AP Statistics

Discrete Probability Distributions
Confidence Interval
&
One Sample Hypothesis Testing

Standards:
CA Standards 7.0:
Students demonstrate an understanding of the standard distributions (normal, binomial, and exponential) and can use the distributions to solve for events in problems in which the distribution belongs to these families.

CA Standards 8.0:
Students determine the mean and the standard deviation of a normally distributed random variable.

CA Standards 14.0:
Students organize and describe distributions of data by using a number of different methods, including frequency tables and histograms.

CA Standards 17.0:
Students determine confidence intervals for a simple random sample from a normal distribution of data and determine the sample size required for a desired margin of error.

CA Standards 18.0:
Students determine the P-value (critical value for rejection region) for a statistic for a simple random sample from a normal distribution.

Last name, First name
Student ID #

Discrete Probability Distributions

Materials:

One bag of regular sized M&M candy, calculator, word processing and data spreadsheet software.
Claim: In each bag of M&M's, the percentage's of each color are:
24% blue, 14% brown, 16% green, 20% orange, 13% red, and 14% yellow

Procedure:

40% of the M&M in a regular size bag is blue. We randomly select 5 pieces of candy each time for 30 times and record the number of Blue M&M observed.

Century Variety Bag
Blue color: 40%
Green color: 50%
White color: 10%

Data

Trial	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Observed	1	2	2	2	1	2	3	2	1	5	3	4	2	2	2	3	1	2	0	4	2	2	2	4	2	3	0	2	2	2

Binomial Distribution

Part A (Theoretical Values)

1. Identify a success and the value of n, p, and q.

   Let x = number of blue M&M in Century's variety bag
   n=5
   p=0.4
   q=0.6

2. Write the distribution and its parameters.

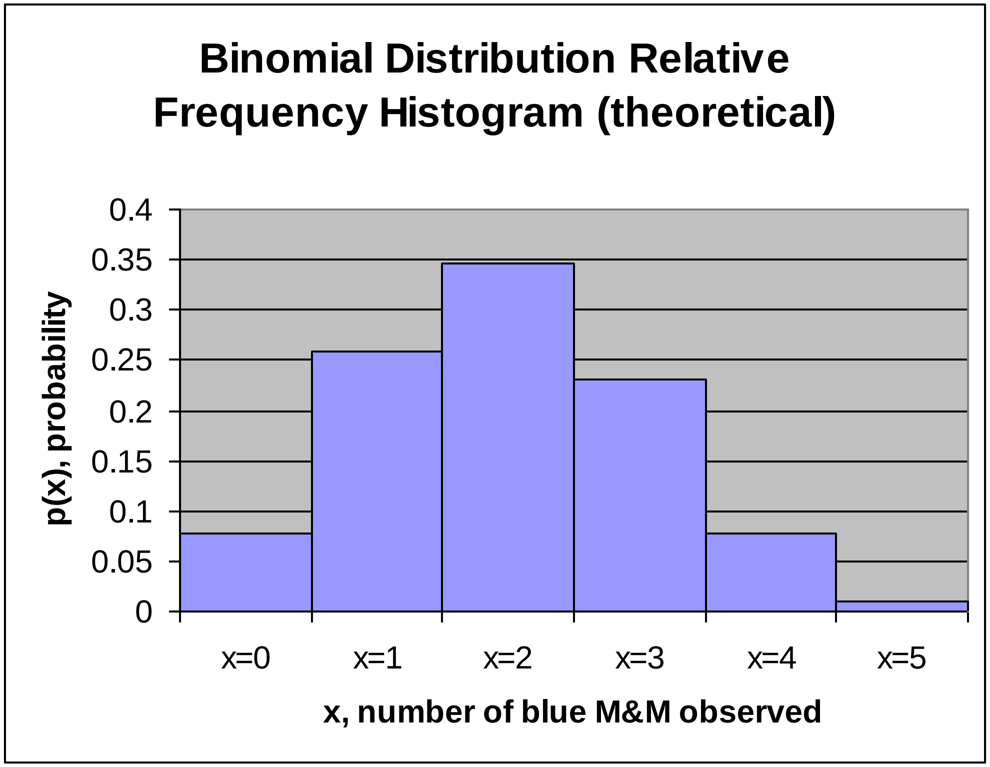
   x ~ Binopdf (5, 0.4)

3. Construct a binomial distribution.

$p(x=0) = {}_5C_0(0.4)^0(0.6)^5 = 0.0778$ Binompdf (5, 0.4, 0)
$p(x=1) = {}_5C_1(0.4)^1(0.6)^4 = 0.2592$ Binompdf (5, 0.4, 1)
$p(x=2) = {}_5C_2(0.4)^2(0.6)^3 = 0.3456$ Binompdf (5, 0.4, 2)
$p(x=3) = {}_5C_3(0.4)^3(0.6)^2 = 0.2304$ Binompdf (5, 0.4, 3)
$p(x=4) = {}_5C_4(0.4)^4(0.6)^1 = 0.0768$ Binompdf (5, 0.4, 4)
$p(x=5) = {}_5C_5(0.4)^5(0.6)^0 = 0.0102$ Binompdf (5, 0.4, 5)

x	0	1	2	3	4	5
p(x)	0.0778	0.2592	0.3456	0.2304	0.0768	0.0102

4. Graph the binomial distribution using a relative frequency histogram.



5. Find the mean, μ

$$\mu = np = (5)(0.4) = 2$$

6. Find the variance, σ^2

$$\sigma^2 = npq = (5)(0.4)(0.6) = 1.2$$

7. Find the standard deviation, σ

$$\sigma = \sqrt{npq} = \sqrt{(5)(0.4)(0.6)} = 1.095$$

8. Interpret the results in the context of the real-life situation.

If we randomly select 5 pieces of M&M from Century variety bag, we can expect to see on an average 2 blue M&M with standard deviation of 1.095 M&M

9. Find the probability that the number of blue observed is exactly three.

$$p(x=3) = {}_5C_3(0.4)^3(0.6)^2 = 0.2304 = 23.04\%$$

Binompdf (5, 0.4, 3)

10. Find the probability that the number of blue observed is at least three.

$$p(x \ge 3) = p(x=3) + p(x=4) + p(x=5) = {}_5C_3(0.4)^3(0.6)^2 + {}_5C_4(0.4)^4(0.6)^1 + {}_5C_5(0.4)^5(0.6)^0 = 0.3174 = 31.74\%$$

Binompdf (5, 0.4, 3) + Binompdf (5, 0.4, 4) + Binompdf (5, 0.4, 5)

11. Find the probability that the number of blue observed is less than three.

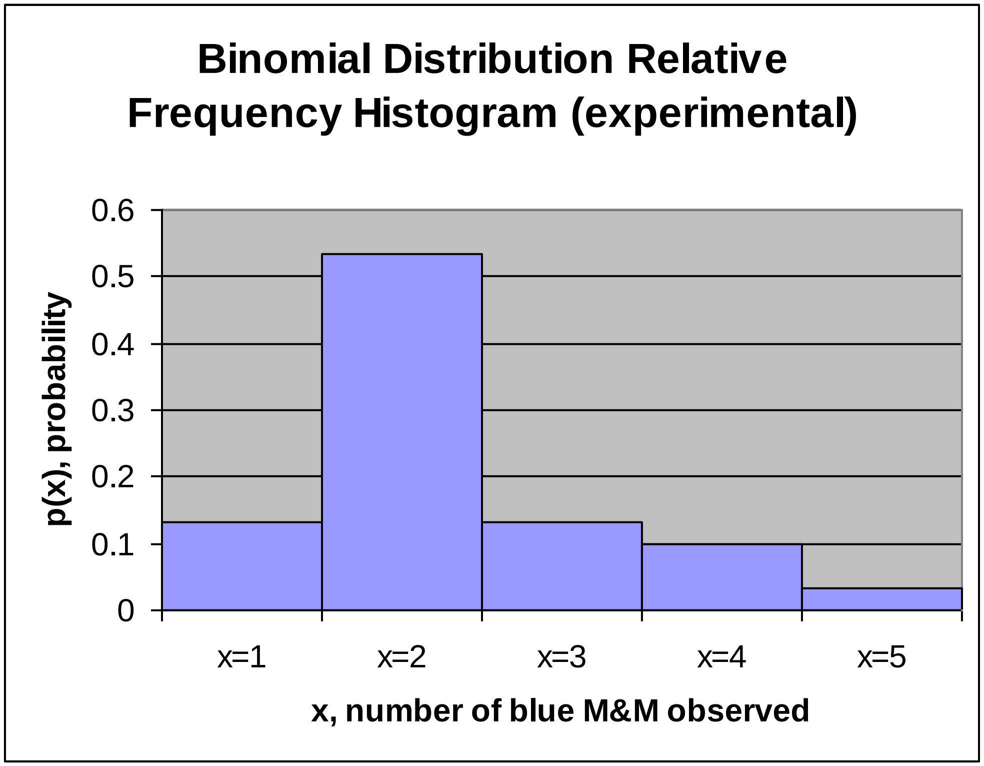
$$p(x<3) = 1 - p(x \ge 3) = 1 - 0.3174 = 0.6826 = 68.26\%$$

Part B (Experimental Value)

1. Use the data recorded (30 times), construct a binomial distribution.

x	0	1	2	3	4	5
p(x)	0.067	0.133	0.533	0.133	0.1	0.033

2. Graph the binomial distribution using a relative frequency histogram.



3. Find the mean, $\bar{X}$

$$\overline{X} = \frac{1+2+2+\ldots+2+2+2}{30} = 2.167 \text{ (1-Var Stats)}$$

$$\mu = E(x) = \sum xp(x)$$

4. Find the variance, s²

$$s^2 = \frac{2(0-2.167)^2 + 4(1-2.167)^2 + 16(2-2.167)^2 + 4(3-2.167)^2 + 3(4-2.167)^2 + 1(5-2.167)^2}{30-1} = 1.2471$$

$$Var(x) = \sigma^2 = \sum (x-\mu)^2 p(x)$$

5. Find the standard deviation, s

$$s = \sqrt{1.2471} = 1.117 \text{ (1-Var Stats)}$$

$$SD(x) = \sigma = \sqrt{Var(x)}$$

Geometric Distribution

Part C (Theoretical Value)

Let x = number of draws until the first blue M&M is observed

X ~ Geometpdf (0.4)

1. Find the probability the first M&M is a blue one will occur on the fourth draw.

$$p(x=4) = (0.4)(0.6)^{4-1} = 0.0864 = 8.64\%$$

Geometpdf (0.4, 4)

2. Find the probability the first M&M is a blue one will occur on the fourth or fifth draw.

$$p(x = 4 or 5) = p(x = 4) + p(x = 5) = (0.4)(0.6)^{4-1} + (0.4)(0.6)^{5-1} = 0.1382 = 13.82\%$$

3. How many M&M can we expect to see until his first blue one.

$$n = \frac{1}{p} = \frac{1}{0.4} = 2.5$$

We can expect to see the first blue M&M after 2.5 draws on average

Confidence Interval

Part D

1. From your experiments, use the mean and standard deviation found in Part B, construct a 90% confidence interval of the population mean numbers of blue M&M. Assuming the population for M&M is normally distributed.

$$\varepsilon = Z_c \frac{\sigma}{\sqrt{n}} = (1.645)\frac{1.117}{\sqrt{30}} \approx 0.335$$

$$\bar{x} - \varepsilon < \mu < \bar{x} + \varepsilon$$

$$2.167 - 0.335 < \mu < 2.167 + 0.335$$

$$1.832 < \mu < 3.999$$

At 90% confidence level, the mean blue M & M in a 5 candy draw is between 1.83 ~ 4.

Z-Interval

2. How many samples do you need if you want to be 95% confident that the sample mean is with in one M&M of the population mean?

$$n = \left(\frac{Z_c \sigma}{\varepsilon}\right)^2 = \left(\frac{1.96 * 1.117}{1}\right)^2 = 4.79 \approx 5$$

Only 5 samples.

Hypothesis Testing

Part E

Conclusion

Write a conclusion using your experiment result as the sample. Apply one sample hypothesis test with $\alpha = 0.05$ to test company's clam.

According to the official number published by the M&M candy maker, if we randomly select 5 pieces of M&M from Century variety bag, we can expect to see on an average 2 blue M&M (Part A). We want to perform a one sample hypothesis testing on the company's claim at $\alpha = 0.05$.

Let $H_0 : \mu = 2$ (claim)

$Ha : \mu \neq 2$

And the test statistics is

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{2.167 - 2}{1.117 / \sqrt{30}} = 0.8189$$

With $\alpha = 0.05$ and a two tailed test, our critical values are -1.51 and1.51.

So, at 5% level of significance, we fail to reject H_0 and conclude that there is enough evidence to support company's claim that if randomly select 5 pieces of M&M from Century variety bag, we can expect to see on an average 2 blue M&M.

Z-Test