17-08-2025

Agenda :— Statistics - 8

Z-statistics → better / worse / direction → Tail test (single - two tailed)

method - 1

How far it is ?

Standard Error

Hypothesis Testing (method to judge the difference using statistics)

method - 2

→ p-value

A — old — expected ↓ Null hypo

B — new — observed ↓ Alternate Hypo

single tail & two tail (significance values)

formula: $(1-\alpha)$

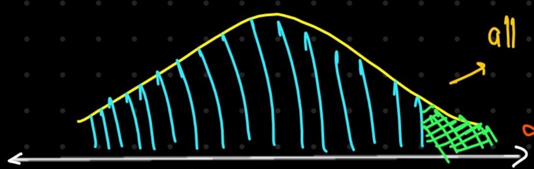
Z-stats: How far away is my sample result from what I expect under the null hypothesis.

$\alpha = 0.05$, 95 % confidence

One tail: → +ve

→ -ve

all in one tail



$\alpha = 0.05$

← 95% → 5%

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
+2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
+2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
+2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
+2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
+3	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
+3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
+3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
+3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
+3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
+3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983

$$1-\alpha = 1-0.05 = 0.95$$

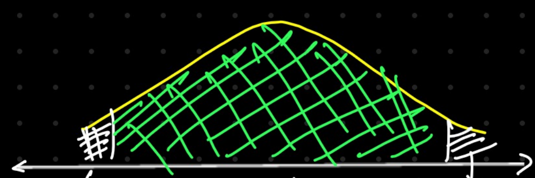
$$z_{(1-\alpha)} = z_{0.95} = 1.645$$

$$\frac{1.64 + 1.65}{2} = 1.645$$

Two tail :—

$1-\alpha$ → alpha is split into two tail

$\alpha = 0.05$ → $0.05/2 \rightarrow 0.025$



-ve 0.025 ← 95% → 0.025 +ve

$$z_{(1-\alpha)} = z_{(1-\alpha/2)} = z_{(1-0.025)}$$

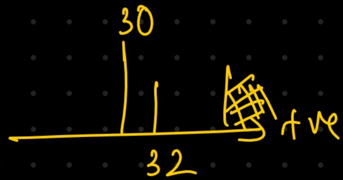
$$z_{(0.975)} = 1.96$$

Intuition of -ve, +ve, single & two tail.

Single tail: (directional hypothesis) (one direction)

population mean = 30
sample mean = 32

} prove mean is bigger



Example:

Mobile battery : Tech A : 24 hrs
: Tech B > 24 hrs
: +ve single tail

p-mean caffine : 200mg
s-mean : 197mg

-ve single tail

Reduce Inflammation: Med A : 15
: Med B < 15
: -ve single tail

Two tailed test : (Non-Directional Hypothesis)

: You care about effects in either direction.
: it could increase or decrease

Students (10) : Classes A : 75%.
↳ : Classes B : improving or not ?

-ve ⟷ +ve

Hypo testing :—

$\rightarrow$ Ho (original claim)

p.1 : pizza shop advertises — Average delivery time = 30 minutes.

A group of customers claims you're slower. $\rightarrow$ H_1 (Alternate Hypothesis)

$n = 36$ deliveries

sample mean = 31 minutes

population σ = 4 minutes

$\alpha = 0.05$ (95% confidence)

Step-1 : Hypotheses

$\rightarrow H_0 : \mu = 30$

$\rightarrow H_1 : \mu > 30$ (one-tailed test)

Step-2 : Significance level

$\alpha = 0.05 \qquad z_{(1-\alpha)} = z_{(1-0.05)} = z_{(0.95)} = 1.645$

Step-3 : Standard Error (SE):

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{36}} = \frac{4}{6} = \frac{2}{3} = 0.6667$$

Step 4:

$$z\text{-score} : \frac{\bar{x} - \mu}{SE} = \frac{31 - 30}{0.6667} \approx 1.4992 \approx 1.5$$

our z-score = 1.5

critical z = 1.645

$1.5 < 1.645$ = Fail to reject H0

→ There isn't enough evidence at 5% significance.
→ we are not able to prove that pizza shop is slower that 30 minutes.

Rough notes on $\alpha = 0.05, 0.01$. when to pick which α?

95% 5% → 80%
0.05
0.01
↓
inurance
CEO → 5% 92, → 80%
1% 2%, 2.5, 0.5
95%
50%
98
99%

Test cancer → B → 0.01 → 99% → P → P →
→ B
→ A → 0.05 →
→ 95% → P → P
P → F

TP, TN, FP, FN (confusing matrix element)

Element	Meaning	Dog-count Actual	Cat-count Actual	Dog-count predicted	Cat-count predicted
True positive (TP)	present, identified	0	1	0	1
True Negative	not present, not identified	0	0	0	0
False Positive	not present, but identified	0	1	1	1
False Negative	present, but not identified	1	1	0	1

	Actual	predict
TP	Car	Car
FP	None	car
TN	None	None
FN	car	None

p-value :-

The p-value is the prob of getting the observered data assumes the null hypothesis is true.

small-p value :- $\leq \alpha$ → significance value

: Reject H_0 → Evidence supports H_1

Large p-value ($> \alpha$)

→ Fail to reject H_0 → Not enough evidence for H_1

p.) : claim Average battery life = 10 hrs

sample : 30 batteries

sample mean : 9.6 hrs

population σ : 1.2 hrs

$\alpha = 0.05$

z stats : SE : $\frac{\sigma}{\sqrt{n}}$

z-score : $\frac{\bar{x} - \mu}{SE} \rightarrow \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

z-score = $\frac{9.6 - 10}{1.2/\sqrt{30}} \approx -1.83$

1.8257

1.8260

1.83

from z-table $P(Z \leq \underbrace{-1.83}_{\text{z-table}}) \approx \underline{0.0336}$ → p-value

compare: p-value with α

$0.0336 \leq 0.05$ → True → Reject H_0

→ Battery life is statistically significantly less than 10 hrs.

Method	Decision Rule
z-score vs critical z	if $z_{stat} > z_{critic}$ ($\downarrow$ $(1-\alpha)$), reject H_0
p-value	if $p \leq \alpha$; p ↓ z_{stat} ↓ z-table; α ↓ single tail ↓ 0.05

z-value : — pass/fail $z_{score} >$ $z_{critical}$ → pass/fail

X strenght of pass & fail

p-value :

- quantity strenght of evidence.

z-value :

$\alpha = 0.05$ two-tail test

zscore = 2.1 compare with 1.96 → reject H_0
→ statistically significant at 5%.

$\alpha = 0.01$
zscore = 2.1 compare with 2.58
→ not enough evidence to reject H_0
→ not stat significant

p-value : — Instead of saying yes or no, we will get a degree of evidence,

(p-value) → 0.0336 tells the exact prob →
compare with any threshold of α (5%, 1%, 10%)

✓ $0.0336 \leq 0.05$
$0.0336 \leq 0.01$
$0.0336 \leq 0.1$

z-stats → Red / Green traffic light
p-value → speedometer → fast we are going

p-value is close to 0 → small p
p-value is close to 1 → large p.

$p \leq 0.05$ (evidence against null)

$p > 0.5 \leq 0.1$ (weak evidence , "maybe"

$p > 0.1 \rightarrow$ no evidence against H_0

|2| → A B C

30 min

↓

31 → 42

34 - 42

reject H_0, fail to reject

p

A B C

↓

strenght ← 0.X 0.X

0.02

0.03

2%, chance result are due to randomness

3% chance result are due to random

>

Doctor :— negative | positive → 2-stars

↓

artificial cutoff

$p <<$ less the error 5.1% | 4.9%

↓

prob of error

(A) (B)

0.1 | 0.6

P.1 : pizza shop advertises – Average delivery time = 30 minutes. → H0

A group of customers claims you're slower. → H_1 (Alternate Hypothesis)

$n = 38$ deliveries

sample mean = 31 minutes

population σ = 4 minutes

$\alpha = 0.05$ (95% confidence)

$Z_{0.95} = 1.645 \rightarrow (.94950)$

p.value :

$(1 - .94950)$

↓

0.0888

$0.0888 > 0.05 \rightarrow$ yes

→ failed to reject H0

-ve → as it is ← zscore

p-value ←

+ve → (1 - value) ← zscore

Note:

$P(\text{observed})$ α (single tail)

$2 \times P(\text{observed})$ α (double tail)

Two-tailed test: average $\neq$ 30 minutes

$Z_{0.975} = 1.96$ (more extreme than 1.645)

$Z_{score} = 1.5$ $\quad$ $1.5 > 1.96$ → fail to reject.

p-value: $\quad 2 \times (P(1.5)) = 0.1336 > 0.05$ fail to reject.

$1.5 \to 0.06681$

$\times 2 = 0.13362$

single tail test.

$Z(0.95) = 1.645$

$Z_{score} = 1.5 > 1.645$ → fail to reject.

p-value $= 0.0668 > 0.05$ → fail to reject.