PSY 201: Statistics in Psychology

Lecture 19

Hypothesis testing of the mean

Why clinical studies use thousands of subjects.

Greg Francis

Purdue University

Fall 2023

SUPPOSE

- we think the mean value of a population of SAT scores is $\mu = 455$
- we can take a sample of $n = 144$ from the population and calculate the sample mean of SAT scores $\overline{X} = 535$ with sample standard deviation $s = 100$

HYPOTHESIS TESTING

- four steps
    1. State the hypothesis and criterion.
    2. Compute the test statistic.
    3. Compute the p value
    4. Make a decision.

RECAP OF LAST TIME

- (1) State the hypotheses and set the criterion

$$H_0 : \mu = 455$$

$$H_a : \mu \neq 455$$

- $\alpha = 0.05$

RECAP OF LAST TIME

- (2) Compute the test statistic

$$t = \frac{\overline{X} - \mu}{s_{\overline{X}}}$$

$$t = \frac{535 - 455}{8.33} = 9.60$$

- (3) Compute the p-value (using the t-distribution calculator with $df = n - 1$):

$$p \approx 0$$

- (4) Make a decision: $p < \alpha$, so reject H_0
  - the found sample mean would be a very rare event if H_0 were true

DIFFERENT MEAN

- suppose we had the same situation as before, but we had instead found

$$\overline{X} = 465$$

- (1) State the hypotheses and set the criterion (unchanged!)

$$H_0 : \mu = 455$$

$$H_a : \mu \neq 455$$

- $\alpha = 0.05$
- (2) Compute the test statistic

$$t = \frac{\overline{X} - \mu}{s_{\overline{X}}}$$

$$t = \frac{465 - 455}{8.33} = 1.20$$

DIFFERENT MEAN

- (3) Compute the p-value (using the t-distribution calculator with $df = n - 1$):

$$p = 0.2301$$

- (4) Make a decision: $p > \alpha$, so do **not** reject H_0
- the found sample mean would not be very rare if H_0 were true
  - if the null hypothesis is true, then the probability that $|\overline{X}| \geq 465$ would be found by random sampling is greater than .05

SAMPLE SIZE

- suppose we had the same situation as before, but we had instead found

$$\overline{X} = 465$$

- with a sample size of $n = 500$
- (1) State the hypotheses and set the criterion

$$H_0 : \mu = 455$$

$$H_a : \mu \neq 455$$

- $\alpha = 0.05$

SAMPLE SIZE

- (2) Compute the test statistic

$$t = \frac{\overline{X} - \mu}{s_{\overline{X}}}$$

- we need to recompute $s_{\overline{X}}$

$$s_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{100}{\sqrt{500}} = 4.47$$

$$t = \frac{465 - 455}{4.47} = 2.24$$

SAMPLE SIZE

- (3) Compute the p-value (using the t-distribution calculator with $df = n - 1 = 499$):

$$p = 0.0251$$

- (4) Make a decision: $p < \alpha$, so do reject H_0
  - the found sample mean would be a rare event if H_0 were true. The probability that $|\overline{X}| \geq 465$ would be found by random sampling is less than .05

CALCULATOR

- you need to understand the math and calculations, but generally you should not **do** it

Enter data:

Sample size n = 500

Sample mean $\overline{X}$ = 465

Sample standard deviation s = 100

Specify hypotheses:

$H_0 : \mu =$ 455

$H_a :$ Two-tails

$\alpha =$ 0.05

Run Test

Test summary	
Null hypothesis	$H_0 : \mu = 455$
Alternative hypothesis	$H_a : \mu \neq 455$
Type I error rate	$\alpha = 0.05$
Sample size	$n = 500$
Sample mean	$\overline{X} = 465.0000$
Sample standard deviation	$s = 100.000000$
Sample standard error	$s_{\overline{X}} = 4.472136$
Test statistic	$t = 2.236068$
Degrees of freedom	$df = 499$
p value	$p = 0.025789$
Decision	Reject the null hypothesis
Confidence interval critical value	$t_{cv} = 1.964729$
Confidence interval	$CI_{95} = (456.213463, 473.786537)$

CLINICAL TRIALS

- often hear about medical studies that track thousands of patients
- why do they need so many people?
- a larger sample makes for less variation in the sampling distribution of the mean

$$s_{\overline{X}} = \frac{s}{\sqrt{n}}$$

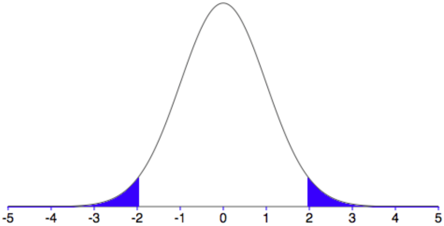
- thus, if the null hypothesis really is false, you are more likely to reject it with a larger sample
- if the null hypothesis is really true, you are not more likely to reject it (no extra mistakes with a larger sample size!)

COMMENTS

- several things are worth noting
  - The α probability is about the *process* of making decisions. It controls Type I error rates, but for any given decision you do not know if you made an error or not.
  - Even when we reject H_0, there is always a chance that it is true.
  - Even when we do not reject H_0, there is always a chance that it is false.
  - The statement $p < 0.05$ is about the **statistic** given the hypothesis, not about the hypothesis. We never conclude that H_0 is false with probability 0.95.
  - Technically, we have done all of this before.
  - These techniques are quantifiable.
  - No inclusion of knowledge about the direction of difference.

DIRECTIONAL HYPOTHESIS

- we choose a significance level, α
- indicates probability of Type I error
- earlier, we split this error across the two tails of the sampling distribution

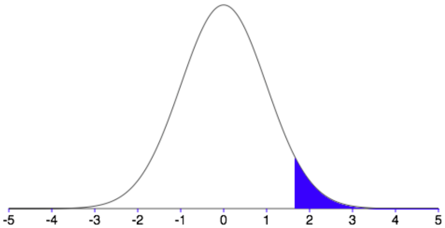


DIRECTIONAL HYPOTHESIS

- suppose we know (or strongly suspect) that if the sample mean $\overline{X}$ is different from the population mean μ, it will be **greater**
- then we don't need to worry about the left-side tail

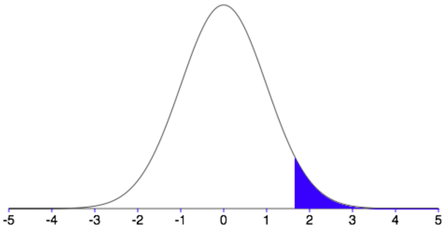
$$H_0 : \mu = 455$$

$$H_a : \mu > 455$$



REGION OF REJECTION

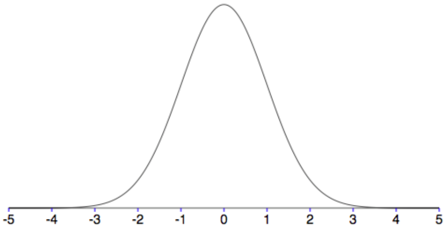
- if we only have to worry about one tail, the region of rejection (in that tail) is larger!
- with $df = 143$, last 5% starts with a t-score of 1.656
- we can reject H_0 when the difference between $\overline{X}$ and μ is smaller!



EXAMPLE

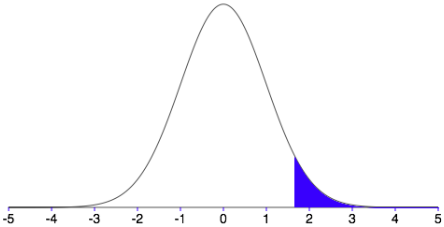
- we know that the sampling distribution of t is:
  - A t distribution with $df = 143$.
  - Has a mean of $\mu = 0$.
  - Has a standard error of the mean

$$s_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{100}{\sqrt{144}} = 8.33$$



REGION OF REJECTION

- area under the curve represents the probability of getting the corresponding t values, given that H_0 is true
- the extreme right tail of the sampling distribution corresponds to what should be very rare t values
- critical t-score value is 1.656



TEST STATISTICS

- we compute test statistic

$$t = \frac{\overline{X} - \mu}{s_{\overline{X}}}$$

$$t = \frac{535 - 455}{8.33} = 9.60$$

- greater than critical value

$$9.60 > 1.656$$

- reject H_0
- The same decision is found by computing the *p*-value

$$p \approx 0 < \alpha = 0.05$$

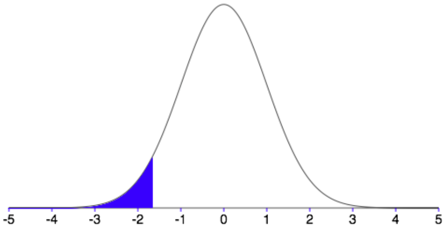
EXAMPLE

- suppose everything was the same, except we had hypotheses:

$$H_0 : \mu = 455$$

$$H_a : \mu < 455$$

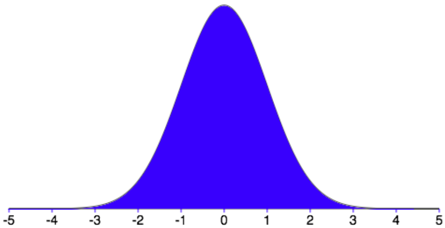
- then we would shift the region of rejection to the **left** tail



EXAMPLE

- the critical t-score value becomes -1.656
- with our sample mean of $\overline{X} = 535$, and $z = 9.60$,
- we **cannot** reject H_0

$$p \approx 1 > \alpha = 0.05$$



CALCULATOR

- you need to understand the math and calculations, but generally you should not **do** it

Enter data:

Sample size $n =$ 144

Sample mean $\overline{X} =$ 535

Sample standard deviation $s =$ 100

Specify hypotheses:

$H_0 : \mu =$ 455

$H_a :$ Negative one-tail

$\alpha =$ 0.05

Run Test

Test summary	
Null hypothesis	$H_0 : \mu = 455$
Alternative hypothesis	$H_a : \mu < 455$
Type I error rate	$\alpha = 0.05$
Sample size	$n = 144$
Sample mean	$\overline{X} = 535.0000$
Sample standard deviation	$s = 100.000000$
Sample standard error	$s_{\overline{X}} = 8.333333$
Test statistic	$t = 9.600000$
Degrees of freedom	$df = 143$
p value	$p = 1.000000$
Decision	Do not the reject null hypothesis
Confidence interval critical value	$t_{cv} = 1.976692$
Confidence interval	$CI_{95} = (518.527565, 551.472435)$

CONCLUSIONS

- hypothesis testing
- sample size
- directional test

NEXT TIME

- Designing experiments
- Power
- Selecting sample size

Plan ahead!