Stephanie and Lola are convinced that the deans are picking on the seniors by giving extra detentions just out of spite and not just because they keep showing up for late passes with Starbucks cups in their hands. Ethan and Ethan, in an effort to calm them, collect samples of all the detentions given over a two week period. They come up with the results below

Class	Frosh	Soph	Junior	Senior	Total
Tickets Written	103	114	106	125	448

Assumptions: Random Samples? Yes

Expected Counts ≥ 5? Yes

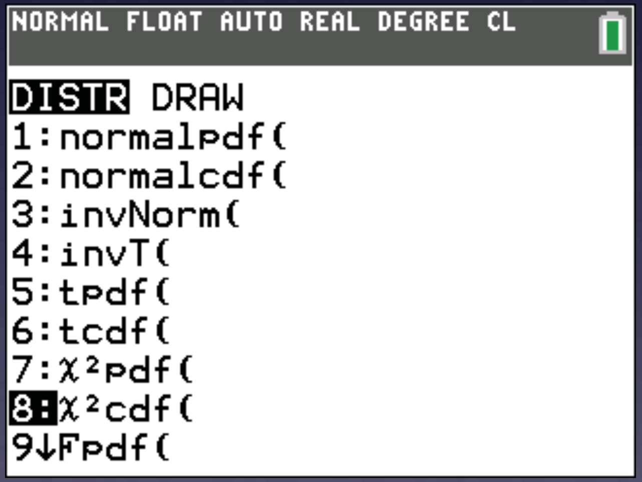
How can we use this data to test Stephanie and Lola's claim?

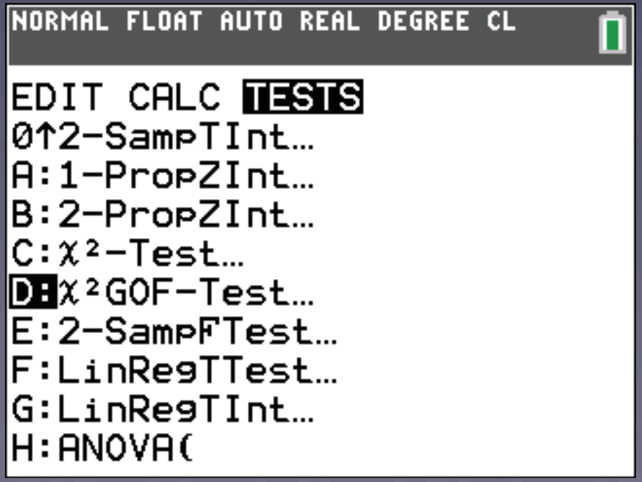
For this we use what is called the Chi Square Distribution symbolized by

$$\chi^2$$

It's available on your calculator as well in both the distribution and test windows

TI-83 users will notice that they only have one test on their menus. I'll show you how to get around that shortly





Observed	Expected
103	448/4 = 112
114	112
106	112
125	112

Chi-Squared Goodness of Fit Hypothesis Test

Note #1: We are now looking at CATEGORICAL DATA

$$H_0: \begin{array}{l} p_1 = \\ p_2 = \\ \vdots \\ p_k = \end{array}$$

$$H_a: H_0 \text{ is not true}$$

Note #2: If all the proportions are the same, use -

$$H_0: p_1 = p_2 = \ldots = p_k = \#$$

Note #3:
$df = k - 1$

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \#$$

As you will see in the examples and checkpoint questions, the Chi-Squared calculation shown above will use categorical data despite our applying the same rules as we do for proportions

Steps in Hypothesis Testing

1. Define the population characteristic (i.e. parameter) about which hypotheses are to be tested.

2. State the null hypothesis H_0.

3. State the alternative hypothesis H_a.

4. State the significance level for the test α.

5. Check all assumptions and state name of test.

6. State the name of the test.

7. State *df* (which applies here even though we are dealing with proportions).

8. Display the test statistic to be used without any computation at this point. $\chi^2 = \sum \frac{(O-E)^2}{E} = \#$

9. Compute the value of the test statistic, showing specific numbers used.

10. Calculate the P – value.

11. Sketch a picture of the situation.

12. State the conclusion in two sentences -
    1. Summarize in theory discussing H_0.
    2. Summarize in context discussing H_a.

Chi-Squared Goodness of Fit Hypothesis Test

Steps in Chi-Squared GOF Hypothesis Testing

1. $p_1 =$ true proportion of ...
$p_2 =$ true proportion of ...
$\vdots$
$p_k =$ true proportion of ...

2. H_0: $p_1 = \#$
$p_2 = \#$
$\vdots$
$p_k = \#$

3. $H_a : H_0$ is not true

4. State α.

5. Assumptions:
   1. Random Samples
   2. Expected Counts ≥ 5

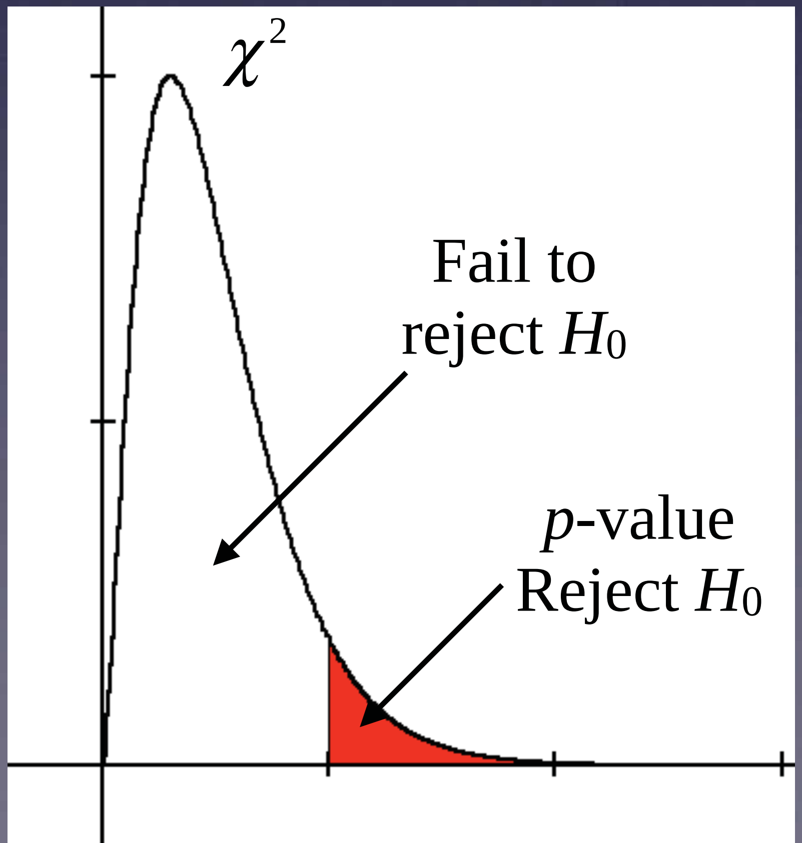
6. χ^2 GOF Test

7. $df = k - 1$

8/9. $$\chi^2 = \sum \frac{(O-E)^2}{E} = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} = \#$$

10. $P-value = P\left(\chi^2 > \#\right) = \chi^2 cdf\left(\#, 1E99, df\right)$

11.



12. State the conclusion in two sentences -
    1. Summarize in theory discussing H_0.
    2. Summarize in context discussing H_a.

Stephanie and Lola are convinced that the deans are picking on the seniors by giving extra detentions just out of spite and not just because they keep showing up for late passes with Starbucks cups in their hands. Ethan and Ethan, in an effort to calm them, collect samples of all the detentions given over a two week period. They come up with the results below

Class	**Frosh**	**Soph**	**Junior**	**Senior**	**Total**
Tickets Written	103	114	106	125	448

Assumptions: Random Samples? Yes

Expected Counts $\geq$ 5? Yes

A chi-square analysis was performed to test the claim that there is a relationship between the week of the month and the number of tickets written. What is the *P*-value of the test?

$$\chi^2 = \frac{(103-112)^2}{112} + \frac{(144-112)^2}{112} + \frac{(106-112)^2}{112} + \frac{(125-112)^2}{112}$$

$$= 2.589$$

$$p\text{-value} = P(\chi^2 > 2.589) = \chi^2\text{cdf}(2.589, 1E99,3)$$

$$= 0.4594$$

Observed	Expected
103	448/4 = 112
114	112
106	112
125	112

A reporter believed that police officers were required to write a specific quota of traffic tickets during a month. In order to meet the alleged quota, he believed officers would need to write more tickets during the last week of the month. To investigate the claim, the reporter collected the number of tickets written by the local police force in a month and organized them by weeks as shown in the table below.

Week	First Week	Second Week	Third Week	Fourth Week	Total
Tickets Written	133	124	154	145	556

Assumptions: Random Samples? Yes

Expected Counts $\geq$ 5? Yes

A chi-square analysis was performed to test the claim that there is a relationship between the week of the month and the number of tickets written. What is the *P*-value of the test?

$$\chi^2 = \frac{(133-139)^2}{139} + \frac{(124-139)^2}{139} + \frac{(154-139)^2}{139} + \frac{(145-139)^2}{139}$$

$$= 3.7554$$

$$p\text{-value} = P(\chi^2 > 3.7554) = \chi^2\text{cdf}(3.7554, 1E99, 3)$$

$$= 0.2891$$

Observed	Expected
133	556/4 = 139
124	139
154	139
145	139

For 1000 shoppers donating blood at a mall, the frequencies of blood types were as shown in the table shown below. Consider this an SRS of all mall shoppers.

Blood Type	Frequency
O	465
A	294
B	196
AB	45
Total	1000

In the general population, the blood type distribution is as follows:

Type O = 45%, Type A = 40%,
Type B = 10%, Type AB = 5%

Do these data provide evidence that the blood type proportions of mall shoppers differ from the blood type proportions of the general public? Test the appropriate hypotheses using $\alpha = 0.01$

We will test the following hypotheses:

H_0: Mall shoppers have the same blood type proportions as the general public

H_a: Mall shoppers DO NOT have the same blood type proportions as the general public

Data was given for $n = 1000$ mall shoppers. If H_0 is true, what are the expected counts?

	Blood Type			
	O	**A**	**B**	**AB**
Observed Count	465	294	196	45
Expected Count	0.45(1000) = 450	0.4(1000) = 400	0.1(1000) = 100	0.05(1000) = 50

Note: A different look for H_0 could be

$$H_0 : p_O = 0.45, p_A = 0.40, p_B = 0.10, p_{AB} = 0.05$$

For 1000 shoppers donating blood at a mall, the frequencies of blood types were as shown in the table shown below. Consider this an SRS of all mall shoppers.

Blood Type	Frequency
O	465
A	294
B	196
AB	45
Total	1000

In the general population, the blood type distribution is as follows:

Type O = 45%, Type A = 40%,
Type B = 10%, Type AB = 5%

Do these data provide evidence that the blood type proportions of mall shoppers differ from the blood type proportions of the general public? Test the appropriate hypotheses using $\alpha = 0.01$

We will test the following hypotheses:

H_0: Mall shoppers have the same blood type proportions as the general public

H_a: Mall shoppers DO NOT have the same blood type proportions as the general public

Let's calculate our χ^2 test statistic.

$$\chi^2 = \frac{(465-450)^2}{450} + \frac{(294-400)^2}{400} + \frac{(196-100)^2}{100} + \frac{(45-50)^2}{50} = 121.25$$

$P(\chi^2 > 121.25) = \chi^2 cdf(121.25, 1E99, 3) = 0$

18

1. A highway superintendent states that five bridges into a city are used in the ratio 2:3:3:4:6 during the morning rush hour. A highway study of a simple random sample of 6000 cars indicates that 720, 970, 1013, 1380, and 1917 cars use the five bridges, respectively. Can the superintendent's claim be rejected at the 2.5% or 5% level of significance?

Assumptions: Random Samples? Yes

(a)There is sufficient evidence to reject the claim at either of these two levels.
(b)There is sufficient evidence to reject the claim at the 2.5% but not at the 5% level.
(c)There is sufficient evidence to reject the claim at the 5% but not at the 2.5% level.
(d)There is not sufficient evidence to reject the claim at either of these two levels.
(e)There is not sufficient information to answer this question.

Expected Counts $\geq$ 5? Yes

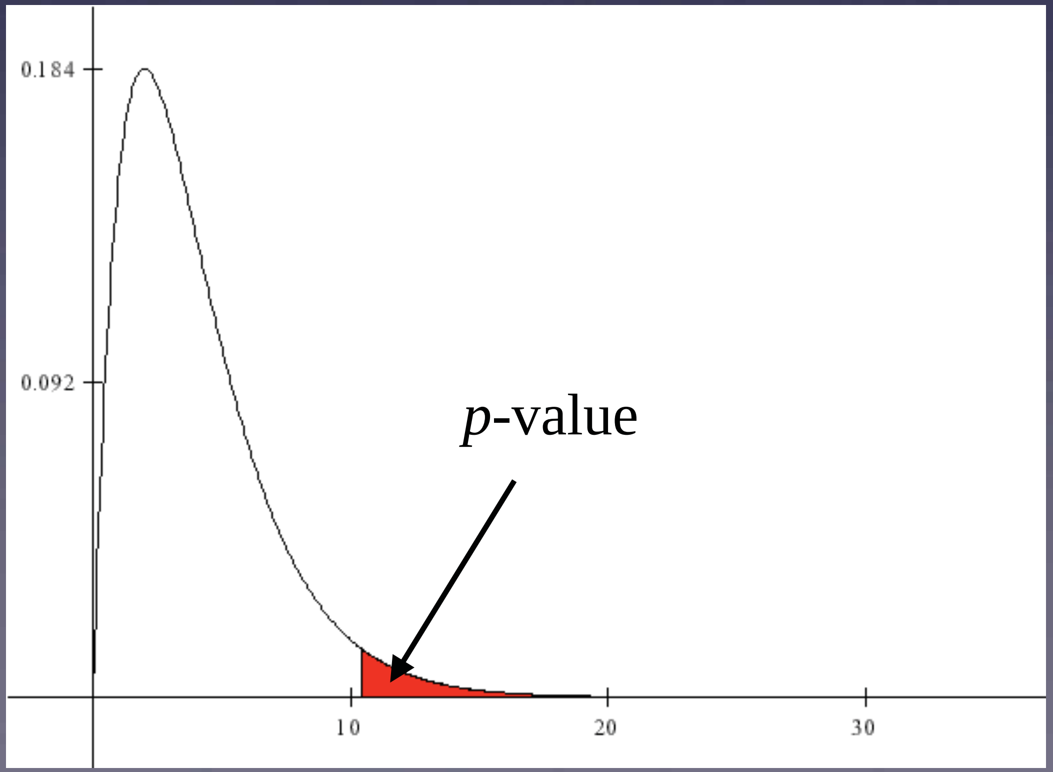
Observed	Expected
720	$6000\frac{2}{18} = 666.\overline{6}$
970	$6000\frac{3}{18} = 1000$
1013	$6000\frac{3}{18} = 1000$
1380	$6000\frac{4}{18} = 1333.\overline{3}$
1917	$6000\frac{6}{18} = 2000$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(720-666.\overline{6})^2}{666.\overline{6}} + \frac{(970-1000)^2}{1000} + \frac{(1013-1000)^2}{1000} + \ldots = 10.414$$

p-value = $P(\chi^2 > 10.414)$

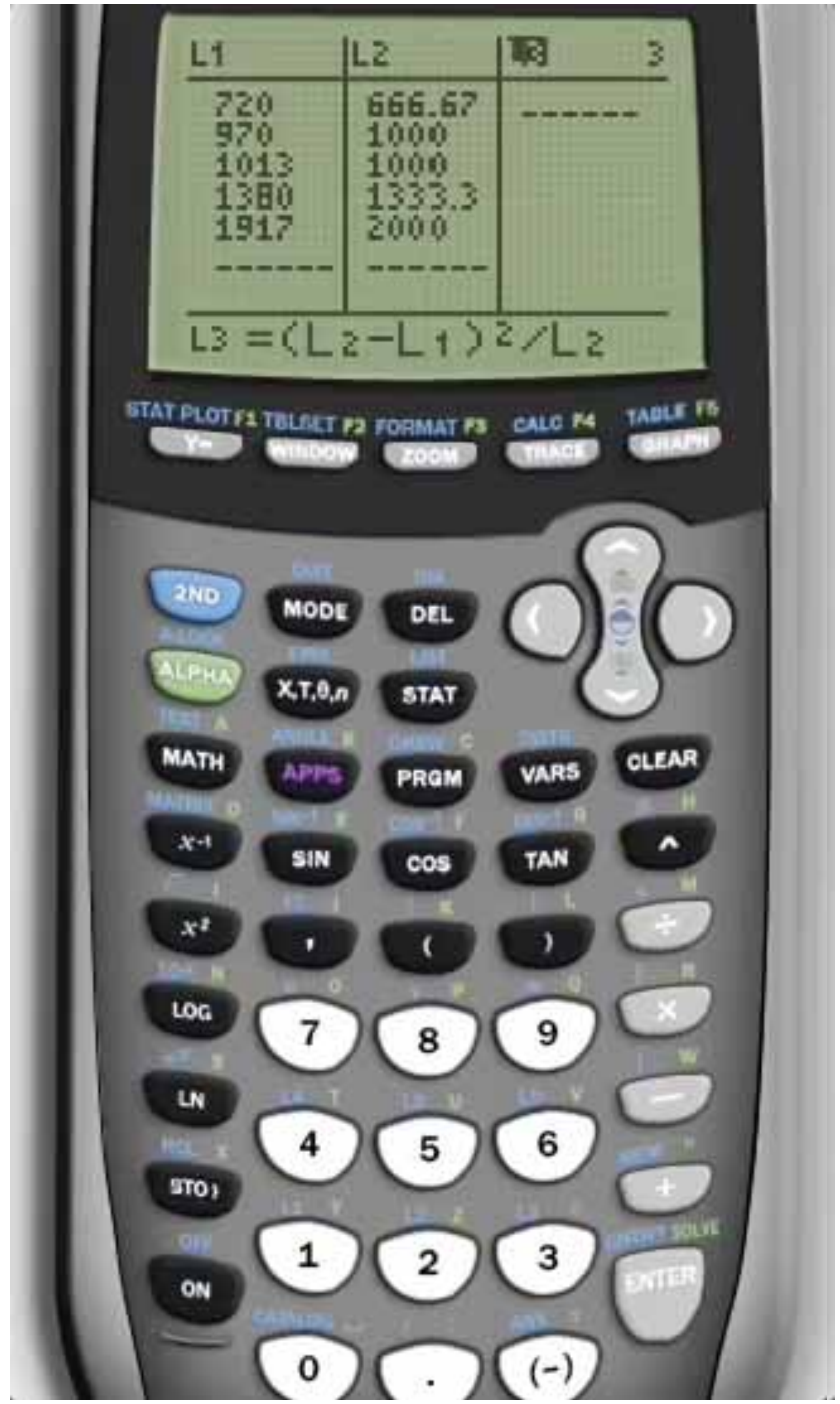
$= \chi^2\text{cdf}(10.414, 1E99, 4)$

How can we do this on the calculator?



$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = \sum \frac{(L_1 - L_2)^2}{L_2}$$



$$\chi^2 = 10.414$$

1. A highway superintendent states that five bridges into a city are used in the ratio 2:3:3:4:6 during the morning rush hour. A highway study of a simple random sample of 6000 cars indicates that 720, 970, 1013, 1380, and 1917 cars use the five bridges, respectively. Can the superintendent's claim be rejected at the 2.5% or 5% level of significance?

(a)There is sufficient evidence to reject the claim at either of these two levels.
(b)There is sufficient evidence to reject the claim at the 2.5% but not at the 5% level.
(c)There is sufficient evidence to reject the claim at the 5% but not at the 2.5% level.
(d)There is not sufficient evidence to reject the claim at either of these two levels.
(e)There is not sufficient information to answer this question.

Observed	Expected
720	$6000\frac{2}{18}=666.\overline{6}$
970	$6000\frac{3}{18}=1000$
1013	$6000\frac{3}{18}=1000$
1380	$6000\frac{4}{18}=1333.\overline{3}$
1917	$6000\frac{6}{18}=2000$

$$\chi^2=\sum\frac{(O-E)^2}{E}=\frac{\left(720-666.\overline{6}\right)^2}{666.\overline{6}}+\frac{\left(970-1000\right)^2}{1000}+\frac{\left(1013-1000\right)^2}{1000}+\ldots$$

$$=10.414$$

$$p\text{-value} = P(\chi^2>10.414)=\chi^2\text{cdf}(10.414,\ 1\text{E}99,4)=0.034 \quad \begin{matrix}> 0.025\\ < 0.05\end{matrix}$$

Answer: C