Ian Mills
Professor Richard Oremus
Math 1040
24 September 2022

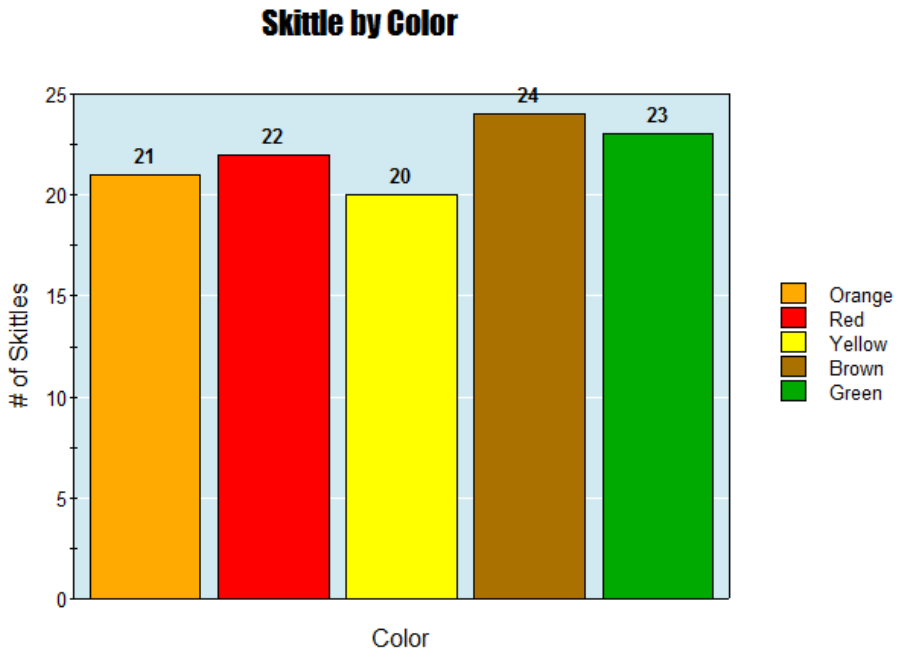
Skittle Project

Introduction

This project looks at the statistical likelihood of colors in a bag of skittles. Since a skittle bag contains 5 different colors the assumption would be that it would be evenly distributed at ⅕. Using my observations and that of my classmates was able to extrapolate and analyze the true breakdown and if that matches the ⅕.

My bag of Skittles by color:

Color	**Frequency**
Red	22
Orange	21
Yellow	20
Green	23
Brown/Purple	23
Total	109



Since each bag of skittles contains 5 colors it's likely that each color would be about 20% or ⅕.

The relative frequency of red skittles in my bag corresponds to P^ = .2018.

This bag of the proportion of red skittles is within expected accuracy of roughly 20%.

The skittles are a random sample, because what was selected to be put into the package by a machine was likely automated and random, as well as which one being selected from the store was random. Because it meets this requirement able to test the null hypothesis.

Trial/Simulation

Simulation of the occurrence of red skittles (.2) in 10,000 trials:

H0:p=0.2

HA:p≠0.2

One Proportion

Describe process:

Probability of success (π): 0.2
Sample size (*n*): 109
Number of samples: 100

☑ Show animation
Draw Samples
Total Samples = 10000

Choose statistic:

○ Number of successes
◉ Proportion of successes

Count samples

As extreme as ≥ .201 Count

Proportion of samples:
(4786 + 5214) / 10000 = 1

Options:

☑ Two-sided (between: ☐)
☐ Exact Binomial
☐ Normal Approximation

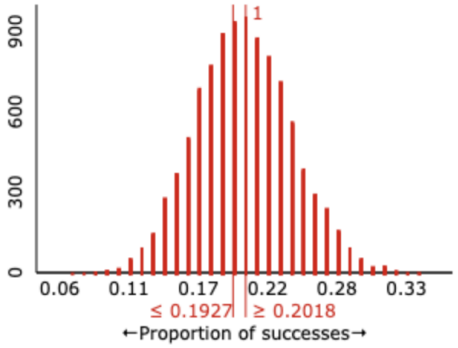
Reset

Most recent results

Number of Successes = 15

Number of Failures = 94

☐ Summary Statistics



☐ Show previous results
☐ Show sliders

Because the p-value is greater than alpha it fails to reject the null hypothesis. There is not sufficient evidence to support that the proportion of red skittles is different then .2

Number of skittles by color for entire class:

Color	Frequency
Red	398
Orange	398
Yellow	405
Green	359

Color	Frequency
Brown/Purple	384
Total	1,944

Using this information to find the interval of these candies with a 95% confidence interval.

Values:

n = 1,944

x (red skittles) = 398

$\hat{p}$ = 398/1,944 = .205

α = 0.05

Mean = 388.8

Z = 1.96

Lower Limit = 0.1868

Upper Limit = 0.2227

Z Estimate of a Proportion

Confidence Level 0.95

Sample

Successes 398

N 1944

Z Estimate of a Proportion

Result

Successes	398
N	1944
SE	0.0091517024187
Lower Limit	0.1867955031502
Upper Limit	0.2226695174259
Interval	0.2047325102881 ± 0.0179370071378

The proportion of skittles in my bag had a p̂ value of .2018. Rounded up to .202 is only a difference of .003 from the classes p̂ value. Using the confidence interval above the proportion of red skittles in my bag was likely.

To verify p=0.2 going to perform a two sided test. This will give the range in which the red skittles ought to be.

Mean = 388.8

α = 0.05

H_0: p = 0.2

H_A: p ≠ 0.2

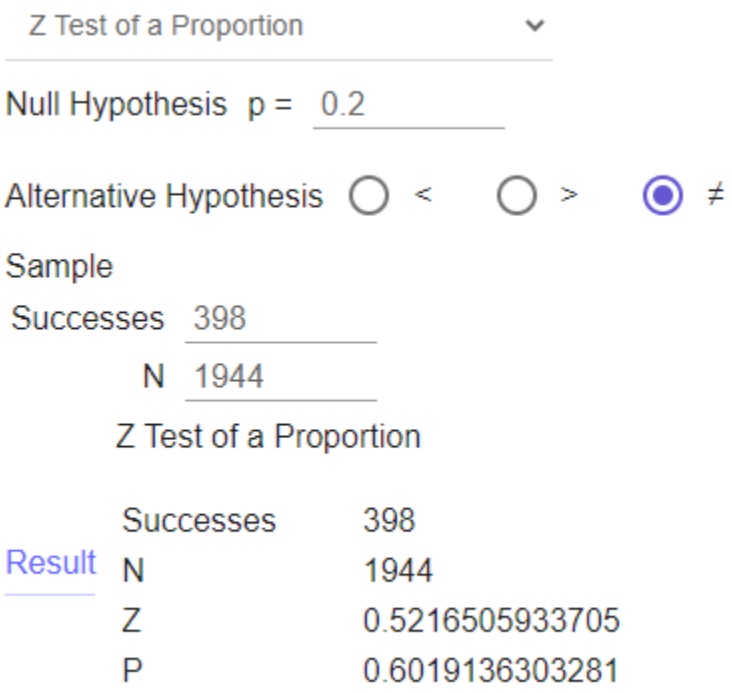
As stated earlier the sample is random and passes that requirement.

$n \leq 0.05N$ ⇒ 1944 ≤ .05N (1944/.05) = 38880

$np_0 \geq 10;\ n(1 - p_0) \geq 10$ ⇒ 1944(.2) ≥ 10 = 388.8; 1944(1 - .2) ≥ 10 = 155.2 ≥ 10

$n \geq Z^{*2} x\, p(1 - p)/ME^2$ ⇒ 1944 ≥ 0.5217^2 x 0.2(1 - 0.2) ÷ $.5127\sqrt{0.2(1 - 0.2) \div 1944}$ =

.2722 x .16 = 9.2664 = 1944 ≥ 9.2664



The p value is larger than alpha which means we reject the null hypothesis.

One Proportion

Describe process:

Probability of success (π): 0.2
Sample size (*n*): 109
Number of samples: 100(

☑ Show animation
Draw Samples
Total Samples = 10000

Choose statistic:

○ Number of successes
◉ Proportion of successes

Count samples

As extreme as ≥ .201 Count

Proportion of samples:
(4786 + 5214) / 10000 = 1

Options:

☑ Two-sided (between: ☐)
☐ Exact Binomial
☐ Normal Approximation

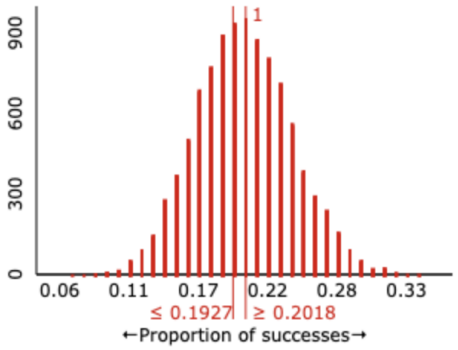
Reset

Most recent results

Number of Successes = 15

Number of Failures = 94

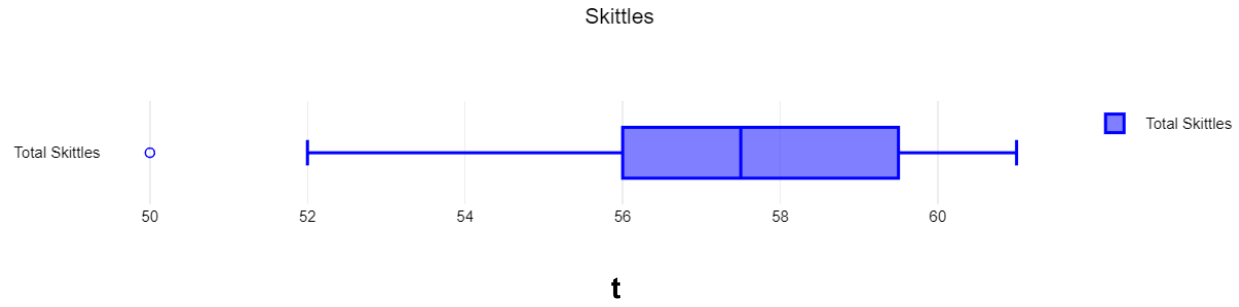
☐ Summary Statistics

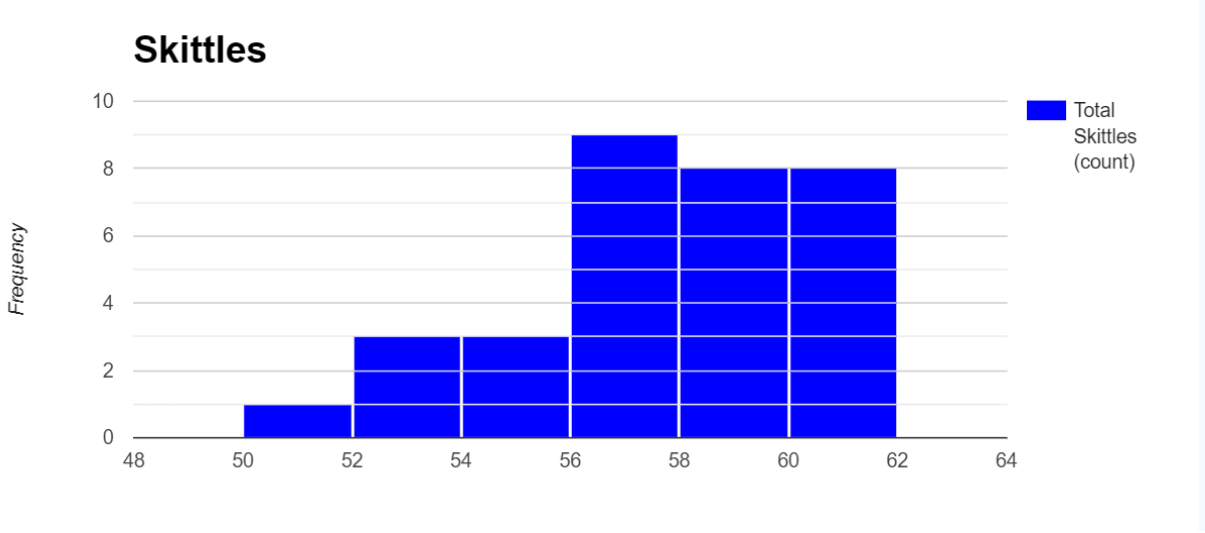


☐ Show previous results
☐ Show sliders

The null hypothesis test and the Z estimate of proportion lower and upper limits do fall in line with the results shown here in the simulation. Continue to reject the null hypothesis and do find that the .2 estimate for red skittles is accurate in most cases.

Can now us this to create graphs to further analyze the information.





The shape of these graphs shows a left skew.
Using the Population Mean formula

$\mu = \frac{\Sigma x}{N}$ = 57.3548

And to find the variance

$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$ = 8.7032

The only outlier was the group of 50.

$Q_3 - Q_1 = IQR$

$59.5 - 56 = 3.5$

IQR = 3.5

$Q_1 - 1.5 * IQR$

56 - 1.5 * 3.5 = 50.75

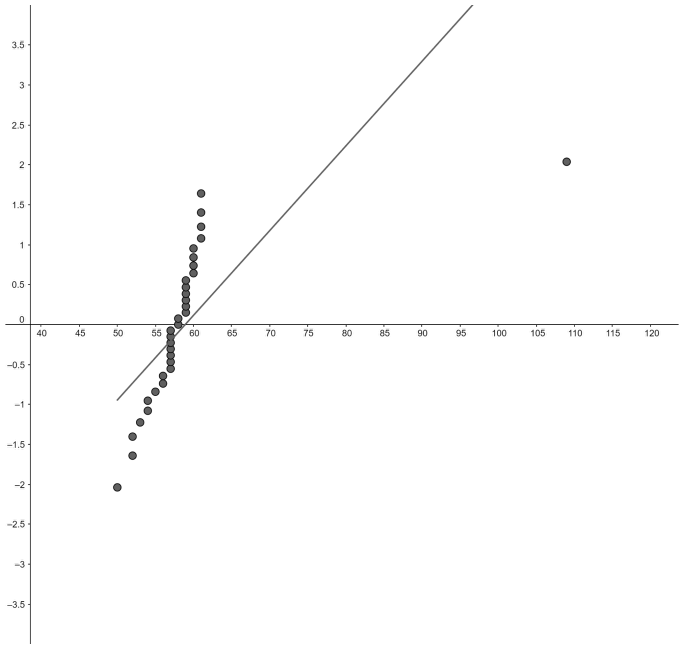
$Q_3 + 1.5 * IQR$

59.5 + 1.5 * 3.5 = 64.75

Lower fence = 50.75

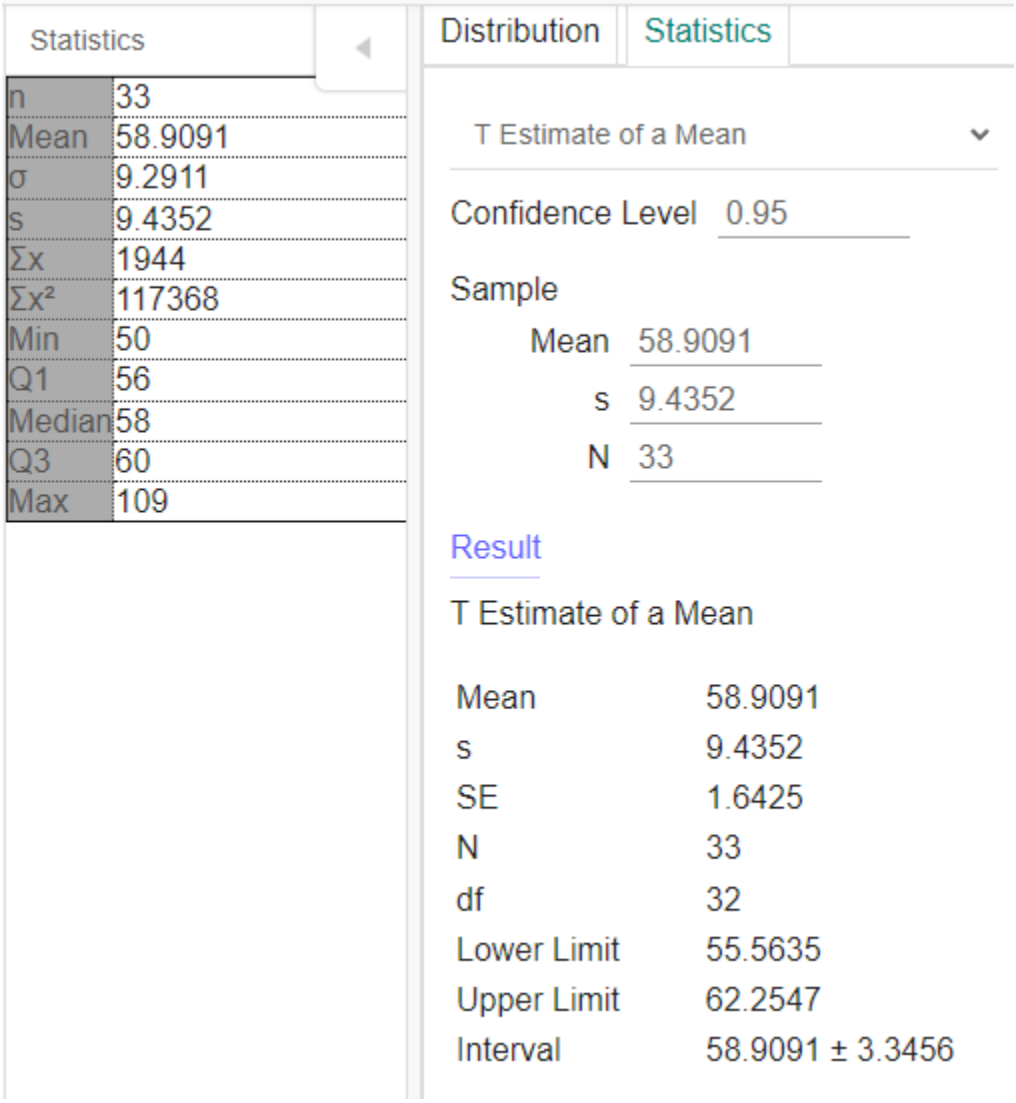
Upper fence = 64.75

Anything falling outside of the range of 50.75 - 64.75 would be considered an outlier which would include 50.



This does not meet the conditions for a normal model to be able to use a hypothesis test. I don't believe the following results will be valid.

Using the class data and a .95% confidence interval were able to calculate the confidence interval.



Statistics

n	33
Mean	58.9091
σ	9.2911
s	9.4352
Σx	1944
Σx²	117368
Min	50
Q1	56
Median	58
Q3	60
Max	109

Distribution | Statistics

T Estimate of a Mean

Confidence Level 0.95

Sample

Mean 58.9091

s 9.4352

N 33

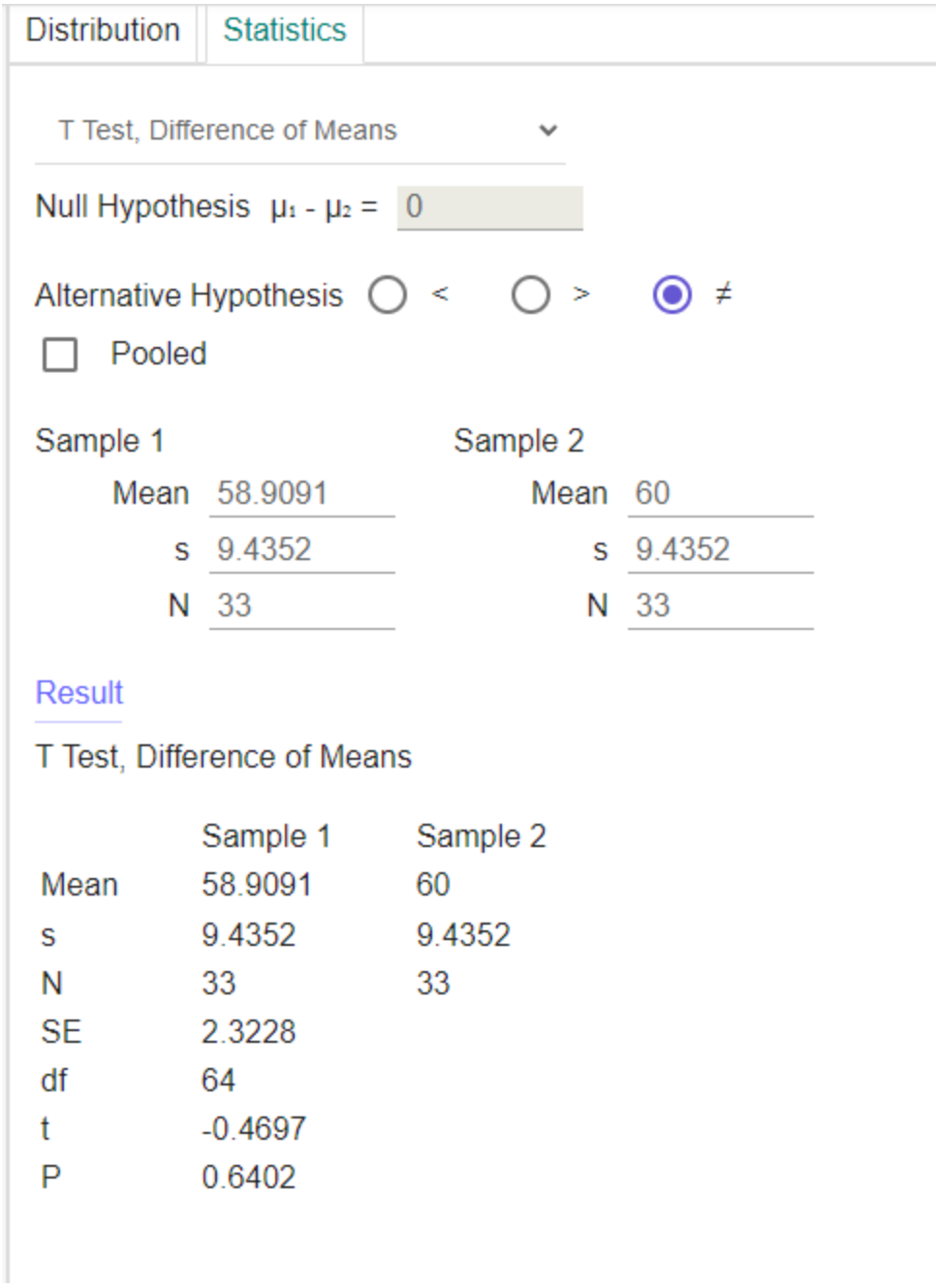
Result

T Estimate of a Mean

Mean	58.9091
s	9.4352
SE	1.6425
N	33
df	32
Lower Limit	55.5635
Upper Limit	62.2547
Interval	58.9091 ± 3.3456

This indicates that with 95% certainty the number of skittles will be between 55.5635 and 62.254.

If we were to assume that the standard mean found in a skittles bag is 60 and compare that to what was found for the classes bags.



Results in a P Value of .6402 which exceeds $\alpha = 0.05$.

This would then mean we would accept the null hypothesis.

Contingency table:

	Red	Orange	Yellow	Green	Purple	Total
Me	22	21	20	23	23	109
Student A	13	12	7	15	11	58
Student B	9	16	14	6	15	60
Total	44	49	41	44	49	227

What is the probability that a randomly selected skittle is red or yellow?

44+41=85

85/227=.3744

37.44 %

What is the probability that a randomly selected skittle is red or from your bag?

44+109-22=131

131/227=.5771

57.71%

What is the probability that a randomly selected skittle is red and from your bag?

22/109=.2018

20.18%

What is the probability that a randomly selected skittle is from your bag, given it is red?

22/227=.0969

9.69%

What is the probability that three randomly selected skittles (without replacement) are all red?

(22/227) * (21/226) * (20/225) = .0008

.08%

Checking that the conditions are met to calculate the difference in proportions of red skittles.

$\sqrt{\frac{22}{109} * \frac{13}{58}}$ = 0.21 (Pooled proportion

$n_1\hat{p} \geq 10$

109(.21) ≥ 10 =
22.89

$n_1(1 - \hat{p}) \geq 10$

109(1-.21) ≥ 10
86.11 ≥ 10

$n_2\hat{p} \geq 10$

58(.21) ≥ 10
12..18

$n_2p(1 - \hat{p}) \geq 10$

58(1-.21) ≥10
45.82 ≥ 10

Independent Samples?
The samples are independent or random.

$n_1 \leq .05N_1 \; and \; n_2 \leq .05N_2$

109/.05 = 2180

58/.05=1160

Meets all of the conditions.

Z Estimate, Difference of Proportions

Confidence Level 0.95

Sample 1		Sample 2	
Successes	22	Successes	13
N	109	N	58

Result

Z Estimate, Difference of Proportions

	Sample 1	Sample 2
Successes	22	13
N	109	58
SE	0.0669	
Lower Limit	-0.1534	
Upper Limit	0.1088	
Interval	-0.0223 ± 0.1311	

Z Test, Difference of Proportions

Null Hypothesis $p_1 - p_2 =$ 0

Alternative Hypothesis ○ < ○ > ◉ ≠

Sample 1		Sample 2	
Successes	22	Successes	13
N	109	N	58

Result

Z Test, Difference of Proportions

	Sample 1	Sample 2
Successes	22	13
N	109	58
SE	0.0662	
Z	-0.3372	
P	0.736	

The p value being 0.736 > .05 there does seem to be a significant difference in proportion between the two bags.