10.2 Tests of Significance

Use a confidence interval when you want to estimate a population parameter μ or p.

Use a Test of Significance when you want to determine if the observed value of a sample statistic $\bar{x}$ or $\hat{p}$ differs significantly from the value you think the population parameter will have (the hypothesized value).

Two competing hypotheses:

the null hypothesis: H_0

the alternative hypothesis: H_a

The big idea: an outcome that would rarely happen if a claim were true is good evidence that the claim is not true.

Ex 10.8 (P559): Free Throws

Will claims that he makes 80% of his free throws. Jamel doesn't believe him, so he asks Will to shoot 20 free throws. Will only makes 8. Jamel says that if Will really makes 80% of his free throws it's almost impossible that he would only make 8/20 (40%). Jamel says Will is a liar; no way does he make 80% of his free throws.

Jamel's reasoning is based on the idea that if the 80% claim were true and the sample of 20 free throws were repeated many times, Will would almost never make as few as 8. The 8/20 outcome is so unlikely, that if it does occur, the 80% claim is probably false.

Jamel supports the evidence against Will's 80% claim with long term probability. If Will really makes 80%, then

$P(X \le 8)$ = binomcdf(20,.8,8)=.0001

The tiny probability of scoring 8/20 convinces the rest of the class that Will was overestimating his free throw probability.

Just like with confidence intervals, we need to repeat a sample many times to develop an idea of what will happen. We will use example 10.9 (P560) to develop the process of a significance test:

Diet colas use artificial sweetener which loses sweetness over time. Companies use trained tasters to sip, compare and score whether the cola has lost sweetness. On a 1-10 scale, the result is 1st - 2nd taste.

2.0	0.4	0.7	2.0	-0.4	2.2	-1.3	1.2	1.1	2.3

+ means loss, - means gain

Does $\bar{x}$ = 1.02 mean there is a real loss of sweetness or is the outcome a product of chance?

1st: identify the population parameter - population mean μ in this case.

next: state the null hypothesis (H_0) - which says there is no effect or no change in the population. If it is true, then the sample result is just chance. Our null hypothesis is that the cola does not lose sweetness. Write as the mean sweetness loss μ in the population is zero:

$$H_0: \mu = 0$$

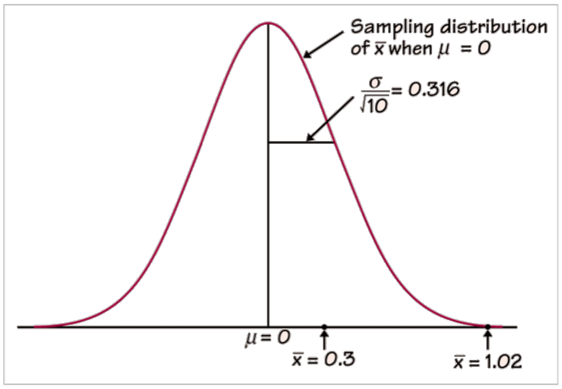
next: state the alternative hypothesis (H_a) - which is that we suspect the cola really does lose sweetness:

$$H_a: \mu > 0$$

then calculate $\sigma_{\bar{x}}$ so we can locate our sample on the normal curve and determine the likelihood of it occurring.

P-value

The p-value is what we use to measure the strength of evidence against H_0. It is the probability under the normal curve to the right of the observed value.



2/17 start here

Small P-values are evidence against H_0.
No fixed rule tells how small p should be to reject H_0, but 0.05 is a good rule of thumb. We call this level α. We will learn that we can test for a specific level α dependent on the level of certainty needed.

statistically significant:

P-values less than 0.05
chance alone is not likely to be responsible for the result.

Which way do I test?

one-sided alternative:

$H_a > H_0$ we think the alternate is bigger than the null

$H_a < H_0$ we think the alternate is smaller than the null

two-sided alternative:

$H_a \neq H_0$ we think the alternate is different than the null

4 Step Process for Significance Test:

1. Identify population and parameter of interest.
   State null and alternative hypotheses in words and symbols
2. Choose procedure and verify conditions
3. - calculate test statistic (observed z)
   - find P-value
4. Interpret results in context

Important notations:

null hypothesis has specified value μ_0 $\quad H_0: \mu = \mu_0$

one-sample z statistic (observed z) $\quad z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

We will use standard normal variable Z (capital Z) when calculating our P-values.

Our notation will look like this:

P = P(Z ≥ z)
or
P = P(Z ≤ z)
for one-sided tests

and

P = 2P(Z ≥ |z|) for two-sided tests

The smaller p is, the stronger our evidence against H_0.

If testing for $H_a > H_0$, then p-value is area to the right of our $\overline{x}$.

$H_a: \mu > \mu_0$ is $P(Z \geq z)$

If testing for $H_a < H_0$, then p-value is area to the left of our $\overline{x}$.

$H_a: \mu < \mu_0$ is $P(Z \leq z)$

If testing for $H_a \neq H_0$, then p-value is area to the right of our $\overline{x}$ PLUS the area to the left of our $\overline{x}$.

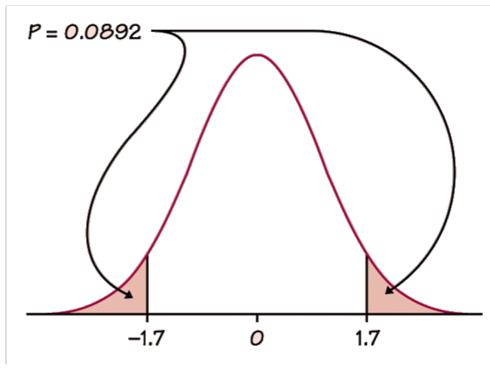
Since they are symmetrical, we will calculate once and then double.

$H_a: \mu \neq \mu_0$ is $2P(Z \geq |z|)$

ex. 10.12 Calculating a two sided P-value:

Given the z test statistic for a two sided test: z = 1.7

That means P(Z ≥ |z|)
= P(Z ≤ -1.7 or Z ≥ 1.7)
= P(Z ≤ -1.7) + P(Z ≥ 1.7)
= 2*P(Z ≥ 1.7)
= 2*(1 - 0.9554)
= 0.0892



ex 10.13: Executives' Blood Pressures (a two-sided test)

National Center for Health Statistics says:

mean systolic blood pressure for males 35-44 is 128 & std dev is 128. Medical director looks at 72 execs. and finds mean of sample is 126.07. Is this evidence that the company's execs have different mean bp from general population? Assume the same std dev as general pop of middle aged males.

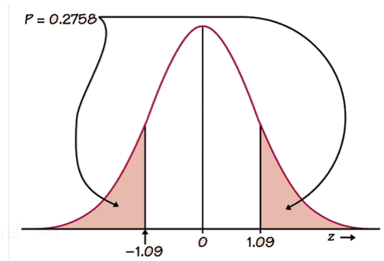
Need to identify all key pieces of information and follow 4-step toolbox for one-sample z statistic on next slide.

1. I want to test the claim that the mean blood pressure of all middle-aged male execs. in this company is different from the national mean for males of the same age.

$H_0: \mu = 128$ $H_a: \mu \neq 128$ different : $\neq$

2. One-sample z-test for population μ.
   - assume SRS
   - sampling distribution is normal b/c n=72 (large enough to use CLT)

3.
   - $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{126.07 - 128}{15/\sqrt{72}} = -1.09$
   - P-value $= 2P(Z \geq 1.09)$
     $= 2(.1379)$
     $= .2758 \Rightarrow$ * .2758 is larger than both common α levels: .05 and .01



4. A sample (SRS) of 72 male executives has a 27% chance of being as far from 128 or farther than any other sample of 72 men from the population. Our observed mean ($\bar{x} = 126.07$) is not good enough evidence to reject the null hypothesis. We fail to reject H_0.

Testing at a Fixed Significance Level

Sometimes we need a very specific degree of evidence to reject the H_0, so we use α to drive the level of evidence we need.

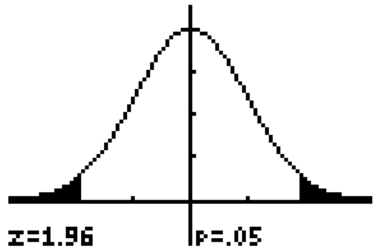
We can use the same z* values that we found for confidence intervals to test for significance. We already know that the central 90% of data for any normal curve is bound by the critical values z*=-1.645 and z*=1.645, so we already know that each tail has p = 0.05.

Once we find the observed z, we can compare it with the appropriate z* from Table C to determine significance.

For one sided testing:
α = 0.05 means z*= -1.645 or 1.645
α = 0.01 means z*= -2.326 or 2.326

But for two-sided testing (≠):

α = 0.05 means there is p = 0.025 in each tail or .95 in the central portion. We can use z* = -1.96 and 1.96



α = 0.01 means there is p = 0.005 in each tail or .99 in the central portion. We can use z* = -2.576 and 2.576