CONCORDIA UNIVERSITY
Department of Economics

ECON 221/2 SECTIONS A, B, C and DD
STATISTICAL METHODS I
FALL 2016 – ASSIGNMENT 1 (SOLUTIONS)
Due: Monday, October 17, before 3:00 pm

Name: **I.D:** **Section:**

Points Total: 100 points

1. (**16 points**) Of the employees that work for Hapolonians Incorporated, 35 percent have an MBA and 40 percent are over the age of 35. Of those employees that have an MBA, 30 percent are over the age of 35.

   Let M be the number of employees that have an MBA and A be the number of employees over the age of 35. Then, $P(M) = 0.35$, $P(A) = 0.40$ *and* $P(A|M) = 0.30$.

   a. (**2 points**) Calculate the probability that a randomly chosen employee has both an MBA and is over the age of 35.

   $$P(M \cap A) = P(A|M)P(M) = 0.30 \cdot 0.35 = 0.105$$

   b. (**2 points**) Calculate the probability that a randomly chosen employee who is over the age of 35 has an MBA.

   $$P(M|A) = \frac{P(M \cap A)}{P(A)} = \frac{0.105}{0.40} = 0.2625$$

   c. (**2 points**) Calculate the probability that a randomly chosen employee has an MBA or is over the age of 35.

   $$P(M \cup A) = P(M) + P(A) - P(M \cap A) = 0.35 + 0.40 - 0.105 = 0.645$$

d. (**2 points**) Calculate the probability that a randomly chosen employee who is not over the age of 35 does not have an MBA.

$$P\left(\bar{M}\middle|\bar{A}\right)=\frac{P\left(\bar{M}\cap\bar{A}\right)}{P\left(\bar{A}\right)}=\frac{1-P\left(M\cup A\right)}{1-P\left(A\right)}=\frac{1-0.645}{1-0.40}=\frac{0.355}{0.60}=0.591\dot{6}$$

e. (**2 points**) Determine whether having an MBA is independent of being over the age of 35.

If they were independent, $P\left(M\cap A\right)=P\left(M\right)P\left(A\right)$. *From part (a),* $P\left(M\cap A\right)=0.105$, *but* $P\left(M\right)P\left(A\right)=0.35\cdot 0.40=0.14$. *Therefore, they are not independent.*

f. (**2 points**) Determine whether having an MBA and being over the age of 35 are mutually exclusive.

If they were mutually exclusive, $P\left(M\cap A\right)=0$. *From part (a),* $P\left(M\cap A\right)=0.105$. *Therefore, they are not mutually exclusive.*

g. (**2 points**) Determine whether having an MBA and being over the age of 35 are collectively exhaustive.

If they were collectively exhaustive, $P\left(M\cup A\right)=1$. *From part (c),* $P\left(M\cup A\right)=0.645$. *Therefore, they are not collectively exhaustive.*

h. (**2 points**) Complete the following table assuming Hapolonians Incoporated employs 200 people.

Number of employees	*A*	$\bar{A}$	*Total*
M	*21*	*49*	*70*
$\bar{M}$	*59*	*71*	*130*
Total	*80*	*120*	*200*

2. (**4 points**) Hapolonians Incorporated's 12-member board of directors is to be selected from a list of 8 men and 8 women.

   a. (**2 points**) Calculate how many different boards of directors are possible.

$$C_{12}^{16} = \frac{16!}{4!12!} = 1820$$

   b. (**2 points**) Calculate the probability that a majority of the board of directors will be men.

$$\Pr(M \geq 7) = \frac{\left(C_8^8 \cdot C_4^8\right) + \left(C_7^8 \cdot C_5^8\right)}{C_{12}^{16}} = \frac{\left(\frac{8!}{0!8!} \cdot \frac{8!}{4!4!}\right) + \left(\frac{8!}{1!7!} \cdot \frac{8!}{3!5!}\right)}{1820} = \frac{(1 \cdot 70) + (8 \cdot 56)}{1820} = 0.2846$$

3. (**12 points**) Delays plague the airport in Flin Flon, Manitoba. The probability distribution for the number of delays on any given day is as follows:

Number of delays	0	1	2	3	4
Probability	0.10	0.26	0.42	0.16	0.06

   a. (**2 points**) Calculate the cumulative probability distribution.

Number of delays	0	1	2	3	4
Cumulative probability	0.10	0.36	0.78	0.94	1.00

   b. (**2 points**) Calculate the probability that no more than three flights are delayed on any given day.

$$\Pr(x \leq 3) = 1 - \Pr(x = 4) = 1 - 0.06 = 0.94 \quad or \quad \Pr(x \leq 3) = F(x = 3) = 0.94$$

   c. (**2 points**) Calculate the expected number of delayed flights on any given day.

$$E(X) = \sum_{x=0}^{4} x \Pr(x) = 0 \cdot 0.10 + 1 \cdot 0.26 + 2 \cdot 0.42 + 3 \cdot 0.16 + 4 \cdot 0.06 = 1.82$$

   d. (**2 points**) Calculate the standard deviation of delayed flights on any given day.

$$S(X) = \sqrt{\sum_{x=0}^{4} x^2 \Pr(x) - E(x)^2} = \sqrt{0^2 \cdot 0.10 + 1^2 \cdot 0.26 + 2^2 \cdot 0.42 + 3^2 \cdot 0.16 + 4^2 \cdot 0.06 - 1.82^2} = 1.0137$$

Each delay costs the airport \$1500.

e. (**2 points**) Calculate the expected cost of delayed flights on any given day.

$$Y = 1500X \Rightarrow E(Y) = E(1500X) = 1500E(X) = 1500 \cdot 1.82 = 2730$$

f. (**2 points**) Calculate the variance of the cost of delayed flights on any given day.

$$Y = 1500X \Rightarrow V(Y) = V(1500X) = 1500^2 \cdot V(X) = 2250000 \cdot 1.0137^2 = 2312100$$

4. (**8 points**) In US football, a team that scores a touchdown must decide whether to try to get one or two bonus points. If it succeeds, it receives the number of bonus points that it attempted to get; if it fails, it gets nothing. One coach believes that the probability of succeeding on a two-point bonus play is 35 percent and that each attempt is an independent event. In one particular game, the coach's team attempted the two-point bonus play four times.

a. (**2 points**) Calculate the probability that at least two of these attempts were successful.

This question involves using a binomial distribution. The cumulative probability distribution is found in Table 3. From there, $\Pr(X \geq 2) = 1 - \Pr(X \leq 1) = 1 - 0.563 = 0.437$.

b. (**2 points**) Calculate the expected total number of points resulting from the four attempts.

$E(X) = nP = 4 \cdot 0.35 = 1.4$. *Since each successful attempt is worth two points,*

$Z = 2X \Rightarrow E(Z) = E(2X) = 2E(X) = 2 \cdot 1.4 = 2.8$.

c. (**2 points**) Calculate the standard deviation of the total number of points resulting from the four attempts.

$S(X) = \sqrt{nP(1-P)} = \sqrt{4 \cdot 0.35 \cdot 0.65} = 0.9539$.

$Z = 2X \Rightarrow S(Z) = S(2X) = 2S(X) = 2 \cdot 0.9539 = 1.9079$.

d. (**2 points**) If the coach were to try the easier one-point bonus play all four times, calculate the required success rate to achieve the same expected outcome obtained in part (b).

To achieve the same expected outcome, $E(X) = nP = 2.8$. *Since* $n = 4$*, then* $P = 0.7$.

5. (**4 points**) Sheep escape Michelle's farm according to a Poisson distribution at an average of three per week.

   a. (**2 points**) Calculate the probability that more than one sheep escapes in any given week.

      This question involves using a Poisson distribution. The cumulative probability distribution is found in Table 6. From there, $\Pr(X > 1) = 1 - \Pr(X \le 1) = 1 - 0.1991 = 0.8009$.

   b. (**2 points**) Michelle can be 99-percent certain that no more than this many sheep escape in any given week.

      The cumulative probability distribution is found in Table 6. From there, $\Pr(X \le 7) = 0.9881$ *and* $\Pr(X \le 8) = 0.9962$. *Therefore, Michelle can be 99-percent certain that no more than seven sheep escape in any given week.*

6. (**12 points**) Corduroy and Tootie each run their own hedge fund, although it is thought that they share investment strategies because they are such close friends. It is found that they both make money 70 percent of the time and they both lose money 30 percent of the time, but Corduroy never makes money when Tootie loses money and Tootie never makes money when Corduroy loses money. Let $X = 1$ represent when Corduroy makes money and $X = 0$ when he loses money. Similarly, let $Y = 1$ represent when Tootie makes money and $Y = 0$ when he loses money.

   a. (**2 points**) Calculate the marginal probability distributions for *X* and *Y*.

      $\Pr(X = 0) = \Pr(X = 0 \cap Y = 0) + \Pr(X = 0 \cap Y = 1) = 0.3 + 0.0 = 0.3$
      $\Pr(X = 1) = \Pr(X = 1 \cap Y = 0) + \Pr(X = 1 \cap Y = 1) = 0.0 + 0.7 = 0.7$
      The marginal probability distribution for Y is identical to X.

   b. (**2 points**) Calculate the expected values of *X* and *Y*.

      $E(X) = \sum_{x=0}^{1} x \Pr(x) = 0 \cdot 0.3 + 1 \cdot 0.7 = 0.7$. *The expected value of Y is identical to X.*

   c. (**2 points**) Calculate the variances of *X* and *Y*.

      $V(X) = \sum_{x=0}^{1} x^2 \Pr(x) - E(X)^2 = 0^2 \cdot 0.3 + 1^2 \cdot 0.7 - 0.7^2 = 0.21$. *The variance of Y is identical to X.*

d. (**2 points**) Calculate the covariance of *X* and *Y*.

$$\operatorname{cov}(X,Y)=\sum_{x=0}^{1}\sum_{y=0}^{1}xy\Pr(x,y)-E(X)E(Y)=0\cdot 0\cdot 0.3+0\cdot 1\cdot 0.0+1\cdot 0\cdot 0.0+1\cdot 1\cdot 0.7-0.7\cdot 0.7=0.21$$

e. (**2 points**) Calculate the correlation for *X* and *Y*.

$$\operatorname{cor}(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sqrt{V(X)V(Y)}}=\frac{0.21}{\sqrt{0.21\cdot 0.21}}=1$$

f. (**2 points**) Based on your answer to part (h), do you believe that Corduroy and Tootie share investment strategies?

A correlation of +1 means that there is a perfect positive linear relationship between their investment strategies. This would support the belief that they share investment strategies.

7. (**10 points**) The random variable *X* follows a normal distribution with $\mu = 0.2$ and $\sigma^2 = 0.0025$.

By converting X into Z, we can use the cumulative distribution function data in Table 1.

a. (**2 points**) Calculate $\Pr(X>0.4)$.

$$\Pr(X>0.4)=\Pr\left(\frac{X-\mu}{\sigma}>\frac{0.4-0.2}{\sqrt{0.0025}}\right)=\Pr(Z>4)=1-\Pr(Z<4)=0.0000$$

b. (**2 points**) Calculate $\Pr(0.15<X<0.28)$.

$$\Pr(0.15<X<0.28)=\Pr\left(\frac{0.15-0.20}{\sqrt{0.0025}}<\frac{X-\mu}{\sigma}<\frac{0.28-0.20}{\sqrt{0.0025}}\right)=\Pr(-1<Z<1.6)$$
$$=\Pr(Z<1.6)-\Pr(Z<-1)=\Pr(Z<1.6)-\Pr(Z>1)=\Pr(Z<1.6)-\left(1-\Pr(Z<1)\right)$$
$$=\Pr(Z<1.6)+\Pr(Z<1)-1=0.9452+0.8413-1=0.7865$$

c. (**2 points**) Calculate $\Pr(X<0.1)$.

$$\Pr(X<0.1)=\Pr\left(\frac{X-\mu}{\sigma}<\frac{0.1-0.2}{\sqrt{0.0025}}\right)=\Pr(Z<-2)=\Pr(Z>2)=1-\Pr(Z<2)=1-0.9772=0.0228$$

d. (**2 points**) Calculate the value of *x* such that $\Pr\left(X > x\right) = 0.025$.

$$\Pr\left(X > x\right) = \Pr\left(\frac{X-\mu}{\sigma} > \frac{x-0.2}{\sqrt{0.0025}}\right) = \Pr\left(Z > z\right) = 0.025 \Rightarrow 1-\Pr\left(Z < z\right) = 0.975 \Rightarrow z = 1.96$$

$$\Rightarrow \frac{x-0.2}{\sqrt{0.0025}} = 1.96 \Rightarrow x = 0.298$$

e. (**2 points**) Calculate the values of x_0 and x_1 that are equidistant from the mean such that $\Pr\left(x_0 < X < x_1\right) = 0.34$.

$$\Pr\left(x_0 < X < x_1\right) = \Pr\left(\frac{x_0-0.2}{\sqrt{0.0025}} < \frac{X-\mu}{\sigma} < \frac{x_1-0.2}{\sqrt{0.0025}}\right) = \Pr\left(-z < Z < z\right) = \Pr\left(Z < z\right) - \Pr\left(Z < -z\right) =$$

$$\Pr\left(Z < z\right) - \Pr\left(Z > z\right) = \Pr\left(Z < z\right) - \left(1-\Pr\left(Z < z\right)\right) = 2\Pr\left(Z < z\right) - 1 = 0.34 \Rightarrow \Pr\left(Z < z\right) = 0.67$$

$$\Rightarrow z = \pm 0.44 \Rightarrow \begin{cases} \dfrac{x_0-0.2}{\sqrt{0.0025}} = -0.44 \Rightarrow x = 0.178 \\ \dfrac{x_1-0.2}{\sqrt{0.0025}} = 0.44 \Rightarrow x = 0.222 \end{cases}$$

8. (**4 points**) The amount of time that Professor Hapoleon spends with students during office hours averages 10 minutes and follows an exponential distribution.

a. (**2 points**) Calculate the probability that he spends more than 20 minutes with a given student.

$$\Pr\left(T > 20\right) = 1-\Pr\left(T < 20\right) = 1-\left(1-e^{-20/10}\right) = e^{-2} = 0.1353$$

b. (**2 points**) Calculate the probability that he spends between 10 and 15 minutes with a given student.

$$\Pr\left(10 < T < 15\right) = \Pr\left(T < 15\right) - \Pr\left(T < 10\right) = \left(1-e^{-15/10}\right) - \left(1-e^{-10/10}\right) = e^{-1} - e^{-1.5} = 0.1447$$

9. (**16 points**) A random sample of 12 statistics students have the following ages (in years):

21	22	27	36	18	19	22	23	22	28	41	33

a. (**10 points**) Calculate the various elements of the five-number summary.

Arranging the ages in order from youngest to oldest gives:

18	*19*	*21*	*22*	*22*	*22*	*23*	*27*	*28*	*33*	*36*	*41*

The minimum is 18, the maximum is 41. The median is the average of values in the 6th and 7th ordered positions: $\frac{22+23}{2}=22.5$. *The 1st quartile is the value in the* $0.25(12+1)=3.25^{th}$ *ordered position:* $21+0.25(22-21)=21.25$. *The 3rd quartile is the value in the* $0.75(12+1)=9.75^{th}$ *ordered position:* $27+0.75(28-27)=27.75$.
The five-number summary is 18, 21.25, 22.5, 27.75, 41.

b. (**2 points**) Calculate the sample mean.

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{12}(18+19+21+22+22+22+23+27+28+33+36+41)=26$$

c. (**2 points**) Calculate the sample standard deviation.

$$s=\sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)}$$
$$=\sqrt{\frac{1}{12-1}\left(18^2+19^2+21^2+22^2+22^2+22^2+23^2+27^2+28^2+33^2+36^2+41^2-12\cdot 26^2\right)}$$
$$=\sqrt{52.\overline{18}}=7.2237$$

d. (**2 points**) Use Chebyshev's theorem to state the range of ages, centered about the mean, that contains at least 75 percent of the observations.

Chebyshev's theorem states that at least 75 percent of the observations are within 2 standard deviations ($0.75=1-\frac{1}{k^2}\Rightarrow\frac{1}{k^2}=0.25\Rightarrow k^2=4\Rightarrow k=2$*) of the mean. Therefore,* $\bar{x}\pm 2s=26\pm 2\cdot 7.2237=(11.5526, 40.4473)$.

10. (**4 points**) A random sample of 25 ECON221 students needed the following number of hours to complete this assignment:

Completion time	0.00 – 3.99	4.00 – 7.99	8.00 – 11.99	12.00 – 15.99	16.00 – 20.00
Number of students	3	7	8	5	2

a. (**2 points**) Calculate the sample mean.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} f_i m_i = \frac{1}{25}\left(3\cdot 2 + 7\cdot 6 + 8\cdot 10 + 5\cdot 14 + 2\cdot 18\right) = 9.36$$

b. (**2 points**) Calculate the sample variance.

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} f_i m_i^2 - n\bar{x}^2\right) = \frac{1}{25-1}\left(3\cdot 2^2 + 7\cdot 6^2 + 8\cdot 10^2 + 5\cdot 14^2 + 2\cdot 18^2 - 25\cdot 9.36^2\right) = 20.90\dot{6}$$

11. (**10 points**) Excel Question.

Consider two random variables: X represents the throw of a fair die and Y satisfies $\Pr(Y=1) = \frac{\text{your birth month}}{15}$ and $\Pr(Y=2) = \Pr(Y=3) = \Pr(Y=4) = \Pr(Y=5) = \Pr(Y=6)$. The sample space for each variable is {1, …, 6}.

a. (**2 points**) Fill in columns 2 and 3 in the following table where each entry is the probability of the random variable having the indicated outcome.

My birthday is in November, so $\Pr(Y=1) = \frac{11}{15}$.

Therefore, $\Pr(Y=2) = \Pr(Y=3) = \Pr(Y=4) = \Pr(Y=5) = \Pr(Y=6) = \frac{1}{5}\cdot\left(1 - \frac{11}{15}\right) = \frac{4}{75}$

b. (**3 points**) In Excel, simulate 100 draws of the random variable *X*; then 100 draws of the variable *Y*. Record the results as proportions in columns 4 and 7. Then repeat the simulations where the number of draws is 1000; record the results in columns 5 and 8. Finally, take 10000 draws of the two variables and record results in columns 6 and 9.

			X variable # of draws			*Y* variable # of draws		
N	P(*X* = n)	P(*Y* = n)	100	1000	10000	100	1000	10000
1	*0.1666*	*0.7333*	*0.14*	*0.15*	*0.17*	*0.75*	*0.72*	*0.73*
2	*0.1666*	*0.0533*	*0.14*	*0.18*	*0.17*	*0.07*	*0.04*	*0.05*
3	*0.1666*	*0.0533*	*0.19*	*0.16*	*0.17*	*0.05*	*0.06*	*0.06*
4	*0.1666*	*0.0533*	*0.18*	*0.15*	*0.16*	*0.06*	*0.05*	*0.05*
5	*0.1666*	*0.0533*	*0.20*	*0.17*	*0.17*	*0.03*	*0.05*	*0.05*
6	*0.1666*	*0.0533*	*0.15*	*0.20*	*0.16*	*0.04*	*0.08*	*0.06*

c. (**2 points**) ***Briefly*** comment on the results.

As the number of draws increase, the proportions should limit to the theoretical outcome.

d. (**3 points**) Attach two pages. On one, place the 3 histograms for the *X* variable; on the other, the 3 histograms for the *Y* variable.