Questions 1-4

Shell Oil Company is prospecting for oil in Northern Alberta. A survey of area A cost $65 million with a 0.21 chance of finding oil. Since they do not know the consistency of the ground there is a 0.7 chance that the drilling would cost $320 million and a 0.3 chance that the drilling would cost $650 million. If the drilling is successful there is a 0.75 change that they will have revenues of $2.7 billion and a 0.25 chance they will have revenues of $3.2 billion.

#1) What is the expected cost of drilling?

a) < $200 million
b) $200 million - $300 million
c) $300 million - $400 million
d) $400 million - $500 million
e) > $500 million

Answer:
Expected Cost = 0.7*320 million + 0.3*650 million = $419 million -> (d)

#2) What is the standard deviation of the cost of drilling?

a) < $200 million
b) $200 million - $300 million
c) $300 million - $400 million
d) $400 million - $500 million
e) > $500 million

Answer:
Variance = $(320 - 419)^2*0.7 + (650-419)^2*0.3 = \$^2$ 22869 million
Standard Deviation = SQRT($\$^2$ 22869) = -> $151.225 million -> (a)

#3) What is the expected revenue from drilling?

a) < $200 million
b) $200 million - $300 million
c) $300 million - $400 million
d) $400 million - $500 million
e) > $500 million

Answer: Assume that we will only proceed with drilling if the survey is successful.

Expected Revenue = 0.75*0.21*2.7 billion + 0.25*0.21*3.2 billion + 0.79*0 = $593.25 million -> (e)

#4) What is the expected profit from drilling?

a) < $200 million
b) $200 million - $300 million
c) $300 million - $400 million
d) $400 million - $500 million
e) > $500 million

Answer:
Expected Profit = $440.26 million -> (d)

					Profit	Probability		
			Profit	Probability	2485	0.25	(Survey Positive, Cost High, Revenue High)	
			2110	0.3	1985	0.75	(Survey Positive, Cost High, Revenue Low)	
	Profit	Probability			2815	0.25	(Survey Positive, Cost Low, Revenue High)	
	2341	0.21	2440	0.7	2315	0.75	(Survey Positive, Cost High, Revenue Low)	
PROFIT								
440.26	-65	0.79	(Survey Negative)					

Questions 5-8

Accidents involving deer in northern Ontario occur at a rate of 4.1 per day during the month of June. The province has built animal crossings in an effort to reduce this rate. One day in June 2013 there was 2 accidents involving deer.

#5) What is the appropriate distribution?

a) Poisson Distribution
b) Sampling Distribution
c) Binomial Distribution
d) Normal Distribution
e) None of the above

Answer: Poisson distribution as we are given a rate of occurrence of an event. -> (a)

#6) The Z-value is:

a) Between -2.0 and -0.6
b) Between -0.6 and 0.4
c) Between 0.4 and 2.4
d) Between 2.4 and 4
e) Not used in this type of question

Answer: The Z-value is not used for poisson. -> (e)

#7) If the June rate is considered accurate, the probability of 2 accidents on a given day:

a) Greater than 0.5
b) Between 0.1 and 0.5
c) Between 0.01 and 0.1
d) Between 0.001 and 0.01
e) Less than 0.001

Answer: $E(x) = \frac{e^{-\lambda t}\lambda t^x}{x!} = \frac{e^{-4.1*1}4.1*1^2}{2!} = 0.1393$ -> (b)

#8) Was the animal crossing effective in reducing the number of accidents involving deer?

a) Yes, because 2 accidents is less than 4.1
b) Yes, because the probability is less than 0.05
c) No, because the probability is less than 0.05
d) No, because the probability is more than 0.05
e) We cannot tell because the probability is more than 0.05

Answer: A probability under 0.05 would be a red flag and suggest that the daily date is not 4.1. Since our probability is greater than 0.05, there is no red flag. This suggests that the rate is not different from 4.1, since the rate has not changed, the crossing were not effective. -> (d)

Questions 9-12

The diagram below shows the average garbage production of 32 households in Nepean.

1	9
2	23788999
3	11112234559
4	1235
5	33
6	26
7	49
8	
9	
10	8
11	1

Stem: Ones Leaf: Tenths

#9) The mode of the date is:

a) The mode is the range of values between 31 and 39
b) The data has no mode
c) The mode is 31
d) The mode is 3.1
e) The mode cannot be calculated

Answer: The mode is the number that occurs most often in the data set which is 3.1 -> (d)

#10) The center of the garbage rate distribution is:

a) The median is a better measure for the center
b) The mean is a better measure for the center
c) The mode is a better measure for the center
d) The IQR is a better measure or the center
e) None of the above is correct.

Answer: The median is a better measure given that it is not affected by outliers -> (a)

#11) What is the IQR on of garbage production?

a) The IQR is between 1.8 and 2.2
b) The IQR is between 15 and 17
c) The IQR is 16
d) The IQR is 1.6
e) It cannot be calculated

Answer: IQR = 75th percentile - 25th Percentile = 4.9 - 2.9 = 2 -> (a)
25th = 0.25* 32 = 8th (take average of 2.9 and 2.9) = 2.9
75th = 0.75* 32 = 24th (take average of 4.5 and 5.3) = 4.9

#12) Are there outliers in the data set?

a) 19 is an outlier
b) 7.0 is an outlier
c) 10.8 is an outlier
d) 111 is an outlier
e) There are no outliers

Answer: Calculate the upper and lower fences. Upper fence = 75th percentile + 1.5* IQR = 4.9 + 1.5 *2 = 7.9
Lower fence = 25th percentile - 1.5* IQR = 2.9 - 1.5 *2 = -0.1 -> (c)

#13) What is the 60th percentile?

a) The 60th percentile is 39
b) The 60th percentile is cannot be calculated
c) The 60th percentile is 21.6
d) The 60th percentile is 3.9
e) The 60th percentile is 20

Answer: 60th = 0.6* 32 = 19.2th (round up to 20th observation) = 3.9 -> (d)

Questions 14-16

The sugar content of two soft drinks was measured by taking 19 observations for each drink. The sample statistics are listed below:

	Sugar in Soft Drink Samples (g)	
Statistic	**Pepsi**	**Coke**
Mean	45	50.63
Median	45	51
Mode	45	62
Standard Deviation	11.84	12.2
Interquartile Range	12	22

#14) What is the shape of each distribution?

a) Both distributions are symmetric
b) Pepsi Distribution is symmetric, Coke distribution is left skewed
c) Pepsi Distribution is left skewed, Coke distribution is symmetric
d) Both Distributions are right skewed
e) Cannot make a conclusion about the shape of the distributions

Answer: Pepsi - Mean is the same as median so it is symmetric. Coke - Mean is less than median, therefore distribution is left skewed. -> (b)

#15) What is the first quartile of the Pepsi Distribution?

a) The Q1 is approximately 6g
b) The Q1 is approximately 39g
c) The Q1 is approximately 33g
d) The Q1 is approximately 18g
e) The Q1 cannot be approximated

Answer: Since this distribution is symmetric, we can approximate the Q1 by using the following formula:
Q1 = Median - (1/2)IQR = 45-(1/2)*12 = 39g -> (b)

#16) The relative variation of both distributions

a) The coefficient of variation is higher in the Pepsi Sample
b) The coefficient of variation is higher in the Coke Sample
c) The coefficients of variation are equal in both samples
d) The coefficient of variation for Pepsi cannot be calculated
e) The coefficient of variation for Coke cannot be calculated

Answer: CV = SD/Mean CV(Pepsi) = 11.84/45*100 = 26.3% CV(Coke) = 12.2/50.63*100 = 24.1%
The CV for Pepsi is higher. -> (a)

Questions 17-21

A computer manufacturer sells computer in 2 markets, Canada and the US. It makes a profit of $2000 per unit in Canada and a profit of $3500 per unit in the US. The sales for the upcoming year will depend on the market conditions. They have chosen to use a probability distribution to estimates the sales. The distributions are as follows:

Units Sold (Canada)	1000	3000	5000	10000
Probability	0.1	0.3	0.4	0.2

Units Sold (US)	300	500	750
Probability	0.14	0.5	0.36

	Expected Value	Standard Deviation
Units Sold (Canada)	5000	2792.85
Units Sold (US)	562	155.74

#17) What is the standard deviation of total sales (in units)?

a) Is between 2900 and 3000
b) Is between 2800 and 2900
c) Is between 2700 and 2800
d) Is around 2949
e) Cannot be calculated

Answer: Variance (Canada + US) = $(2792.85)^2 + (155.74)^2 = 7824266.1$ SD (Canada + US) = SQRT(7824266.1) = 2797.189 -> (c)

#18) What assumption is needed to calculate the SD of total sales in units?

a) No assumptions are needed and the SD cannot be calculated
b) Canada and US sales vary independently
c) Canada and US sales do not vary independently
d) No assumptions are necessary
e) The Canada and US sales must both follow a normal distribution

Answer: In order to combine the variances and calculate the combined SD, we must assume that the Canada and US sales are independent of one another. -> (b)

Assume now that the correlation between Canada and US sales is 0.21.

#19) What is the expected profit?

a) Between $1,000,000 and $2,000,000
b) Between $5,000,000 and $6,000,000
c) Between $6,000,000 and $7,000,000
d) Between $11,000,000 and $12,000,000
e) It cannot be calculated

Answer: Expected Profit (Canada + US)= $2000*5000 + $3500*562 = $11,967,000 -> (d)

#20) What assumption is needed to calculate the Expected Value of profits?

a) No assumptions are needed since the expected profit cannot be calculated
b) Canada and US sales vary independently
c) Canada and US sales do not vary independently
d) No assumptions are necessary
e) The Canada and US sales must both follow a normal distribution

Answer: Expected profits are calculated the same way whether the two samples are dependent or independent, so there are not assumptions needed. -> (d)

#21) What is the SD of Profit in millions?

a) Between $1,000,000 and $2,000,000
b) Between $5,000,000 and $6,000,000
c) Between $6,000,000 and $7,000,000
d) Between $11,000,000 and $12,000,000
e) It cannot be calculated

Answer:

Variance = $2000^2*2792.85^2+3500^2*155.74^2+ 2*(2000)*(3500)*(0.21)*(2792.85)*(155.74) = 32775945467560$

SD = SQRT(32775945467560) = $5,725,027.99 -> (b)

Questions 22-25

You take a survey of Ottawa U students. They are asked to identify their favorite vacation spot and whether or not they will visit this location during reading week. The results show that 81% of students prefer Mexico. It also shows that 38% of students will visit their favorite destination during reading week. Finally, it shows that 4% of students indicated that they prefer a location other than Mexico and will go on vacation to their preferred destination.

#22) The events "Favorite Place is Mexico" and "Will go on vacation to favorite location":

a) Independent and Mutually Exclusive
b) Dependent and Mutually Exclusive
c) Independent and Not Mutually Exclusive
d) Dependent and Not Mutually Exclusive
e) It cannot be determined from the information given

Answer: The two events can occur simultaneously so they are not mutually exclusive.
The two events are independent if P(Will go on vacation) = P(Will go on vacation | Prefer Mexico)
P(Will go on vacation) = 0.38 P(Will go on vacation | Prefer Mexico) = 0.34/0.81 = 0.42 (you need to draw tree)
Therefore the two events are not independent. -> (d)

#23) If you pick one person at random, what is the probability that a prefers Mexico and will not go on vacation?

a) 0.15
b) 0.34
c) 0.47
d) 0.81
e) 0.94

Answer: P(M and NV) = 0.47 -> (c) (you need to draw tree)

#24) You have picked a student and they prefer to vacation in Mexico. What is the probability that this person intends to go on vacation?

a) 0.21
b) 0.34
c) 0.42
d) 0.89
e) Cannot be calculated

Answer: P(V|M) = 0.42 =(0.34/0.81) -> (c) (you need to draw tree)

#25) If you pick 3 people at random, that is the probability that you choose 3 in a row that either prefer Mexico or will go on vacation?

a) 0.28
b) 0.61
c) 0.85
d) 0.91
e) Cannot be calculated

Answer: P(V or M) = P(V) + P(M) - P(V and M) = 0.38 + 0.81 - 0.34 = 0.85
The probability of 3 in a row $P(V \text{ or } M)^3 = (0.85)^3 = 0.614$ -> (b) (you need to draw tree)

Questions 26-28

Color blindness affects 3% of the population. A simple test is 89% accurate in identifying color blindness in those that are known to have the disease. This same test falsely identifies the disease 6% of the times in those who don't have the disease.

#26) What is the probability an individual has the disease given that the test came out positive? (Approximately)

a) 0.3%
b) 0.6%
c) 3%
d) 30%
e) 40%

Answer: P(CB | +) = P(CB and +) / P(+) = 0.0267/(0.0267+0.0582) = 0.314488 or 31.44% -> (d) (draw tree)

#27) What is the probability and individual has the disease given that the test came out negative? (Approximately)

a) 0.3%
b) 0.6%
c) 3%
d) 30%
e) 40%

Answer: P(CB | -) = P(CB and -) / P(-) = 0.0033/(0.0033+0.9918) = 0.0036 or 0.36% -> (a) (draw tree)

#28) What concepts are most appropriately used to complete the calculations in question 26 and 27?

a) Binomial Distribution
b) Normal Distribution
c) Sampling Distribution
d) Bayes Theorem
e) None of the Above

Answer: Bayes Theorem is appropriate when calculating conditional probabilities. -> (d)