# Lecture 20 (ch. 8)

We now have a method for testing hypotheses with p-values.
The method involves the prob of getting more extreme (than obs.) events, and whether that prob is sufficiently small.

Q More extreme?
Depends on $(H_0, H_1)$:

Because of the blue note, above, it is sufficient to test $\mu = \mu_0$

1-sided

If $H_0: \mu \leq \mu_0$ (or $\mu = \mu_0$)
$H_1: \mu > \mu_0$

p-value $= pr(\bar{x} > \bar{x}_{obs} \mid \mu < \mu_0)$ = right area

If $H_0: \mu \geq \mu_0$ (or $\mu = \mu_0$)
$H_1: \mu < \mu_0$

p-value $= pr(\bar{x} < \bar{x}_{obs} \mid \mu > \mu_0)$ = left area

2-sided

If $H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$

p-value = sum of tail areas, or twice one tail area.

In Summary:

1) You choose the value of $\alpha$.

2) Compute p-value from the above procedure.

3) If p-value $< \alpha$, Then Reject $H_0$ in favor of $H_1$.

Else, cannot reject " " " " "

---

Yesterday, a student (who we will call Sugar) asked a good question:

Note: consider p-value $= pr(\bar{x} > \bar{x}_{obs} \mid \overbrace{\mu < \mu_0}^{H_0})$ which measures evidence from data in favor of $H_1: \mu > \mu_0$.

switched

One may think That (1 - pvalue) measures evidence for $H_1: \mu < \mu_0$.

But it doesn't because (1 - pvalue) $= pr(\bar{x} < \bar{x}_{obs} \mid \mu < \mu_0)$

This is switched, not.

In prev. example, we had $n = 64$, $\bar{x}_{obs} = 34.4$, $s = 1.1$, and asked "Does data provide evidence to support $\mu > 34$?" Then

$H_0: \mu \le 34$
$H_1: \mu > 34$

I always write these so that $H_0$ and $H_1$ have opposite directions, because it's logical. The book does not. The "equality" in $H_0$ just reminds us that it's sufficent to test $H_0: \mu = 34$. (The Blue note).

$\therefore$ p-value $= pr(\bar{x} > \bar{x}_{obs}) = prob(t > t_{obs}) = pr(t > 2.91) \overset{df = 64-1}{=} 0.0025$.

$$t = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}} = \frac{34.4 - 34}{1.1/\sqrt{64}} = 2.91$$

Since p-value $< \alpha$, then there is evidence to support $\mu > 34$.

It is tempting to say the above "conclusion" (at $\alpha = .05$), that $\mu > 34$, is obvious and trivial. After all the sample gave $\bar{x}_{obs} = 34.4$, which is greater than 34 already.

It's NOT obvious! Suppose the sample/data gave $\bar{x}_{obs} = 34.1$, ie. still larger than 34 ($\mu_0$). Then

$$t_{obs} = \frac{34.1 - 34}{1.1/\sqrt{64}} = .73 \Rightarrow \text{p-value} = prob(t > 0.73) = 0.24$$

This p-value is larger than any reasonable $\alpha$. So, we cannot reject $H_0$ in favor of $H_1$ even though the obs. sample mean is bigger than 34. 34.1 is larger than 34, but just not enough (in units of standard error, $\frac{s}{\sqrt{n}}$) to justify rejecting $H_0$ ($\mu < 34$) in favor of $H_1$ ($\mu > 34$).

A-priori !

There are many ways to rephrase the statement/question in a problem. Here are some of them:

$\alpha = .05$  Data says: $n = 64$, $\bar{x} = 34.4$, $s = 1.1$

$t_{obs} = \frac{34.4 - 34}{1.1/\sqrt{64}} = 2.91$

Does data support $\mu > 34$? $\Longrightarrow$ "prior claim": $H_0: \mu \leq 34$

$H_0: \mu \leq 34$  p-value $= prob(\bar{x} > \bar{x}_{obs}) = prob(t > t_{obs})$

$H_1: \mu > 34$  $= prob(t > 2.91) = .0025 < \alpha$

$\therefore$ Reject $H_0$ ($\mu \leq 34$) in favor of $H_1$ ($\mu > 34$).

$\therefore$ Data does support $\mu > 34$.

---

Does data support $\mu < 34$?

$H_0: \mu \geq 34$  p-value $= prob(\bar{x} < \bar{x}_{obs}) = prob(t < t_{obs})$

$H_1: \mu < 34$  $= prob(t < 2.91) = 1 - pr(t > 2.91) = 0.998 > \alpha$

$\therefore$ Cannot Reject $H_0$ ($\mu \geq 34$) in favor of $H_1$ ($\mu < 34$).

$\therefore$ Data does not support $\mu < 34$.

---

Does data contradict $\mu \geq 34$? $\leftarrow$ prior claim: $H_0: \mu \geq 34$

$H_0: \mu \geq 34$  p-value $= prob(\bar{x} < \bar{x}_{obs}) = prob(t < t_{obs})$

$H_1: \mu < 34$  $= prob(t < 2.91) = 1 - pr(t > 2.91) = 0.998 > \alpha$

$\therefore$ Cannot Reject $H_0$ ($\mu \geq 34$) in favor of $H_1$ ($\mu < 34$).

$\therefore$ Data does not contradict $\mu \geq 34$.

---

Does data contradict $\mu \leq 34$?

$H_0: \mu \leq 34$  p-value $= prob(\bar{x} > \bar{x}_{obs}) = prob(t > t_{obs})$

$H_1: \mu > 34$  $= prob(t > 2.91) = .0025 < \alpha$

$\therefore$ Reject $H_0$ ($\mu \leq 34$) in favor of $H_1$ ($\mu > 34$).

$\therefore$ Data does contradict $\mu \leq 34$.

Now, given the similarity between C.I. and the hypothesis testing approach (ie. with p-value) guess what the hypotheses for a 2-sample test are:

$H_0: \mu_2 \square \mu_1 \qquad H_1: \mu_2 \square \mu_1 \quad (\text{ie. } \mu_2-\mu_1 \square 0)$

It turns out we can solve a more general problem:

$H_0: \mu_2-\mu_1 \square \Delta \qquad H_1: \mu_2-\mu_1 \square \Delta$

I.e. Instead of zero, use $\Delta$, the null parameter.
You can always set it to zero, if desired.
YOU choose $\Delta$! Not Data.

⇒ If 2-samples are indep., then assuming $H_0 = T$,

$$z = \frac{(\bar{x}_2-\bar{x}_1)-\Delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$$t = \frac{(\bar{x}_2-\bar{x}_1)-\Delta}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \sim t\text{-dist. with } df = \text{Welch}.$$

Then, p-values are computed just as before:

$$\text{p-value} = \begin{cases} \text{prob}(t > t_{obs}) & \text{if } H_1: \mu_2-\mu_1 > \Delta \\ \text{prob}(t < t_{obs}) & \text{if } H_1: \mu_2-\mu_1 < \Delta \\ \text{twice "tail"} & \text{if } H_1: \mu_2-\mu_1 \neq \Delta \end{cases}$$

(Table VI)

⇒ If the two samples are paired: Make a new column:

| $x_1$ | $x_2$ | $d = x_1 - x_2$ |
|---|---|---|
| . | . | . |
| . | . | . |
| . | . | . |
| | | $\bar{d}, s_d$ |

$$t = \frac{\bar{d}-\Delta}{s_d/\sqrt{n}} \sim t\text{-dist.} \quad df = n-1$$

p-value computed as before.

Reconsider This example from a past lecture:

Example: 82 students have picked-up Their Test, but 30 have not, even 1 week after the test was returned.

Call These 2 groups "Attenders" and "Non-attenders".

| | | $n$ | $\bar{x}$ | $s$ |
|---|---|---|---|---|
| 1 | Non-attend | 30 | 11.8 | 3.32 |
| 2 | Attend | 82 | 13.25 | 3.04 |

$\}$ sample

$\mu_1$ = mean of test1 for Non-attend students who have ever taken 390.

$\mu_2$ = " " Attend students " " " " ".

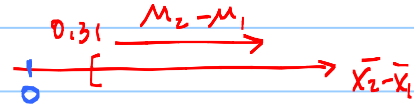
Is There evidence from data That $\mu_2 > \mu_1$?

We need to build The LOWER conf. bound for $\mu_2 - \mu_1$:

$$(\bar{x}_2 - \bar{x}_1) - 1.645\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(13.25 - 11.8) - 1.645\sqrt{\frac{(3.32)^2}{30} + \frac{(3.04)^2}{82}} = 1.45 - 1.645(.693)$$

$$1.45 - 1.14 = 0.31 \implies$$




Corollary: Zero is not included in that interval. So There is evidence That attending students have a higher pop. mean than Non-attend.

---

Now, In Chapter 8's way:

$H_0: \mu_2 - \mu_1 \le 0$

$H_1: \mu_2 - \mu_1 > 0$

$$t_{obs} = \frac{1.45 - 0}{0.693} = 2.1$$

Table VI

p-value = prob$(t > 2.1) \doteq \underline{0.0205} \implies$ At $\boxed{\alpha = .05}$, p-value $< \alpha$.

$\therefore$ Reject $H_0$ in favor of $H_1$

$\mu_2 < \mu_1$ $\quad$ $\mu_2 > \mu_1$

$$df = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{1}{n_1 - 1}\left[\frac{s_1^2}{n_1}\right]^2 + \frac{1}{n_2 - 1}\left[\frac{s_2^2}{n_2}\right]^2} = 47.91$$

"In English": There is evidence for $\mu_2 > \mu_1$.

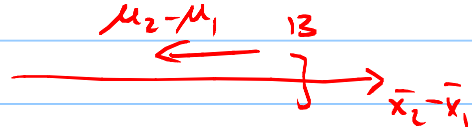
Q1: In a certain article, The 95% upper conf. bound for $\mu_2-\mu_1$ is reported as 13.0. What would be The appropriate $H_0, H_1$ for addressing The same question asked in The article.

| A | B | C |
|---|---|---|
| $H_0: \mu_2-\mu_1 > \Delta$ | $H_0: \mu_2-\mu_1 < \Delta$ | $H_0: \mu_2-\mu_1 < 13$ |
| $H_1: \mu_2-\mu_1 < \Delta$ | $H_1: \mu_2-\mu_1 > \Delta$ | $H_1: \mu_2-\mu_1 > 13$ |

If The paper reports The upper conf. bound, Then it wants to know how big $\mu_2-\mu_1$ can possibly get. Then The appropriate question is "Does data provide evidence for $\mu_2-\mu_1$ < some number?" So

$H_0: \mu_2-\mu_1 > \Delta$

$H_1: \mu_2-\mu_1 < \Delta$




C is wrong because The inequality is wrong.

And because The 13 comes from data!

$\boxed{\alpha = ?}$

In the procedure we have learned, the last step involves comparing the p-value with $\alpha$. That practice is (slowly) become "old style". More recently, one reports the p-value itself, because by itself it's useful - it reflects the evidence from data against $H_0$.

But, $\alpha$ does have an important interpretation nevertheless. We know that it is the largest prob at which we are confident to reject $H_0$ in favor of $H_1$. But there is more to it!

Suppose we are testing $H_0: \mu \le \mu_0$ vs. $H_1: \mu > \mu_0$.

We assume $H_0$ = True (ie. $\mu = \mu_0$), then compute a p-value.

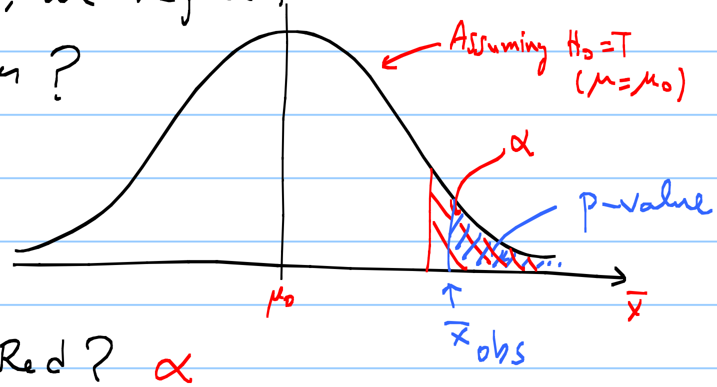
If p-value $< \alpha$, then Reject $H_0$ in favor of $H_1$.

So, every time p-value $< \alpha$, we reject.

$\underline{Q}$ How often will that happen?

For $H_0, H_1$ given here

p-value = prob$(\bar{x} > \bar{x}_{obs})$




$\underline{Q}$ How frequently is $\bar{x}$ in the Red? $\alpha$

$\alpha$ = prob(p-value $< \alpha \mid H_0$=T)

So, $\alpha$ = prob$(\underbrace{\text{Data Reject } H_0 \text{ in favor of } H_1 \mid H_0 = T}_{\text{Type I error}})$ ⟹ "Bad" error / "False Alarm Rate" / (Convicting an innocent person.)

This is how you decide the value of $\alpha$. You ask
"How much bad error can I tolerate in the long run?"

The other error is called $\underbrace{\text{Type II}}$, and it's not as bad!

(Data cannot reject $H_0$ in favor of $H_1$ | $H_0$ = False)

(Releasing a guilty person.)

# Summary

We are done with 1-sample and 2-sample, z and t-tests, for paired and unpaired data, but all of that has dealt with the pop. means. What about pop. proportions?

Easy! Follow the pattern:

## 1-Sample

**C.I. for $\mu_x$:**

$$\bar{x} \pm z^* \frac{\sigma_x}{\sqrt{n}}$$

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

$df = n-1$

**C.I. for $\pi_x$:**

$$p \pm z^* \sqrt{\frac{p(1-p)}{n}}$$

t-version does not exist

**test for $\mu_x$:**

$H_0: \mu \,\square\, \mu_0 \qquad H_1: \mu \,\square\, \mu_0$

$$z_{obs} = \frac{\bar{x}_{obs} - \mu_0}{\sigma_x / \sqrt{n}}$$

p-value = ...

$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s_x / \sqrt{n}}$$

$df = n-1$

**Test for $\pi_x$:**

$H_0: \pi \,\square\, \pi_0 \qquad H_1: \pi \,\square\, \pi_0$

$$z_{obs} = \frac{p_{obs} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

No p! Because we assumed $H_0 = T$, ie. $\pi = \pi_0$.

p-value = ...

## 2-Sample

**C.I. for $\mu_2 - \mu_1$:**

$$\bar{x}_2 - \bar{x}_1 \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

df = Welch

**Test for $\pi_2 - \pi_1$:**

$$p_2 - p_1 \pm z^* \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Again No t!

**Test for $\mu_2 - \mu_1$:**

$H_0: \mu_2 - \mu_1 \,\square\, \Delta \qquad H_1: \dots$

$$z_{obs} = \frac{(\bar{x}_2 - \bar{x}_1)_{obs} - \Delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$t_{obs} = \frac{(\bar{x}_2 - \bar{x}_1)_{obs} - \Delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

df = Welch

**Test for $\pi_2 - \pi_1$:**

$H_0: \pi_2 - \pi_1 \,\square\, \Delta \quad \dots$

$$z_{obs} = \frac{(p_2 - p_1) - \Delta}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

hw-lect 20-1

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 C.I's and p-values, let's ask "Does data provide evidence for $\mu_x <$ observed 95% upper confidence bound for $\mu_x$?"

a) Set-up $H_0, H_1$, b) compute The p-value

Hint: Recall The defn of The 95% upper conf. bound, and note That The $t^*$ That appears in That formula satisfies $pr(t > -t^*) = 0.95$

hw-lect 20-2

hw-lect 17-2 ~~hw-lect 18-1~~ asked does it appear That $\pi_x$ (The true proportion of defective screws) is at most 2.5%?

There, The appropriate interval is The upper conf. Bound for $\pi_x$.

a) Set-up The appropriate pair of hypotheses.

b) Compute The p-value (using The data in hw-lect 17-2 ~~hw-lect 18-1~~)

c) At $\alpha = .05$, is The conclusion consistent with The conclusion from The CI approach in ~~hw-lect 18-1~~ hw-lect 17-2?