Business Statistics 41000 Winter Quarter 2013 Practice Midterm Exam

1. (a) Answer TRUE or FALSE. A binomial distribution with $n = 100$ and $p = 0.65$ is best approximated by a normal distribution with mean $\mu = 65$ and variance $\sigma^2 = 18$.

   FALSE. The best approximation has mean $pn = 65$ and variance $np(1-p) = 22.75 \neq 18$.

   (b) Answer TRUE or FALSE. The expectation of X minus $2Y$ is the expectation of X minus twice the expectation of Y.

   TRUE, $\mathrm{E}(X - 2Y) = \mathrm{E}(X) - 2\mathrm{E}(Y)$ by the linearity of expectation.

   (c) Answer TRUE or FALSE. Suppose X and Y are independent random variables and $\mathrm{V}(X) = 6$ and $\mathrm{V}(Y) = 6$. Then $\mathrm{V}(X + Y) = \mathrm{V}(2X)$.

   FALSE, $\mathrm{V}(X + Y) = \mathrm{V}(X) + \mathrm{V}(Y) = 2\mathrm{V}(Y)$, while $\mathrm{V}(2X) = 4\mathrm{V}(X)$.

   (d) Answer TRUE or FALSE. If two events A and B are independent, then $\mathrm{P}(A \mid B) = \mathrm{P}(A)$ and $\mathrm{P}(B \mid A) = \mathrm{P}(A)$.

   FALSE, because if A and B are independent, then $\mathrm{P}(A \mid B) = \mathrm{P}(A)$ and $\mathrm{P}(B \mid A) = \mathrm{P}(B)$ and $\mathrm{P}(A) \neq \mathrm{P}(B)$ in general.

   (e) Answer TRUE or FALSE. If A and B are mutually exclusive events, then $P(A \mid B) = 0$.

   TRUE, because $\mathrm{P}(A \mid B) = \mathrm{P}(A \text{ and} B)/\mathrm{P}(B)$ and $\mathrm{P}(A \text{ and } B) = 0$ by mutual exclusivity.

(f) Answer TRUE or FALSE. If P(A and B) ≥ 0.10 then $P(A) \geq 0.10$.

TRUE, because the law of total probability says P(A and B)+P(A and not$-B$) = P(A), so P(A) $\geq$ P(A and B) ≥ 0.10.

(g) Answer TRUE or FALSE. Let investment X have mean return 5% and a standard deviation of 5% and investment Y have a mean return of 10% with a standard deviation of 6%. Suppose that the correlation between returns is zero. Then it is possible to find a portfolio with higher mean and lower variance than X.

TRUE. Define the random variable $T = 0.5X + 0.5Y$. Then $\text{E}(T) = 7.5 > 5$ and $\text{V}(T) = 0.25\text{V}(X) + 0.25\text{V}(Y) = (25 + 36)/4 = 15.25 \leq 25$.

(h) Answer TRUE or FALSE. Suppose that two investments A and B have the same mean, but different standard deviations. Then a diversified portfolio that includes both A and B will have a higher expected return and lower variance than either investment alone.

FALSE, because $\text{E}(aX + (1 - a)Y) = \text{E}(X) = \text{E}(Y)$ for any value of a between 0 and 1, so the mean cannot be increased.

(i) Answer TRUE or FALSE. Historically, 15% of chips manufactured by a computer company are defective. The probability of a random sample of 10 chips containing exactly one defect is 0.15.

FALSE, because the number of defects X is a binomial random variable with parameters $N = 10$ and $p = 0.15$. So $\text{P}(1 \text{ defect}) = 10(0.15)(1 - 0.15)^9 = 0.347 \neq 0.15$.

2. The following table shows survey results relating household income and level of education of the head-of-household. The cells of this table may be interpreted as the probability that a randomly selected household satisfies the criteria denoted by the cell's position in the table.

	Under \$50K	\$50-75K	Over \$75K
Not HS graduate	0.114	0.019	0.012
HS graduate	0.161	0.068	0.072
Some College	0.109	0.063	0.099
College degree	0.055	0.056	0.182

(a) What is the probability of the head-of-household having *at least* some college?

Summing over the bottom two rows we calculate $0.109 + 0.063 + 0.099 + 0.055 + 0.056 + 0.182 = 0.564$.

(b) If the head-of-household has a bachelor's degree or more, what is the probability that his/her income is not under \$50K?

We sum up all the cells satisfying both conditions and then divide by the sum of those cells satisfying the "if" condition:

$$\frac{0.056 + 0.182}{0.055 + 0.056 + 0.182} = 0.812.$$

(c) Knowing that a household income is between \$50K and \$75K, what is the probability that the household head has a bachelor's degree or more?

Again, we sum up all the cells satisfying both conditions, which in this case is just the second column of the last row, 0.056. We then divide by the "given" condition of having income between \$50 and \$75K, which is the total of the second column:

$$\frac{0.056}{0.019 + 0.068 + 0.063 + 0.056} = 0.27.$$

(d) Are education level and household income independent?

Summing across the bottom row we get $0.055 + 0.056 + 0.182 = 0.293$. Summing down the final column we get $0.012 + 0.072 + 0.099 + 0.182 = 0.365$. If education and household income were independent we would have Pr(college degree and income over 75K) = Pr(college degree)Pr(income over 75K), but in fact we see that $0.293 \times 0.365 = 0.107$ which is not equal to 0.182.

3. An oil company has purchased an option on land in Alaska. Preliminary geologic studies have assigned the following probabilities of finding oil:

event	high quality oil (H)	medium quality oil (M)	no oil(N)
Probability	0.50	0.20	0.30

After 200 feet of drilling on the first well, a soil test is taken which shows a certain soil type. The probabilities of finding this particular soil type (event S), given the type of oil present (if any) are:

$$\mathrm{P}(S \mid H) = 0.20 \qquad \mathrm{P}(S \mid M) = 0.80 \qquad \mathrm{P}(S \mid N) = 0.20.$$

(a) What are the revised (post-soil-test) probabilities of finding the three types of oil?

$$\begin{aligned}
\mathrm{P}(S) &= \mathrm{P}(S \mid H)\mathrm{P}(H) + \mathrm{P}(S \mid M)\mathrm{P}(M) + \mathrm{P}(S \mid N)\mathrm{P}(N) \\
&= 0.2(0.5) + 0.8(0.2) + 0.2(0.3) = 0.32, \\
\mathrm{P}(H \mid S) &= \frac{\mathrm{P}(S \mid H)\mathrm{P}(H)}{\mathrm{P}(S)} = 0.2(0.5)/0.32 = 0.3125, \\
\mathrm{P}(M \mid S) &= \frac{\mathrm{P}(S \mid M)\mathrm{P}(M)}{\mathrm{P}(S)} = 0.8(0.2)/0.32 = 0.5, \\
\mathrm{P}(N \mid S) &= \frac{\mathrm{P}(S \mid N)\mathrm{P}(N)}{\mathrm{P}(S)} = 0.2(0.3) = 0.1875.
\end{aligned} \tag{1}$$

(b) How should the firm interpret the soil test?

The overall probability oil has gone up from 70% to 81.25%, but the probability of high quality oil has gone down from 50% to 31.25%.

4. Cooper Realty is a small real estate company located in Albany, New York, specializing primarily in residential listings. They have recently become interested in determining the likelihood of one of their listings being sold within a certain number of days. Based on historical data, they produced the following figures based on the past 800 homes sold.

Days Listed until Sold	Under 20	31-90	Over 90	Total
Under $50K	50	40	10	100
$50-$100K	20	150	80	250
$100 - $150K	20	280	100	400
Over $150K	10	30	10	50

(a) What is the probability that a randomly selected home is listed over 90 days before being sold?

Summing down the Over 90 column, we get P(listed over 90 days) $= \frac{200}{800} = 25\%$.

(b) What is the probability that a randomly selected initial asking price is under $50K?

Summing across the Under $50K row, we get P(asking price less than $50K) $= \frac{100}{800} = 12.5\%$.

(c) What is the probability of both the previous two events happening? Are these two events independent?

Only 10 of 800 homes had asking prices of less than $50K and were listed over 90 days, so the probability is $\frac{1}{80}$. This does not equal $0.25(0.125) = 0.03125$ implied by independence, so the events are not independent.

(d) Assuming that a contract has just been signed to list a home that has an initial asking price less than $100K, what is the probability that the home will take Cooper Realty more than 90 days to sell?

We take the total number of homes satisfying both conditions, over the total number of homes satisfying the price condition to get

$$\text{P(listed} < 90 \text{ days} \mid \text{asking price } < \$100K) = \frac{10+80}{10+80+40+150+50+20} = \frac{9}{35}.$$

5. Consider the following probability distribution over two outcomes:

	$Y=1$	$Y=2$	$Y=3$
$X=1$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{1}{14}$
$X=2$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{3}{14}$
$X=3$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{1}{14}$
$X=4$	$\frac{1}{14}$	$\frac{1}{14}$	$\frac{1}{14}$

(a) Calculate the probability distribution for $\Pr(X)$.

Summing across each row individually we find that $\Pr(X=1) = \Pr(X=3) = \Pr(X=4) = \frac{3}{14}$ and $\Pr(X=2) = \frac{5}{14}$.

(b) Calculate $\Pr(X < Y)$.

$\Pr(X<Y) = \Pr(X=1, Y=2) + \Pr(X=1, Y=3) + \Pr(X=2, Y=3) = \frac{5}{14}$.

(c) Calculate $\Pr(X < Y \mid Y = 3)$.

$\Pr(X<Y \mid Y=3) = \frac{\Pr(X=1,Y=3)+\Pr(X=2,Y=3)}{\Pr(X=1,Y=3)+\Pr(X=2,Y=3)+\Pr(X=3,Y=3)+\Pr(X=4,Y=3)} = \frac{2}{3}$.

(d) Calculate $\Pr(X < 3)$.

Summing over all the cells where $X < 3$ (the top half of the table) we get $\frac{4}{7}$.

(e) Are X and Y independent? Why or why not?

Because the answers in (b) and (c) above are different X and Y cannot be independent.

6. Consider a professional baseball player with a drinking problem. His probability of getting a hit on any given at bat depends on which of three mutually exclusive physical states he happens to be in: drunk, sober, or hung-over. Assume that he must be in one of these states. When he's drunk he only gets a hit 5% of the time. When he's sober he gets a hit 30% of the time and when he's hung-over he gets a hit 21% of the time. Over the course of a season he plays drunk 15% of the time and hung-over 5% of the time.

   (a) What is his overall probability of getting a hit?

   Using the law of total probability:

$$\begin{aligned} \text{P(hit)} &= \text{P(hit} \mid \text{sober)P(sober)} + \\ &= \text{P(hit} \mid \text{hung over)P(hung over)} \\ &= \text{P(hit} \mid \text{drunk)P(drunk)} \\ &= 0.3(0.8) + 0.21(0.05) + 0.05(0.15) = 0.258. \end{aligned}$$

   (b) What is the probability that he is hung-over, given that he got a hit?

$$\frac{\text{P(hit} \mid \text{hung over)P(hung over)}}{\text{P(hit)}} = 0.21(0.15)/0.258 = 0.041.$$

(c) In a game where he has 3 at-bats, what is his probability of getting exactly 2 hits (assuming the probability of a hit in each at bat is independent of the other at bats)?

We answer this using the law of total probability, being careful to note that he does not switch between states of drunkenness within a given game. Let X denote the random variable recording his number of hits in 3 at bats. Then, given his drinking status, the distribution of X is a binomial random variable, from which we compute

$$
\begin{aligned}
\mathrm{P}(X=2) = {} & \mathrm{P}(X=2 \mid \text{sober})\mathrm{P}(\text{sober}) + \\
& \mathrm{P}(X=2 \mid \text{hung over})\mathrm{P}(\text{hung over}) + \\
& \mathrm{P}(X=2 \mid \text{drunk})\mathrm{P}(\text{drunk}) \\
= {} & 3(0.3)^2(0.7)(0.8) + 3(0.21)^2(0.79)(0.05) + 3(0.05)^2(0.95)(0.15) \\
= {} & 0.1575.
\end{aligned}
$$

(d) Given that in a game where he had 3 at-bats he got exactly 2 hits, what is the probability that he was sober? (Hint: you computed the necessary denominator for this calculation in part c above.)

Following from above,

$$\mathrm{P}(X=2 \mid \text{sober}) = \frac{\mathrm{P}(X=2 \mid \text{sober})\mathrm{P}(\text{sober})}{\mathrm{P}(X=2)} = 96\%.$$

7. Answer the following questions using the table below, which shows a breakdown of race *given* differing educational levels.

	ELEMENTARY LEVEL		HIGH SCHOOL		COLLEGE		
	<7	7 or 8	1 to 4	Graduate	Some college	Bachelor's Associate	Master's Professional Doctorate
WHITE	28.2	67.4	62.9	78.3	78.3	86.5	88.6
BLACK	11.8	12.4	18.5	12.9	13.3	8.2	6.5
HISPANIC	60.0	20.2	18.6	8.8	8.4	5.3	4.9
TOTAL	100.0	100.0	100.0	100.0	100.0	100.0	100.0

(a) Are ethnicity and education independent based on the table? Why or why not?

The conditional distribution of race clearly varies across education level, so the variables are not independent.

(b) Suppose that the proportions of the various levels of education are (in order): 0.03, 0.03, 0.11, 0.34, 0.2, 0.22, and 0.07. Is it more likely for a randomly selected person to be [white with less than seven years of elementary level education] or [hispanic with a bachelor's/associate degree]? Show your work.

We compare

$$\Pr(\text{white } | < 7 \text{ years of education })\Pr(< 7 \text{ years of education })$$

to

$$\Pr(\text{hispanic } | \text{ bachelor's/associate degree})\Pr(\text{bachelor's/associate degree})$$

to find

$$(0.282)(0.03) < (0.053)(0.22)$$
$$0.00846 < 0.01166.$$

8. At a certain private college students concentrate in one of three areas: pre-med, pre-law, or theater-arts. Only 15% of the students are theater-arts majors, with the remaining population divided equally among pre-med and pre-law. The most popular intramural sport at this college is ultimate frisbee, but participation differs between concentrations. Only 10% of pre-med students play on a frisbee squad, while 20% of pre-law students do, and 80% of theater-arts students do.

   (a) Suppose that your new summer intern shows up wearing an ultimate frisbee t-shirt with the school logo on it (which you take as certain evidence that she played ultimate frisbee there in college). What is the probability that this student was a theater-arts major?

$$\Pr(\text{theater} \mid \text{frisbee}) = \frac{\Pr(\text{frisbee} \mid \text{theater})\Pr(\text{theater})}{\Pr(\text{frisbee})},$$
$$= \frac{0.8(0.15)}{\Pr(\text{frisbee})} = \frac{0.8(0.15)}{0.2475} = .485.$$

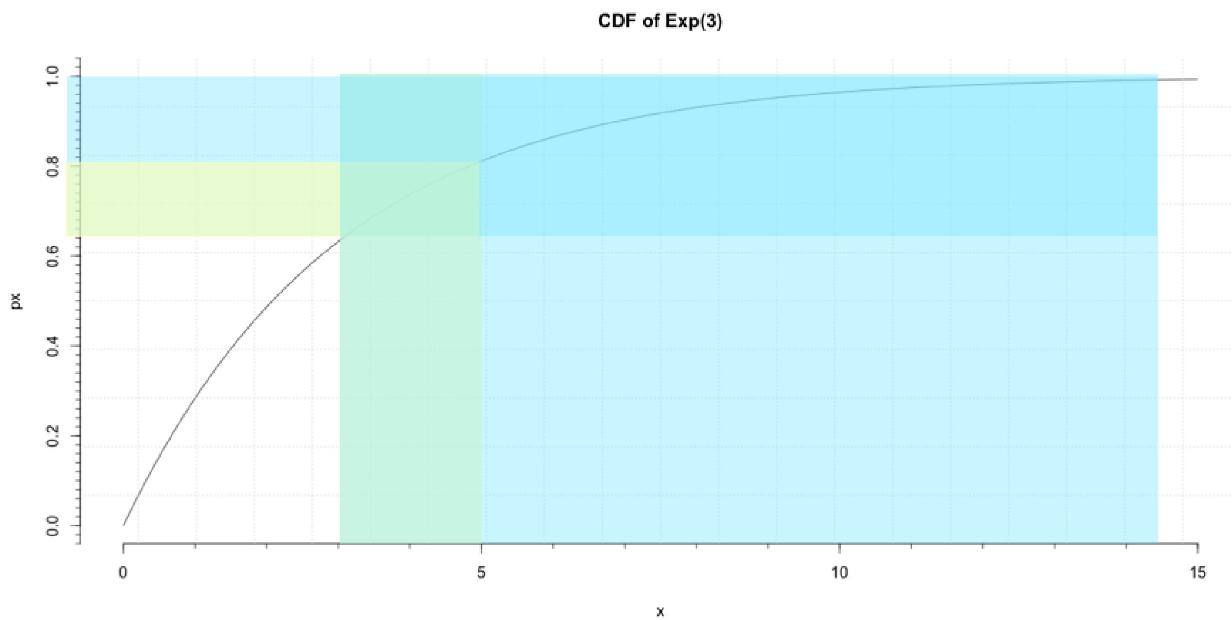
$$\begin{aligned}\Pr(\text{frisbee}) &= \Pr(\text{frisbee} \mid \text{theater})\Pr(\text{theater}) + \\ &\quad \Pr(\text{frisbee} \mid \text{pre-med})\Pr(\text{pre-med}) + \\ &\quad \Pr(\text{frisbee} \mid \text{pre-law})\Pr(\text{pre-law}) \\ &= 0.8(0.15) + 0.1(0.425) + 0.2(0.425), \\ &= 0.2475.\end{aligned}$$

(b) Suppose we know that pre-med students go on to have a loan-adjusted average post-graduation salary of \$60K, pre-law students go on to have a loan-adjusted average post-graduation salary of \$80K and theater-arts students have a loan-adjusted average post-graduate salary of -\$5K. Given that we think the student did play ultimate frisbee in college, what is the student's expected salary?

$$\begin{aligned}\text{E}(\text{salary} \mid \text{frisbee}) &= \text{E}(\text{salary} \mid \text{pre-med})\Pr(\text{pre-med} \mid \text{frisbee}) + \\ &\quad \text{E}(\text{salary} \mid \text{pre-law})\Pr(\text{pre-law} \mid \text{frisbee}) + \\ &\quad \text{E}(\text{salary} \mid \text{theater})\Pr(\text{theater} \mid \text{frisbee}), \\ &= (\$60K)\Pr(\text{pre-med} \mid \text{frisbee}) + \\ &\quad (\$80K)\Pr(\text{pre-law} \mid \text{frisbee}) + \\ &\quad (-\$5K)\Pr(\text{theater} \mid \text{frisbee}), \\ &= (\$60K)\frac{0.1(0.425)}{\Pr(\text{frisbee})} + (\$80K)\frac{0.2(0.425)}{\Pr(\text{frisbee})} - (\$5K)\frac{0.8(0.15)}{\Pr(\text{frisbee})}, \\ &= \$35.3K.\end{aligned}$$

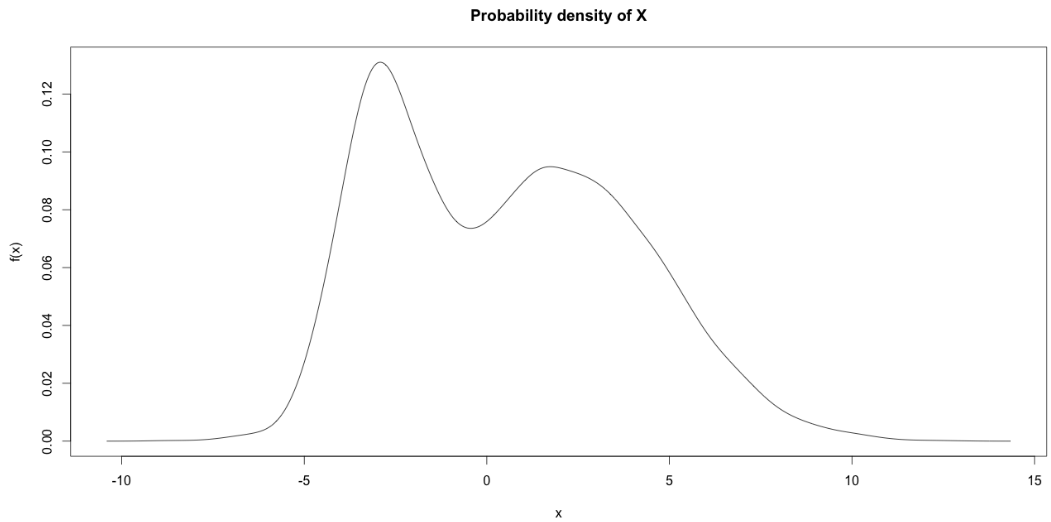
9. Use the corresponding plots to answer the following questions.

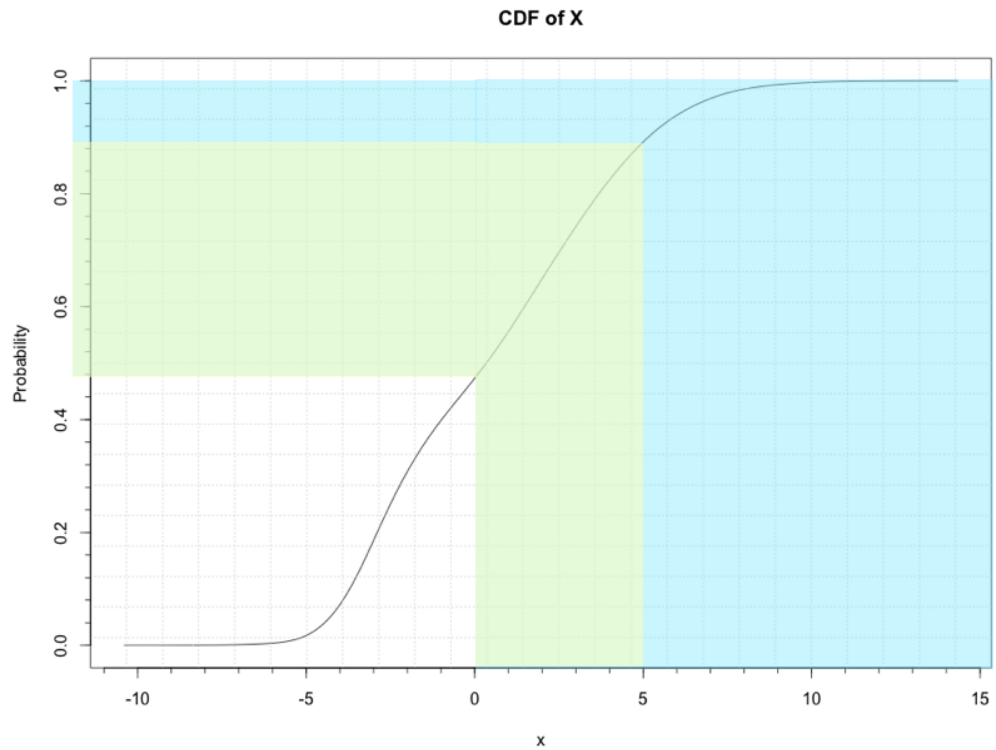
(a) Consider an exponentially distributed random variable X with mean parameter $\lambda = 3$. Use the CDF below to compute $\Pr(2 < X < 5|X > 3)$.



If $A = \{2 < X < 5\}$ and $B = \{X > 3\}$, then A and $B = \{3 < X < 5\}$. The answer to the question is then the ratio of the green area on the horizontal axis to the total shaded area, or roughly 0.48.

(b) Consider the random variable X with the density function and cdf shown below. Compute $\Pr(X \leq 5 \mid X > 0)$.





If $A = \{X \leq 5\}$ and $B = \{X > 0\}$, then A and $B = \{0 < X \leq 5\}$. The answer to the question is then the ratio of the green area on the horizontal axis to the total shaded area, or roughly $10/13$ or 77%.