Beads in the Tin – Measuring Discrepancy

In the tin are some beads. Your teacher knows what the color distribution of beads in the tin is, and has made the following claim:

Claim: The beads in the tin follow the following color distribution:
35% are blue, 35% are green, 15% are yellow, and 15% are red

Your group should select ten beads from the tin, record their color, and put them back. DO NOT LOOK IN THE TIN. Then fill in the chart with the number you got of each color, and what you would have expected to get of each color to be exactly consistent with the claim (It does not matter if these "expected" values are possible or not).

Color	Number in Sample	Number Expected	How different?
Blue			
Green			
Yellow			
Red			

It is possible that *some* of the colors had numbers close to what you would have expected, while others did not. Is this what you found, or were all the numbers you obtained in your sample very different from the expected, or were they all very close to the expected?

Your groups job is to come up with a way to measure *overall discrepancy* (we will call this statistic D). That is, a single statistic for this sample that takes into account how far off each category was. Discuss this statistic, write its formula and its value for your sample below.

$D =$

So, you obtained a value for your D-statistic. Is this value a high value or a low value? What is the LOWEST possible value D could have taken in your sample of 10?

What is the HIGHEST possible value D could have taken in your sample of 10?

We will perform a simulation to see where your value of D compares to other samples that were taken from the population in question. Your teacher has plastic bags filled with beads that contain beads *in the claimed proportions*. You can see this is true by looking in the bags.

Select ten samples of ten beads. For each sample, calculate the value of D. Plot those values on the dotplot below, and collaborate with your group members to obtain 30 dots.

← D →

Place a vertical line at the value of D we obtained from our sample of beads from the tin. What proportion of dots falls to the right of this value?

Interpret this proportion of dots.

Given this, do you think the original claim about the beads in the tin is plausible, or not?

In statistics, we have an accepted measure of "discrepancy" for categorical data, and it is called Chi-Square (χ^2). It is calculated this way:

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

Calculate the χ^2 statistic for your sample of beads from the tin.

Color	Number in Sample	Number Expected	$\frac{(Observed - Expected)^2}{Expected}$
Blue			
Green			
Yellow			
Red			
		χ^2 = Sum =	

The test we performed on the beads was called a "Goodness-of-Fit" test.

We will talk about these procedures in greater detail next class period. For homework, watch the videos "Chi Square Statistic" and "Chi-Square Goodness-of-Fit" on Canvas.