19.0 Goodness-of-Fit Tests

- Answer Questions
- Review Significance Tests
- Goodness-of-Fit Tests

19.1 Review of Significance Tests

Example 1: You want to show that, on average, men weigh at least 40 pounds more than women. You collect a random sample of 100 men and find $\bar{X}_M = 180$; the sample standard deviation is $sd_M = 20$. You also collect a random sample of 200 women and find $\bar{X}_W = 130$ and $sd_W = 30$.

What are the hypotheses, the test statistic, and the conclusion?

The hypotheses are:

$$\mathrm{H}_0\colon\ \mu_M - \mu_W \leq 40 \qquad \text{vs.} \qquad \mathrm{H}_A : \mu_M - \mu_W > 40$$

or, in words, H_0: The average weight of men is at least 40 pounds more than the average weight of women.

H_A : The average weight of men is less than 40 pounds more than the average weight of women.

The test statistic is:

$$\begin{aligned} ts &= \frac{\bar{X}_M - \bar{X}_W - 40}{\sqrt{\frac{sd_M^2}{n_M} + \frac{sd_W^2}{n_W}}} \\ &= \frac{180 - 130 - 40}{\sqrt{\frac{20^2}{100} + \frac{30^2}{200}}} \\ &= 3.42997 \end{aligned}$$

The test statistic is 3.43. We want the significance probability. From the table, this is:

$$P - value = P[z > 3.43] = .00028.$$

This means that the chance of observing a difference as large as the one seen in this data set is very unlikely if the null hypothesis is true.

We strongly reject (at the .05 level, the .01, level, the .001 level) the null hypothesis. There is lots of evidence that the mean weight of men is <u>more</u> than 40 pounds larger than the mean weight of women.

Example 2: You want to show that more men like “Kill Bill Vol. 2” thn women. You sample 100 random men—50 liked it. Of 200 random women, 110 liked it.

What are your hypotheses? In symbols, it is:

$$\mathrm{H}_0\colon\ p_M - p_W \leq 0 \qquad \text{vs.} \qquad \mathrm{H}_A : p_M - p_W > 0$$

or, in words, H_0: The proportion of men who like KB2 is less than or equal to the proportion of women who like it.

H_A: The proportion of men who like KB2 is greater than the proportion of women who like it.

Do we really need to do a hypothesis test here? Isn’t the answer obvious?

The test statistic is:

$$\begin{aligned} ts &= \frac{\hat{p}_M - \hat{p}_W - 0}{\sqrt{\frac{\hat{p}_M(1-\hat{p}_M)}{n_M} + \frac{\hat{p}_W(1-\hat{p}_W)}{n_W}}} \\ &= \frac{.5 - .55 - 0}{\sqrt{\frac{.5*(1-.5)}{100} + \frac{.55*(1-.55)}{200}}} \\ &= -.8179 \end{aligned}$$

The significance probability comes from a z-table:

$$P - value = P[z \geq -.8179] = .78815.$$

When the null hypothesis is true, about 79% of the time you would expect this kind of result. There is no evidence to support the alternative.

19.2 Goodness-of-Fit Tests

Recall Darwin's experiment with peonies. He crossed red and white and got all pink. He crossed pinks with pinks and got some red, some white, and some pink.

If he had read Mendel's paper, he would have predicted what?

Mendel and Darwin needed a way to assess the statistical significance of such predictions. Are the observed numbers too far from the numbers predicted by Mendelian genetics, thus making it very unlikely, or are the numbers close enough that there is no reason to doubt Mendel's model for inheritance?

For this example, we want to know whether the counts of red, white and pink peonies agree closely with the 1/4: 1/4: 1/2 ratios that are predicted.

In this type of test, the null and alternative do not follow the usual pattern. They are:

H_0: The model holds   vs. H_A: The model fails.

In particular applications one can be more specific; e.g.:

H_0: The ratios of red, white and pink are 1/4: 1/4/: 1/2

H_A: The ratios differ from 1/4: 1/4/: 1/2.

Note that we can only reject the model. We cannot prove it, since we never "prove" the null hypothesis. All we do is fail to reject it.

The test statistic is also different. It is:

$$ts = \sum \frac{(O_i - E_i)^2}{E_i}$$

where the sum is taken over all categories (i.e., red, white, and pink). The O_i is the observed count in category i, and E_i is the count predicted in that category by the model.

To be concrete, suppose Darwin had made 100 crosses of pink with pink and had gotten 22 red, 29 white, and 49 pink. So $O_1 = 22$, $O_2 = 29$, and $O_3 = 49$.

The expected counts are those predicted by the model. Thus $E_1 = 25$, $E_2 = 25$, and $E_3 = 50$.

The test statistic is

$$\begin{aligned} ts &= \frac{(22-25)^2}{25} + \frac{(29-25)^2}{25} + \frac{(49-50)^2}{50} \\ &= 1.02. \end{aligned}$$

The significance probability comes from a chi-squared table. Let W be a chi-squared random variable with

$$k = \#\textbf{categories} - 1$$

degrees of freedom. In this example, $k = 3 - 1 = 2$.

The significance probability is:

$$P - value = \mathbf{P}[W \geq ts] = \mathbf{P}[W \geq 1.02].$$

From the table, this is between .7 and .5. So the null is not rejected. The data support Mendel.