Ztest

Monday, July 29, 2024 9:05 PM

DAv3

HT - 1

- Type 1 and Type 2 error ✓
- Ho and Ha. ✓

Agenda

- HT → Implement
  - ↳ Ztest
    - Critical
    - confidence Interval

Default assumption → Null Hypothesis

P → Market ↑ sales - Claim → No change after Mater

No relationship

No virus

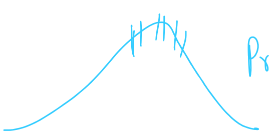
If there is a strong evidence → Reject Null hypothesis

↳ Alternate hypothesis.

Not guilty

P (guilty | innocent)

Claim → Murder (guilty)



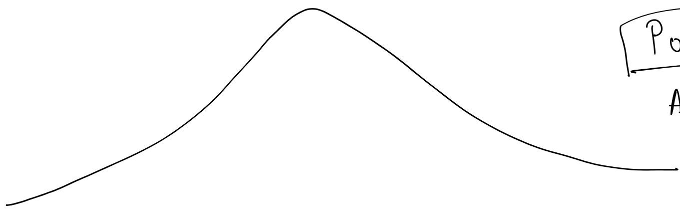
Prob
1. knife
2. Blood

} common Null

3. finger
4. Camera

} rare ↓ Prob low ↓ $\boxed{\begin{matrix} 0.05 \\ 0.01 \end{matrix}}$ reject Null hypothesis

Ztest - Central Limit Theorem



Population

Avg height – 65 inc
Std dev – 2.5 inch

std dev

The average height is 65 inches with std dev 2.5. We take a sample of 50 people.
Let "m" represent the sample mean. What distribution does "m" follow?

From <https://www.scaler.com/instructor/meetings/i/z-test-90/>

Sample – 50
m –

Normal

The average height is 65 inches with std dev 2.5. We take a sample of 50 people.
Let "m" represent the sample mean.
What is the average or expected value of "m"

From <https://www.scaler.com/instructor/meetings/i/z-test-90/>

65 – pop } 50 ~ Mean
2.5

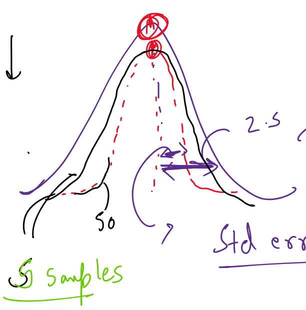
Average height is 65 inches with std dev 2.5. We take a sample of 50 people.
Let "m" represent the sample mean.
What is the standard deviation of "m"

From <https://www.scaler.com/instructor/meetings/i/z-test-90/>

$\mu = 65$
$\sigma = 2.5$ } 50 Samp

M = 50 samples
M = 5 sample

Thinn



Population distri

Std error = std dev = sample

$$\frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{50}}$$

Retail case study : 2000 Stores JIO
: Shampoo bottle

Mean – 1800 bottles per week
std – 100

Marketing   Marketing

Marketing A

Marketing B

Pay

⊙ Deploy this Marketing Stragut in 2000 stores

✓ ⊙ Deloy on small sample, verify the sales, then Scale

Pin ball

store

Company A

1860 — 2000

100 — std

Populati

50 stores } 1850 samples on average

? Marketing

Randomness in the data

SRNo	No of Bed	Price	Location
1	3	30L	R
2	7	95L	R
3	2.5	2CR	MUM

Randomness

Company B

5 sample → 1900 sales

{ Company A → sales → 1850 → Marketing

{ Company A →

sales → 1850 → Marketing

Sample → 50 → Randomness

Null → No effect of Marketing

Alternate → Marketing has significant effect

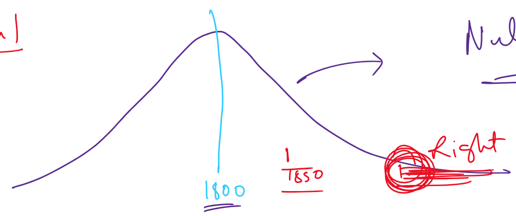
Null $H_0 = u_A = 1800$

Alternate → $\boxed{H_a = u_A > 1800}$

Claim 1800>

< left

> # two tail



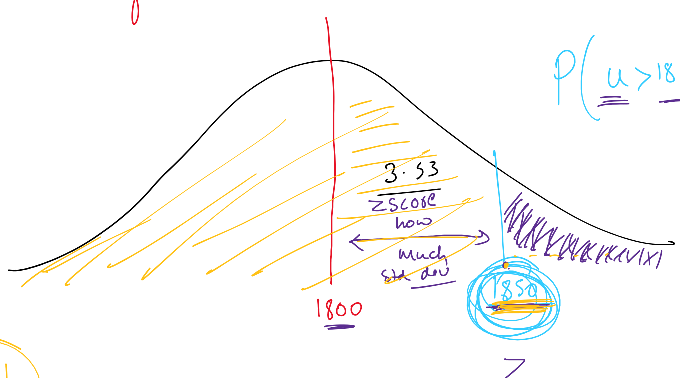
→ Null hypothesis

Data is observes

$P(M > 1800 \mid H_0 \text{ is true})$ Marketing No effect

$u > 1800$

Right Tail — one tail



$P(u > 1800 \mid H_0 \text{ is true})$

cdf

$$Z = \frac{1850 - 1800}{\text{std error}} \rightarrow \sigma / \sqrt{n} \quad 100/\sqrt{50}$$

$$= \frac{1850 - 1800}{100/\sqrt{50}} = 3.53$$

norm.cdf(3.53) = 0.0002

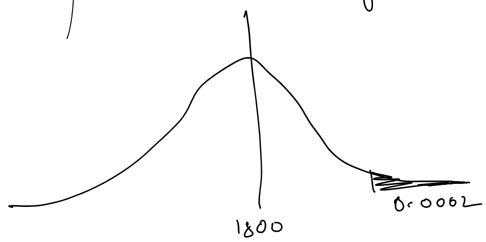
$1 - \text{norm.cdf}(3.53) = 0.0002$

$P(\text{guilty} \mid \text{innocent})$ } rare low

$P(u > 1850 \mid H_0 \text{ true})$ } rare low

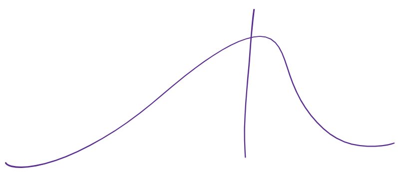
$\boxed{0.01}$

0.0002



1800 — 1850 — rare

~~Chance~~

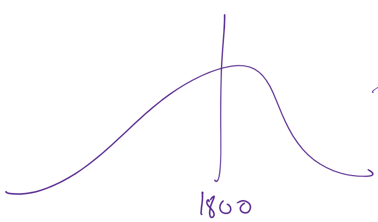


0.01

99% confidency

$p = 0.0002 < 0.01$ } Reject the Null hypothesis

~~No effect~~

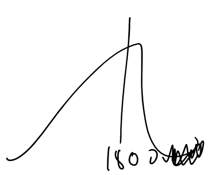


$u = 1800$

H_0

H_a

$P(> 1850 \mid H_0 \text{ is true}) = 0.0002$

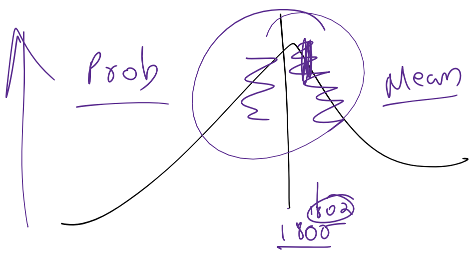


$0.0002 < 0.01$ 1%

Reject

Z test

1. z score → away
2. Prob → 0.002 norm.cdf
3. Prob < threshold reject null



$\leq$

Break

A fitness App claims that its users walk an average of 8,000 steps per day.

A random sample of 30 users showed an average of 7,600 steps per day with a standard deviation of 1,200 steps.

Conduct a left-tailed Z-test at a 5% significance level to determine if the App's claim is supported.

What is the p-value?

From <https://www.scaler.com/instructor/meetings/i/z-test-90/>

z-score: -1.8257418583505536 p-value: 0.033944577430914516 Reject the null hypothesis.

From <https://www.scaler.com/instructor/meetings/i/z-test-90/>

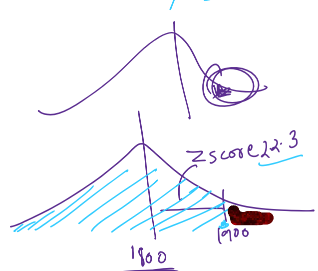
5 stores

M — 5 Stores

$SE \rightarrow \frac{100}{\sqrt{5}} — \frac{\sigma}{\sqrt{n}} = \frac{100}{\sqrt{5}}$

$u = 1900$

$Z\ score = \frac{1900 - 1800}{100/\sqrt{5}} = 2.23$



norm.cdf (2.23)

1 − norm.cdf (2.23)

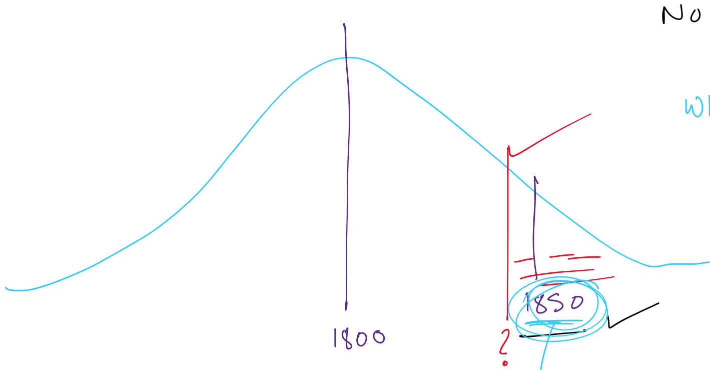
= 0.012

$\boxed{0.012 < 0.01}$ NO

$0.012 < 0.05$

$H_0 \rightarrow \boxed{1900}$ ✓

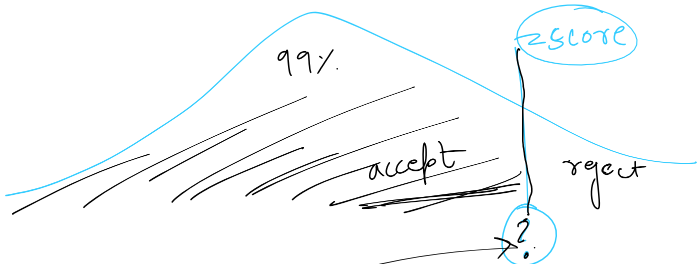
Marketing company No effect



What zcore will be there
$p = 0.01$

Zscore

$\frac{1850-1800}{100/\sqrt{50}} = 3.35$



norm.cdf(z) = 99%

norm.ppf (0.99) = $\boxed{2.32}$

Z score

$1800 + 2.32\left(\frac{100}{\sqrt{50}}\right) = \boxed{1832.8}$

upper limit

critical value

1850.

Team B

$u = 1800$

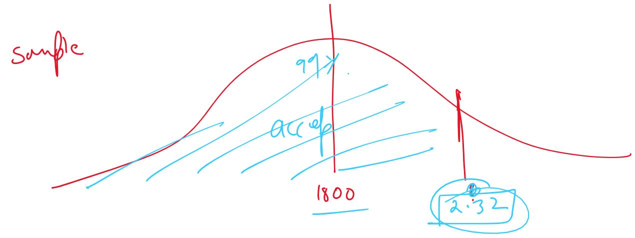
$u > 1800$

$\boxed{1\%}$

1900 ✓

1963 — 99%

95%



$$norm.ppf(0.99) = 2.32$$

$$1800 + 2.32 \times \frac{100}{\sqrt{5}} = \boxed{1903.7}$$

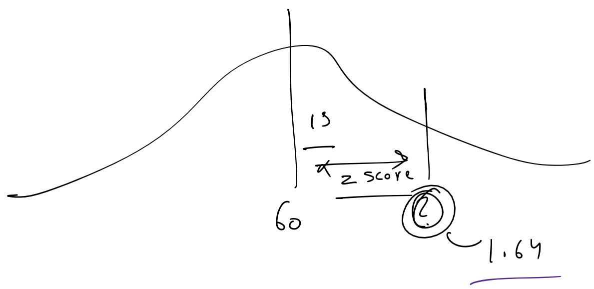
$\boxed{1903.7}$

In a dataset of exam scores with a mean of 60 and a standard deviation of 15, What is the critical value for the corresponding Z-score at a 95% confidence level?

From <https://www.scaler.com/instructor/meetings/i/ztest-90/>

$u = 60$

$\sigma = 15$



$$norm.ppf(0.95 = 1.64$$

$$60 + 1.64 \times (15)$$

$$= 84.6$$

P mean = 1800

sample = 1850

} so sample

pop std = 100

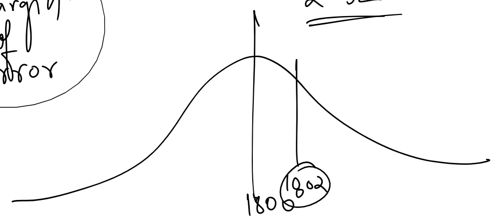
size = 50

alpha = 0.01

99%

$Z = norm.ppf(0.99)$

$2.32 \times$



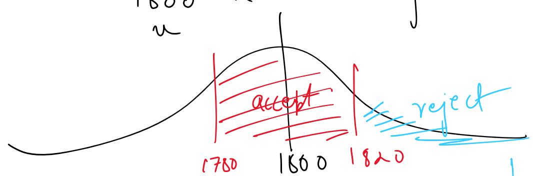
Margin of error

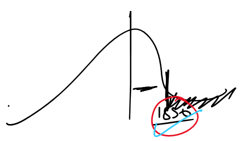
$\rightarrow$ Zscore $\times \left(\frac{\sigma}{\sqrt{n}} \right)$

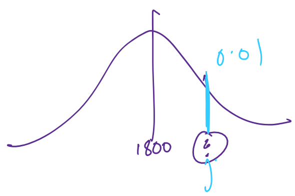
$$ME = 2.32 \times \frac{100}{\sqrt{50}}$$

20

μ
$1800 + 20$
$1800 - 20$
μ
} Confidence Interval



① 



① Zscore
② Prob
③ Prob < Threshold
Reject

Z test

find zsore

① norm.ppf(0.99)
$\boxed{\text{zscore}}$

② $\underset{\mu}{1800}$ + zscore × (std error)

critical

① Zscore
norm.ppf(0.99)

② Margin of error

Zscore × $\left(\frac{\sigma}{\sqrt{s}} \right)$
std

Zscore x $\left(\frac{\sigma}{\sqrt{s}}\right)$

std error

$u + ME$

$u - ME$

Z test continue

A country has
a **population average** height
of **65** inches with a **standard deviation** of **2.5**. A person feels people from his state are shorter. He takes the **average** of **20** people and sees that it is **64.5.**
At a **5%** significance level (or 95% confidence level), can we conclude that people from his state are shorter, using the Z-test? What is the p-va

From <https://www.scaler.com/hire/test/problem/53791/>

$u = 65$

$\sigma = 2.5$

$H_0 = u = 65$

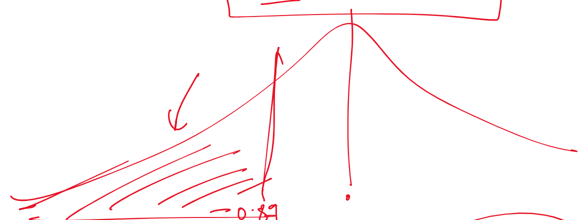
$H_a = u < 65$

$\alpha = 0.05$

sample = 20

(1) z value $= \dfrac{64.5 - 65}{2.5/\sqrt{20}} = \dfrac{ob - mean}{\sigma/\sqrt{n}}$

$\boxed{= -0.89}$



norm.cdf(-0.89) = 0.185

$\boxed{0.185 < 0.05}$ No

$H_0 \rightarrow$ People are not shorter

A fitness App claims that its users walk an average of 8,000 steps per day.

A random sample of 30 users showed an average of 7,600 steps per day with a standard deviation of 1,200 steps.

Conduct a left-tailed Z-test at a 5% significance level to determine if the App's claim is supported.

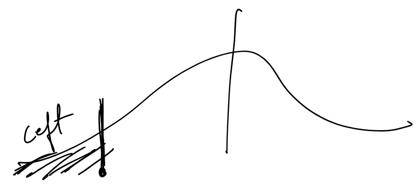
What is the p-value?

From <https://www.scaler.com/instructor/meetings/i/z-test-90/>

Claim → 8000

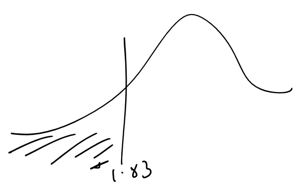
sample = 30 — 7600 (u)

1200'



$H_a < 8000$

~~$H_0 = \mu_0 = 8000$~~



0.05

$$Z = \frac{7600 - 8000}{1200/\sqrt{30}} = -1.83$$

$$\text{norm.cdf}(-1.83) = 0.033$$

$\boxed{0.03 < 0.05}$ reject

guilty → Positive

innocent (Not guilty) → Negative

Fp → False Positive

$\boxed{\text{False - wrong - Decision}}$ → Positive

guilty

Yogesh → ~~you are guilty~~

Yeqoh → you ~~are guilty~~

— innocent → ~~guilty~~ FP

α Type 1 error

FN → Falsly Negative

guilty → ~~inno~~ ← inncocent Type 2

β

H_0 → innocent

Positive — guilty

Negative — innocent.

$\boxed{\text{Type 1 } H_0}$ and $\boxed{\text{Type 2 } H_a}$

G

Q5 Assignment Questions (- Not Creditworthy should be considered Negative - Case A = Should be True Negative) but it is given as True Positive , why ?

From <https://www.scaler.com/instructor/meetings/i/z-test-90/>