AP Statistics Name__________________________

Chi-Square Test for Goodness of Fit Introduction

Suppose you have a six-sided die numbered in the usual way. Is this die fair? How would you determine its fairness? If the die is fair and each of your classmates rolled the die 60 times, how many 1's, 2's, 3's, etc would you expect your classmates to get, on average? How much deviation from the averages would you expect to get in your own set of 60 rolls? Most importantly, how would you use data from your 60 rolls to make a statistical determination about the fairness of the die?

This handout will introduce you to another test of significance to add to your repertoire of statistical inference tools—the chi-square test for Goodness of Fit.

1. Write a null and alternative hypothesis (in words) that would be appropriate for testing if the die is fair.

To get an idea of how this chi-square test works, you will invent three data sets concerning the die.

2. In the tables below, invent the following data sets

   - In the left most table, create a distribution of 60 rolls that you think would not be unusual for a fair die to produce. Make sure your distribution adds up to 60 rolls.
   - In the center table, create a distribution of 60 rolls that would make you question the fairness of your die. Make sure your distribution adds up to 60 rolls.
   - In the right most table, create a distribution of 60 rolls that would give you strong, almost overwhelming, evidence that your die is unfair. Make sure your distribution adds up to 60 rolls.

Not Unusual	
Die Number	Number of Rolls
1	
2	
3	
4	
5	
6	

Questionable	
Die Number	Number of Rolls
1	
2	
3	
4	
5	
6	

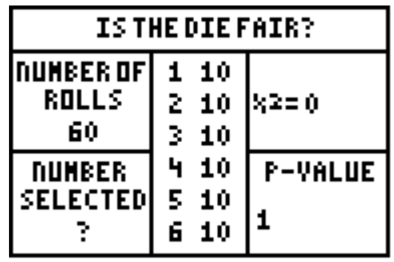
Likely Unfair	
Die Number	Number of Rolls
1	
2	
3	
4	
5	
6	

The general pattern of any test of significance can be simplified as follows:

- Make null and alternative hypotheses about a situation.
- Collect data.
- Obtain a *P*-value based on these data.
- Use the *P*-value to determine the strength of evidence against the null hypothesis

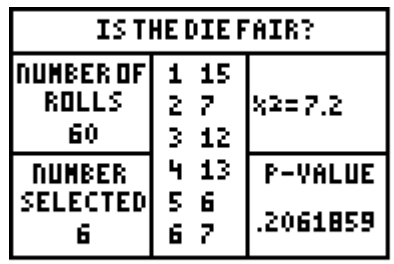
You will use a program on the calculator to skip to the end of this process and calculate the *P*-value for each of your three data sets.

In your StatsApp application, run the program CHISQDIE. You will get the screen given at right.



The 10's in each category represent what you would obtain, on average, in each category if the die is fair.

You will now use the program to input each of your three distributions and obtain the *P*-value for each distribution. To use the program, hit the number (1 through 6) on the keypad of the number you would like to change. Then, use the UP and DOWN arrows on your keypad to get the correct number for the selected category. An example is given at right.



3. Use the program to obtain the *P*-value for your "Not Unusual" distribution. In addition, record the value above the *P*-value. This value is called "chi-square" and is abbreviated as χ^2 -- more about this character later in the handout. Record these values in the table below.

Not Unusual	
χ^2	*P*-value

Questionable	
χ^2	*P*-value

Likely Unfair	
χ^2	*P*-value

Press "ENTER" to have the option of resetting the frequencies to 10 in each category.

4. Use the program to obtain the *P*-value and χ^2 for your "Questionable" distribution and record these values in the table above.

5. Use the program to obtain the *P*-value and χ^2 for your "Likely Unfair" distribution and record these values in the table above.

6. For each of the three data sets, write a sentence or two describing what the *P*-value indicates about the strength of the evidence against the null hypothesis (the die is fair).

A Real Data Set and What χ^2 Represents

A simple random sample of 175 students from among students at a school from the past 10 years was conducted. The day of the week on which each of the students was born is given in the table at right.

Day	Observed Count
Mon	29
Tue	30
Wed	33
Thu	27
Fri	26
Sat	19
Sun	11
Total	**175**

In most cases childbirth happens on its own timetable, and one of the great mysteries of late pregnancy is when this momentous event will occur.

For some mothers-to-be, however, circumstances dictate that they will have a more planned delivery, either due to Cesarean section or induced labor. Is there evidence in our sample that the distribution of days of the week on which this school's students are born is not uniform?

The key to answering the above question is χ^2. What is χ^2? Essentially, it is one number that summarizes the variation in the data. How is it calculated? Answer the questions below to work through the calculation.

7. Assuming that each day of the week is equally likely for a birth, how many of the 175 births would you expect, on average, to occur on each day? Fill in the numbers in the "Expected Count" column below.
8. Calculate the difference between the Observed and Expected counts and fill in the "Difference" column—you should have some positive and some negative numbers.
9. Square each of your differences and fill in the "Squared Difference" column.
10. Finally, take each of your squared differences and divide by the expected count for that day (25 in each case here). These "Scaled Squared Differences" account for the sample size in this problem and brings things down to a common scale.

Day	Observed Count	Expected Count	Difference	Squared Difference	Scaled Squared Difference
Mon	29				
Tue	30				
Wed	33				
Thu	27				
Fri	26				
Sat	19				
Sun	11				
Total	175				

11. Find the sum of these scaled squared differences. This number is χ^2!

Simulation of the Birth Days

If it is true that a randomly selected student is just as likely to be born on any day of the week, then among the 175 students sampled you would expect to see some variation from 25 each day simply due to random variation. The sum of the scaled squared differences that you just found, known as the χ^2 test statistic, would be zero if the counts were all exactly 25. χ^2 measures how different the observed counts are from 25, and the essential question we'd like to answer is whether the value we obtained—which should have been 13.68—is among the typical totals if the uniform model for birthdays is true.

One way to see if 13.68 is an unusual χ^2 value is to simulate the random selection of 175 students and see what value of χ^2 the resulting data set produces.

12. You will simulate this situation using your calculator. To create a random list of days of birth for 175 students, run the command below to put 175 birth days in list L1.

randInt(1,7,175)-> L_1

13. To make it easier to count the number of each "day" that occurs, construct a histogram with the data in list L1. Then use TRACE to find the number in each category. For your histogram, suggested window settings are given below.

Xmin=-.5   Xmax=7.5   Xscl=1   Ymin=0   Ymax=40

14. Record your numbers for each category in the table below—then calculate the columns as before to obtain your χ^2 value. Share your total with the class.

Day	Observed Count	Expected Count	Difference	Squared Difference	Scaled Squared Difference
Mon					
Tue					
Wed					
Thu					
Fri					
Sat					
Sun					
Total	175				

15. Where does 13.68 fall among the values in your class distribution? Does 13.68 appear to be an unusual value?

The Chi-Square Test for Goodness of Fit

Follow the steps below to carry out a chi-square test for Goodness of Fit using the original data set of birth days. (Note the similarities to previous tests of significance you have performed.)

1. **Conditions**:

   This is a random sample of the students. There is a categorical variable. All expected counts are greater than 5. This is true, since ____.

2. **Hypotheses:**
   Let $p_{Mon}, p_{Tue}, \ldots, p_{Sun}$ denote the true proportions of students born on each of the days of the week.

   $H_0: p_{Mon} = \frac{1}{7},\ p_{Tue} = \frac{1}{7}, \ldots,\ p_{Sun} = \frac{1}{7}.$

   $H_a: H_0$ is not true.

3. **Test Statistic**: $\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp}$ = ___________

4. ***P*-value**:
   This is calculated using the χ^2 distribution on your calculator. (If certain assumptions are met, the χ^2 distribution is the theoretical distribution of totals for **all** tables like the one you simulated.) The graph of the χ^2 distribution is given at right and the *P*-value is the shaded area. This shaded area is the probability—if the birth days follow a uniform model—of a random sample of 175 students producing a value of $\chi^2 \geq 13.68$.

   To obtain this probability, the command you would enter is [2nd] [DISTR] 8: χ^2 cdf(13.68, 1000, 6). The last value, 6, is the degrees of freedom, one less than the number of categories in the categorical variable. The calculator requires an upper bound in order to calculate the *P*-value, and 1000 is used in this case since the χ^2 distribution is extremely close to the *x*-axis by this point.

   ***P*-value = ___________**

5. **Conclusion:**
   Use the *P*-value as you have in previous tests: small *P*-values cause you to doubt the null hypothesis and reject H_0; larger *P*-values do not provide enough evidence to reject H_0. Give an explanation in the context of this situation.