**Statistics 500.002: Applied Statistics**

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

**SOLUTION**

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
6. (4 pts) Classify the following random variables: (underline or highlight the right answer)
7. Weight of a 12 year old child.
   * 1. **quantitative, continuous**
     2. quantitative, discrete
     3. qualitative
8. 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.
9. quantitative, continuous
10. quantitative, discrete
11. **qualitative**
12. (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.

**If the population distribution is normal, the distribution of the sample mean is normal and has standard deviation less than the standard deviation of the population. Thus the distributions are not the same.**

**If the population distribution is not normal (skewed), the distribution of the sample mean of a sample size over 30 approaches normal. The larger the sample size, the closer to normal it will be. And, again the standard deviations are very different from the population. The distributions are not the same**

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**

* 1. (4 pts) What is the median and range of this data set?

**Median 10, Range 25 – 0 = 25**

* 1. (3 pts) Later on another observation 64 was added to this data set. What is the median of the modified data set?

**Median for modified data is (10+10)/2 = 10.**

**(If you have just written 10 without any justification you have lost 1 point here.)**

1. (8 pts) If P(A) = 0.4, P(B) = 0.6, P(A  B) = 0.2, calculate the following:
2. P(A | B)

**P(A | B) = P(A  B) / P (B) = 0.2 / 0.6 *= 0.33***

1. P (|B), where is A complement

**Since P (**  ***B*) = P(B) – P(A**  ***B) = 0.6 – 0.2 = 0.4***

**P (**  **| *B*) = 0.4/0.6=0.667**

1. Are A and B independent? Justify.

**A and B are not independent.**

**Because P(A**  **B) = 0.2 does not equal to P (A) x P (B) = 0.24.**

1. Are A and B complementary events? Justify.

**A and B are not complementary events even though P(A) + P(B) = 1**

**Because P(A  B) = 0.2 and they are not mutually exclusive**

1. (4 + 3 pts)
2. Find P( Z < - 0.5 or Z > 2.1)

P(Z>2.1)=0.0179

P(Z<-0.5)=0.309

So P(Z<-0.5 or Z>2.1)=0.0179+0.309=0.326

1. Use the z-table of standard normal values to find a value for Z, say z0 such that

P ( Z < z0 ) = 0.75.

Z0=0.674

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?

**More than half means more than 5, i.e. 6 or 7 or 8 or 9 or 10**

**P(X > 5, n = 10, pi = 0.35) = 1 – P(X < = 5) = 1- 0.905066 = 0.094934**

**Cumulative Distribution Function**

Binomial with n = 10 and p = 0.35

x P( X <= x )

5 0.905066

1. What is the probability that at least one of the 10 trees sampled is either an oak or a pine?

**Pi = 0.20 + 0.25 = 0.45**

**At least one means 1 or MORE (i.e. X > = 1)**

**P(X > = 1) = 1 – P(X < 1) = 1 – P(X = 0)**

**P(X = 0) = 10! / (10!0!) (0.45)0 (0.55)10 = 0.0025330**

**P(X > = 1) = 1 – 0.0025 = 0.9975**

**Probability Density Function**

Binomial with n = 10 and p = 0.45

x P( X = x )

0 0.0025330

1. 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?

**Mean of Y = n \* pi = 10 \* 0.25 = 2.5**

This number is not to be rounded up. This mean is the mean of sampling of 10 trees over the long run. There is no reason this sampling mean should be a whole number

1. What is the standard deviation of Y?

**Sd = √ (n \* pi \* (1-pi)) = √(10 \* 0.25 \* 0.75 ) = √ 1.875 = 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?

4/15=0.267

1. What is the probability that both balls are blue?

4/15 \* 3/14=0.057

1. Is Y a binomial random variable? Justify your answer.

Y is NOT a binomial variable. For binomial distribution the success probability remains constant in successive trials. In this case, since the drawings are WITHOUT REPLACEMENT, the success probability changes over trials

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

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

**Z = (3.7 - 3.2) /0.4 = 1.25 P(X < 3.7) = P(z < 1.25) = 0.8943**

1. Find the probability that the mean weight of 36 oranges is less than 3.0 oz.

**Z = (3.0 - 3.2) /(0.4/sqrt(36)) = -3 P(Xbar < 3) = P(z < -3) = 0.0013**

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?

P(Not high school graduate) = P(less than high school) = 0.20

1. 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)?

P(college graduate ∩ not likes job)

= P(college graduate ∩ dissatisfied) + P(college graduate ∩ very dissatisfied)

= 0.026 + 0.038 = 0.064

1. 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?

**A: education high school or above**

**B: job satisfaction (2, 3 or 4)**

**Need P(B|A)**

**P(B|A) = P(B∩A)/P(A)**

**P(B∩A) = 0.682 (sum of the green cells)**

**P(A) = 0.45+0.35 = 0.80**

**P(B|A) = 0.682/0.8 = 0.8525**

1. Given that a worker’s job satisfaction level is at most 2 what is the probability that the worker is not a college graduate?

A: job satisfaction at most 2, i.e. 0, 1 or 2

B: not college graduate (i.e. either less than high school or high school)

Consider the first three columns together.

**P(B|A) = P(B∩A)/P(A)**

**P(A) = 0.669 (sum of the blue cells in the margin)**

**P(B ∩ A) = 0.48 (sum of the cells in red font)**

**P(B |A) = 0.717**