + (0.03)(0.95)} \\ &\approx 0.5512 \end{aligned}$$

2. (*Bayes Rules!* exercise 2.10) A recent study of 415,000 Californian public middle school and high school students found that 8.5% live in rural areas and 91.5% in urban areas. Further, 10% of students in rural areas and 10.5% of students in urban areas identified as Lesbian, Gay, Bisexual, Transgender, or Queer (LGBTQ). Consider one student from the study.

   (a) What's the probability they identify as LGBTQ?

   Let Q be the event that the student identifies as LGBTQ and let U be the event that the student lives in an urban area. The total probability is

$$\begin{aligned} P(Q) &= P(Q|U) \cdot P(U) + P(Q|U^c) \cdot P(U^c) \\ &= (0.105)(0.915) + (0.10)(0.085) \\ &\approx 0.1046 \end{aligned}$$

   (b) If they identify as LGBTQ, what's the probability that they live in a rural area?

By Bayes' Rule,

$$\begin{aligned} P(U^c|Q) &= \frac{P(Q|U^c) \cdot P(U^c)}{P(Q|U^c) \cdot P(U^c) + P(Q|U) \cdot P(U)} \\ &= \frac{(0.10)(0.085)}{(0.105)(0.915) + (0.10)(0.085)} \\ &\approx 0.0813 \end{aligned}$$

(c) If they do not identify as LGBTQ, what's the probability that they live in a rural area?

By Bayes' Rule,

$$\begin{aligned} P(U^c|Q^c) &= \frac{P(Q^c|U^c) \cdot P(U^c)}{P(Q^c|U^c) \cdot P(U^c) + P(Q^c|U) \cdot P(U)} \\ &= \frac{[1 - P(Q|U^c)] \cdot P(U^c)}{[1 - P(Q^c|U^c)] \cdot P(U^c) + [1 - P(Q^c|U)] \cdot P(U)} \\ &= \frac{(1 - 0.10)(0.085)}{(1 - 0.105)(0.915) + (1 - 0.10)(0.085)} \\ &\approx 0.0854 \end{aligned}$$

3. A tattoo enthusiast website[1] claims that

- 47% of Millennials have tattoos
- 36% of Generation X have tattoos
- 13% of Boomers have tattoos

whereas the population proportions are 22%, 20%, and 22% for those generations respectively.[2] Compute the probability that a person is a Millennial given that they have tattoos. (For homework brevity, let us assume that no one in other age groups have tattoos.)

Note that there will be 3 terms in the denominator.

$$\frac{(0.47)(0.22)}{(0.47)(0.22) + (0.36)(0.20) + (0.13)(0.22)} \approx 0.5069$$

4. A contractor is has hired a team of engineers to build a prototype contraption of a balloon detection sensory machine. Let W be the presence of a weather balloon, and let T be the event that the device claims that a flying object is a weather balloon.

[1]Source: https://www.reddit.com/r/todayilearned/comments/dwy925/til_47_of_millennials_ages_18_to_29_have_tattoos/

[2]Source: https://www.statista.com/statistics/797321/us-population-by-generation/

Suppose that we know that 32 percent of flying objects are weather balloons. The team says that the specificity of the device is

$$P(T^c|W^c) = 0.98$$

What must the sensitivity $P(T|W)$ be so that the value of $P(W|T)$ is greater than 95 percent?

Via Bayes' Rule with $P(W) = 0.32$, and maybe let $x = P(T|W)$,

$$\begin{aligned}
0.95 &< P(W|T) \\
0.95 &< \frac{P(T|W) \cdot P(W)}{P(T|W) \cdot P(W) + P(T|W^c) \cdot P(W^c)} \\
0.95 &< \frac{x \cdot P(W)}{x \cdot P(W) + [1 - P(T^c|W^c)] \cdot [1 - P(W)]} \\
0.95 &< \frac{0.32x}{0.32x + (1 - 0.98)(1 - 0.32)} \\
0.95 &< \frac{0.32x}{0.32x + 0.0136} \\
0.304x + 0.01292 &< 0.32x \\
0.01292 &< 0.016x
\end{aligned}$$

and we have that the sensitivity

$$P(T|W) > 0.8075$$

(It's okay if the student got $\frac{0.0130}{0.016} = 0.8125$ at the end.)

5. (*Bayes Rules!* exercise 2.11) A student applies for six equally competitive data science internships. They have the following prior model for their chances of getting into any given internship, π

π	**0.3**	**0.4**	**0.5**	**Total**
$f(\pi)$	0.25	0.60	0.15	1

(a) Let Y be the number of internship offers that the student gets. Specify the model for the dependence of Y on π and the corresponding pmf, $f(y|\pi)$.

Since the outcomes are either offers or rejections for the internships, we can try a binomial model with $n = 6$ applications:

$$f(y|\pi) = \binom{6}{y} \pi^y (1 - \pi)^{6-y} \text{ for } y \in \{0, 1, 2, 3, 4, 5, 6\}$$

(b) The student got some pretty amazing news. They were offered four of the six internships! How likely would this be if $\pi = 0.3$?

The likelihood with $y = 4$ internship offers is

$$L(\pi = 0.3|y = 4) = \binom{6}{4}(0.3)^4(0.7)^2 \approx 0.0595$$

(c) Construct the posterior model of π in light of the student's internship news.

The normalizing constant is

$$\begin{aligned} f(y = 4) &= f(0.3)L(0.3|y = 4) + f(0.4)L(0.4|y = 4) + f(0.5)L(0.5|y = 4) \\ &= (0.25)\tbinom{6}{4}(0.3)^4(0.7)^2 + (0.60)\tbinom{6}{4}(0.4)^4(0.6)^2 + (0.15)\tbinom{6}{4}(0.5)^4(0.5)^2 \\ &\approx 0.1330 \end{aligned}$$

and the posterior distrubtion is

$$\begin{aligned} f(\pi = 0.3|y = 4) &= \frac{(0.25)\tbinom{6}{4}(0.3)^4(0.7)^2}{0.1330} &\approx 0.1119 \\ f(\pi = 0.4|y = 4) &= \frac{(0.60)\tbinom{6}{4}(0.4)^4(0.6)^2}{0.1330} &\approx 0.6236 \\ f(\pi = 0.5|y = 4) &= \frac{(0.15)\tbinom{6}{4}(0.5)^4(0.5)^2}{0.1330} &\approx 0.2643 \end{aligned}$$

6. (*Bayes Rules!* exercise 2.18) Lactose intolerance is an inability to digest milk, often resulting in an upset stomach. A lab tech wants to learn more about the proportion of adults who are lactose intolerant, π. Their prior model for π is:

π	**0.4**	**0.5**	**0.6**	**0.7**	**Total**
$f(\pi)$	0.1	0.2	0.44	0.26	1

The lab tech surveys a random sample of 80 adults and 47 are lactose intolerant. Use simulation to approximate the posterior model of π. Simulate data for 10,000 people.

```
library("janitor")
library("tidyverse")
patients <- data.frame(pi = c(0.4, 0.5, 0.6, 0.7))
prior <- c(0.1, 0.2, 0.44, 0.26)
patient_sim <- sample_n(patients, size = 10000,
  weight = prior, replace =TRUE)
patient_sim <- patient_sim |>
```

```
  mutate(y = rbinom(10000, size = 80, prob = pi))
y_47 <- patient_sim |> filter(y == 47)
y_47 |> tabyl(pi) |> adorn_totals("row")

    pi   n    percent
   0.5  51 0.10759494
   0.6 395 0.83333333
   0.7  28 0.05907173
 Total 474 1.00000000
```

7. (*Bayes Rules!* exercise 3.11) A university wants to know what proportion of students are regular bike riders, π, so that they can install an appropriate number of bike racks. Since the university is in sunny Southern California, staff think that π has a mean of 1 in 4 students, and a mode of $\frac{5}{22}$.

   (a) Specify and plot a Beta model that reflects the staff's prior ideas about π.

   Matching the expected value and mode of a beta distribution with the provided information,

   $$\frac{\alpha}{\alpha+\beta} = \frac{1}{4} \quad \text{and} \quad \frac{\alpha-1}{\alpha+\beta-2} = \frac{5}{22}$$

   suggest that we can try the hyperparameters $\alpha = 6$ and $\beta = 18$

   ```
   R: bayesrules::plot_beta(6,18)
   ```

   (b) Among 50 surveyed students, 15 are regular bike riders. What is the posterior model for π?

   With $\alpha = 6$, $\beta = 18$, $y = 15$, and $n = 50$; we can get an updated beta model

   $$\pi|(Y = y) \sim \text{Beta}(\alpha + y, \beta + n - y) = \text{Beta}(21, 53)$$

   (c) What is the mean, mode, and standard deviation of the posterior model?

   ```
   bayesrules::summarize_beta_binomial(alpha = 6, beta = 18,
                                       y = 15, n = 50) |>
     mutate_if(is.numeric, round, digits = 4)

      model alpha beta   mean   mode    var     sd
   1  prior     6   18 0.2500 0.2273 0.0075 0.0866
   ```

```
2 posterior    21    53 0.2838 0.2778 0.0027 0.0521
```

8. (*Bayes Rules!* exercise 3.12) A 2017 Pew Research survey found that 10.2% of LGBT adults in the U.S. were married to a same-sex spouse. Now it's the 2020s, and Bayard guesses that π, the percent of LGBT adults in the U.S. who are married to a same-sex spouse, has most likely increased to about 15% but could reasonably range from 10% to 25%.

   (a) Identify and plot a Beta model that reflects Bayard's prior ideas about π.

   It is probably too tedious to match the beta distributiom formulas to the variance, but we can start with tuning the parameters to align with the mean

   $$\frac{\alpha}{\alpha+\beta} = 0.15 = \frac{3}{20} \quad \rightarrow \quad \alpha = 3, \quad \beta = 17$$

   `R: bayesrules::plot_beta(3,17)`

   (b) Bayard wants to update his prior, so he randomly selects 90 US LGBT adults and 30 of them are married to a same-sex partner. What is the posterior model for π?

   With $\alpha = 3$, $\beta = 17$, $y = 30$, and $n = 90$; we can get an updated beta model

   $$\pi|(Y = y) \sim \text{Beta}(\alpha + y, \beta + n - y) = \text{Beta}(33, 77)$$

   (c) Calculate the posterior mean, mode, and standard deviation of π.

```
bayesrules::summarize_beta_binomial(alpha = 3, beta = 17,
                                    y = 30, n = 90) |>
  mutate_if(is.numeric, round, digits = 4)

      model alpha beta mean   mode    var     sd
1     prior     3   17 0.15 0.1111 0.0061 0.0779
2 posterior    33   77 0.30 0.2963 0.0019 0.0435
```

Here are some of the answers. Note that answers may slightly vary depending on when and where rounding took place, and due to randomization in code.

1. 0.5512
2. (a) 0.1046
   (b) 0.0813
   (c) 0.0854
3. 0.5069
4. $P(T|W) > 0.8075$
5. (a)
   (b) 0.0595
   (c)
6.
7.
8.