STA 3032 - Practice set modules 1&2

True or False

1. A quality engineer in a factory is interested in the proportion of all computer chips that her assembly line produces that meet a particular quality requirement. She selects a sample of 100 chips and finds that 85 pass the test. This means that $p = 0.85$ is the proportion of all chips meeting the requirement.
   False

2. The (large-sample) sampling distribution for the sample mean is normally distributed ONLY if the underlying population of measurements is normal.
   False

Open Ended Problems

1. A plant researcher has measured the amount of growth among five plants that received a growth formula. He finds that the amounts of growth were: 4, 8, 6, 7, 10 cm. Give the sample mean, median, standard deviation and 33rd percentile of the amounts of growth.
   Answer:

   (a) mean:
   $$\bar{x} = \frac{4+8+6+7+10}{5} = 7$$

   (b) median: located at $0.5(5+1) = 3$ position of *ordered data*, so 7

   (c) standard deviation: 2.236

   (d) 33rd percentile: located at $0.33(5+1) = 1.98$ position of *ordered data*, so $0.02(4) + 0.98(6)$

2. Among male birds of a species, 20% have a particular gene. Among females of the species, 10% have the gene. The males comprise 40% of all the birds of the species (thus, females comprise 60%).

   (a) What is the probability a randomly selected bird of this species has the gene?
   Answer:

$$\begin{aligned} P(\text{gene}) &= P(\text{gene}|\text{male})P(\text{male}) + P(\text{gene}|\text{female})P(\text{female}) \\ &= (0.2)(0.4) + (0.1)(0.6) \\ &= 0.14 \end{aligned}$$

   (b) What is the probability the bird is male, given it has the gene?
   Answer:

$$P(\text{male}|\text{gene}) = \frac{P(\text{gene}|\text{male})P(\text{male})}{P(\text{gene})} = \frac{(0.2)(0.4)}{0.14}$$

3. A sample of 5 animals of a particular species is selected at random from the population being managed in a wildlife refuge. If 5% (0.05) of the population have a particular trait, what is the probability that none of the 5 tested have the trait?
   Answer: Let X stand for number of animals with trait, hence $X \sim \text{Binomial}(5, 0.05)$, so

$$P(X = 0) = \binom{5}{0}(0.05)^0(0.95)^{5-0} = 0.95^5$$

   Computer can find this directly using in `R: dbinom(0,5,0.05)`

4. You are going to a foreign nation to conduct your research. On a weather website you see that the average high temperature during the period you will be there has been historically 20 degrees Celsius and variance 5 degrees^2. What is the mean and variance of the dataset in degrees Fahrenheit? Hint: $°\text{F} = 32 + (9/5)°\text{C}$
   Answer:

   (a) $E(F) = E(32 + (9/5)C) = 32 + (9/5)E(C) = 32 + (9/5)(20) = 68$
   (b) $V(F) = (9/5)^2V(C) = (9/5)^2(5) = 81/5$

5. In a population of 100 watt light bulbs manufactured by a company, 80% (0.80 as a proportion) have lifetimes exceeding 800 hours. An inspector samples 10 bulbs at random. What is the probability at least 2 of the bulbs lifetimes exceed 800 hours?
Answer: Let X stand for number of bulbs with a lifetime that exceed 800 hours, hence $X \sim \text{Binomial}(10, 0.80)$, so

$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) \\ &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - \binom{10}{0}(0.8)^0(0.2)^{10} - \binom{10}{1}(0.8)^1(0.2)^9 \\ &= 0.999996 \end{aligned}$$

Computer can find this directly using in `R`: `1-pbinom(1,10,0.8)`

6. Among students taking a standardized exam, scores are normally distributed with a mean of 550 and standard deviation 100.

   (a) What proportion of the students score above 700?
   Answer:

$$\begin{aligned} P(\text{score} > 700) &= P\left(Z > \frac{700 - 550}{100}\right) \\ &= P(Z > 1.5) \\ &= 1 - P(Z \leq 1.5) \\ &= 1 - 0.9332 = 0.0668 \end{aligned}$$

   Computer can find this directly using in `R`: `1-pnorm(700,550,100)`

   (b) What is 15th percentile?
   Answer: The 15th percentile of Z is -1.036433. Hence, the 15th percentile of scores is $100(-1.036433) + 550 = 446.3567$. Computer can find this directly using in `R`: `qnorm(0.15,550,100)`

7. Based on the following contingency table, complete the following parts:

	Concussion	No Concussion	Total
Male	40	25596	25636
Female	60	27107	27167
Total	100	52703	52803

(a) What is the probability of being female and have no concussion?
Answer: 27107/52803

(b) Among Males, what is Probability of concussion?
Answer: 40/25636

(c) Among individuals with concussions, what is the probability of being male?
Answer: 40/100

(d) Among Females, what is Probability of concussion?
Answer: 60/27167

8. A random variable has p.d.f.

$$f(x) = \frac{1}{10}x^3 \qquad 1 < x < 41^{1/4}.$$

Find the mean, variance, 85th percentile and $P(X \leq 2)$.
Answer:

(a) mean:

$$E(X) = \int_1^{41^{1/4}} x\frac{1}{10}x^3 \, dx \approx 2.055$$

In wolframalpha input

```
integrate 1/10*x^4 dx from 1 to 41^(1/4)
```

(b) variance: Note that $E(X^2) \approx 4.3588$, so $V(X) \approx 4.3588 - 2.055^2$

(c) 85th percentile. Denote it by t.

$$F(t) = \int_1^t \frac{1}{10}x^3 \, dx = \frac{1}{40}(t^4 - 1)$$

Hence we need so solve

$$\frac{1}{40}(t^4 - 1) = 0.85$$

so $t = 2.4323$. In wolframalpha input

```
(integrate 1/10*x^3 dx from 1 to t) equals 0.85
```

(d)

$$P(X \leq 2) = \int_1^2 \frac{1}{10}x^3 \, dx = 0.375$$

In wolframalpha input

```
integrate 1/10*x^3 dx from 1 to 2
```

9. Consider the following discrete distrubution

x	1	5	7	8	9	10
p(x)	0.13	0.17	0.07	0.22	0.11	0.30

Find the mean, variance, 65th percentile and $P(X \leq 2)$.
Answer:

(a) mean: $\sum xp(x) = 7.22$

(b) variance: Note that $E(X^2) = \sum x^2 p(x) = 60.8$. Hence, $V(X) = 60.8 - 7.22^2 = 8.6716$

(c) Obtaining the c.d.f.

x	1	5	7	8	9	10
$p(x)$	0.13	0.17	0.07	0.22	0.11	0.30
$F(x)$	0.13	0.30	0.37	0.59	0.70	1

Hence the 65th percentile lies between 8 and 9.

$$\frac{0.70 - 0.65}{0.70 - 0.59}(8) + \frac{0.65 - 0.59}{0.70 - 0.59}(9) = 8.5455$$

(d) $P(X \leq 2) = P(X = 1) = 0.13$

10. Five applicants for a job are ranked according to ability, with applicant 1 being the best, number 2 being the second best, and so on. These ranking are unknown to an employer, who simply hires two applicants (for two identical positions) at random. What is the probability that this employer hires exactly one of the two best applicants? Hint: First find the total number of applicant selection possibilities and then calculate how many possibilities result in exactly one of the two best being selected.
Answer: From the best 2 choose 1, times, from the worst 3 choose 1, over the total number of ways.

$$\frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} = \frac{6}{10}$$

11. Let

$$f(x_1, x_2) = 3x_1, \quad 0 \leq x_2 \leq x_1 \leq 1$$

Find

(a) $P(0.2 < X_2 < 0.4)$
Answer:

$$\int_{0.2}^{0.4} \int_{x_2}^{1} 3x_1 dx_1 dx_2 = 0.272$$

In WolframAlpha:
`integrate (integrate 3x dx from y to 1) dy from 0.2 to 0.4`

(b) $P(X_2 < 0.2 | X_1 = 0.5)$
Answer: First we need to find the conditional distribution

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)} = \frac{3x_1}{\int_0^{x_1} 3x_1 dx_2} = \frac{3x_1}{3x_1^2} = \frac{1}{x_1}, \quad 0 \leq x_2 \leq x_1$$

Hence,

$$P(X_2 < 0.2 | X_1 = 0.5) = \int_0^{0.2} (1/0.5) dx_2 = 0.4$$

(c) $\text{Cov}(X_1, X_2)$
Answer: Recall that $\text{Cov}(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$

$$E(X_1 X_2) = \int_0^1 \int_0^{x_1} (x_1 x_2) 3x_1 dx_2 dx_1 = 3/10$$
$$E(X_1) = \int_0^1 \int_0^{x_1} x_1 3x_1 dx_2 dx_1 = 3/4$$
$$E(X_2) = \int_0^1 \int_0^{x_1} x_2 3x_1 dx_2 dx_1 = 3/8$$

Hence, $\text{Cov}(X_1, X_2) = 3/10 - (3/4)(3/8) = 0.01875$

12. For a certain manufacturing industry, the number of industrial accidents average 3 per (6 work day) week. Find the probability that two accidents will occur in a given day.
Answer: We have a Poisson(3) when considering a week as our interval of time for occurrences to happen. Assuming the rate is fixed, for a 1 work day we have a $X \sim$Poisson(0.5). Hence,

$$P(X = 2) = \frac{e^{-0.5} 0.5^2}{2!}$$