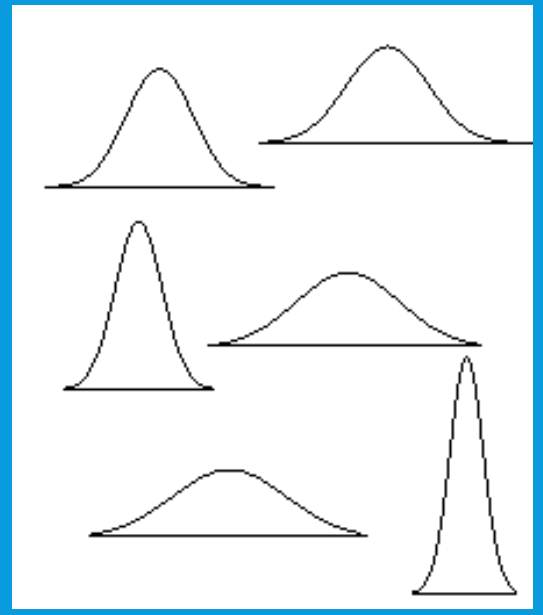
RESEARCH METHODOLOGY

z-test and t-test

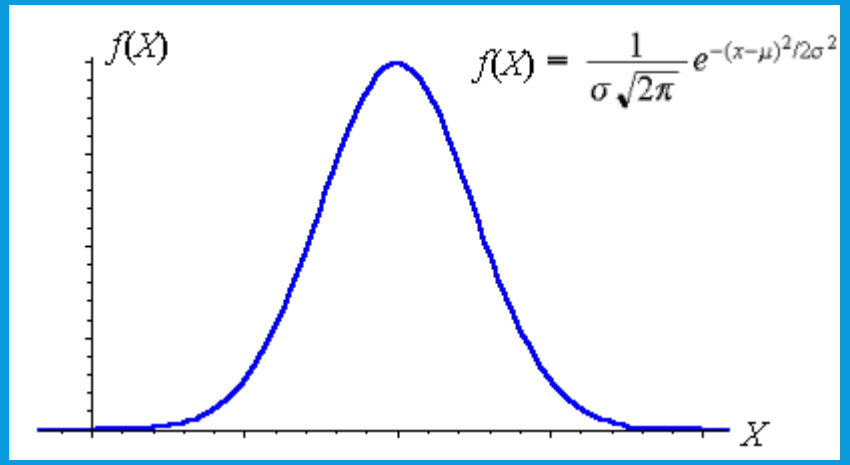
NORMAL DISTRIBUTION

- $N(\mu, \sigma)$

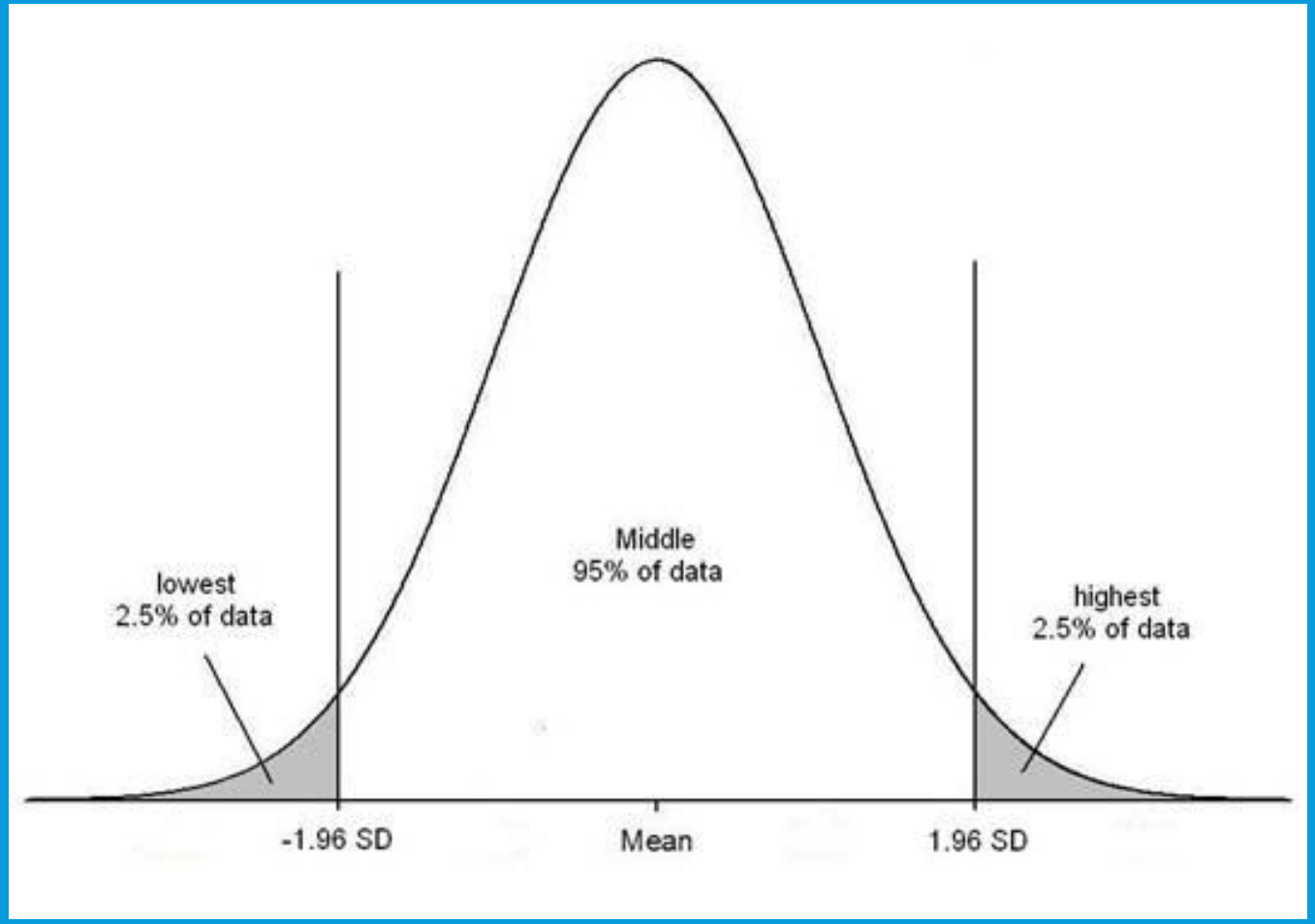


Standard Normal Distribution

- $N(0,1)$



90%, 95%, 99%


lowest
2.5% of data
Middle
95% of data
highest
2.5% of data
-1.96 SD
Mean
1.96 SD

STANDARD NORMAL DISTRIBUTION

- $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

- $z = \frac{x-\mu}{\sigma/\sqrt{n}}$
- $t = \frac{x-\mu}{s_1/\sqrt{n}}$
- $s_1 = \frac{\sum(x_i-\bar{x})^2}{n-1}$

INTERVAL ESTIMATION

- Let θ be the unknown parameter and let T be the unbiased point estimate E(T)= θ
- Now, fix a desired level of confidence denoted as
- (1-α) × 100 % . Here 'α' is the probability of type –I errors.
- Usually , confidence level is fixed as 90%, 95 %, or 99%.
- Thus , for 95 % confidence level α = 0.05.
- The interval estimate of θ is T-h, T+h
- $P(T - h \leq \Theta \leq T + h) = 1 - \alpha$

CRITICAL VALUES FOR SOME COMMONLY USED LEVELS

Confidence Level	Confidence Coefficient	alpha	Critical Value
80%	0.8	0.2	1.28
90%	0.9	0.1	1.645
95%	0.95	0.05	1.96
98%	0.98	0.02	2.33
99%	0.99	0.01	2.58

FORMULAE

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Z-TESTS REQUIRE ONE OF TWO CONDITIONS: EITHER THE POPULATION IS NORMALLY DISTRIBUTED WITH A KNOWN VARIANCE, OR THE SAMPLE SIZE IS LARGE.

- For small n

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right)\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

EXAMPLE 9.1 (KOTHARI)

- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is 0.35 ohms. Determine a 95% confidence interval for the true mean resistance of the population.

- Caution: Even though $n < 11$, but population's standard deviation is given.

Solution — Example 9.1 (σ known → z)

$z = (\bar{x} - \mu) / (\sigma/\sqrt{n})$; for 95%: $z = 1.96$

$\mu = 2.20 \pm 1.96 \times 0.35/\sqrt{11} = 2.20 \pm 0.21$

CI: 1.99 Ω to 2.41 Ω

EXAMPLE 9.2 (KOTHARI)

- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. Population standard deviation is not known. Sample standard deviation (s_1) is 0.35 ohms. Determine a 95% confidence interval for the true mean resistance of the population.

EXAMPLE 9.9 (KOTHARI)

- From a random sample of 36 New Delhi civil service personnel, the mean age and the sample standard deviation were found to be 40 years and 4.5 years respectively. Construct a 95% confidence interval for the mean age of civil servants in New Delhi.

EXAMPLE 9.11 (KOTHARI)

- The foreman of a company has estimated the average quantity of iron ore extracted to be 36.8 tons per shift and the sample standard deviation to be 2.8 tons per shift, based upon random selection of 4 shifts. Construct a 90 % confidence interval around this estimate.

TESTING OF HYPOTHESIS

- $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

- $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

EXAMPLE 10.5 (KOTHARI)

- In a survey of buying habits 400 women shoppers are chosen at random in supermarket 'A'. Their average weekly food expenditure is Rs 250 with a standard deviation of Rs 40. For 400 women shoppers chosen at random in another supermarket 'B', the average weekly food expenditure is Rs 220 with a standard deviation of Rs 55. Do these two populations have similar shopping habits? Is the average weekly food expenditure of two populations of shoppers equal? Test at 5% level of significance.

EXAMPLE 10.6(KOTHARI)

- You are a finance analyst for a brokerage firm. Is there a difference in dividends by NYSE and NASDAQ ? You collect the following data:
- Assuming both populations are approximately normal with equal variances, is there a difference in average yield?

Number of stocks	21	25
Sample Mean	3.27	2.53
Sample Std. Deviation	1.30	1.16

A PROFESSOR IS MARKING THE TEST OUT OF 60. HERE ARE THE RESULTS:

20 15 26 32 18 28 35 14 26 22 17

Most of the students did not get 30 out of 60 and most will fail. The test must have been really hard. The professor decides to standardize and detain only the students who are 1 standard deviation below the mean. Which of the students will be detained?

$\mu = 23,\ \sigma = 6.6$

Score	Standard Score
20	-0.45
15	-1.21 (Detain)
26	0.45
32	1.36
18	-0.76
28	0.76
35	1.82
14	-1.36(Detain)
26	0.45
22	-0.15
17	-0.91

INTERPRETING Z-VALUES — Z = 1.96

- Hypothesis: the mean score of a group is different from the national average.
- $z = 1.96 \rightarrow$ two-tailed $p = 0.05$.
- At $\alpha = 0.05$: $p = 0.05$ sits exactly at the threshold — reject H_0; the difference is statistically significant.
- Meaning: a 5% chance of a result at least this extreme if H_0 were true.

Z = 2.58 — STRONG EVIDENCE

- Hypothesis: the mean height of a sample differs from the population average.
- $z = 2.58 \rightarrow$ two-tailed $p = 0.01$.
- Only a 1% chance of a result this extreme under H_0 — strong evidence; reject H_0 even at the 1% level.

Z = 1.28 — NOT ENOUGH EVIDENCE

- Hypothesis (one-tailed): employees work more hours than the national average.
- $z = 1.28 \rightarrow$ one-tailed $p = 0.10$.
- $p = 0.10 > \alpha = 0.05 \rightarrow$ fail to reject H_0: insufficient evidence that hours are higher.

Z = 0 — NO EVIDENCE AT ALL

- Hypothesis: a medication changes blood pressure vs a placebo.
- $z = 0 \rightarrow$ the sample mean equals the population mean exactly; $p = 1.0$.
- No evidence against H_0 — fail to reject.

Z = −2.33 — SIGNIFICANT, IN THE OTHER DIRECTION

- Hypothesis (one-tailed): mean weight loss in the sample is LESS than in the population.
- $z = -2.33 \rightarrow$ one-tailed $p = 0.01$.
- $p = 0.01 \leq \alpha = 0.05 \rightarrow$ reject H_0. The sign of z tells the direction; its size tells the strength.

EXAMPLE — IS THE CHEMICAL TOXIC TO BEETLES?

- A researcher suspects a chemical interferes with beetle reproduction. One group is fed the chemical; a control group is not. After 2 weeks she counts the eggs produced per beetle.

Group 1 (fed chemical): 33, 31, 34, 38, 32, 28

Group 2 (control): 35, 42, 43, 41

- H_0: the chemical has no effect on mean egg count.
- H_1: the chemical changes the mean egg count.
- Two small independent samples, σ unknown ⟶ two-sample t-test with pooled variance.

BEETLES — THE WORKED T-TEST

- $\bar{x}_1 = 32.67$, $s_1^2 = 11.07$ ($n_1 = 6$); $\bar{x}_2 = 40.25$, $s_2^2 = 12.92$ ($n_2 = 4$)
- Pooled variance: $s_p^2 = (5 \times 11.07 + 3 \times 12.92) / 8 = 11.76 \rightarrow s_p = 3.43$
- $SE = 3.43 \times \sqrt{(1/6 + 1/4)} = 2.21$

$t = (32.67 - 40.25) / 2.21 = -3.43$, $df = 8$

- Critical t ($\alpha = 0.05$, two-tailed, $df = 8$) = 2.306.
- $|-3.43| > 2.306 \rightarrow$ reject H_0 ($p \approx 0.009$).

The chemical significantly reduces egg production.