**Statistics 500.002: Applied Statistics**

**Summer 2015 Midterm 1 (World Campus)**

Name: \_\_\_\_Daljeet Maken\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Instructions:

1. From the time you first access the test you have **4 hours** to solve and upload the exam. Another hour may be used to earn partial credit.
2. The value of each part of each question is given in parentheses.
3. Work must be shown to earn full credit.
4. Provide justifications for your answers to earn partial credit.
5. Write your answers on the test booklet.
6. You can access all your lesson materials, text, handouts etc. But you may not take help from another person.
7. Calculators may be used and you may use Minitab whenever you want.
8. **The work handed in must be your own.**

***Good Luck!***

1. (8 pts) Circle the correct answer for the following five problems:

A and B are independent events. Which of the following statements are true? (There are more than one answer):

1. P(A  B) = P(A) + P(B) - P(A B)
2. P(A B) = P(A) + P(B)
3. P(A  B) = P(A) × P(B)
4. P(A  B) = P(A) × P(B)
5. P(A  B) = 0

The following statements are true:

P(A  B) = P(A) + P(B) - P(A B)

P(A  B) = P(A) × P(B)

1. (4 pts) Classify the following random variables: (underline or highlight the right answer)
2. Weight of a 12 year old child.
   * 1. quantitative, continuous
     2. quantitative, discrete
     3. qualitative
3. For the final project, there are 15 teams in our class designated as Team 1 to Team 15. The team a student is assigned to.
4. quantitative, continuous
5. quantitative, discrete
6. qualitative
7. (6 pts) Circle the correct answer in parentheses:
   1. If outliers are present in the data

(mean / median / mode)

is the most preferable measure of central tendency

* 1. Probability of an event which is sure to occur is

(0 / 1 / any number between 0 and 1)

* 1. Between the first quartile (Q1) and the third quartile (Q3) there is always

(25% / 50% / 75%)

of the data.

1. (4 pts) Which of the following statements is correct?
   1. Values of parameters vary from sample to sample but values of statistics do not.
   2. Values of statistics vary from sample to sample but values of parameters do not.
   3. Values of parameters and statistics both may vary from sample to sample.
   4. Values of neither parameters nor statistics can vary from sample to sample.
2. (4 pts) True or False: If the sample size is 30 or more, then the sampling distribution of sample mean is the same as the population distribution. Justify your answer.

False

The sampling distribution of the sample mean becomes normal for sample sizes >= 30. This is true for populations that originally are not normally distributed. While it is true that for populations that are highly skewed you may need a high number of samples, the principle is that the sampling distribution of the mean will become normally distributed for a sample size >= 30 (30 is used as a thumb rule).

For a large sample size (rule of thumb: n ≥ 30), *y*Bar is approximately normally distributed, regardless of the distribution of the population one samples from. If the population has mean μ and standard deviation σ, then *y*Bar has mean μ and standard deviation *σ*/√*n*

1. (10pts) The Minitab Stem and Leaf output that corresponds to a set of measurements is as follows:

Stem-and-Leaf Display:

Stem-and-leaf

Leaf Unit = 1

6 0 013334

8 0 59

(4) 1 0022

5 1 5678

1 2

1 2 5

Use this stem-and-leaf to answer the following questions

* 1. (3 pts) How many data points are in this data set? 17
  2. (4 pts) What is the median and range of this data set? Median: 10, Range: 25 – 0 = 25
  3. (3 pts) Later on another observation 64 was added to this data set. What is the median of the modified data set? Median: 10

1. (8 pts) If P(A) = 0.4, P(B) = 0.6, P(A  B) = 0.2, calculate the following:
2. P(A | B)
3. P (|B), where is A complement
4. Are A and B independent? Justify.
5. Are A and B complementary events? Justify.
6. P(A|B) = P(A ∩ B)/P(B) = 0.2 / 0.6 = 1/3 = 0.333
7. P (|B) = 1 - P (|B) = 1 – 0.333 = 0.667
8. A and B are not independent because:
   1. P(A ∩ B) = 0.2 ≠ P(A) P(B) = 0.4 \* 0.6 = 0.24
   2. The knowledge of one of the events has an effect on the conditional probability of the other. So if we know that event B occurred that changes the probability of A i.e. P(A|B) = 0.333 ≠ P(A)
9. A and B are not complementary events since complementary events are mutually exclusive by their very definition. That would require the intersection of the two to be empty. However we have been given P(A ∩ B) = 0.2 and therefore A and B are not complementary events.
10. (4 + 3 pts)
11. Find P( Z < - 0.5 or Z > 2.1)
12. Use the z-table of standard normal values to find a value for Z, say z0 such that

P ( Z < z0 ) = 0.75.

1. P( Z < - 0.5 or Z > 2.1) = P(Z < - 0.5) + P(Z > 2.1) = P(Z < - 0.5) + (1 - P(Z <= 2.1)) = 0.3085 + (1 - 0.9821) = 0.3264
2. From the normal table we see P(Z < 0.67) = 0.7486 (≈ 0.75). Minitab can be used to get higher accuracy:

P( X ≤ x ) x

0.75 0.674490

1. (4 + 4 + 3 + 2 pts) In a forest, the probability that a randomly sampled tree is oak is 0.20, pine is 0.25, maple is 0.35 and 0.20 is any other type trees.

If 10 trees are sampled randomly from a forest

(Hint: You may also think of the problem as probability of a tree being oak is 0.1 and non-oak is 0.9).

1. What is the probability that more than half of the 10 trees sampled is maple?
2. What is the probability that at least one of the 10 trees sampled is either an oak or a pine?
3. Let Y be a random variable denoting the number of pine trees in the sample of 10. What is the expected value of Y (mean value of Y)? If this value is not a whole number, will you round it up to a whole number?
4. What is the standard deviation of Y?
5. P(M > 5) = 1 – P(M<=5) = 1 - 0.905066 (binomial with n = 10 and π = 0.35) = 0.094934
6. P(Oak or Pine) = Probability of Oak or Pine = P(O) + P(Pine) = 0.2 + 0.25 = 0.45

Further P(O or Pine >= 1) = 1 – P(Oak or Pine = 0) = 1 - 0.0025330 = 0.997467

1. Expected value for n=10 and π = 0.25 = n π = 10 \* 0.25 = 2.5

No, we will not round up 2.5 to 3 since while Y can only take values 0, 1, 2, ..., n, but the expected value (mean) of Y may be some value other than those that can be assumed by Y.

1. Standard Deviation for binomial with n and π is given by: 

= sqrt(10 \* 0.25 \* 0.75) = 1.369

1. (2 + 4 + 2 pts) A carton of 15 balls in an urn and we know 4 are blue balls and rest are red balls. We pick 2 balls out one by one and without replacement. If Y denotes total number of blue balls picked.
2. What is the probability that the first ball is blue?
3. What is the probability that both balls are blue?
4. Is Y a binomial random variable? Justify your answer.
5. Probability is 4 / 15 = 0.267
6. Probability 4/15 \* 3/14 = 0.267 \* 0.214 = 0.0572
7. Y is not a binomial random variable since it doesn’t meet the following conditions:

* Condition: The probability of success, denoted π, remains the same from trial to trial.

In our case the probability of success changes with every draw since the balls are drawn without replacement.

* Condition: The *n* trials are independent. That is, the outcome of any trial does not affect the outcome of the others.

Again in our case the trials are not independent since the balls are drawn without replacement and therby change the probabilities for subsequent events.

11. (8 pts) The weight of an orange follows a normal distribution with mean 3.2 oz and standard deviation 0.4 oz.

(a) Find the probability that a randomly selected orange has weight less than 3.7 oz.

(b) Find the probability that the mean weight of 36 oranges is less than 3.0 oz.

1. P(Y < 3.7) = P(Z < (3.7 – 3.2)/0.4)) = P(Z < 1.25) = 0.8944
2. In the case of sample of size 36, Sample mean = 3.2 and SD = mu / sqrt(36) = 0.067

P(yBar < 3) = P(Z < (3 – 3.2)/0.067) = P(Z < -2.985) = 0.0014

12. (20 pts) A survey of 10,000 workers gave the following probability table for job satisfaction and workers’ education level. Job satisfaction level is quantified on a 5 point scale.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Education level | Job satisfaction level | | | | | Total |
|  | Very dissatisfied  (0) | Dissatisfied  (1) | Indifferent  (2) | Satisfied  (3) | Very satisfied  (4) |  |
| Less than high school | 0.051 | 0.117 | 0.02 | 0.011 | 0.001 | 0.20 |
| High school graduate | 0.019 | 0.035 | 0.238 | 0.088 | 0.07 | 0.45 |
| College graduate | 0.026 | 0.038 | 0.125 | 0.101 | 0.06 | 0.35 |
| Total | 0.096 | 0.19 | 0.383 | 0.20 | 0.131 |  |

Use the probabilities to answer the following questions:

1. What is the probability that a worker selected at random will not be a high school graduate?
2. What is the probability that a worker selected at random is a college graduate but does not like his/her job (i.e. will be dissatisfied or very dissatisfied)?
3. Given that a worker’s education level is high school or above what is the probability that his/her job satisfaction level will be at least 2?
4. Given that a worker’s job satisfaction level is at most 2 what is the probability that the worker is not a college graduate?
5. P(less than high school) = 0.2
6. P(College ∩ Very Dissatisfied) + P(College ∩ Dissatisfied) = 0.026 + 0.038 = 0.064
7. Let J denote job satisfaction with J2, J3 and J4 as Job Satisfaction level of 2, 3 and 4 respectively. Also, let HC denote high school and above.

We have P(HC) = 0.45 + 0.35 = 0.8

We have to find P(J2 | HC) + P(J3 | HC) + P(J4 | HC)

= P(J2 ∩ HC) / P(HC) + P(J3 ∩ HC) / P(HC) + P(J4 ∩ HC) / P(HC)

= (0.238 + 0.125) / 0.8 + (0.088 + 0.101) / 0.8 + (0.07 + 0.06) / 0.8

= 0.45375 + 0.23625 + 0.1625

= 0.8525

1. Let C denote a college graduate and therefore C^ will be non college graduate. Also let J denote job satisfaction with J0, J1 and J2 as Job Satisfaction level of 0, 1, 2 respectively

P(J0 or J1 or J2) = 0.096 + 0.19 + 0.383 = 0.669

We have to find P(C^ | J0 or J1 or J2) = 1 – P(C | J0 or J1 or J2)

= 1 – P(C ∩ J0 or J1 or J2) / P(J0 or J1 or J2)

= 1 – (0.026+ 0.038+ 0.125) / 0.669

= 1 – 0.283

= 0.717

***Congratulations! You have completed Midterm 1!!!***