

Chapter 8-Hypothesis Testing

8.1 Last night's hockey game:

- a) Null hypothesis: The game was actually an NHL hockey game.
- b) On the basis of that null hypothesis I expected that each team would earn somewhere between 0 and 6 points. I then looked at the actual points and concluded that they were way out of line with what I would expect if this were an NHL hockey game. I therefore rejected the null hypothesis. Notice that I haven't drawn a conclusion about what type of game it actually was, because that is not what I set out to test.

8.3 A Type I error would be concluding that I was shortchanged when in fact I was not.

8.5 The rejection region is the set of outcomes for which we would reject the null hypothesis. The critical value would be the minimum amount of change below which I would reject the null. It is the border of the rejection region.

8.7 There really isn't anything to show here—either it works or it doesn't.

8.9 Guessing the height of the chapel.

- a) The null hypothesis is that the average of two guesses is as accurate as one guess. The alternative hypothesis is that the average guess is more accurate than the single guess.
- b) A Type I error would be to reject the null hypothesis when the two kinds of guesses are equally accurate. A Type II error would be failing to reject the null hypothesis when the average guess is better than the single guess.
- c) I would be tempted to use a one-tailed test simply because it is hard to image that the average guess would be *less* accurate, on average, than the single guess.

8.11 A sampling distribution is just a special case of a general distribution in which the thing that we are plotting is a statistic which is the result of repeated sampling.

I have found that M&Ms, though they have precious little to do with psychology, are a good way for students to create their own sampling distribution. Each student can calculate the percentage of red M&Ms in a handful, and that distribution can be plotted. We then have a good idea what kinds of results we can expect in the future.

8.13 Magen et al (2008) study

- a) The null hypothesis is that the phrasing of the question will not effect the outcome—the means of the two groups are equal in the population. The alternative hypothesis is that the mean will depend on which condition the person is in.
- b) I would compare the two group means.

c) If the difference is significant I would conclude that the phrasing of the choice makes a real difference in the outcome.

8.15 Rerunning Exercise 8.14 for $\alpha = .01$:

We first have to find the cutoff for $\alpha = .01$ under a normal distribution. The critical value of $z = 2.33$ (one-tailed), which corresponds to a raw score of 42.69 (from a population with $\mu = 59$ and $\sigma = 7$).

We then find where 42.69 lies relative to the distribution under H_1 :

$$z = \frac{X - \mu}{\sigma} = \frac{42.69 - 50}{7} = -1.04$$

From the appendix we find that .85.08% of the scores fall above this cutoff. Therefore $\beta = .851$.

It should be clear that students should think about what happens to α and β as we change the questions we ask. I am giving them a lot of problems here so that they will become comfortable with solving simple equations.

8.17 To determine whether there is a true relationship between grades and course evaluations I would find a statistic that reflected the degree of relationship between two variables. (The students will see such a statistic (r) in the next chapter.) I would then calculate the sampling distribution of that statistic in a situation in which there is no relationship between two variables. Finally, I would calculate the statistic for a representative set of students and classes and compare my sample value with the sampling distribution of that statistic.

8.19 Allowances for fourth-grade students:

- a) The null hypothesis in this case would be the hypothesis that boys and girls receive the same allowance on average.
- b) I would use a two-tailed test because I want to reject the null whenever there is a difference in favor of one gender over the other.
- c) I would reject the null whenever the obtained difference between the average allowances were greater than I would be lead to expect if they were paid the same in the population.
- d) I would increase the sample size and get something other than a self-report of allowances.

This might be a good example to lead into our choice of a sample statistic to test. Tradition would have us test the mean, but perhaps a good case can be made for the median. Compiling a list of things that need to be controlled would give the students a sense of what goes into an experiment. It would be ideal if we could control the variance, but I don't see how to do that, since it is part of the experimental results.

8.21 Hypothesis testing and the judicial system

The judicial system operates in ways similar to our standard logic of hypothesis testing. However, in a court we are particularly concerned with the danger of convicting an innocent person. In a trial the null hypothesis is equivalent to the assumption that the accused person is innocent. We set a very small probability of a Type I error, which is far smaller than we normally do in an experiment. Presumably the jury tries to set that probability as close to 0 as they reasonably can. By setting the probability of a Type I error so low, they knowingly allow the probability of a Type II error (releasing a guilty person) to rise, because that is thought to be the lesser evil.