Presented by Daniel Kula

ID: 215977497

Introduction to Data Science - 2020 Semester A Final Project

מחברת 1 - הסתברות, חוק בייס

1. **Tweens and Bowls**
   a. 5/17 . The explanation is as follow:

      The probability for none-identical tweens out of all he births is 1/125 (given) and for identical tweens out of all the births is 1/300 (given).

      There for, the probability for tweens out of all the births should be:

      $^1/_{125} + ^1/_{300} = \frac{300+125}{300\times125} = \frac{425}{37{,}500} = \mathbf{\frac{17}{1{,}500}}$

      And the probability for none-identical tweens out of all the tween's births should be: $\frac{^1/_{300}}{^{17}/_{1{,}500}} = \frac{1}{300} \times \frac{1{,}500}{17} = \frac{1{,}500}{300\times17} = \frac{1{,}500}{5{,}100} = \mathbf{\frac{5}{17}}$

      The probability that Alvis dead tween was a none-identical tween is: $\mathbf{\frac{5}{17}} \approx \mathbf{0.29}$

   b. Due to the fact the choice of one of the tow bowls was random (given), the probability should be: $\mathbf{^1/_2 = 50}\%$

2. **M&M's**
   The probability of yellow from the 1994's bag is $\frac{20}{100}$ and from the 1996's bag is $\frac{14}{100}$, so the probability of yellow from the tow bags should be: $\frac{20}{100} + \frac{14}{100} = \frac{34}{100}$
   The probability that the yellow M&M came from the 1994's bag may be reflected as:
   $\frac{^{20}/_{100}}{^{34}/_{100}} = \frac{20}{34} = \frac{10}{17} \approx 0.5882$

3. **Swine Flu**
   a. It is given that the positive test accuracy is 99%, which mean that, if the test was positive, the probability that it is not positive false is 99%.
   b. The population dose not effect the accuracy of the test and therefor if the test was positive, the probability that it is not positive false is 99% as in a. situation.

Random Variables:

1. Here are the combinations of a sum of 2 six-sided dices (painted in green are the divisible by 3 sum):

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The chance of a divisible by 3 sum is $^{12}/_{36}$, and the chance of a none-divisible by 3 sum is $^{24}/_{36}$, the payoffs of those tow situations are \$6 and -\$3, respectively (given).

So, the expected value of playing this game should be:

$$^{12}/_{36} \times \$6 - {}^{24}/_{36} \times \$3 = \$2 - \$2 = \$\mathbf{0}$$

2. Here are the combinations of a sum of Marker Mixup (painted in green are the "more than 12 sum" and in red the "12 sun"):

	1	2	3	4	5
6	7	8	9	10	11
7	8	9	10	11	12
8	9	10	11	12	13
9	10	11	12	13	14
10	11	12	13	14	15

The chance of a more than 12 sum is $^{6}/_{25}$, The chance of a exactly 12 sum is $^{4}/_{25}$ and the chance of a less than 12 sum is $^{15}/_{25}$, the payoffs of those three situations are \$5.5, \$0 and -\$3, respectively (given).

So, the expected value of playing Marker Mixup should be:

$$^{6}/_{25} \times \$5.5 + {}^{4}/_{25} \times \$0 - {}^{15}/_{25} \times \$3 = \$1.32 + \$0 - \$1.8 = -\$\mathbf{0.48}$$

3. The male portion, of the division's employees, is 40% (80 employees) and the female portion is 60% (120 employees) (given).

<u>Mean</u>:

Suppose that the selection of the 8 employees happens together, the approximate average of selected male each month may be reflect as 40%*8 = 3.2.

Standard Deviation:

There are 9 possible combinations as follows (M=Male and F=Female):

combination	Probability	Number of Combinations	$\sqrt{(X-\bar{X})^2}$
FFFFFFFF	$60\%^8$	1	3.2
MFFFFFFF	$40\%X60\%^7$	8	2.2
MMFFFFFFF	$40\%^2X60\%^6$	28	1.2
MMMFFFFF	$40\%^3X60\%^5$	56	0.2
MMMMFFFF	$40\%^4X60\%^4$	70	0.8
MMMMMFFF	$40\%^5X60\%^3$	56	1.8
MMMMMMFF	$40\%^6X60\%^2$	28	2.8
MMMMMMMF	$40\%^7X60\%$	8	3.8
MMMMMMMM	$40\%^8$	1	4.8

The general formula for Standard Deviation is: $\sigma = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$

And in our situation:

$$\sigma = \sqrt{\frac{3.1x1 + 2.2x8 + 1.2x28 + 0.2x56 + 0.8x70 + 1.8x56 + 2.8x28 + 3.8x8 + 4.8x1}{256-1}} \sim \mathbf{1.318}$$

4. If the distribution presented with a mean and standard deviation of 26,000$ and 2,000$, respectively (given).
   The P(26<X<30) may be described as P(26<X)- P(X<30).
   The P(X<30) =1- P(X>30)
   The P(26<X) = 50% and the P(X>30) = 2.275% and therefore:
   P(26<X<30) = 50%-2.275% = 47.725%

5. Since the distribution is represented in the triangle, the area below the triangle from x = 3 to x = 5 represents the probability of x> 3.
   The area of a triangle will be measured using the formula: $\frac{h\times b}{2}$
   Were:

   h = 0.4 and

   b = 2 (5-3)

   $\frac{h\times b}{2} = \frac{0.4\times 2}{2} = \mathbf{0.4}$

6. In general, in rounded calculation, there are 5 possible combinations as follows (C=with children and 0=without children):

combination	Probability (P)	Number of Combinations (N)	N X (P)
CCCC	$60\%^4$	1	12.96%
CCC0	$40\%X60\%^3$	4	34.56%
CC00	$40\%^2X60\%^2$	6	34.56%
C000	$40\%^3X60\%$	4	15.36%
0000	$40\%^4$	1	2.56%
			100.00%

Under the rounded calculation, exactly 3 of the 4 employees selected have children equal to the second line in the table = 34.56%

the accurate calculation may be described as the following (the red are employees that have children):

$\mathrm{CCC0} \times \text{Number of Combinations} = \left[\frac{300}{500} \times \frac{299}{499} \times \frac{298}{498} \times \frac{200}{497}\right] \times 4$=**34.63%**

7. The expected Value of X is presented as:

$$E(X) = \sum (\mathrm{P}_x)\mathrm{x}$$

And in our situation, based on the information from the graph:

$$10\% \times (-10) + 35\% \times (-5) + 10\% \times 0 + 35\% \times 5 + 10\% \times 10 = \mathbf{0}$$