Here, our null hypothesis will be that the population follows a certain distribution. The alternative hypothesis is that it does *not*.

Statistics
Class Notes
Goodness-of-Fit Test: Testing a Population's Distribution (Section 12.1)

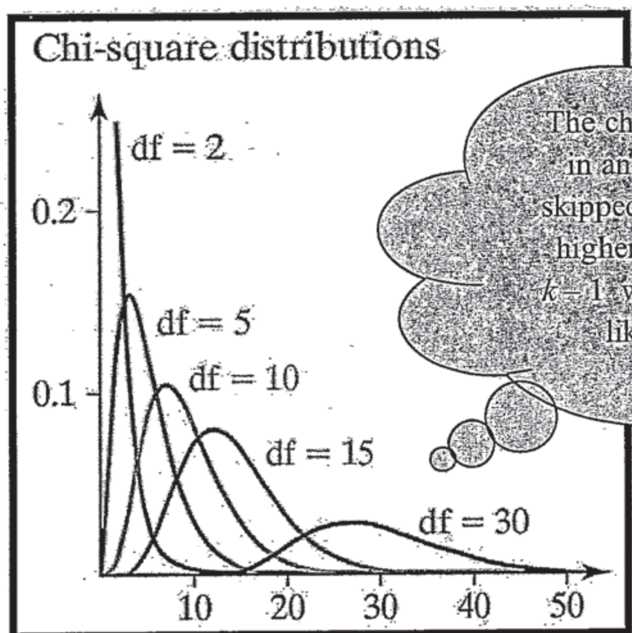
The M&Ms company wants to ensure that the distribution of colors in every bag roughly follows this distribution. They sample a random bag (counting each color) and use that data to perform a goodness-of-fit test to see if the distribution is what they want.
Let's see how that works.

Color	Percent
brown	13%
yellow	14%
red	13%
orange	20%
blue	24%
green	16%

Definition: Goodness-of-Fit Test: A **goodness-of-fit** test is an inferential procedure used to determine whether a frequency distribution follows a specific distribution. We use the **Chi-Square (χ^2) Distribution.**

Characteristics of the Chi-Square Distribution:

1. It is *not* symmetric.
2. Its shape depends on the degrees of freedom, just like Student's t-distribution.
3. As the number of degrees of freedom increases, it becomes more symmetric, as illustrated in the figure below.
4. The values of χ^2 are non-negative (greater than or equal to 0).



The chi-square distribution was introduced in an earlier section of the book that we skipped. Notice how its shape changes with higher degrees of freedom (which will be $k-1$, where k is the number of categories, like colors in the above example).

Again, we find critical values or use P-values. Table VIII will give us those critical values.

In practice, what we will do is compare the actual observations (from the sample) to what we would *expect* them to be, given that the null hypothesis is true. If a significant difference exists between the observed and expected counts, we have evidence against the null hypothesis.
So, how do we do that?

Expected Counts:
This is related to the expected values for binomial distributions and so will look familiar.

Suppose that there are n independent trials of an experiment (or individuals in a survey) with $k \geq 3$ mutually exclusive possible outcomes (categories). Let p_1 represent the (assumed) probability of observing the first outcome, p_2 represent the (assumed) probability of observing the second outcome, and so on. Then the **expected counts** for each possible outcome are given by $E_i = \mu_i = np_i$ for $i = 1, 2, \ldots, k$.

> These p_i probabilities are those *assumed* in the null hypothesis.

Goodness-of-Fit Test Statistic:

Let O_i represent the *observed* count of category i (data from sample or experiment) and E_i represents the *expected* count of category i. The number of categories is denoted by k and the number of independent trials of the experiment is n. Then

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

> We use sigma notation here to indicate addition of these terms for $i = 1, 2, \ldots, k$.

approximately follows the
chi-square distribution with $k - 1$ degrees of freedom, provided that
1. all expected counts are greater than or equal to 1 (all $E_i \geq 1$), and
2. no more than 20% of the expected counts are less than 5.

Again, we calculate $E_i = \mu_i = np_i$ for $i = 1, 2, \ldots, k$.

> If these are *not* met, one option is to combine low-frequency categories into a single category.

> This is the only option for step 1.

The Goodness-of-Fit Test:

Step 1: Determine the null and alternative hypotheses.

H_0: The random variable follows a certain distribution.
H_1: The random variable does *not* follow this distribution.

Step 2: Select a level of significance, α, depending on the seriousness of making a Type I error.

Step 3a: Calculate the expected counts, $E_i = \mu_i = np_i$ for $i = 1, 2, \ldots, k$. Recall k is the number of categories, n is the number of trials, and p_i is the *assumed* probability of the i^{th} category from the null hypothesis.

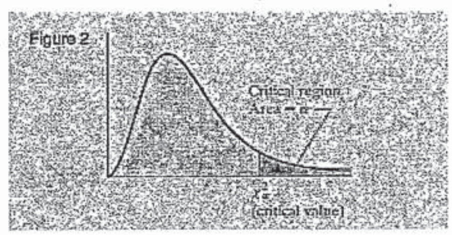
Step 3b: Verify that

1. all expected counts are greater than or equal to 1 (all $E_i \geq 1$), and
2. no more than 20% of the expected counts are less than 5.

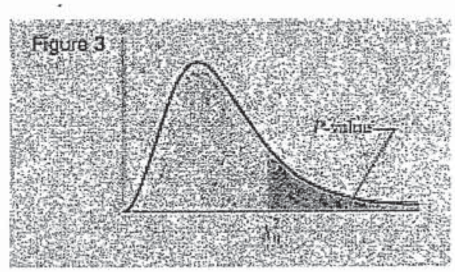
Step 3c:

Compute the test statistic $\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$ where O_i is the *observed* count for category i.

Step 4 (Classical Approach):
Determine the critical value using Table VIII. **All goodness-of-fit tests are right-tailed**, so the critical value is χ_α^2 with $k-1$ degrees of freedom (Figure 2). Compare the test statistic to the critical value. If $\chi_0^2 > \chi_\alpha^2$, reject the null hypothesis.



Step 4 (*P*-value Approach):
Use Table VIII or technology to approximate the *P*-value by determining the area under the chi-square distribution with $k-1$ degrees of freedom to the **right** of the test statistic (Figure 3). If the *P*-value $< \alpha$, reject the null hypothesis.



Step 5: State the conclusion.

> Instructions for technology are given on the next page.

Remember, if we do *not* reject the null hypothesis, we are *not* saying that it *must* be true. We simply state that we do *not* have evidence to conclude it is as stated in the alternative hypothesis.

Instructions for TI-84 Plus Calculators:

1. Enter the observed counts in **L1** and the expected counts in **L2** in the **STAT** editor.
2. Press STAT > TESTS > D: χ^2 GOF-Test.
3. Of course, tell it that **L1** and **L2**, respectively, are where the **Observed** and **Expected** counts are located. Calculate degrees of freedom as one less than the number of categories and enter it for **df**.
4. Highlight **Calculate** and press **ENTER**.
5. The calculator displays the test statistic, the *P*-value, and the degrees of freedom (df). The bottom line of output is labeled **CNTRB**. That shows the individual terms of the sum that makes up the test statistic. You can scroll through them with the right and left arrows.

TI-83 and older models of the TI-84 *cannot* do this test. However, a workaround can be found on the internet.

Instructions for StatCrunch:

1. Enter the observed counts in the first column and the expected counts in the second column. Name the columns.
2. Select **Stat** > **Goodness-of-fit** > **Chi-Square Test**.
3. Tell it the column that contains the **Observed** counts. Tell it the column that contains the **Expected** counts. Then hit **Compute!**.
4. Besides posting various bits for checking (like sample size, degrees of freedom, and the data), it will display the test statistic (**Chi-Square**) and the ***P*-value**.

expl 1: Is wearing a helmet on your motorcycle necessary? The National Highway Traffic Safety Administration publishes reports on motorcycle safety. The distribution shows the proportion of fatalities by location of injury for motorcycle accidents.

Location of Injury	**Multiple locations**	**Head**	**Neck**	**Thorax**	**Abdomen/ Lumbar/Spine**
Proportion	0.57	0.31	0.03	0.06	0.03

The following data show the location of injury and number of fatalities for 2,068 riders who were *not* wearing a helmet.

Location of Injury	**Multiple locations**	**Head**	**Neck**	**Thorax**	**Abdomen/ Lumbar/Spine**
Count (out of 2,068)	1036	864	38	83	47

Continued on next page…

expl 1 (continued):

Use a 0.05 level of significance to test if the distribution of fatal injuries for riders *not* wearing a helmet follows the distribution for all riders. Follow these steps.

a.) Use the proportions given in the first table to calculate the *expected* counts for each category for the 2,068 riders who died while *not* wearing a helmet. Record them below. Do *not* round.

Location of Injury	Multiple locations	Head	Neck	Thorax	Abdomen/ Lumbar/Spine
Observed Count	1036	864	38	83	47
Assumed Proportion	0.57	0.31	0.03	0.06	0.03
Expected Count (out of 2,068)	1178.76	641.08	62.04	124.08	62.04

b.) Write down the null and alternative hypotheses.

Ho: The distribution for helmet-less riders is same as for all accidents

H1: The dist. is not the same for helmeted vs helmetless riders.

c.) Perform your Chi-Square test at the 0.01 level. Fill out the information below and state your conclusion.

Test statistic: $\chi^2_0 \approx 121.37$

P-value: $p \approx 2.7 \times 10^{-25} \approx 0$

Conclusion: Pvalue $< \alpha = 0.01$ — so reject null hyp. We have sufficient evidence to say the distribution for helmet-less riders is not the same as for all riders.

d.) Our conclusion is not surprising, except possibly to those who do not believe in wearing helmets. Look through the CNTRB output (if you used the calculator) to find the category whose term contributed most to the test statistic. Is it surprising? (You can also compare observed and expected counts directly.)

~~multiple locations 17.3~~
head 77.5 ← It's not surprising people w/out helmets hurt their head lots.
neck 9.3
thorax 13.6
abdomen/spine 3.6

Ho: All sides would have prob. of 1/6

H_1: Ho is not true.

expl 2: I have made a six-sided die out of cardboard and must test it to see if it is fair. How would we do this?

Roll it, say, 100 times. Come up with sample data. For instance, it may look like

(die roll) output	1	2	3	4	5	6
trials out of 100	15	15	20	17	23	10

Compare these numbers to the expected counts. For each die number on a fair die, you'd expect $\frac{1}{6} * 100 \approx 16.667$ rolls.

If we put those #s in calculator for L_1, L_2, we get $\chi_0^2 \approx 6.08$ and a pvalue of 0.29. So we'd not reject the null hypothesis at the $\alpha = 0.01, 0.05,$ or 0.10 levels.

Conclusion: We do not have sufficient evidence to say the die is not fair.

Table VIII

Chi-Square (χ^2) Distribution

Area to the Right of Critical Value

Degrees of Freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169