AP STATISTICS
Chi-Squared Tests

PODUNK MATH DEPARTMENT

1. **THE M&Ms PROBLEM I** M&M/Mars claims that the distribution of proportions of colors for their "plain" milk-chocolate M&M's candies follows the information in the table below. A class of AP Statistics students sifted through a large bag of plain M&M's (which we will assume qualifies as a random sample of plain M&M's) and recorded the colors that they observed, to see if the claimed proportions are accurate. The students expect that the observed proportions will be consistent with the company's claimed proportions.

Color	Blue	Orange	Green	Yellow	Red	Brown	Total
"True" Proportions for plain M&Ms (according to M&M / Mars company)	0.24	0.20	0.16	0.14	0.13	0.13	**1.00**
Number of observed M&M's	124	85	99	48	59	74	**489**

Expected counts: 117.36 (489(0.24)) 97.8 (489(0.2)) 78.24 ...etc 68.46 63.57 63.57 (489(0.13))

Based on the data collected by these students, is there evidence at the 10% level of significance to state that the distribution of proportions of colors of plain M&M's is not consistent with the company's claims? Test an appropriate hypothesis and give statistical evidence to support your conclusion. *(hint: in order to do this problem, you should start by calculating the expected number of M&M's in each color category)*

χ^2 goodness of fit test

☆ It is NOT okay to say "is CONSISTENT with"...

H_0: the distribution of proportions of colors is the same as the company's claims.

H_A: NOT the same.

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp}$$ ← on the bottom-right of the formula chart!

$$= \frac{(124 - 117.36)^2}{117.36} + \cdots + \frac{(74 - 63.57)^2}{63.57}$$

CONDITIONS
- The students took a random sample of M&M's
- All expected counts are > 5.

$\chi^2 = 15.7138$ $df = 5$ (6 − 1)

p-value = 0.00771
α = 0.10

since $p < \alpha$, we reject H_0.

We have evidence that the proportions of colors does NOT match the company's claims.

☆ CALCULATOR INSTRUCTIONS ☆ ON PAGE 611!

2. **THE M&Ms PROBLEM II** The M&Ms company claims that the manufactured proportions of colors for Peanut M&Ms differs slightly from Plain M&Ms. What does our data say?

A group of statistics students from another school across town counted the colors in a large bag of peanut M&M's, and those numbers are added to our data in the table below. We will assume that each of the separate samples is a "representative" of its type of M&M.

	Brown	Yellow	Red	Blue	Orange	Green	TOTAL
Plain	124	85	99	48	59	74	**489**
Peanut	16	36	20	55	47	28	**202**
TOTAL	**140**	**121**	**119**	**103**	**106**	**102**	**691**

Is there evidence of a difference in the distributions of proportions of colors between plain and peanut M&Ms? Test an appropriate set of hypotheses at the 10% significance level.

Since we are comparing the relative breakdown of one variable (color) across different "populations" (plain vs. peanut), this is a...

χ^2 test for homogeneity, $df = (6-1)(2-1) = 5$ (# of columns, # of rows)

H_0: The distributions of proportions of colors are the same for plain & peanut M&Ms.

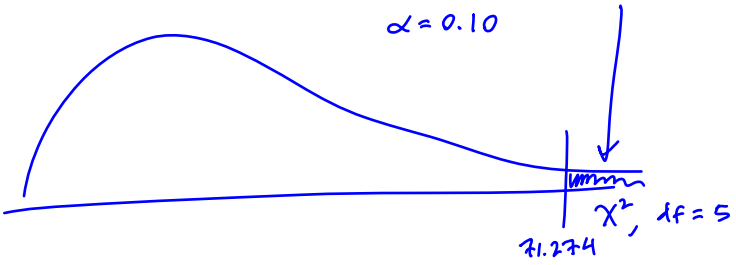
H_A: The distributions of proportions of colors are different for plain vs. peanut M&Ms.

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp}$$

$$= \frac{(124 - 99.074)^2}{99.074} + \ldots + \frac{(28 - 29.82)^2}{29.82}$$

[from calculator... refer to page 623 in textbook]

$\chi^2 = 71.275$ p-value $= 5.562 \times 10^{-14}$



Since p-value $< \alpha$, we reject Ho.

We have evidence that the relative distributions of colors are different for the two types of M&M's (plain vs. peanut).

CONDITIONS:

- We have separate representative samples from each population of M&M's.
- Expected counts are all ≥ 5:

	Br	Y	R	Bl	O	G
plain	99.07	85.63	84.21	72.89	75.01	72.18
peanut	40.93	35.37	34.79	30.11	30.99	29.82

* Calculating expected counts:

(Row total) x (column total) / TOTAL

Example:

Plain/yellow: $\frac{(489)(121)}{691} = 85.628$

3. **THE EXTRACURRICULAR PROBLEM** An administrator at a large university is interested in determining whether the residential status of a student is associated with level of participation in extracurricular activities. Residential status is categorized as on campus for students living in university housing and off campus otherwise. A simple random sample of 100 students in the university was taken, and each student was asked the following two questions.

- Are you an on campus student or an off campus student?
- In how many extracurricular activities do you participate?

The responses of the 100 students are summarized in the frequency table shown.

	Residential Status		
Level of Participation in Extracurricular Activities	On campus	Off campus	Total
No activities	9 (12.87)	30 (26.13)	39
One activity	17 (13.86)	25 (28.14)	42
Two or more activities	7 (6.27)	12 (12.73)	19
Total	33	67	100

(Expected counts in parentheses)

Do the survey results provide statistical evidence of an association between residential status and level of participation in extracurricular activities among the students at the university?

χ^2 test of independence, $df = 2$

H_0: There is NO association between residential status and level of participation in extracurricular activities among students at this university.

H_A: There IS an association.....

Conditions:
- We have a SRS of students at this university
- Expected counts ≥ 5 ✓

[from calculator... refer to page 623 in textbook]

$\chi^2 = 2.925$

$p = 0.2316$

$\alpha = 0.05$

Since $p > \alpha$, we fail to reject H_0.

We lack evidence of an association between residential status and level of participation in extracurricular activities among students at this university.