MATH 130: 4/23 WORKSHEET
STATISTICS: HYPOTHESIS TESTING

An important use of statistics is to test predictions. Imagine you are a medical scientist testing a new drug. You want to know whether it is effective. To do this, you run trials and collect data on how the drug affects, and also collect data on a *control group* given a placebo. Once you have all this data you want to know, does the evidence support rejecting the *null hypothesis* that the drug is no better than a placebo? This is an example of a *hypothesis test*.

Hypothesis testing example.

Suppose you want to know whether a coin is fair. If it were fair—the null hypothesis—then on average you expect about the same number of heads and tails. To test this, you flip the coin 100 times, and it comes up with 65 heads and 35 tails. It is possible that this happens with a fair coin, so you calculate that given a fair coin, the probability that the number of heads is off the expected mean by at least 15 is 0.4%. This outcome is thus very unlikely if the coin is actually fair. So you feel justified in rejecting the null hypothesis.

- Suppose you got 53 heads and 47 tails. The probability a fair coin is off by the expected mean of 50 heads by at least 3 is 62%. This is reasonably likely, so you wouldn't have enough evidence to reject the null hypothesis.

Hypothesis testing in general.

You want to know something about a *test statistic* T, such as the mean or standard deviation, of a real world data set. You have a null hypothesis that the value of T should be t_0. You do a data sample and calculate the value t of the statistic for that sample. (Note: t is almost certainly not equal to T!) You then calculate the likelihood that your sample t would be off by at least that much, under the null hypothesis $T = t_0$. If this probability is low, that is evidence to reject the null hypothesis.

Z-tests.

This test is used for a test statistic T which is normally distributed. Determine the mean value t_0 for T under the null hypothesis and estimate the standard deviation as s. Also determine the value t for your sample. Then calculate the z-score

$$z = \frac{t - t_0}{s}.$$

Given this z-score you can use computer tools to calculate the probability of being at least z standard deviations from the mean.[a] This probability is called a *p-value*. A common standard is, if the p-value is < 0.05 then that is evidence to reject the null hypothesis, whereas if the p-value is ≥ 0.05 then there is not enough evidence.

[a]On desmos: get a normal distribution with mean 0 and standard deviation 1. Click the "Cumulative Probability" dropdown and select "Outer" for region. Then type the values $\pm z$ in the two boxes.

The central limit theorem.

If you take a large number of samples of a value and average them to get a sample mean $\bar{x}$, then this sample mean is a random variable. Depending on how the sampling goes the value will likely change a little. The *central limit theorem* tells us that if n is large enough (> 30 is usually enough) then this random variable is approximately following a normal distribution.

More precisely, if the true mean is μ and the true standard deviation is σ, then $(\bar{x} - \mu)$ is approximately normally distributed with mean 0 and standard deviation $\sigma/\sqrt{n}$. This value $\sigma/\sqrt{n}$ is called the *standard error*.

Z-test example.

You are an inspector at a factory which produces bags of frozen peas. The bags, labeled as 400 grams in weight, really have a mean weight of 401 grams with a standard deviation of 6 grams. To test one of the machines, you have it produce 100 bags of peas and calculate their mean weight is 399 grams. Is there sufficient evidence to reject the null hypothesis that the machine is functioning as usual?

- Since you computed a sample mean, it follows roughly a normal distribution. Thus you can do a Z-test.
- To compute the z-score, first compute the standard error
$$SE = \frac{\sigma}{\sqrt{n}} = \frac{6}{10} = 0.6.$$
- Then plug that into the formula for the z-score
$$z = \frac{\bar{x} - \mu}{SE} = \frac{399 - 401}{0.4} \approx -3.3.$$
- Finally, use desmos to calculate the probability of being ± 3.3 standard deviations from the mean:
$$0.001.$$
- This p-value is 0.1%, so we reject the null hypothesis.

Practice problems

(1) You are a marine biologist. You know from last year's data gathering that the eels in a local reef have a mean weight of 3.1 pounds with a standard deviation of 0.9 pounds. You want to know if that has changed, so you grab 40 eels and measure their weights. You calculate a sample mean of 3.5 pounds. Perform a Z-test to determine whether you should reject the null hypothesis that the mean weight is still 3.1 pounds. What p-value do you get?

(2) Read the following xkcd comic, about p-values: `https://xkcd.com/882/`. Should you believe that green jelly beans cause acne? Why or why not?

(3) Read the wikipedia page on p-hacking (`https://en.wikipedia.org/wiki/Data_dredging`), also called data dredging. Explain how the xkcd comic illustrates p-hacking.

(4) Read part of the wikipedia page for statistical hypothesis tests (`https://en.wikipedia.org/wiki/Statistical_hypothesis_test`), from History through to Education. Write a paragraph or two summarizing the controvesy over null hypothesis significance testing.