My M&M OCD

Yoni

13 04, 2025

Intro

Objective of Simulation

The goal of this simulation is to test the statistics of M&M and other stacks even Chocolate lentils by color.

I wanted to know what is the chance of my finishing the package of M&M without mixing any color in one bite, eating them 2 by 2

In addition, here are some BI incite that needed to be checked:

1. What is the probability of M&M packages packaged fairly?
2. What is the probability of M&M packages packaged without one color?
3. How does the size of the package or number of colors affect this probability?

Method

I do not have an inner information of how does (M&M)[https://www.mms.com/en-us] make their delicious snacks nor we know how they make sure each package have fair amount of each color. Therefore, the method I chose is based of simulation of some M&M bags, according to the most common sizes of packages.

Each time we sample x lentils(units of M&M), name them by colors (V1,V2…), and see the results for many packages as a statistic data.

my hypothesis is that the probability of perfect package (aka a package with all colors number been even) is very small, at least for a standard 6 colors pack.

Parameters

Basic parameters:

```
#parameters
n<- 800            #numbers of bags per sample
n_color<- 6         #unique colors of M&M
gram<- 0.91         #weight of one M&M
bag_g<- 250         #common weight of M&M package
n_unit<- bag_g/gram #M&M per packagenm,
av_per_color= n_unit/n_color
paste0("The avarage number of lentils per color is ", round(av_per_color,2))
```

```
## [1] "The avarage number of lentils per color is 45.79"
```

Creating of the Sample

General Sample

In order to test the theoretical data, we need to simulate it using customize functions. here are there:

- Create_bag- function to create one snack package for chosen package size and number of colors.
- sample_MnM- function to create n bags from the Create_bag function.

```
## [1] "One bag:"

##       1  2  3  4  5  6
## [1,] 10 19 14 20 18 20

## [1] "3 bags:"

##       Red Blue Green Orange Yellow Brown
## Bag_1   1    0     1      5      3     0
## Bag_2   2    1     2      1      2     2
## Bag_3   1    1     1      1      4     3
```

Preview Graph

Now will be creating n bugs of M&M

columns:

1. V1:V6- the number of lentils per color
2. even_count- how many evens colors there are
3. even_evens- are the uneven colors even
4. Variance- variance of lentils per color
5. low_col- sum true if one color's count is lower than $\frac{2}{3}$ of expected value
6. min- the lowest color in each row

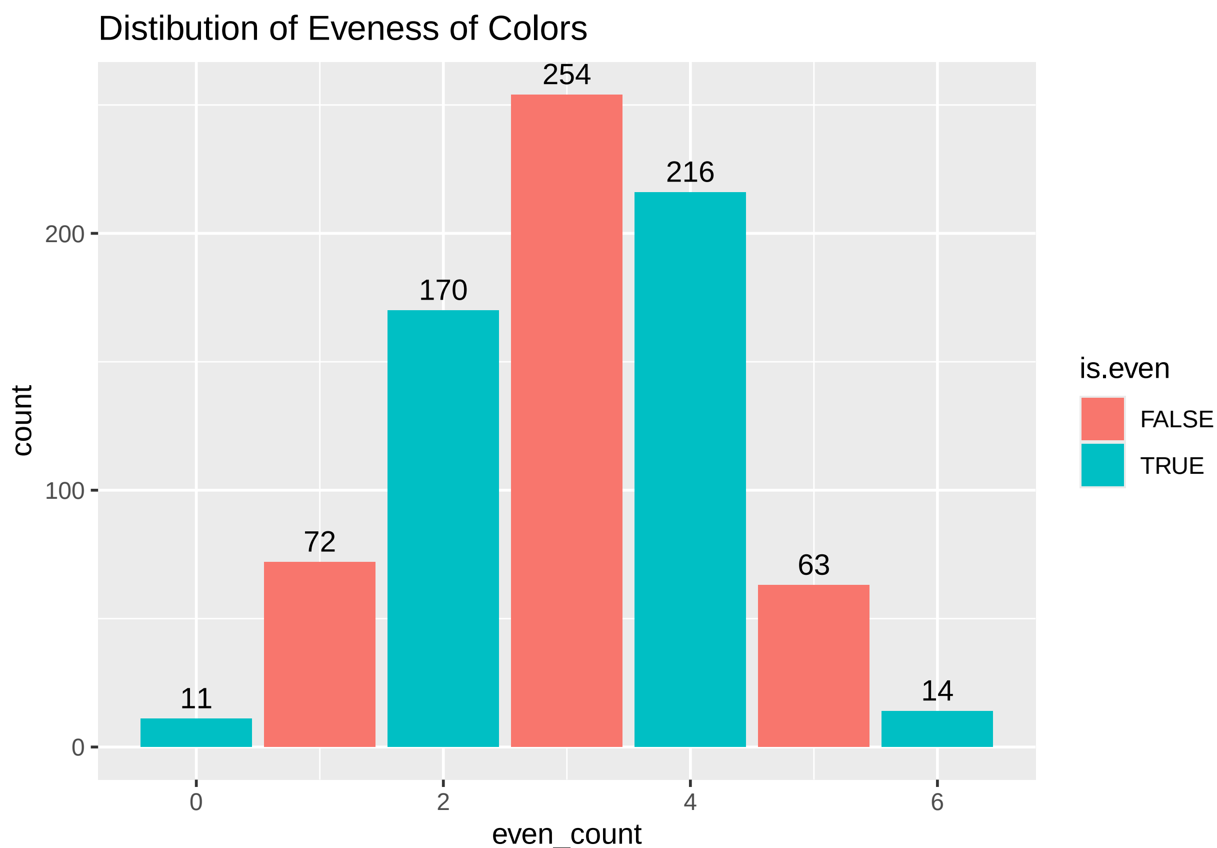
here are the first rows:

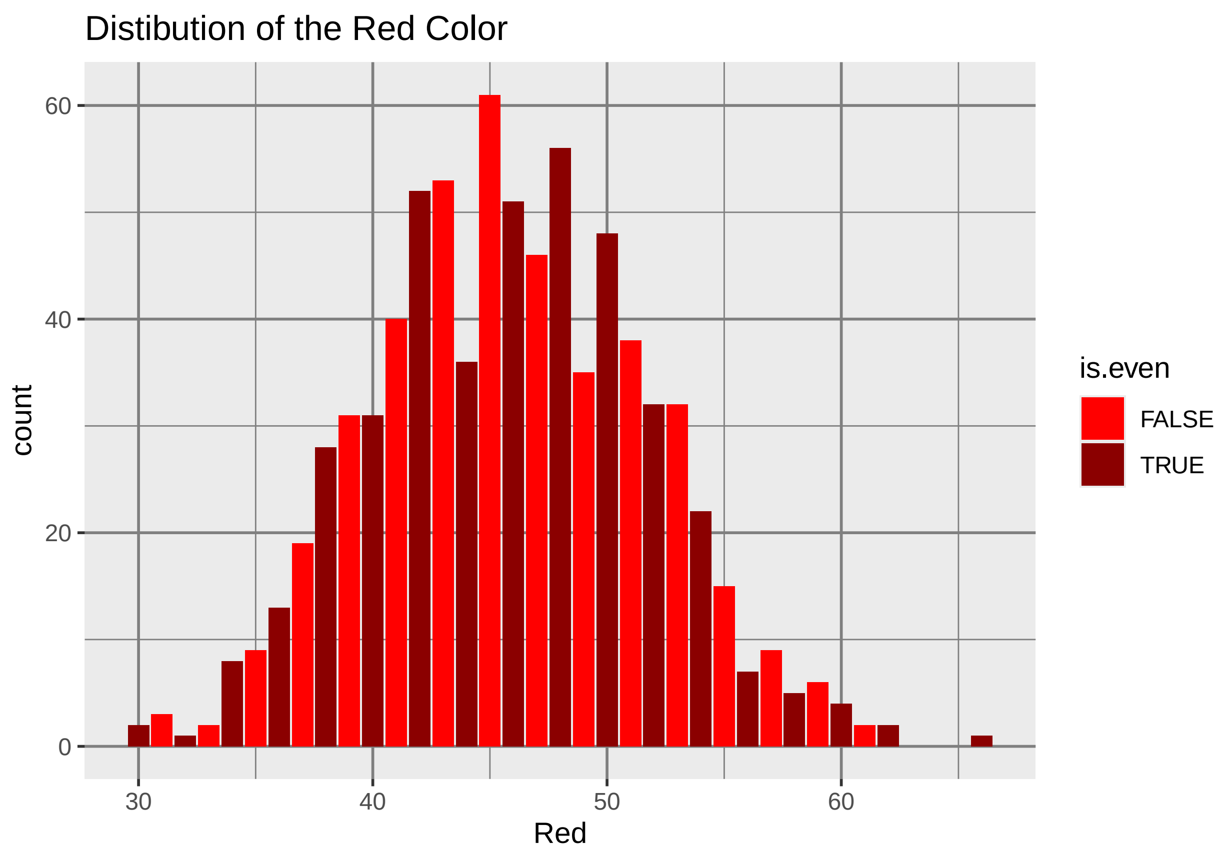
```
## # A tibble: 6 x 12
##     Red  Blue Green Orange Yellow Brown even_count even_evens low_col Variance
##   <int> <int> <int>  <int>  <int> <int>      <dbl> <lgl>        <dbl>    <dbl>
## 1    42    50    53     41     44    45          3 FALSE            0     22.2
## 2    43    47    53     44     46    42          3 FALSE            0     15.8
## 3    32    52    44     57     47    43          3 FALSE            0     73.4
## 4    39    56    46     49     37    48          3 FALSE            0     48.6
## 5    48    39    45     36     50    56          4 TRUE             0     53.9
## 6    44    42    50     53     39    47          3 FALSE            0     27.0
## # i 2 more variables: min <int>, all_even <lgl>
```

plot the M&M sample sample

```
## [1] "summary of all colors Distibution:"
```

```
##         Min. 1st Qu. Median     Mean 3rd Qu. Max.      Var
## Red       30      42     46 45.84625      50   66 33.83991
## Blue      24      42     46 46.10125      51   62 36.84455
## Green     26      42     46 45.74750      50   69 41.31038
## Orange    24      41     45 45.54375      50   67 38.92424
## Yellow    28      41     46 45.74750      50   66 39.48310
## Brown     26      41     46 45.50000      50   67 37.80225
```





Statisics Checking of the Simullation

Test Expected Value

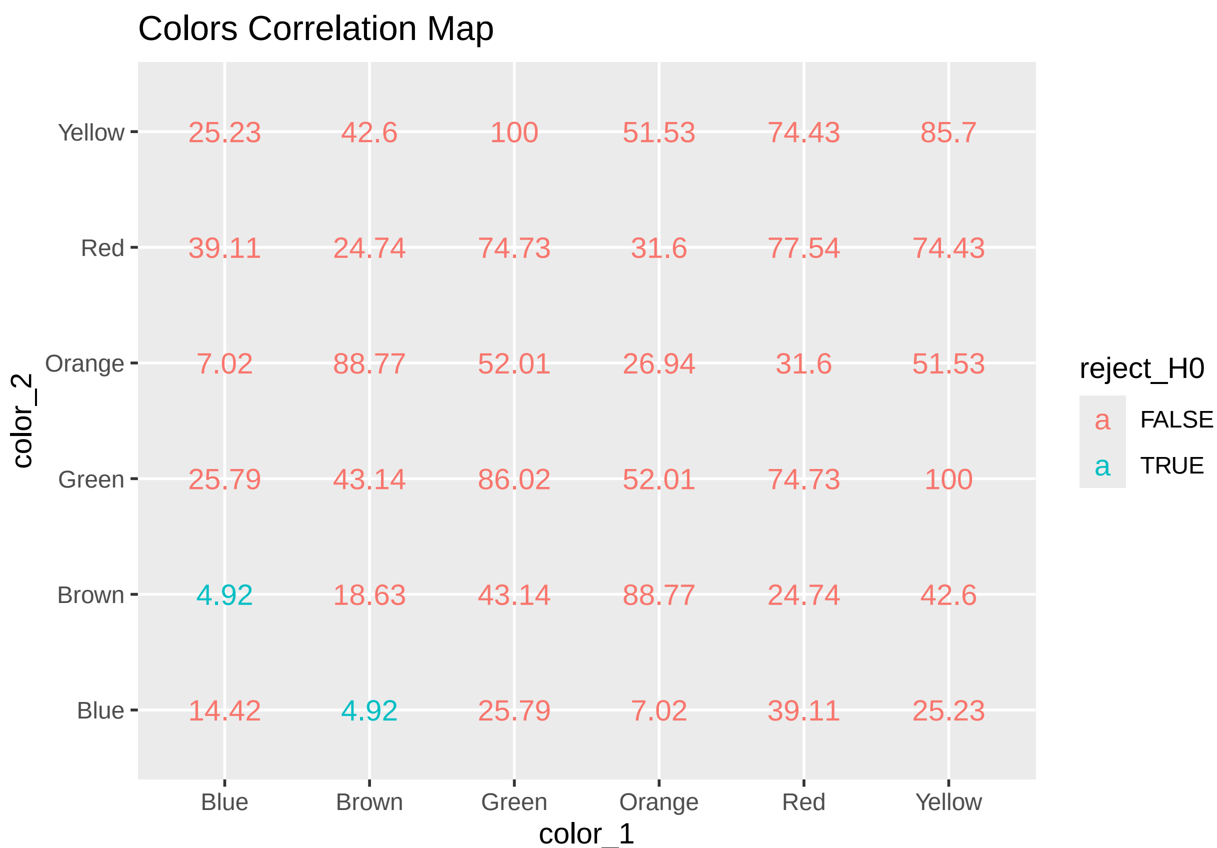
to see is the μ of the lentils per color are fair, we will test it per column with t.test for each color.

Here is the result, none of them bellow 5% P. value

```
##      Red     Blue    Green   Orange   Yellow    Brown
## "77.54%" "14.42%" "86.02%" "26.94%" "85.70%" "18.63%"
```

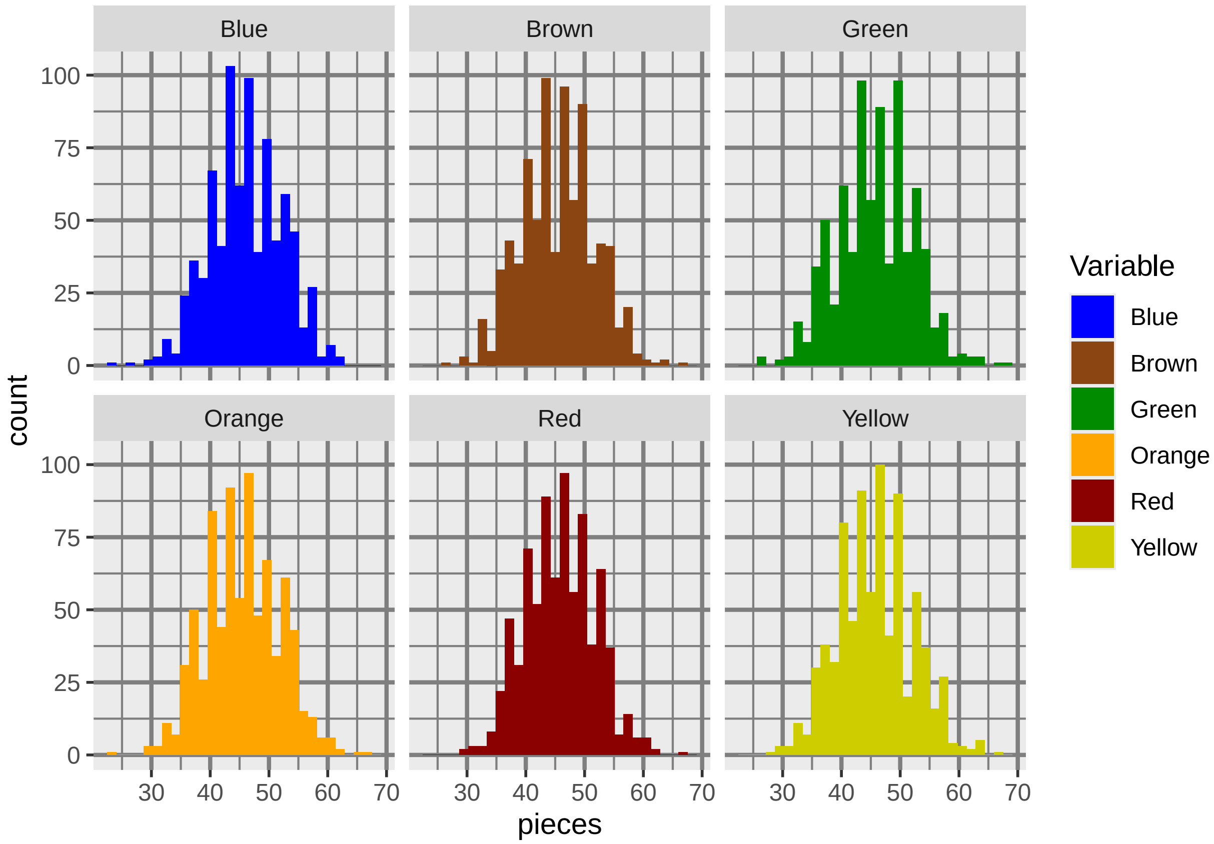
Now we will do the same checking for 2 samples, to see whether there is correlation between each 2 colors distribution.

for each row i and column j, 1) if i==j, this it the check from before of the expected value to n_unit/n_color 2) if i!=j, this is two samples test of same expected value hypothesis



Now here Is visualization of the actual data per color

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



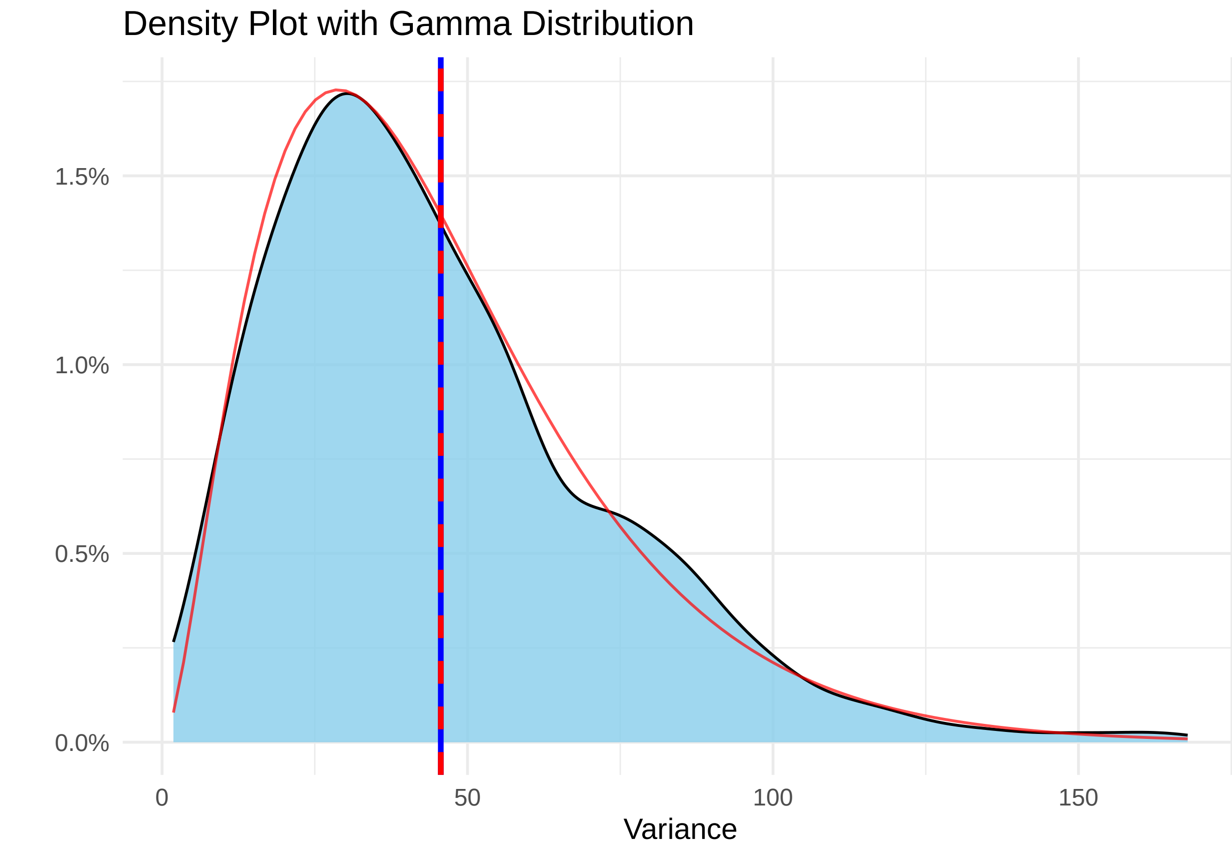
Variance Distribution Checking

We know that the distribution of variance is approximately Gamma distribution:

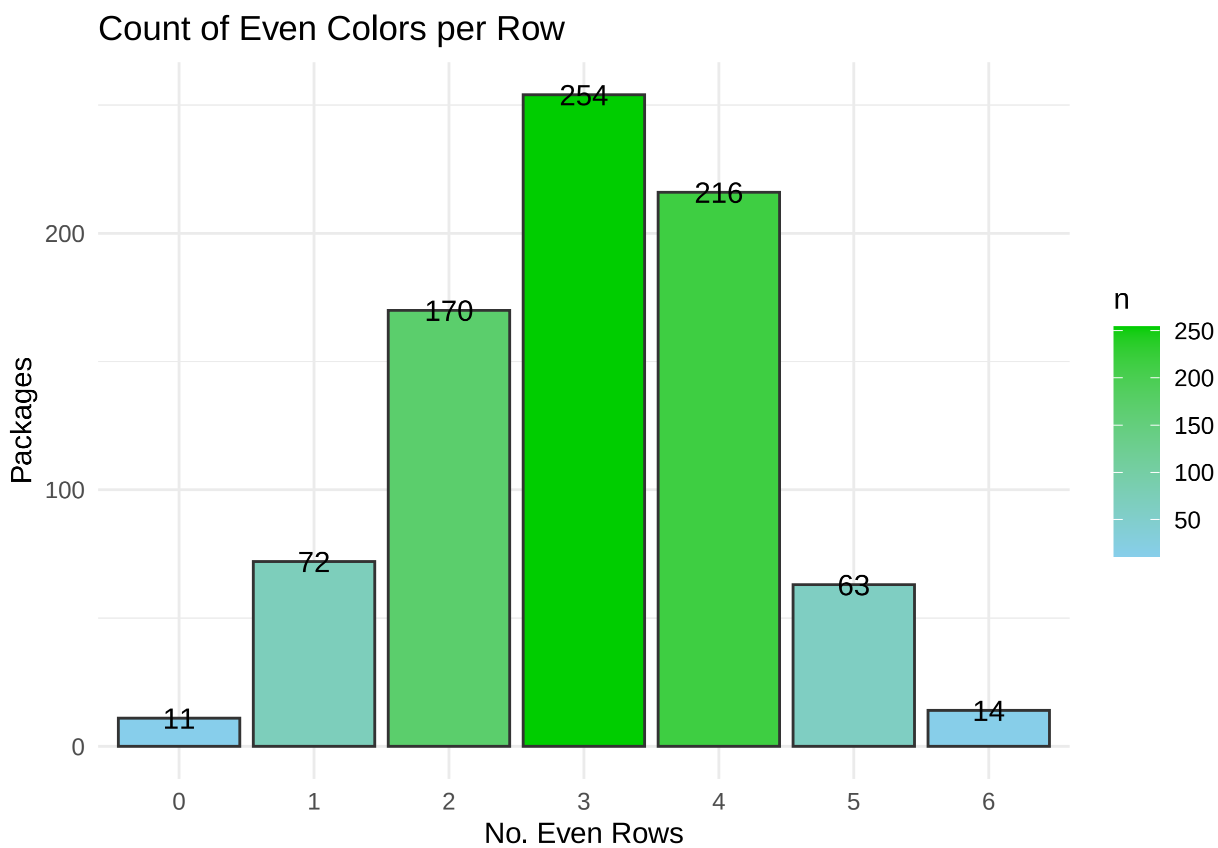
$$f(x) = \frac{1}{(\Gamma(\alpha)\theta^{\alpha})} x^{\alpha-1} e^{-x/\theta}$$

We can see that the variance distribution is Gamma like with shape and rate as seen below

```
## [1] "The parameters of the gamma shaped variance is shape 2.716 and rate 0.06"
```



Are All Even in the Sample?



n*m types of snacks

I will create a function that create sample for each number of colors and package size we want, and then calculate some interesting parameters

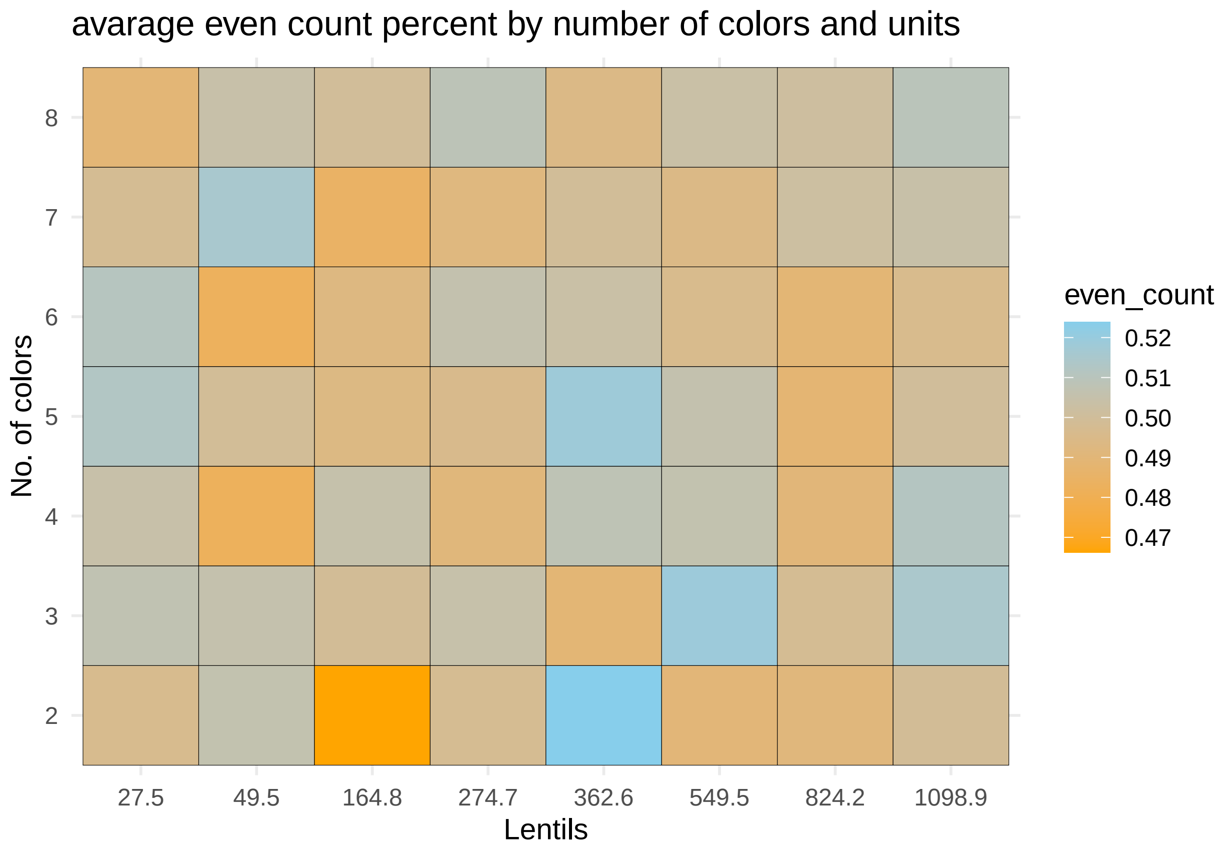
```
color_op<- 2:8
grams_op<- c(25,45,150,250,330,500,750,1000)
n_unit_op<- grams_op/gram
nn=800
```

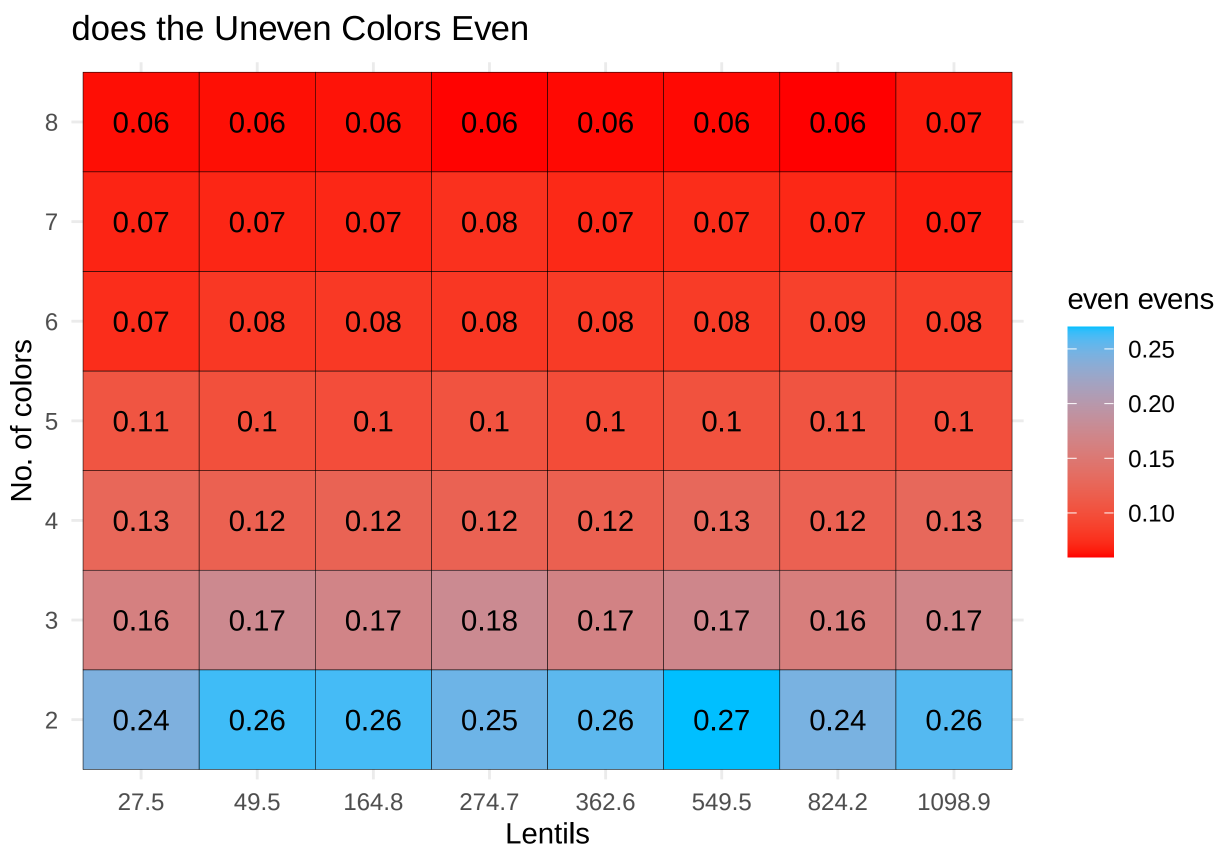
We will make the multiple sample. Here is some random rows:

```
##   n_unit n_color even_count even_evens   var_col all_even low_color
## 1  164.8       7  0.4846429 0.07107143  23.20125  0.00625   0.06125
## 2  274.7       4  0.4909375 0.12093750  67.88813  0.04750   0.00125
## 3   49.5       3  0.5054167 0.17500000  17.03750  0.14000   0.05250
## 4 1098.9       4  0.5115625 0.12875000 284.28500  0.06875   0.00000
## 5  362.6       3  0.4891667 0.16541667 116.87188  0.10875   0.00000
##   smallest_col
## 1            9
## 2           44
## 3            5
## 4          226
## 5           94
```

Deep Insight on the Data

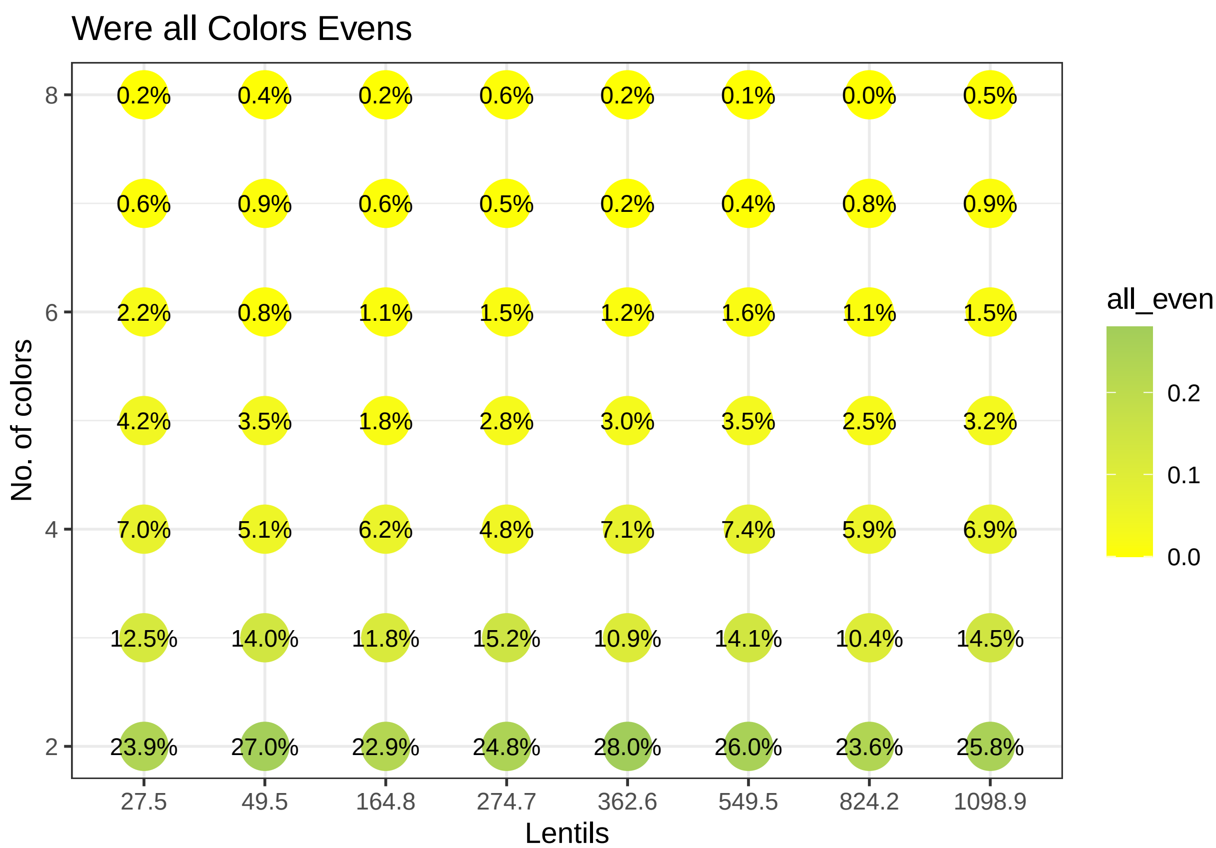
here are some insights:

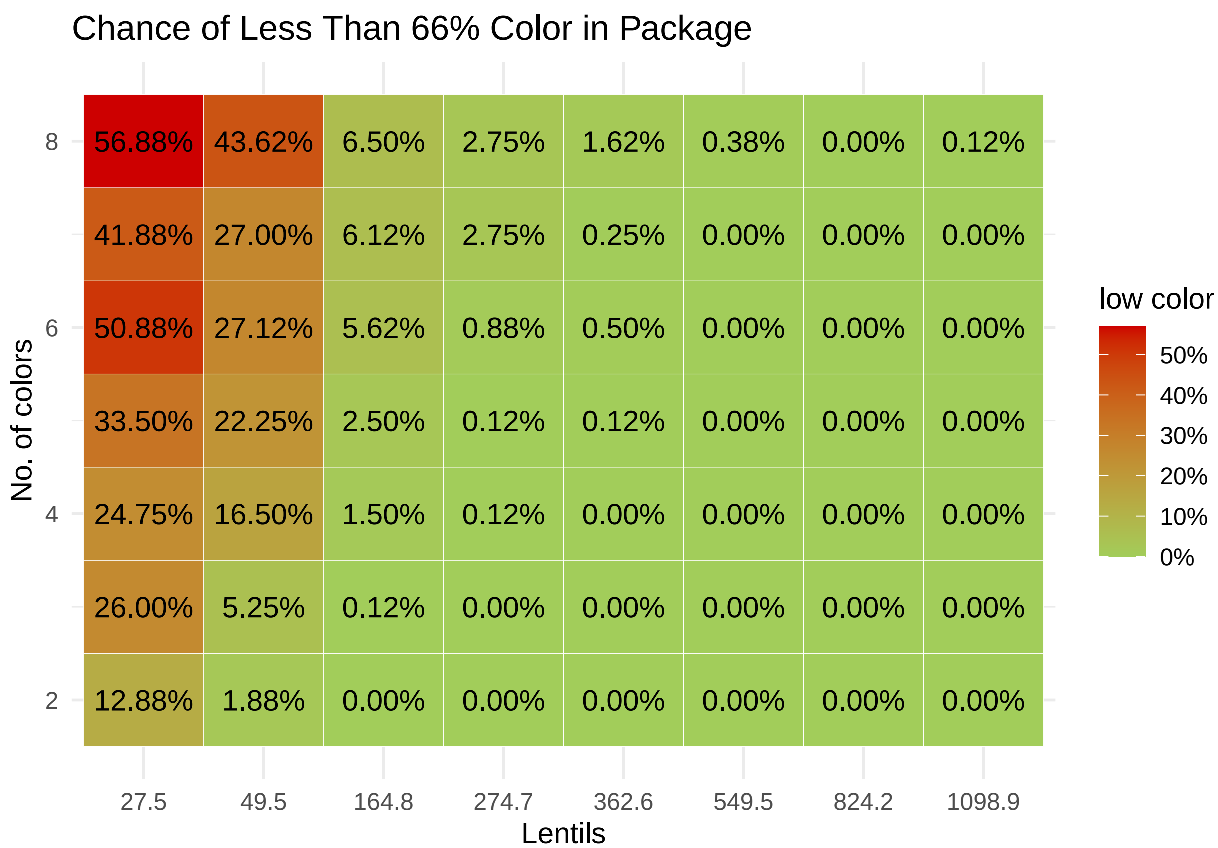


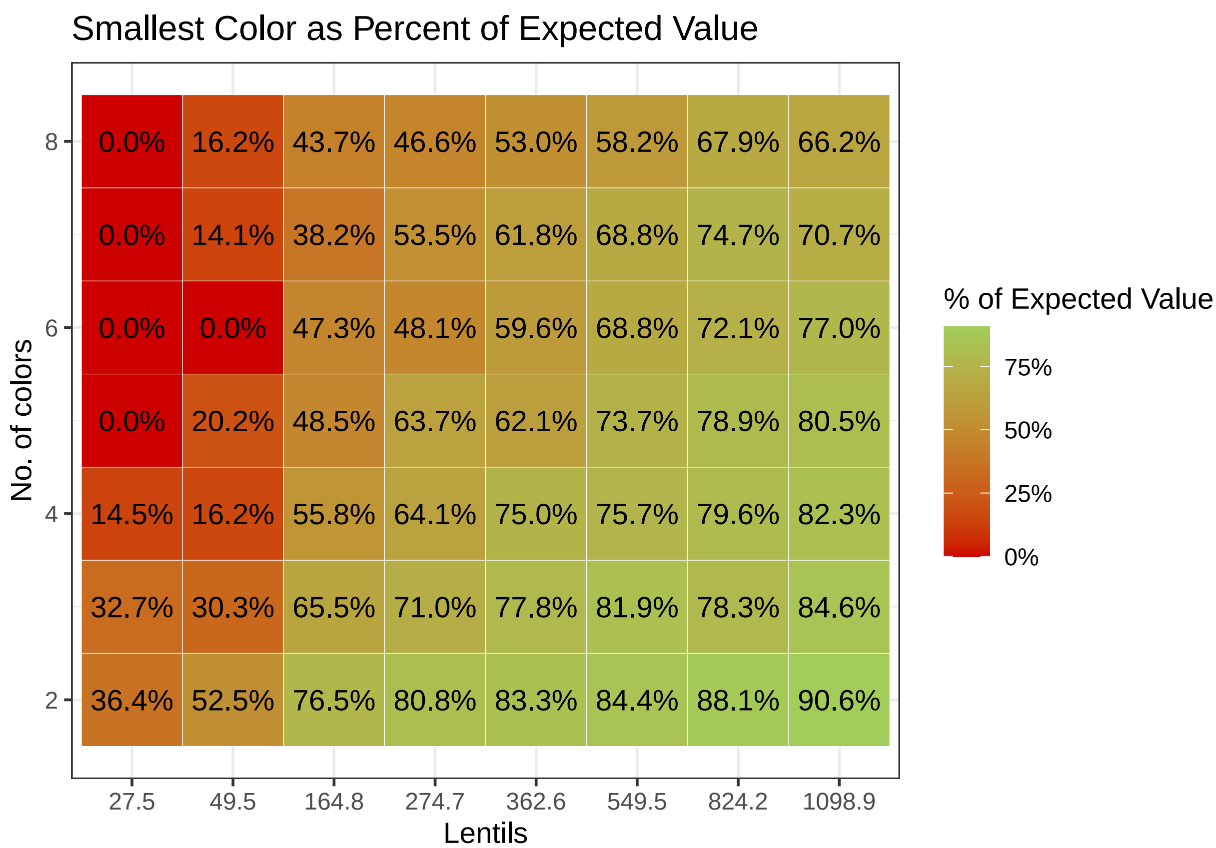


Here is probability of all even, and whether there is pattern.

```
## <Guides[1] ggproto object>
##
## colour : "none"
```


Were all Colors Evens
No. of colors
Lentils
all_even
0.0
0.1
0.2
2
4
6
8
27.5
49.5
164.8
274.7
362.6
549.5
824.2
1098.9
0.2%
0.4%
0.2%
0.6%
0.2%
0.1%
0.0%
0.5%
0.6%
0.9%
0.6%
0.5%
0.2%
0.4%
0.8%
0.9%
2.2%
0.8%
1.1%
1.5%
1.2%
1.6%
1.1%
1.5%
4.2%
3.5%
1.8%
2.8%
3.0%
3.5%
2.5%
3.2%
7.0%
5.1%
6.2%
4.8%
7.1%
7.4%
5.9%
6.9%
12.5%
14.0%
11.8%
15.2%
10.9%
14.1%
10.4%
14.5%
23.9%
27.0%
22.9%
24.8%
28.0%
26.0%
23.6%
25.8%


Chance of Less Than 66% Color in Package
No. of colors
8
6
4
2
Lentils
27.5
49.5
164.8
274.7
362.6
549.5
824.2
1098.9
56.88%
43.62%
6.50%
2.75%
1.62%
0.38%
0.00%
0.12%
41.88%
27.00%
6.12%
2.75%
0.25%
0.00%
0.00%
0.00%
50.88%
27.12%
5.62%
0.88%
0.50%
0.00%
0.00%
0.00%
33.50%
22.25%
2.50%
0.12%
0.12%
0.00%
0.00%
0.00%
24.75%
16.50%
1.50%
0.12%
0.00%
0.00%
0.00%
0.00%
26.00%
5.25%
0.12%
0.00%
0.00%
0.00%
0.00%
0.00%
12.88%
1.88%
0.00%
0.00%
0.00%
0.00%
0.00%
0.00%
low color
50%
40%
30%
20%
10%
0%



As we can see, only the small package (less than 50 lentils) have high probability of at least one color to appear severely lower.

Therefore, splitting package by color on the big ones should be relatively even.

using regression for correlation check

```
##
## Call:
## lm(formula = mega_snack_2$even_count ~ mega_snack_2$n_color +
##     mega_snack_2$n_unit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.52516 -0.16197 -0.00482  0.14984  0.52417
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           5.327e-01  2.476e-02  21.510   <2e-16 ***
## mega_snack_2$n_color -3.254e-03  4.216e-03  -0.772    0.440
## mega_snack_2$n_unit  -3.693e-05  2.359e-05  -1.565    0.118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2444 on 837 degrees of freedom
## Multiple R-squared:  0.003625,   Adjusted R-squared:  0.001244
```

```
## F-statistic: 1.523 on 2 and 837 DF,  p-value: 0.2187
```

```
##
## Call:
## lm(formula = mega_snack_2$all_even ~ mega_snack_2$n_color + mega_snack_2$n_unit +
##     mega_snack_2$color_No2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.15171 -0.06540 -0.02814  0.01530  0.88555
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.440e-01  1.878e-02   7.671 4.75e-14 ***
## mega_snack_2$n_color      -2.589e-02  2.975e-03  -8.703  < 2e-16 ***
## mega_snack_2$n_unit       -3.478e-05  1.665e-05  -2.089    0.037 *
## mega_snack_2$color_No2TRUE  6.042e-02  1.202e-02   5.025 6.17e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1725 on 836 degrees of freedom
## Multiple R-squared:  0.1119, Adjusted R-squared:  0.1087
## F-statistic: 35.12 on 3 and 836 DF,  p-value: < 2.2e-16
```

Conclusions

Data Structure

The simulation created a random samples of snack packs, which was proven to be statistically random. We created with "sample_MnM" one sample with specific size and numbers of colors, and then "mega_snack" that create costume samples and check the relevand indicators fot this project.

we I out that:

- the smallest package can barely have one color or even to not have one.
- The more colors there are, the less chance there is for the uneven numbers in colors to be even
- For any medium or bigger package, the probability of one color to be unfairly small (less than $\frac{2}{3}$ than expected) is nearly 0%

Main Q: Eating M&M by Two

Although there is no clear pattern to the right M&M package for all the colors to have even count, maybe different approach can find a clear reason for more or less couples of M&M.

The general probability of all colors to be even in 6 colored pack is 1.5% for small 50g package 2.1% for big 1000g package, and overall 1.5%, which is more than I expected.

For 5 colored pack like Skittles the average is about 2.9%

For 2 colored pack the average is 24.5%, so for most 2 colored marshmallow bag the method of eating by 2 can be relatively available.

Summery

To sum it up, for each medium pack the probability of all even colors is 1.4%, or 1 in a 73 packs of 250g. So I might need to change my snack preference to marshmallow if I want to keep this method!