**Chi-square goodness of fit, chi-square tests for independence,
one-sample proportion problems, and 2-sample proportion problems**

Goodness of fit problems can look a lot like chi-square tests for independence in some cases.

Ex 1: Test the claim that M & M colors occur in equal frequency.
H_0: $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}$.

Ex 2: Test the claim that at 6 different colleges, the same proportion of students are dog-owners.
H_0: $p_1 = p_2 = p_3 = p_4 = p_5 = p_6$.

Null hypotheses look almost the same, huh? Example 1 is definitely a goodness of fit. But example 2 is not. Let's look closer at the p's.

In Example 1, each p stands for the proportion of all M & M's that are a specific color. These proportions must add up to 100%, so if they're equal, we know they all equal 1/6.

In Example 2, each p stands for the proportion of students at a particular school who are dog-owners. These 6 proportions are not from the same population. They are from 6 different groups, so there is no reason why their sum should be 100%. In fact, these 6 percents could all be equal to any number, say 57%. But we're not testing whether they are all equal to 57%, only are they all equal to each other. So there is no known number in H_0.

Look closer at example 2. We could display our sample data in a 2-way table, like the example on age and education, or the example on education and smoking

	School 1	School 2	School 3	School 4	School 5	Sch. 6
Dog-owners						
Not dog-owners						

Now it looks like a chi-square test for independence. We could equivalently write H_0 as
H_0: There is no relationship between dog-ownership and school, OR
H_0: Dog-ownership and school are independent OR
H_0: All of the schools have the same proportion of dog-owners (much like the original H_0 above)
Notice that in this 2-way table, there are only 2 categories for one of the variables, so we can talk about only the proportion of dog-owners. Knowing this proportion would automatically give us the proportion of non-dog-owners.

Conclusion: In a chi-square test for independence, if either variable has only 2 categories, we can write our null hypothesis with a bunch of proportions, and call this a "test for homogeneity of proportions."

Note also that if the number of categories gets small enough, we obtain problems previously encountered in previous chapters. Observe Example 1:

Ex 1: Test the claim that M & M colors occur in equal frequency.
H_0: $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}$.
H_a: Not

Suppose M&Ms come in only 2 colors. Now we have:
H_0: $p_1 = p_2 = \frac{1}{2}$.
H_a: Not

But H_0 is redundant, because if we know p_1, then we also know p_2. This goodness-of-fit problem becomes a 1-sample proportion hypothesis test:
H_0: $p = \frac{1}{2}$

Ex 2: Test the claim that at 6 different colleges, the same proportion of students are dog-owners.
H_0: $p_1 = p_2 = p_3 = p_4 = p_5 = p_6$.
H_a: Not

Suppose that there are only 2 schools being compared. Now we have:
H_0: $p_1 = p_2$
H_a: Not. (In other words, p_1 is not equal to p_2.)

And this is exactly a 2-sample proportion hypothesis test with a two-tailed alternative hypothesis.

Conclusions:

If a goodness of fit problem has only 2 categories, it becomes a Chapter 9 problem.

If a chi-square test for independence has 2 categories in both variables
(dog-ownership = yes or no, school = school #1 or school #2), then it becomes a Chapter 11 problem.