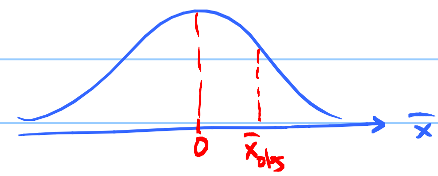
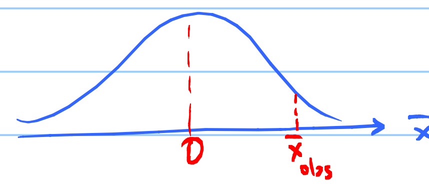
hw-lect20-1

Consider a problem wherein someone claims that $\mu_x < 0$, and you want to see if data provide evidence to its contrary. Consider the following 2 hypothetical $\bar{x}_{obs}$ values.

I)



II)

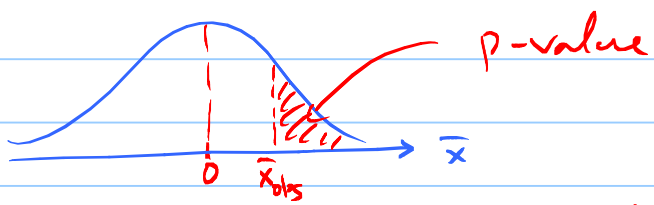


a) Forget p-values and all that! Which of the 2 situations (I or II) provides more evidence (from data) against $\mu_x < 0$?

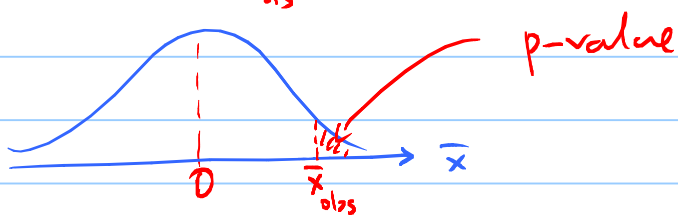
II provides more evidence against $\mu_x < 0$. If $\bar{x}_{obs} = 10^6$ that would constitute a lot of evidence against $\mu_x < 0$.

b) Now, we would like to find the p-values for these 2 cases, but I haven't given you any numbers for $\bar{x}_{obs}$. Instead, shade the p-values as areas under those 2 figs.

I)



II)



c) Which case (I or II) has the <u>smaller</u> p-value? II

Moral:

The situation with the <u>smaller</u> p-value provides <u>more evidence against $\mu_x < 0$</u>. This simple example illustrates how counter-intuitive (how confusing) p-values can be. In short:

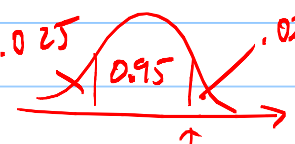
<u>small</u> p-value $\Rightarrow$ <u>more</u> evidence against H_0 (in favor of H_1)

hw_lect20_2

Consider the following sample observations: 2781, 2900, 3013, 2856, and 2888.
Suppose we want to test whether there is evidence contrary to the belief that mu < 3000.
a) Compute the observed 95% 2-sided confidence interval (CI) for mu.
b) Based on the above CI, is there evidence that mu is greater than 3000?
c) Write the appropriate null hypotheses.
d) Compute the p-value, recalling that it measures evidence from data contrary to the null hypothesis.
e) At alpha=0.05, state the conclusion "In English" (i.e., is there evidence that mu is greater than 3000?)

a) 95% CI $\quad \bar{x} \pm t^* \frac{s}{\sqrt{n}} = 2887.6 \pm 2.8 \frac{84.03}{\sqrt{5}}$

$[2782.4,\ 2992.8]$



df = 5-1

Table 6

FYI: Technically we should look at the UCB

b) The entire CI is below 3000, and so There is NO evidence from our data That $\mu > 3000$.

c) The belief goes in H_0. I.e. $H_0: \mu < 3000$.

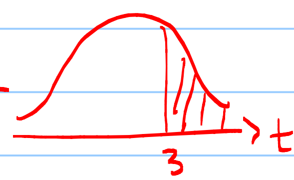
d) As explained in the lecture, evidence contrary to H_0 is sufficient. (Blue note).

$$\text{p-value} = pr(\bar{x} > \bar{x}_{obs} \mid \mu_x = 3000)$$

$$= pr\left(\frac{\bar{x}-\mu_x}{s/\sqrt{n}} > \frac{\bar{x}_{obs}-\mu_x}{s/\sqrt{n}}\right) = pr(t > t_{obs}) \qquad t_{obs} = \frac{2887.6 - 3000}{(84.03)/\sqrt{5}} = -3.0$$

pvalue $= pr(t > -3.0) =$

$= 1 -$



$= 1 - 0.02$

df=4
Table 6

$= 0.98$

e) At $\alpha = .05$, we have p-value $> \alpha \Rightarrow$ Cannot Reject H_0 $(\mu < 3000)$

In English: There is no evidence That $\mu > 3000$

Note: even Though The numbers between CI and p-value are different, The conclusion is The same.

hw_lect21_1

Toothpaste tubes may be wasteful because there is always some amount of toothpaste that one cannot extract. To find out how much toothpaste is wasted, 5 discarded tubes are selected, cut open, and the amount of remaining toothpaste is recorded. The data are : 0.52, 0.65, 0.46, 0.50, 0.37 (in ounces). Is there evidence that the true average of the wasted toothpaste is less than 0.55 ounces? Apply the hypothesis testing procedure as follows:

a) What is the pop. parameter being tested?
Write the symbol for it, AND explain it in words.
μ_x (i.e. the true mean of the wasted tooth paste)

b) Restate the question as "Does data provide evidence ... "

Does data provide evidence
$\underbrace{\text{that the true mean of the wasted tooth paste is at most 0.55 ounces?}}_{\text{ie. that } \mu_x < 0.55}$

c) Which of the following pairs of hypotheses is appropriate?
Check the solns later to see the explanation/thinking.

$H_0 : \mu_x < 0.55$ $\quad$ $H_0 : \mu_x > 0.55$ $\quad$ $H_0 : \mu_x = 0.55$

$H_1 : \mu_x > 0.55$ $\quad$ $H_1 : \mu_x < 0.55$ $\quad$ $H_1 : \mu_x \neq 0.55$

Explanation: We know the answer is one of the 1st two because the question is a 1-sided question. To answer the question

Does data provide evidence for $\mu_x < 0.55$?

we must not assume $\mu_x < 0.55$. Our procedure assumes H_0 is True. So, our H_0 must not be $H_0: \mu_x < 0.55$. Hence $H_0: \mu_x > 0.55$.

d) In our procedure we must assume H_0 = True. Assuming H_0 = T, what is the "worse" value that μ_x can take?
Hint: values in the direction of H_1 are "worse" for H_0.

$\mu_x = 0.55$

e) Assuming The "worse-case" scenario of part d, compute The p-value. Hint: remember That The p-value measures evidence against (Contrary to Ho, or in favor of H_1); consult p. 6 of lect 22 if you need more help. Use Table VI.

$\bar{x}_{obs} = \frac{1}{5}(0.52 + \cdots) = 0.50$

$s^2_{obs} = \frac{1}{5-1}\left[(0.52-0.5)^2 + \cdots\right] = 0.0104 \longrightarrow s_{obs} = 0.102$

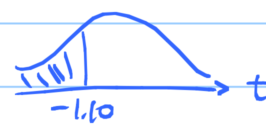
p-value $= pr(\bar{x} < \bar{x}_{obs} \mid \mu_x = 0.6)$

optional.

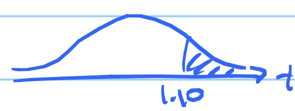
$= pr\left(\frac{\bar{x}-\mu_x}{s_x/\sqrt{n}} < \frac{\bar{x}_{obs}-\mu_x}{s_x/\sqrt{n}} \;\middle|\; \mu_x = 0.6\right)$

$= pr(t < t_{obs})$ where $t_{obs} = \frac{0.5 - 0.55}{0.102/\sqrt{5}} = -1.10$

$= pr(t < -1.10)$ =



$= pr(t > 1.10)$ =



$= 0.167$ (df = 5-1 in Table VI)

f) Is The p-value you have computed small (less Than 0.05 $= \alpha$) or large (larger Than 0.05).

p-value = large (ie. larger Than $\alpha = .05$).

g) Based on your answer to part f), should you reject Ho in favor of H_1? No.

Because p-value > α, we cannot reject Ho ($\mu > 0.55$) in favor of H_1 ($\mu < 0.55$)

h) What is The conclusion (In English)?

There is no evidence That $\mu < 0.55$

hw-lect21-2

I Suppose you are asked if there is evidence that $\mu_x > \bar{x}_{obs}$?

a) Set-up the appropriate H_0/H_1

b) Compute the p-value.

II Suppose you are asked if there is evidence that $\mu_x > \bar{x}_{obs} - 1.645\frac{s}{\sqrt{n}}$?

FYI: The right-hand side is the 95% LCB (which we are skipping).

c) Set-up the appropriate H_0/H_1

d) Compute the p-value. Use $pr(t > 1.645) = 0.05$, i.e. $df = n-1 = \infty$

Look at the soln later to see the moral of this hw.

a) $H_0: \mu_x < \boxed{\bar{x}_{obs}} = \mu_0$ sufficient to test $H_0: \mu_x = \mu_0$

$H_1: \mu_x > \boxed{\bar{x}_{obs}}$

b) $\text{p-value} = pr(\bar{x} > \bar{x}_{obs} \mid \mu_x = \mu_0) = pr\left(\frac{\bar{x}-\mu_x}{s/\sqrt{n}} > \frac{\bar{x}_{obs}-\mu_0}{s/\sqrt{n}}\right)$

$= pr\left(t > \frac{\bar{x}_{obs}-\bar{x}_{obs}}{\cdots}\right) = pr(t > 0) = \boxed{0.5}$

c) $H_0: \mu_x < \overbrace{\bar{x}_{obs} - 1.645\, s/\sqrt{n}}^{\mu_0}$

$H_1: \quad >$

d) $\text{p-value} = pr(\bar{x} > \bar{x}_{obs} \mid \mu_x = \mu_0) = pr\left(\frac{\bar{x}-\mu_x}{s/\sqrt{n}} > \frac{\bar{x}_{obs}-\mu_0}{s/\sqrt{n}}\right)$

$= pr(t > t_{obs})$ $\qquad t_{obs} = \frac{\bar{x}_{obs}-\mu_0}{s/\sqrt{n}}$

$= pr(t > 1.645) = \boxed{0.05}$ $\qquad = \frac{\cancel{\bar{x}_{obs}} - \cancel{\bar{x}_{obs}} + 1.645\, s/\sqrt{n}}{s/\sqrt{n}} = 1.645$

1- pt(1.645, 10000) FYI

Moral: This 0.05 is related to the 95% conf. level associated with the LCB given above. All of this shows how the p-value approach and the confidence interval approach are intimately linked.

hw_lect22_1

In hw_lect17_1 we used a CI to answer the question Does it appear that the true proportion of detective screws is not 2.5%. Here, answer the same question with the p-value approach. Specifically,

a) Set-up the appropriate hypotheses.

b) Compute the p-value (using the data in that hw)

c) At alpha =0.01, state the conclusion? Is it consistent with the conclusion from the CI approach?

a) The question above translates to:

$H_0: \pi_x = 2.5\%$

$H_1: \pi_x \neq 2.5\%$

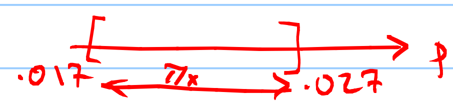
b) $z_{obs} = \dfrac{p_{obs} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} = \dfrac{.022 - .025}{\sqrt{\dfrac{.025(1-.025)}{2000}}} = \dfrac{-0.003}{0.0035} = -0.86$

Not p

p-value $= 2\, pr(z < z_{obs}) = 2\, pr(z < -0.86) = 2(0.19) = \boxed{0.38}$

c) p-value > α => cannot reject $H_0: \pi_x = 2.5\%$ in favor of $H_1: \pi_x \neq 2.5\%$

I.e. There is no evidence for $\pi_x \neq 2.5\%$

In The CI approach, we found 

The 2 conclusions are consistent because They both say That π_x could be 2.5%.

hw_lect22_2

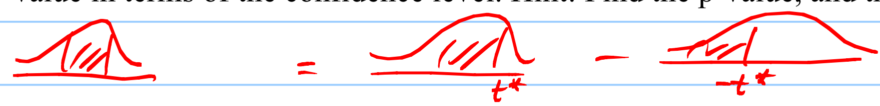
We are supposed, to transform our question into "Does data provide evidence for ...?" Usually the "..." is specified by you, the scientist. But just for practice, and to better understand the relationship between CIs and p-values, let's ask "Does data provide evidence that the difference (mu_2 - mu_1) is less than the upper side of the observed (2-sided) 2-sample CI?" Do not fix the Conf. level, i.e. don't pick a number for t*.

a) First, recall that the t* that appears in the formula for the CI is designed to satisfy pr(-t* < t < t*) = confidence level. Then, starting from that "self-evident fact," show that t* also satisfies pr(t< -t*) = (1-conf level)/2.

b) Setup H0, H1. Hint: look-up our formula for the observed 2-sample CI, and select the upper side. This step will also tell you what is delta for this problem.

c) Find the p-value in terms of the confidence level. Hint: Find the p-value, and then use part a.

a)



$$pr(-t^* < t < t^*) = pr(t < t^*) - pr(t < -t^*) = 1 - 2\,pr(t < -t^*)$$

Total area = 1 → $pr(t<t^*) = 1 - pr(t > t^*)$

t-dist = symmetric → $1 - pr(t > t^*) = 1 - pr(t < -t^*)$

$$\therefore\ pr(t < -t^*) = \frac{1 - \text{Conf. level}}{2}$$

b) The question asked in The problem "translates" to:

H_0: $\mu_2 - \mu_1 <$ upper side of The observed CI for $\mu_2 - \mu_1$

H_1: $\mu_2 - \mu_1 >$ " " " "

Observed 2-sample CI for $\mu_2 - \mu_1$ is $\bar{x}_2 - \bar{x}_1 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

so
$$\left\{ \begin{array}{l} H_0:\ \mu_2 - \mu_1 > (\bar{x}_2 - \bar{x}_1)_{obs} + t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\ H_1:\ \mu_2 - \mu_1 < (\bar{x}_2 - \bar{x}_1)_{obs} + t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \end{array} \right\} \Rightarrow \Delta \equiv (\bar{x}_2 - \bar{x}_1)_{obs} + t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

c) p-value $= pr\left((\bar{x}_2 - \bar{x}_1) < (\bar{x}_2 - \bar{x}_1)_{obs} \mid H_0 = T \right)$

$$= pr\left(\frac{\bar{x}_2 - \bar{x}_1 - (\mu_2 - \mu_1)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} < \frac{(\bar{x}_2 - \bar{x}_1)_{obs} - (\mu_2 - \mu_1)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \,\middle|\, \mu_2 - \mu_1 = \Delta \right) = pr(t < t_{obs})$$

where
$$t_{obs} = \frac{(\bar{x}_2 - \bar{x}_1)_{obs} - \Delta}{\sqrt{\cdots}} = \frac{\cancel{(\bar{x}_2 - \bar{x}_1)_{obs}} - \cancel{(\bar{x}_2 - \bar{x}_1)_{obs}} - t^* \cancel{\sqrt{\cdots}}}{\cancel{\sqrt{\cdots}}} = -t^*$$

$$= pr(t < -t^*) = \frac{1 - \text{Conf. level}}{2}$$