1. On the M&M web page, it claims that they produce 13% brown, 14% yellow, 13% red, 24% blue, 20% orange, and 16% green milk chocolate M&M's. Suppose we buy a bag of milk chocolate M&M's and come up with the following numbers of each color:

	Color	Number	Expected		$(obs-exp)^2/exp$
13%	brown	50	13% of 381	= 49.5	.005
14%	yellow	47	14% "	53.3	.745
13%	red	41		49.5	1.460
24%	blue	94	etc	91.4	.074
20%	orange	102		76.2	8.735
16%	green	47		61.0	3.213
100%		381		380.9 ≈ 381	14.232

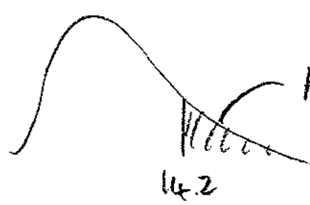
Test the hypothesis that our bag of M&M's is like a simple random sample of M&M's from a population with the specified percentages of each color. You must state a null and an alternative hypothesis, find a test statistic and a P-value and clearly state your conclusions in terms of what you have learned about the color of milk chocolate M&M's.

null: this is like a simple random sample from a population with the stated percentages

alt: this is not like a simple random sample from a population with the stated percentages

$\chi^2 = 14.2$ from above

$df = 6 - 1 = 5$



P-value is between 1% and 5%

We reject the null & conclude this bag is not like a random sample from a population with the stated percentages.

2. In a large city, there are 5 electoral precincts. There are two mayoral candidates, A and B. A political science student takes a simple random sample of 1870 voters from this city and asks them which precinct they live in and whether they voted for candidate A or B. She makes the following table:

	Candidate		
Precinct	A	B	Total
1	400 337	229 292	629
2	184 181	154 157	338
3	101 161	200 140	301
4	182 197	186 171	368
5	136 127	98 107	234
Total	1003	867	1870

Note: you will get 125 if you take 53.6% of 234. The difference is due to roundoff. Either method gives a huge χ^2 but if it's near the cutoff, calcs should be redone with more precision.

We want to test the hypothesis that voting for mayoral candidates A and B is independent of precinct for voters from this city.

(a) Clearly state the null and the alternative hypotheses.

null: voting behavior is independent of precinct

alt: " " " dependent on precinct

(b) Compute a test statistic.

$\frac{1003}{1870} \times 100\% = 53.6\%$

$\frac{53.6}{100} \times 629 = 337$

$\frac{53.6}{100} \times 338 = 181$

$\frac{53.6}{100} \times 301 = 161$

$\frac{53.6}{100} \times 368 = 197$

The others are obtained by subtraction

obs	exp	$(\text{obs}-\text{exp})^2/\text{exp}$
400	337	11.8
184	181	.0
101	161	22.4
182	197	1.1
136	127	.6
229	292	13.6
154	157	.1
200	140	25.7
186	171	1.3
98	107	.8

$\chi^2 = 77.4$

(c) Find the degrees of freedom.

$df = (5-1)(2-1) = 4$

(d) What can you say about the P-value?

The P-value is less than 1%

(e) Do you reject the null hypothesis? Explain why or why not.

We reject the null because the P-value is small

(f) Clearly state your conclusions.

Voting behavior depends on precinct for this population.

3. In a study on snoring and nightmares, researchers found the following results:

	Frequency of nightmares				
	Never	Seldom	Occasionally	Frequently	Total
Nonsnorers	22 21.6	45 43.2	35 35.2	11 13	113
Snorers	16 16.4	31 32.8	27 26.8	12 10	86
Total	38	76	62	23	199

Assuming this is a random sample of people, is there evidence that the frequency of nightmares (never, seldom, occasionally, frequently) is different for snorers and nonsnorers? You must clearly state a null and alternative hypothesis, compute a test statistic and a P-value and state your conclusions.

null: frequency of nightmares & snoring are independent

alt: frequency of nightmares & snoring are dependent

$\frac{113}{199} \times 100\% = 56.8\%$

$\frac{56.8}{100} \times 38 = 21.6$

$\frac{56.8}{100} \times 76 = 43.2$

$\frac{56.8}{100} \times 62 = 35.2$

obs	exp	$(obs-exp)^2/exp$
22	21.6	.007
45	43.2	.075
35	35.2	.001
11	13.0	.308
16	16.4	.010
31	32.8	.099
27	26.8	.001
12	10.0	.400

$\chi^2 = .9$

$df = (2-1)(4-1) = 3$

The P-value is between 70% and 90% so we fail to reject the null & conclude that there is no evidence that frequency of nightmares & snoring are dependent.

4. The average age of all 43 presidents when they entered office is 55.3 years, and the SD is 6.2 years. Explain why it would be inappropriate to use these numbers to conduct a significance test on the hypothesis that the average age of entering presidents is 50 years.

We cannot perform a test because we have no sample - we have the whole population. We can say that the average age is not 50, it's 55.3.

5. Three psychiatrists set out to identify observed characteristics that distinguish schizophrenic patients from nonschizophrenic ones. They considered 77 different characteristics, and did 77 two-sample z-tests. Two of the tests turned up statistically significant at the 5% level (i.e. the P-value was less than 5%). Explain why this is not convincing evidence that these two characteristics are useful for identifying schizophrenic patients. What should the investigators do next to help decide whether or not these characteristics should be used for this purpose?

We would expect 5% to be significant just due to chance error, even if nothing at all were truly important. 5% of 77 is $3.85 \approx 4$. We only got 2 being significant, and our results could simply be due to chance error. If their science really tells them something should be happening here, they need to <u>get more data</u>.