Reminders: HW #8 due Monday November 7
Reading: 7.2 for Wednesday.
Exam 2: 1 week from this Friday! Sample tests now posted on the website. Keys posted outside my office.

Problem review: HW #6

Hypothesis Testing Continued
Population mean and standard deviation unknown

Last week: Z-test for a sample mean (review problems from Friday)

\[ z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \]

Today: More realistic circumstances
Population mean and standard deviation unknown

What could we substitute into \( z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \) to leave \( \mu \) as the only unknown parameter?

In \( z \), the numerator is standardized by the standard deviation of the sampling distribution of the sample mean, \( \frac{\sigma}{\sqrt{n}} \).

A related quantity, but one that we can actually estimate from data in a sample, is called the standard error of the sample mean, \( SE_{\bar{x}} = \frac{s}{\sqrt{n}} \).

If we substitute the standard error of \( \bar{x} \) for its standard deviation, we meet today’s new and improved test statistic, \( t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \)

This test statistic is a random variable, but it is NOT standard Normal. In fact, it is not Normal at all.
It’s sampling distribution is cleverly called the t-distribution. The single parameter for this distribution is called the degrees of freedom, equal to \( n-1 \).
Let’s look at it
Review and rederivation: Level C one-sample t confidence interval for a sample mean, for a sample of size n.

1. Find the “critical value” $t^*$, where $t^*$ is a value from the t-distribution $t(n-1)$ such that the area under the curve between $-t^*$ and $t^*$ equals $C$. Or put another way, $t^*$ is the score for which $P(t > t^*) = \frac{1-C}{2}$ (so once we pick $C$, $z^*$ is a fixed constant)

2. Remember that in the t case, we standardize the sample mean $\bar{x}$ by the standard error $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ and the confidence interval can then be defined as

$$P(-t^* \leq \frac{\bar{x} - \mu}{s/\sqrt{n}} \leq t^*) = C$$

3. Now we just do some 9th grade algebra (see Friday’s notes) to get the level C one-sample t C.I. for the population mean

$$P\left(\bar{x} - t^* \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t^* \frac{s}{\sqrt{n}}\right) = C$$

Put another way the level C one-sample t C.I. is

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

where $t^* \frac{s}{\sqrt{n}}$ is the margin of error for level C

So now, even if I don’t know the population standard deviation $\sigma$ I can construct an interval that includes the population mean $\mu$ for any given level of confidence. Of course, since I usually do not know the population standard deviation, this is terribly useful…

Hypothesis testing procedure: still applies for the t

1. State the null hypothesis $H_0$, and the alternative hypothesis $H_1$.
2. Specify the significance level $\alpha$.
3. Calculate the value of the test statistic. $t$ instead of $z$.
4. Find the $p$-value for the observed data, based on the $t(n-1)$ distribution, rather than the
Today’s hypothesis test: Hypothesis test for a population mean: the One-Sample t-Test.
To test the null hypothesis $H_0: \mu = \mu_0$ based on a SRS of size $n$, we compute the $z$ test statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

This statistic is a standard Normal random variable $T$ with the sampling distribution $t(n-1)$

One-tailed vs. two-tailed hypothesis tests also still apply
Remember: The P-value for two-tailed tests is twice that of either one-tailed test. So it is “easier” to reject a null hypothesis if you have good a priori reason to use a one-tailed (directional) alternative hypothesis.

Revisit our 4 problems from Friday. Ignore information about population standard deviation $\sigma$ and answer the questions based on the $t$-approach. Is this appropriate in all cases?