HYPOTHESIS TESTS

We start with a **null hypothesis** H_0, which we'll assume to be true until we have evidence to the contrary. Exactly what constitutes contrary evidence is determined by our choice of **alternative hypothesis** H_1. We **reject** H_0 in favor of H_1 if an appropriate test statistic falls in a **critical region** (or **rejection region**). If the test statistic does not fall in the critical region, then we **fail to reject** H_0.

When testing a hypothesis, there are two ways to be wrong:

- **Type I error**: reject H_0 when H_0 is actually true;
- **Type II error**: fail to reject H_0 when H_0 is actually false.

The probability of a type I error is α; the probability of a type II error is β. Choose the largest acceptable value for α since this will minimize β. The critical region is chosen so that the test statistic lands in the critical region with probability α when H_0 is true. It may also be useful to find the P**-value** (or **observed significance level**) of your data: this is the smallest value for α that leads you to reject H_0 with your data.

1. Test Statistics

For **tests about the mean** ($H_0 : \mu = \mu_0$) test statistics are:

- $z = \dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ (known variance σ^2, all sample sizes if the pop. is normal, otherwise just large samples)
- $z = \dfrac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ (large samples)
- $t = \dfrac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ (small samples from a normally distributed population, $n - 1$ degrees of freedom)

For **tests about a population proportion** ($H_0 : \theta = \theta_0$) we can use the sample proportion $\hat{\Theta} = X/n$ or the sample total $X = n\hat{\Theta}$ and the test statistics are:

- x (X is binomial with parameters n and θ_0);
- $z = \dfrac{\hat{\theta} - \theta_0}{\sqrt{\frac{1}{n}\theta_0(1-\theta_0)}} = \dfrac{x - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}$ (large samples, both $n\theta_0 \geq 10$ and $n(1-\theta_0) \geq 10$.

For **tests about the difference of two means** ($H_0 : \mu_1 - \mu_2 = \delta_0$) some test statistics are:

- $z = \dfrac{\overline{x}_1 - \overline{x}_2 - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ (known variances, all sample sizes if pops are normal, otherwise just large samples)
- $z = \dfrac{\overline{x}_1 - \overline{x}_2 - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ (large samples)
- $t = \dfrac{\overline{x}_1 - \overline{x}_2 - \delta_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ (normally distributed populations with the same variance, $n_1 + n_2 - 2$ d.f.).

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Problems in which hypotheses aren't explicitly stated should be completed in three steps:

(1) State your null and alternative hypotheses. (Does your null hypothesis make sense as a default assumption?)
(2) Use the data to test H_0 against H_1, testing either at a specified significance level or giving a P-value.
(3) State your conclusion clearly (e.g. reject H_0 at significance level 0.05 or fail to reject H_0 at significance level 0.05).

Date: October 30, 2019.

2. Problems

1. The EPA has determined that the Maximum Contaminant Level Goal (MCLG, "The level of a contaminant in drinking water below which there is no known or expected risk to health") for nitrates in drinking water is 10mg/L. Imagine you have been hired by the City of Spokane to monitor drinking water safety. Your plan is to collect a random sample of water from 25 different sources in and test the water for nitrate levels.

a) State your null and alternative hypotheses.

b) State in plain language what a Type I error would be (e.g. if you were explaining your results to the Mayor).

c) Repeat for a Type II error.

d) What significance level do you think makes sense for the test?

e) The EPA statement on the effect of nitrate contamination: "Infants below the age of six months who drink water containing nitrate in excess of the MCL could become seriously ill and, if untreated, may die. Symptoms include shortness of breath and blue-baby syndrome." Does this change your significance level? Does this change how you want to set up the null and alternative hypotheses?

2. The sample average unrestrained compressive strength for 45 specimens of a particular type of brick was 3107 psi, and the sample standard deviation was 188 psi.

a) Does the data strongly indicate that the true average unrestrained compressive strength is less than the design value of 3200? Test using $\alpha = 0.001$.

b) Why use such a small value for α?

3. The article "Analysis of Reserve and Regular Bottlings: Why Pay for a Difference Only the Critics Claim to Notice?" reported on an experiment to determine if wine tasters could correctly distinguish between reserve and regular versions of a wine. In each trial tasters were given 4 indistinguishable containers of wine, two of which contained the regular version and two of which contained the reserve version of the wine. The taster then selected 3 of the containers, tasted them, and was asked to identify which one of the 3 was different from the other 2. In 855 trials, 346 resulted in correct distinctions. Does this provide compelling evidence that wine tasters can distinguish between regular and reserve wines?

4. Are sons taller than their fathers? A past Math 321 class contributed data to help answer this question (`http://web02.gonzaga.edu/faculty/axon/321/heights.csv`). There are two ways to use this data to answer the question:

(1) Test $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 > 0$ where μ_1 is the mean height of sons and μ_2 is the mean height of fathers;

(2) Test $H_0 : \mu = 0$ against $H_1 : \mu > 0$ where μ is the **mean difference** between the height of a son and the height of his father.

a) Use method 1 to answer the question: $n_1 = 24$, $\overline{x}_1 = 72.21875$, and $s_1 = 2.280961$; $n_2 = 24$, $\overline{x}_2 = 71.41667$, and $s_2 = 2.602953$.

b) Use method 2 to answer the question: $n = 24$, $\overline{x} = 0.8030833$, $s = 3.009928$.

c) Suppose the two methods gave different conclusions. Which one is a better answer to the original question?