CONCORDIA UNIVERSITY
Department of Economics

ECON 221/2 SECTIONS A, B and BB
STATISTICAL METHODS I
FALL 2014 – ASSIGNMENT 1 (SOLUTIONS)
Due: Friday, September 26, before 4:00 pm

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

Points Total: 100 points

1. **(25 points)** A small grocery store has both an express checkout line and a super-express checkout line. Customers with 12 or fewer items are allowed to use the express line; those with 5 or fewer items can use the super-express line. Let X and Y be the number of customers in the express checkout line and the super-express checkout line, respectively. The joint probability distribution of X and Y is given in the table below.

	$Y=0$	$Y=1$	$Y=2$	$Y=3$	$Y\geq 4$
$X=0$	0.00	0.04	0.06	0.07	0.06
$X=1$	0.08	0.06	0.05	0.00	0.00
$X=2$	0.07	0.15	0.04	0.03	0.01
$X\geq 3$	0.04	0.05	0.10	0.04	0.05

a. Find the marginal distribution of X. What does this distribution tell you? (5 points)

$$p(X=x)=p(x,0)+p(x,1)+p(x,2)+p(x,3)+p(x,\geq 4)$$

X	$X=0$	$X=1$	$X=2$	$X\geq 3$
$p(X=x)$	0.23	0.19	0.30	0.28

b. Find the marginal distribution of Y. What does this distribution tell you? (5 points)

$$p(Y=y)=p(0,y)+p(1,y)+p(2,y)+p(\geq 3,y)$$

y	$Y=0$	$Y=1$	$Y=2$	$Y=3$	$Y\geq 4$
$p(Y=y)$	0.19	0.30	0.25	0.14	0.12

The marginal distribution shows the probability of observing a particular number of customers in the express checkout line (X) and super-express checkout line (Y).

c. On average, how many customers would you expect to see in each of these two lines at the grocery store? (5 points)

$$E(X) = \sum_{x=0}^{3} x \cdot p(x) = 0 \cdot 0.23 + 1 \cdot 0.19 + 2 \cdot 0.30 + 3 \cdot 0.28 = 1.63$$

$$E(Y) = \sum_{x=0}^{3} y \cdot p(y) = 0 \cdot 0.19 + 1 \cdot 0.30 + 2 \cdot 0.25 + 3 \cdot 0.14 + 4 \cdot 0.12 = 1.70$$

d. What is the probability that no more than two customers are waiting in both lines combined? (5 points)

$$\begin{aligned} p(X \le 2, Y \le 2) &= p(0,0) + p(0,1) + p(0,2) + p(1,0) + p(1,1) + p(2,0) \\ &= 0.00 + 0.04 + 0.06 + 0.08 + 0.06 + 0.07 = 0.31 \end{aligned}$$

e. By inspection of the probabilities P($X = 2$), P($Y = 2$), and P($X = 2$; $Y = 2$), are X and Y independent random variables? Explain. (5 points)

Independence means that $p(x,y) = p(x) \cdot p(y)$. *From part (a),* $p(X = 2) = 0.30$. *From part (b),* $p(Y = 2) = 0.25$. *From the table,* $p(X = 2, Y = 2) = 0.04$. *Since* $p(X = 2, Y = 2) = 0.04 \neq p(X = 2) \cdot p(Y = 2) = 0.30 \cdot 0.25 = 0.075$, *we conclude that X and Y are not independent random variables.*

2. **(20 points)** An instructor develops a four-option, multiple-choice exam for an introductory statistics course. A student decides to guess randomly on each question. Given that the probability of guessing correctly on each question is 0.25, answer the following questions.

   a. Suppose the exam has 20 questions;

      i. What is the probability that the student will correctly guess exactly six (6)? (2 points)

         This question involves using a binomial distribution. The probability function for the binomial distribution can be found in Table 2. From that, we can calculate $p(X=6)=C_6^{20}\cdot 0.25^6\cdot 0.75^{14}=0.1686$.

      ii. What is the probability that the student will correctly guess between 6 and 11 (inclusive)? (3 points)

         This question involves using a binomial distribution. Cumulative binomial probabilities can be found in Table 3. From that, we can calculate $p(6\le X\le 11)=p(X\le 11)-p(X\le 5)=0.999-0.617=0.382$.

      iii. What is the probability that the student will correctly guess at most 10 questions? (3 points)

         This question involves using a binomial distribution. Cumulative binomial probabilities can be found in Table 3. From that, we can calculate $p(X\le 10)=0.996$.

iv. What is the probability that a student will correctly guess at least 3 questions? (3 points)

This question involves using a binomial distribution. Cumulative binomial probabilities can be found in Table 3. From that, we can calculate $p(X \geq 3) = 1 - p(X \leq 2) = 1 - 0.091 = 0.909$.

v. What is the average grade that you would expect the student to obtain by guessing? (3 points)

This question involves using a binomial distribution. The mean of the distribution is $\mu = E(X) = np = 20 \cdot 0.25 = 5$.

vi. What is the standard deviation of the grade to be obtained by guessing? (3 points)

This question involves using a binomial distribution. The standard deviation of the distribution is $\sigma = \sqrt{\sigma^2} = \sqrt{np(1-p)} = \sqrt{20 \cdot 0.25 \cdot (1 - 0.25)} = 1.94$.

b. Suppose there are only 5 questions, what is the probability of scoring 100% by guessing? (3 points)

This question *involves using a binomial distribution. The probability function for the binomial distribution can be found in Table 2. From that, we can calculate* $p(X = 5) = C_5^5 \cdot 0.25^5 \cdot 0.75^0 = 0.001$.

3. **(20 points**)

a. A student committee has 10 members: 7 undergraduates and 3 graduates. A subcommittee of 3 members is to be chosen randomly, so that each possible combination of 3 of the 10 students is equally likely to be chosen. What is the probability that there are no graduate students on the subcommittee? (10 points)

The number of combinations that involve a subcommittee of 3 members chosen from 10 available members is $C_3^{10} = \frac{10!}{3!(10-3)!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120$ *and the number of combinations that involve a subcommittee of 3 members chosen only from 7 undergraduates is* $C_3^7 = \frac{7!}{3!(7-3)!} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 35$. *Therefore, the probability that there are no graduate students on the subcommittee is* $= \frac{C_3^7}{C_3^{10}} = \frac{35}{120} = 0.2917$.

b. A bank classifies borrowers as either high risk or low risk. Only 20% of its loans are made to those in the high-risk category. Of all its loans, 10% are in default, 50% of those in default were made to high-risk borrowers. What is the probability that a high-risk borrower will default? (10 points)

	High Risk	Low Risk	Total
Default	0.05 (0.50 x 0.10)		0.10
No Default			
Total	0.20		

$$p(D|HR) = \frac{p(HR \cap D)}{p(HR)} = \frac{0.05}{0.20} = 0.25$$

4. (**20 points in total; 4 points each**) Let the random variable X follow a normal distribution with $\mu = 0.5$ and $\sigma^2 = 0.07$.

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

a. Find the probability that X is greater than 0.4.

$$p(X > 0.4) = p\left(Z > \frac{0.4 - 0.5}{\sqrt{0.07}} = -0.38\right) = p(Z < 0.38) = 0.6480$$

b. Find P($0.37 < X < 0.56$).

$$p(0.37 < X < 0.56) = p\left(-0.11 = \frac{0.37 - 0.40}{\sqrt{0.07}} < Z < \frac{0.56 - 0.40}{\sqrt{0.07}} = 0.04\right)$$
$$= p(Z < 0.04) - (1 - p(Z < 0.11)) = p(Z < 0.04) + p(Z < 0.11) - 1$$
$$= 0.5160 + 0.5398 - 1 = 0.0558$$

c. Find P($X < 0.25$).

$$p(X < 0.25) = p\left(Z < \frac{0.25 - 0.40}{\sqrt{0.07}} = -0.57\right) = p(Z > 0.57) = 1 - p(Z < 0.57)$$
$$= 1 - 0.7157 = 0.2843$$

d. The probability is 0.30 that X is greater than what number?

$$p(X > x) = p\left(Z > \frac{x - 0.4}{\sqrt{0.07}}\right) = 0.3 \Rightarrow p\left(Z < \frac{x - 0.4}{\sqrt{0.07}}\right) = 0.7$$
$$\frac{x - 0.4}{\sqrt{0.07}} \approx 0.525 \Rightarrow x \approx 0.5389$$

e. Find x, such that P($-z < Z < z$) = 0.8690.

$$p(-z < Z < z) = 0.8690 \Rightarrow p(0 < Z < z) = 0.4395 \Rightarrow p\left(Z < \frac{x - 0.4}{\sqrt{0.07}}\right) = 0.9395$$
$$\frac{x - 0.4}{\sqrt{0.07}} \approx 1.55 \Rightarrow x \approx 0.8101$$

5. (**15 points**) Be sure to follow the instructions carefully.

   a. Consider a random variable X such that represents the throw of a fair die.

      Let Y be a random variable where P(Y = 1) = (Your birth month)/15; P(Y = 2), P(Y = 3), P(Y = 4), P(Y = 5) and P(Y = 6) are equally probable. The sample space for each variable is {1, …, 6}.

      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 $p(Y=1)=\frac{11}{15}$.

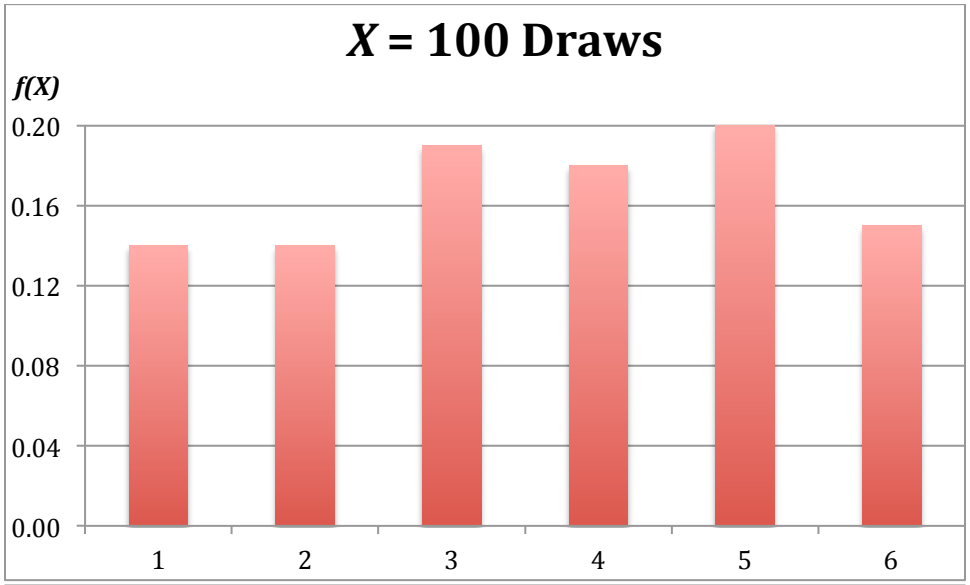
      Therefore, $p(Y=2)=p(Y=3)=p(Y=4)=p(Y=5)=p(Y=6)=\frac{1}{5}\cdot\left(1-\frac{11}{15}\right)=\frac{4}{75}$

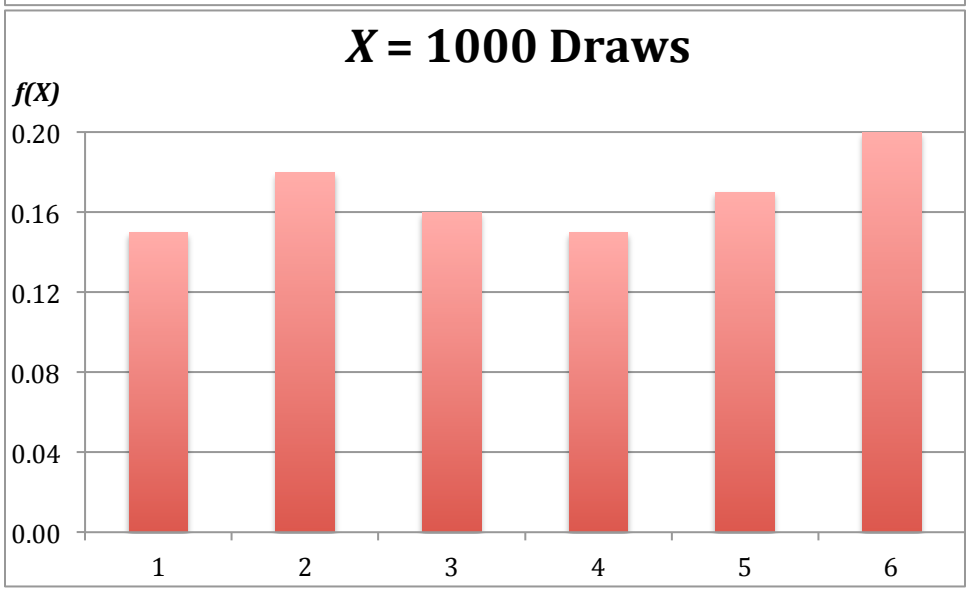
   b. In Excel, simulate 100 draws of the random variable X; then 100 draws of the variable Y. Record the results as proportions in columns 4 and 7. Then repeat the simulations where the number of draws is 1000; record the results in columns 5 and 8. Finally, take 10000 draws of the two variables and record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

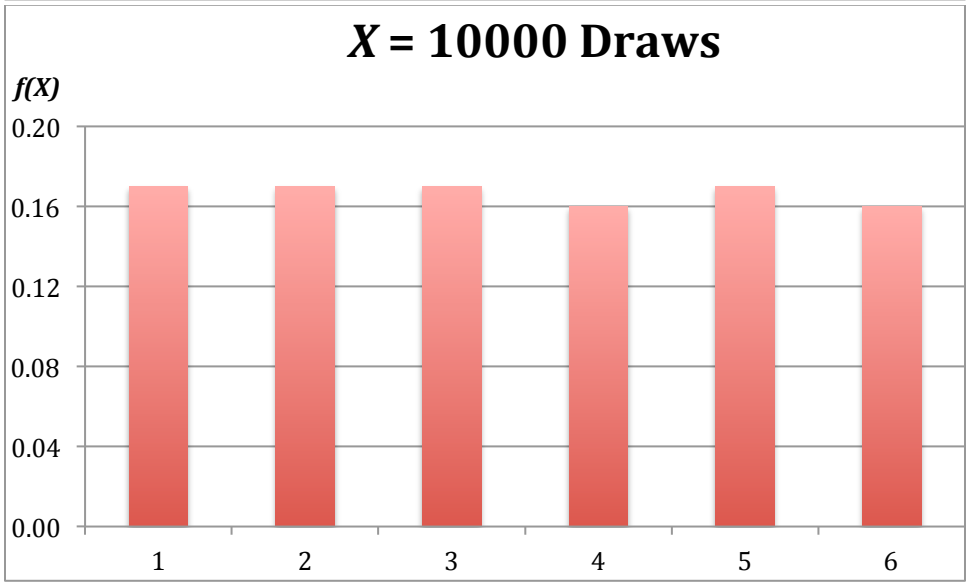
			X variable # of draws			Y Variable # of draws		
n	P(X = n)	P(Y = n)	100	1000	10000	100	1000	10000
1	*0.1666*	*0.7333*	*0.14*	*0.15*	*0.17*	*0.75*	*0.72*	*0.73*
2	*0.1666*	*0.0533*	*0.14*	*0.18*	*0.17*	*0.07*	*0.04*	*0.05*
3	*0.1666*	*0.0533*	*0.19*	*0.16*	*0.17*	*0.05*	*0.06*	*0.06*
4	*0.1666*	*0.0533*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

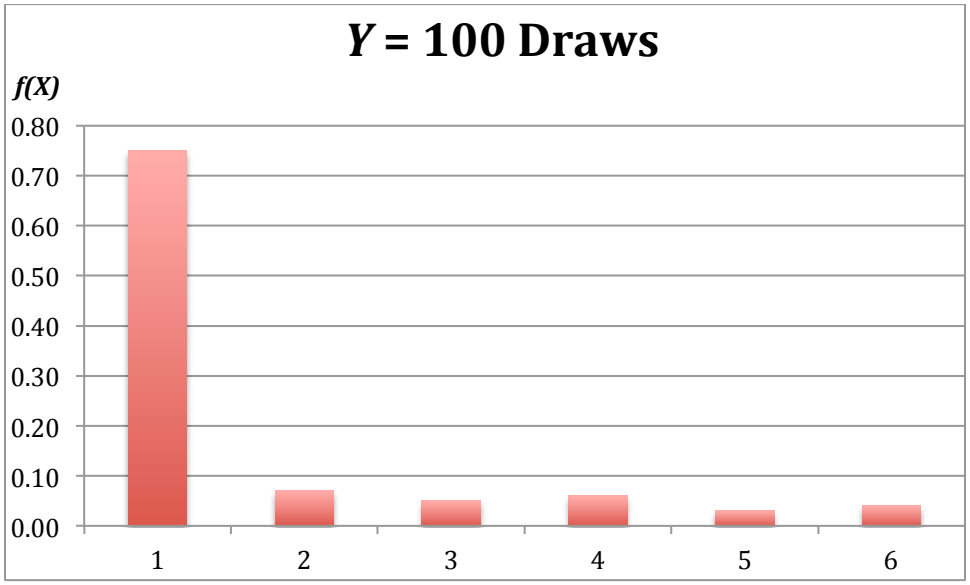
      Comment on the results.

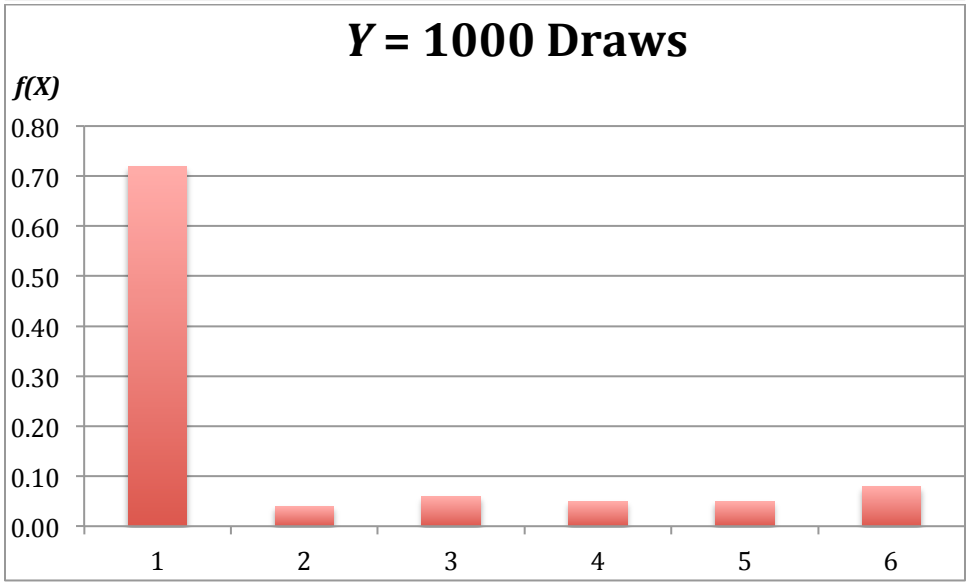
   c. Attach two pages. On one, place the 3 histograms for the X variable; on the other, the 3 histograms for the Y variable.

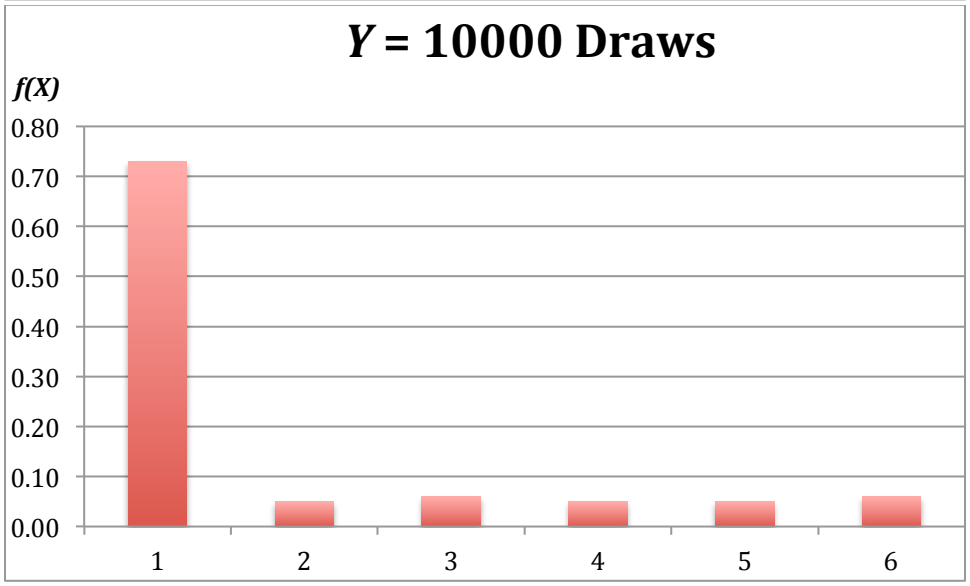

X = 100 Draws
f(X)
0.20
0.16
0.12
0.08
0.04
0.00
1
2
3
4
5
6


X = 1000 Draws
f(X)
0.20
0.16
0.12
0.08
0.04
0.00
1
2
3
4
5
6


X = 10000 Draws
f(X)
0.20
0.16
0.12
0.08
0.04
0.00
1
2
3
4
5
6


Y = 100 Draws
f(X)
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1
2
3
4
5
6


Y = 1000 Draws
f(X)
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1
2
3
4
5
6


Y = 10000 Draws
f(X)
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
1
2
3
4
5
6