Name:
OUID:

Please answer **all 7** questions. You have **50 minutes** to complete this exam. You can answer in any order you wish. Please hand in all the exam questions and answers at the end of the class. Circle the questions you have attempted in the table below once the exam is over.

Question	Marks earned	Out of
Q1		10
Q2		10
Q3		10
Q4		10
Q5		10
Q6		10
Q7		10
Total	Out of 7 questions	70

Formulae you might need:

RV	$p(y)$	μ	σ^2	$m(t) = \sum_y e^{yt}p(y)$
Bernoulli	$p(y) = p^y q^{1-y}, y = 0, 1$	p	pq	$pe^t + q$
Binomial	$p(y) = \binom{n}{y} p^y q^{n-y}, y = 0, 1, \ldots, n$	np	npq	$(pe^t + q)^n$
Neg. Bin.	$p(y) = \binom{y-1}{r-1} p^r q^{y-r}, y = r, r+1, r+2, \ldots$	$\frac{r}{p}$	$\frac{rq}{p^2}$	Not given
Multinomial	$p(y_1, y_2, \ldots, y_k) = \frac{n!}{y_1! y_2! \ldots y_k!} (p_1)^{y_1} (p_2)^{y_2} \cdots (p_k)^{y_k}$	np_i	$np_i(1-p_i)$	Not given
Hyper-geom.	$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$	$\frac{nr}{N}$	Not given	Not given
Poisson	$p(y) = \frac{\lambda^y e^{-\lambda}}{y!}$	λ	λ	$e^{\lambda(e^t - 1)}$

Table 1: Formulae

Bayes rule:

$$p(A_i|B) = \frac{p(A_i)p(B|A_i)}{p(B)}$$

where $p(B) = \sum_{j=1}^{n} p(A_j)p(B|A_j)$

$$E((X-\mu)^2) = E(X^2) - \mu^2$$

1. Suppose that X is a random variable and $Y = aX + b$

Prove the following are true using the three Expectation properties discussed in class:

(a) $E(Y) = aE(X) + b$

(b) $V(Y) = a^2V(X)$

(c) Suppose $Y = -2X + 3$ and $H = 3Y - 2$ and $X \sim N(0, \sigma = 2)$ find:

- (i) $E(Y)$
- (ii) $V(Y)$
- (iii) $V(H)$

2. A Statistician at a prestigious university was looking at past records of students in order that she could better prepare a science teacher who wished to recruit students to more advanced courses. The researcher made the following plot (see figure 1 on the following page) using the code below the data.

```
> head(course.df)
  Grade Pass Exam Degree Gender Attend Assign Test  B  C MC Colour Stage1 Years.Since
1     C  Yes   42    BSc   Male    Yes   17.2  9.1  5 13 12   Blue      C         2.5
2     B  Yes   58   BCom Female    Yes   17.2 13.6 12 12 17 Yellow      A         2.0
3     A  Yes   81  Other Female    Yes   17.2 14.5 14 17 25   Blue      A         3.0
4     A  Yes   86  Other Female    Yes   19.6 19.1 15 17 27 Yellow      A         0.0
5     D   No   35  Other   Male     No    8.0  8.2  4  1 15   Blue      C         3.0
6     A  Yes   72   BCom Female    Yes   18.4 12.7 15 17 20   Blue      A         1.5
  Repeat
1    Yes
2     No
3     No
4     No
5     No
6     No

library(ggplot2)
data("course.df")
head(course.df)
windows()
g = ggplot(course.df, aes(x = Degree, y = Exam, fill = Degree)) + #A
geom_boxplot() + geom_point()
g
```

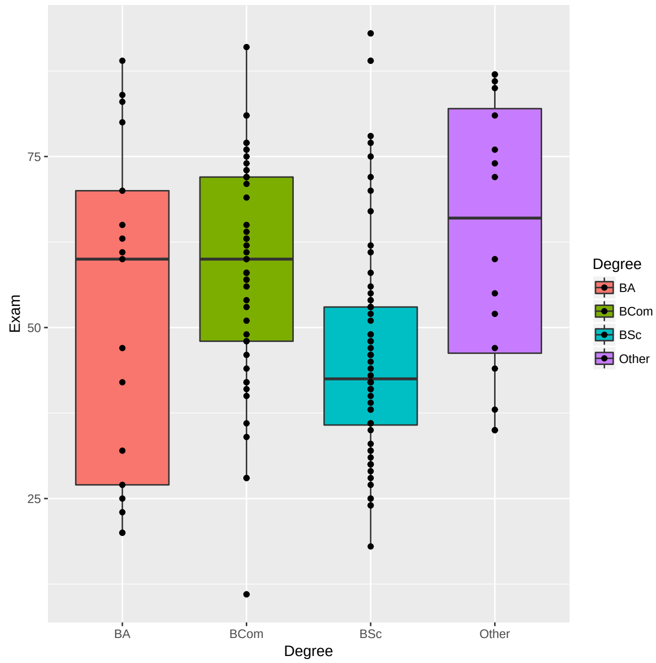

Figure 1: Boxplot of Exam Vs Degree

(a) What are the **quantitative** variables?

(b) What are the **qualitative** variables?

(c) What **Degree** shows the smallest median Exam grade? see figure 1

(d) What **Degree** has the smallest Exam Interquartile range? see figure 1

(e) What does `aes` stand for in Line A ?

3. If $Y \sim Bern(\theta)$ then:

$$P(y) = \theta^y (1-\theta)^{1-y}, \quad y \in \{0, 1\}$$

(a) Show that
$$M_Y(t) = q + pe^t$$
where $M_Y(t) = E(e^{yt})$

(b) Using $\mu_Y = E(Y) = \sum_{i=1}^{n} y_i p(y_i)$ (Not the MGF) find μ_Y.

(c) Using $\sigma^2 = E((Y-\mu)^2) = \sum_{i=1}^{n} (y_i - \mu)^2 p(y_i)$ (Not the MGF) find σ_Y^2.

(d) Use the moment generating function $M_Y(t)$ to find μ_Y.

(e) Use the moment generating function $M_Y(t)$ to find σ_Y^2.

4. To standardize a data set, x, we often use a z transformation, where

$$z_i = \frac{x_i - \overline{x}}{s}$$

To standardize a population with variable X we often use a Z transformation, where

$$Z = \frac{X - \mu}{\sigma}$$

Answer the following:

(a) Predict the output from Line A below!

```
> head(ddt)
  RIVER MILE  SPECIES LENGTH WEIGHT DDT
1   FCM    5 CCATFISH   42.5    732  10
2   FCM    5 CCATFISH   44.0    795  16
3   FCM    5 CCATFISH   41.5    547  23
4   FCM    5 CCATFISH   39.0    465  21
5   FCM    5 CCATFISH   50.5   1252  50
6   FCM    5 CCATFISH   52.0   1255 150
> x=ddt$WEIGHT
> z=(x-mean(x))/sd(x)
> round(mean(z)^2,4) # Line A
```

(b) Predict the output of Line B below!

```
> head(ddt)
  RIVER MILE  SPECIES LENGTH WEIGHT DDT
1   FCM    5 CCATFISH   42.5    732  10
2   FCM    5 CCATFISH   44.0    795  16
3   FCM    5 CCATFISH   41.5    547  23
4   FCM    5 CCATFISH   39.0    465  21
5   FCM    5 CCATFISH   50.5   1252  50
6   FCM    5 CCATFISH   52.0   1255 150
> x=ddt$DDT
> z=(x-mean(x))/sd(x)
> round(var(z),4) # Line B
```

(c) Prove $E(Z) = 0$ and $V(Z) = 1$

(d) Prove $\bar{z} = 0$ and $s_z^2 = 1$

(e) Suppose $Y = 3Z + 5$ and $X = 2Y$ find $V(X)$ where $Z \sim N(0, 1)$.

Figure 2: Fuses

5. From MS Ch. 4 page 161,2. A manufacturer uses electrical fuses in an electronic system. The fuses are purchased in large lots and tested sequentially until the first defective fuse is observed. Assume that the lot contains 10% defective fuses. Please note that there is some R code below in which one or more lines might be helpful in answering one of the parts.

(a) What is the name of the distribution appropriate for this problem?

(b) What is the probability that the first defective fuse will be one of the first five fuses tested?

(c) Suppose Y is the number of fuses tested until the first defective fuse is observed.

  (i) Find the mean of Y

  (ii) Find the variance of Y

  (iii) Find the standard deviation of Y

```
myd = function(y,r,p){
  choose(y-1,r-1)*p^r*(1-p)^(y-r)
}

myp = function(y,r,p){
  y=1:y
  sum(choose(y-1,r-1)*p^r*(1-p)^(y-r))
}

> dbinom(1:5,5,p=0.1)
[1] 0.32805 0.07290 0.00810 0.00045 0.00001
> dbinom(1:5,5, p=0.9)
[1] 0.00045 0.00810 0.07290 0.32805 0.59049
> myd(1:5,1, 0.9)
[1] 9e-01 9e-02 9e-03 9e-04 9e-05
> myd(1:5,1,0.1)
[1] 0.10000 0.09000 0.08100 0.07290 0.06561
> myp(5,1,0.1)
[1] 0.40951
```

6. Testing Problem: Suppose a drug test is 98% sensitive and 97% specific. That is, the test will produce 98% true positive results for drug users and 97% true negative results for non-drug users. Suppose that 0.4% of people are users of the drug. We need to find the solution to **the question**: *If a randomly selected individual tests positive (+), what is the probability he or she is a User (U)*

Bayes' rule:

$$p(A_i|B) = \frac{p(A_i)p(B|A_i)}{p(B)}$$

(a) Write down Bayes' rule needed to answer the above **question** in terms of $+, U, \overline{U}$.

(b) **In the case of the above testing problem** write down the expression for $p(B)$ in terms of a summation.

(c) Identify the prior in the testing problem by writing down its algebraic expression.

(d) If a randomly selected individual tests positive (+), what is the probability he or she is a User (U)?

(e) If a randomly selected individual tests positive (+), what is the probability he or she is not a User $(\overline{U})$

7. A new drug was administered to 10 randomly selected people. Suppose the number of people showing a positive effect is X and the number of people who experienced no positive effect be Y. The drug is positively effective with probability 0.7. Find the following by looking at the appropriate output

(a) $P(X = 10)$

(b) $P(X \leq 2)$

(c) $P(X < 2)$

(d) $P(X > 5)$

(e) $P(8 \leq Y \leq 10)$

```
round(dbinom(0:10,size=10,prob=0.7),4)
 [1] 0.0000 0.0001 0.0014 0.0090 0.0368 0.1029 0.2001 0.2668 0.2335 0.1211 0.0282
> round(dbinom(0:10, size=10,prob=0.3),4)
 [1] 0.0282 0.1211 0.2335 0.2668 0.2001 0.1029 0.0368 0.0090 0.0014 0.0001 0.0000
> round(pbinom(0:10, size=10,prob=0.7),4)
 [1] 0.0000 0.0001 0.0016 0.0106 0.0473 0.1503 0.3504 0.6172 0.8507 0.9718 1.0000
> round(pbinom(0:10, size=10,prob=0.3),4)
 [1] 0.0282 0.1493 0.3828 0.6496 0.8497 0.9527 0.9894 0.9984 0.9999 1.0000 1.0000
```