Name ____

Fall 2021 **STAT 587-1** **Exam I**

(70 points)

Instructions:

- Write your name on the top, but do not open the exam.
- You are allowed to use any resource except aid from another individual.
- A total of 4 pages with a front and back.
- For full/partial credit, show all your work.

1. The United States has identified a previously unknown piece of space debris and is trying to determine the probability it is from a US space craft.

   Answer:

```r
prevalence <- c("Russia"=4/10, "US"=3/10, "China"=2/10, "Other"=1/10)
probability <- c("Russia"=1/100, "US"=1/10, "China"=5/100, "Other"=2/100)
```

   For identified space debris, we have the following proportions by country

```
## Russia      US   China   Other
##    0.4     0.3     0.2     0.1
```

   The space debris is giving off a signature in the infrared spectrum. The following table provides the proportion of existing space debris for each country that has this particular signature, i.e. the conditional probability given the country:

```
## Russia      US   China   Other
##   0.01    0.10    0.05    0.02
```

   Given that the space debris is giving off this signature, determine the probability that the debris belongs to Russia, US, China, and Other. (10 points)

   Answer: This is a Bayes Rule question, but instead of just calculating one probability we must calculate all the probabilities. Assume the following notation:

   - R is the event the debris is from Russia
   - U is the event the debris is from US
   - C is the event the debris is from China
   - O is the event the debris is from Other
   - S is the event the debris is giving off the signature

   We have the marginal probability the debris is from each country, e.g. $P(R)$, as

```r
prevalence
```

```
## Russia      US   China   Other
##    0.4     0.3     0.2     0.1
```

   and the conditional probability the debris is giving off the signature given the country of origin, e.g. $P(S|R)$, is

```r
probability
```

```
## Russia      US   China   Other
##   0.01    0.10    0.05    0.02
```

Now for a particular country we use Bayes Rule to calculate

$$P(R|S) = \frac{P(S|R)P(R)}{P(S|R)P(R) + P(S|U)P(U) + P(S|C)P(C) + P(S|O)P(O)}.$$

Note that the denominator will be the same for every calculation. The purpose the denominator serves is to ensure that the following probabilities sum to 1:

$$P(R|S) + P(U|S) + P(C|S) + P(O|S) = 1.$$

The final conditional probabilities can be computed in R via

```
probability*prevalence/sum(probability*prevalence)

##     Russia         US      China      Other
## 0.08695652 0.65217391 0.21739130 0.04347826
```

2. For the following variables, identify the most appropriate distribution to model the variable from this list: binomial, Poisson, uniform, and normal. (1 point each)

   (a) strength of #3 carbon steel rebar
   Answer: normal

   (b) weight of dust in my office
   Answer: normal

   (c) number of occupied chairs in Gilman 2272 during this exam (there are a total of 70 chairs)
   Answer: binomial

   (d) number of car accidents in Story County in August 2021
   Answer: Poisson

   (e) number of visitors to a website who click on an advertisement out of the next one thousand visitors
   Answer: binomial

   (f) elasticity of a polymer network
   Answer: normal

   (g) time it will take to read The Black Swan
   Answer: normal

   (h) number of student organizations at Iowa State University
   Answer: Poisson

   (i) number of books on my office bookshelf, out of the 74 books on my bookshelf, with the word "Statistics" in the title
   Answer: binomial

   (j) thickness of my office window
   Answer: normal

3. As a road is constructed, there is an average of 1.5 defects per mile segment of the road. Let Y_i be the number of defects on segment i. Assume $Y_i \sim Po(1.5)$ and that the number of defects in each segment is independent.

   Answer:

```r
rate <- 1.5
```

   (a) The following questions are about the first road segment.

   - What is the expected number of defects in the first segment? (2 points)
     Answer:

```r
rate
## [1] 1.5
```

   - What is the **standard deviation** of the number of defects in the first segment? (2 points)
     Answer:

```r
sqrt(rate)
## [1] 1.224745
```

   - Calculate the probability of no defects in the first segment. (3 points)
     Answer: $P(Y_1 = 0)$ is

```r
dpois(0, lambda = rate)
## [1] 0.2231302
```

   - Calculate the probability of two or more defects in the second segment. (3 points)
     Answer: $P(Y_2 \geq 2) = 1 - P(Y_2 < 2) = 1 - P(Y_2 \leq 1)$ is

```r
1 - ppois(1, lambda = rate)
## [1] 0.4421746
```

(b) The following questions are about a 10 mile stretch of road. Assume the number of defects in each mile stretch is independent. (3 points each)

- What is the expected number of total defects in a 10 mile stretch of the road? (2 points)
  Answer: Let S be the number of defects in a 10 mile stretch. Then $S = \sum_{i=1}^{10} Y_i \sim Po(10 * 1.5)$.

```
n <- 10
rate10 <- n*rate
rate10
## [1] 15
```

- What is the **standard deviation** in the number of total defects in a 10 mile stretch of the road? (2 points)
  Answer:

```
sqrt(rate10)
## [1] 3.872983
```

- What is the probability of exactly 14 defects in a 10 mile stretch of the road? (3 points)
  Answer: $P(S = 14)$ is

```
dpois(14, lambda = rate10)
## [1] 0.1024359
```

- What is the probability of 12 or fewer defects in a 10 mile stretch of the road? (3 points)
  Answer: $P(S \leq 12)$ is

```
ppois(12, lambda = rate10)
## [1] 0.267611
```

4. Assume the compressive strength of residential concrete follows a normal distribution with a mean of 17 MPa and a standard deviation of 2 MPa. (2 points each)

   Answer:

```
m = 17
s = 2
```

   (a) What is the expected compressive strength?

   Answer:

```
m
## [1] 17
```

   (b) What is the **variance** on the compressive strength?

   Answer:

```
s^2
## [1] 4
```

   (c) The coefficent of variation (cv) is the standard deviation divided by the mean. What is the coefficient of variation for compressive strength of residential concrete?

   Answer:

```
s/m
## [1] 0.1176471
```

   Note that this quantity is unitless.

   (d) What is the probability the compressive stength of a particular sample of residential concrete is below 15 MPa?

   Answer:

```
pnorm(15, mean = m, sd = s)
## [1] 0.1586553
```

   (e) Provide an interval such that 95% of all residential concrete samples will have compressive strength within that interval.

   Answer:

```
m + c(-1,1)*2*s
## [1] 13 21
```

A company has indicated that they can double the compressive strength plus 5 Mpa by adding a proprietary ingredient. Specifically, if Y is the normal compressive strength, the company can make it $2Y + 5$.

(f) What is the expected compressive stength of the concrete with this ingredient? (3 points)

Answer:

```
(m_new <- 2*m+5)

## [1] 39
```

(g) What is the **variance** in compressive strength of the concrete with this ingredient? (3 points)

Answer:

```
(s2_new <- 2^2*s^2)

## [1] 16
```

(h) What is the probability the new strength will be greater than 40? (4 points)

Answer:

```
1-pnorm(40, mean = m_new, sd = sqrt(s2_new))

## [1] 0.4012937
```

5. An extruding machine has a screw that propels plastic through a barrel to be extruded. Each turn of the screw pushes, on average, a 5 millimeter (mm) length of plastic with a standard deviation of 0.3 mm.

   Answer:

```
m = 5
s = 0.3
```

   (a) After 70 turns of the screw, what is the expected length of the extruded plastic? (2 points)

   Answer:

```
n <- 70
(m70 <- n*m) ## mm

## [1] 350
```

   (b) After 70 turns of the screw, what is the **standard deviation** of the length of the extruded plastic? (2 points)

   Answer:

```
(s70 <- sqrt(n*s^2)) # mm

## [1] 2.50998
```

   (c) What is the probability the extruded plastic after 70 turns is less than 355 mm in length? (3 points)

   Answer:

```
pnorm(355, mean = m70, s = s70)

## [1] 0.9768171
```

   (d) The machine must be adjusted to ensure that after 70 turns the probability the extruded plastic is less than 355 mm is, at most, 0.99. What is the maximum standard deviation for a single turn required to ensure this probability? (3 points)

   Answer: Let $X \sim N(70 \times 5, 70 \times \sigma^2)$ and our goal is to find σ such that $P(X < 355) \geq 0.99$. The z-statistic that corresponds to 0.99 is

```
(z = qnorm(0.99))

## [1] 2.326348
```

$$
\begin{aligned}
0.99 &= P(X < 355) \\
&= P\left(\frac{X-70*5}{\sqrt{70\sigma^2}} < \frac{355-70*5}{\sqrt{70\sigma^2}}\right) \\
&= P\left(Z < \frac{(355-350)/\sqrt{70}}{\sigma}\right)
\end{aligned}
$$

```r
(355-350)/sqrt(70) / z

## [1] 0.2568895
```

Thus if σ is this value or smaller, we are guaranteed that the probability of being less than 355 mm is at least 0.99.