AP BIOLOGY
Evolution/heredity Unit
Unit 1 part 5
Activity #7

NAME_____

DATE_____PERIOD_____

M&M STATISTICS – A CHI SQUARE ANALYSIS

INTRODUCTION

Have you ever wondered why the package of M&Ms you just bought never seems to have enough of your favorite color? Or, why is it that you always seem to get a package with mostly brown M&Ms? What's going on at the Mars Company? Is the number of the different colors of M&Ms in a package really different from one package to the next? Or, does the Mars Company do something to ensure that each package gets a certain number of each M&M color? The following information is from the Mars Company web site.

	
% color **___ Blue ___Red ___Brown** **___Green ___Yellow ___Orange**	**% color** **___ Blue ___Red ___Brown** **___Green ___Yellow ___Orange**

Source: **http://www.mmmars.com/cai/mms/faq.html#what_percent**

One way we can determine if the Mars Company is true to its claims is to sample a package of M&Ms and complete a statistical test known as a "goodness to fit" test. This type of statistical test allows us to determine if any differences between our observed measurements (count of each color from the M&M package) and our expected (claims posted at the M&M web page) are due to chance or some other reason (the Mars Company sorters aren't doing a good job of putting the correct number of M&Ms in each package.) The goodness of fit test we will be using is called a chi-square analysis. This test is generally used when dealing with discrete data (i.e. count data or discontinuous data.) We will be calculating a statistical value and using a table to determine the probability that the differences between observed data and expected data is due to chance alone.

The formula for calculating chi-square *(x^2)* is**:**

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

That is chi-square is the sum of the squared difference between observed (o) and expected (e) data divided by the expected data in all possible categories.

HYPOTHESIS:

If the Mars Company sorters are working properly then any difference between the color percentage in an actual package of M&Ms and the color percentage posted on the web site should be due to random chance.

PROCEDURE:

1. Wash your hands. You will be handling food you may want to munch on later.
2. Put paper towels down on your desk. You will be counting the M&Ms on your desk.
3. Open the bag of M&Ms.
4. Do NOT eat any of the M&Ms right now. Separate the M&Ms into color categories and count the number of each color.
5. Record your M&M color totals in the data table.
6. Determine the total number of M&Ms in your package and record this number in the data table.
7. Calculate the expected number of M&Ms in your package by multiplying the total number of M&Ms in the package by the color percent listed on page 1 of the activity. For example, if your package contains 500 M&Ms and you want to find the expected number of red M&Ms you will need to multiply 500 by 20% (500 x 0.20). Record your calculations in the data table.
8. Calculate the difference between the observed and expected numbers for each M&M color. Record your calculations in the data table.
9. Square the difference between the observed and expected. Record your calculations in the data table.
10. Divide the square of the difference by the expected. Record your calculations in the data table.
11. Total all the answers from step 10 to determine the chi-square (x^2) value. Record the chi-square (x^2) in the data table.

DATA TABLE

	Yellow	Red	Orange	Blue	Green	Brown	Total
Observed (o)							
Expected (e)							
Difference (o-e)							
Difference squared (d^2) $d^2 = (o-e)^2$							
d^2/e							
$\Sigma (d^2/e) = x^2$							

12. Now you must determine the probability that the difference between the observed and expected values (as summarized by the calculated value of chi square) occurred simply by chance. To do this you will need to compare the calculated value of chi-square with the appropriate value from the Chi Square Distribution Table on the next page. Examine the table. Note the term "degrees of freedom." For this statistical test the degrees of freedom is equal to the number of classes (color categories) minus one. Complete the following to determine the degrees of freedom for the M&M analysis:

# of color categories [ ]

– 1

degrees of freedom [ ]

13. The reason why it is important to consider degrees of freedom is that the value of the chi-square statistic is calculated as the sum of the squared differences for all classes. The natural increase in the value of chi-square with an increase in classes must be taken into account. Scan across the row corresponding to 5 degrees of freedom. Values of the chi-square are given for several different probabilities ranging from 0.95 on the left to 0.001 on the right. Note that the chi-square increases as the probability increases. Notice that a chi-square value of 1.63 would be expected by chance in 95% (0.95) of the cases, whereas one of 12.59 would be expected in 5% (0.05) of the cases. Use the chi-square value calculated and recorded on the data

table to determine the probability for the M&M analysis. If the exact chi square value is not listed in the table estimate the probability. Record your answer below.

Chi-square value	
Probability	

CHI-SQUARE DISTRIBUTION TABLE

	Accept Hypothesis →								Reject Hypothesis →		
	Probability (p)										
Degrees of Freedom	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	**0.05**	0.01	0.001
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	**3.84**	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	**5.99**	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	**7.82**	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	**9.49**	13.38	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	**11.07**	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	**12.59**	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	**14.07**	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	**15.51**	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	**16.92**	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	**18.31**	23.21	29.59

14. Scientists, in general, are willing to accept a hypothesis if the probability that the difference between the observed and expected results is greater than 5% (0.05). If the probability determined in question 13 is greater than 5% (0.05) then any differences between the observed color counts and the claims of the Mars Company (posted at their web site) is due to chance alone. Five percent! That's not much … but it's good enough for scientists! If, however, the probability determined in question 13 is less than 5% (0.05) then any differences between the observed color counts and the claims of the Mars Company is not due to chance. Some other factor caused the differences. Based on your group's results, should you accept or reject the hypothesis? Explain your answer.

_____

_____

Chi Square Practice Problems

Solve all problems using a chi square analysis.
You must use statistics to support your answers.

1. A zookeeper hypothesizes that changing the intensity of the light in the primate exhibits will reduce the amount of aggression between the baboons. In exhibit A, with a lower light intensity, he observes 36 incidences of aggression over a one month period. In exhibit B, with normal lights, he observes 42 incidences of aggression. Should he support or reject his hypothesis?

2. A high school, students can choose to enter one of three doors. Custodians noticed that door #3 was always getting broken and suggested that more students use that door because it has a hands-free opener. Science minded students counted the number of students entering each door to see if the custodians were right.

Door #1 had 60 students enter | Door #2 had 66 students enter | Door #3 had 80 students enter.

Were the custodians right?

3. A scientist predicts that the kittens born with a congenital birth defect will be 25% based on the hypothesis that it is caused be a recessive gene in that breed of cat. After surveying several litters, he found that 44 out of 125 kittens had the defect. Is his hypothesis correct?

4. Suppose you take a random sample of 30 students who are using a new math text and a second sample of 30 students who are using a more traditional text. You compare student achievement on the state test given to all students at the end of the course. Based on state test performance, would you recommend the new math book?

	Passed State Test	Failed State Test
New Textbook	26	4
Old Textbook	22	8

Degrees of Freedom	good fit					poor fit		
	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086

End of Chapter Synthesis and Evaluation Problems

Study Guide/ISN (20 points)

1. In your study guide book, review pages 326 to 328. In your ISN, go back to the following: **Chi Square Must Know!** In one color, copy down the must know item listed on page 326 in study guide leaving space underneath to include in an different color a brief description, diagram, model, or pneumonic device that will help you study for the unit test and more importantly the AP Test in May.
2. Do the questions 1-3 (pg 328-329) in study guide. Check your answers pg 398 when done
   a. _____
   b. _____
   c. _____

Bozeman Science/ AP Biology/ISN (See Syllabus for format) **(20 points each)**

1. Chi-squared test (AP Biology labs)
2. Data Collection Strategies (AP Biology Practices)