

**Z score,  
Z-test,  
Standard Error**

# Relative or standard normal deviate Z

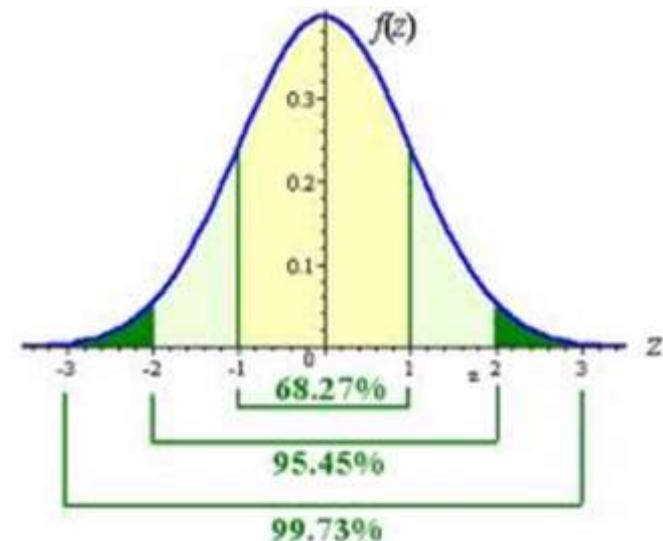
- Deviation from the mean in normal distribution is called relative or standard normal deviate Z.
- Indicates how much an observation (x) is bigger or smaller than mean in units of SD.

$$Z = \frac{X - \bar{X}}{SD} = \frac{\textit{Observation} - \textit{Mean}}{SD}$$

**E.g.** Height distribution, Observation (x)-165 cm,  
Mean-160cm, SD-5 cm

- Means 165 cm observation lied at 1 SD above the mean.
- In other words probability of height above 165 cm is 16%, ( $P= 0.16$ ).

To find the proportion of the individuals who will exceed any particular observation(X) in standard universe, refer to Z value.



# Z-test

- When Z test is applied to sampling variability where samples are larger than 30.
- Means of samples in population will also follow normal distribution.
- So, observer difference is calculated in terms of SE instead of SD.

$$Z = \frac{\bar{X} - \mu}{SE \text{ of mean}}$$

=

Observed difference between sample mean and population mean  
SE of mean

# Prerequisites of applying Z-test

1. The sample must be randomly selected
2. The data must be quantitative
3. The variable is assumed to follow normal distribution.
4. The sample size must be **larger than 30**.

# Standard Error

# SE of Mean

- Population parameter- $\mu$  (mean),  $\sigma$ , P
- Sample statistics-  $\bar{X}$ , s, p
- In most of the cases sample statistics will differ from the population parameters even if care taken while the selecting the samples because of chance or biological variability.
- Such, Difference b/w sample and population values is measured by statistics known as sampling error or standard error (SE).

- Standard error is measure of chance variation and it dose not mean ~~error of mistake~~.

- SE of mean

$$SE_{\bar{x}} = \frac{SD}{\sqrt{N}}$$

- Greater the SD greater will be the SE as occur in small samples.
- Chance variation SE ↓  
if we reduce the SD ↓  
can be done only by ↑ the sample size.

- E.g height of 190 cm in sample of 10 and height of 190 cm in sample of 100.
- In which sample SD will be more?
- Mean and SD will be influenced more in smaller sample as compared to larger sample size.
- So as SE will also reduced with increased sample size.
- If sample is increased 100 times, the SE is reduced to almost  $1/10^{\text{th}}$ .

- height of 190 cm in sample of 10
- Mean  $\uparrow$
- SD  $\uparrow$
- SE  $\uparrow$

- height of 190 cm in sample of 100.
- Mean  $\downarrow$
- SD  $\downarrow$
- SE  $\downarrow$

## Application and uses of SE of mean

1. To work out the limits of desired confidence within which the population mean would lie.
2. To determine whether the sample is drawn from a known population or not? When it population mean ( $\mu$ ) is known.

If sample mean ( $\bar{X}$ ) is larger than the known population mean ( $\mu$ )  $\pm 1.96$  SE, 95% chances are that the sample is not drawn from the same population.

Probability of difference observed by Z test

$$Z = \frac{\bar{X} - \mu}{SE}$$

### **3. To find the SE of difference between two means.**

To know if the observed difference between the mean of two samples is real and statistically significant. Or it is apparent and insignificant due to chance.

### **4. To calculate the size of sample** in order to have desired confidence limits.

e.g. systolic BP of 566 males. Mean-128mm, SD-13.05 mm.

$$SE = \frac{SD}{\sqrt{n}} = \frac{13.05}{\sqrt{566}} = 0.55$$

$$\text{Mean} \pm 2SE = 128.8 \pm 2(0.55)$$

So, Range will be 129.9 to 127.7

- Means Confidence limit for population mean will be between the given range in 95% of cases.
- We could be wrong In 5% cases(5 out of 100 cases).
- If mean BP of another sample is beyond this range sample is unlikely to be drawn from the same universe or some external factor is working on the samples.

- E.g In large population sample of children  
mean Height=66cm, SD= 2.7 cm,

Can a sample of 100 with mean height of 67 cm occurs easily? Find the probability. And what dose is indicate?

# STANDARD ERROR OF DIFFERENCE BETWEEN TWO MEANS

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{SE(\bar{X}_1 - \bar{X}_2)}$$

# STANDARD ERROR OF DIFFERENCE BETWEEN TWO MEANS

$$SE (\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{SD1^2}{n1} + \frac{SD2^2}{n2}}$$

# Application of SE of difference

- To compare the means of normally distributed variable in the two like or unlike group such as of Ht, Wt, BP, Pulse rate
- To check the action of drug on BP, Pulse etc.
- The action of two different drugs or two different dose of same drugs can be compared

# Steps in Test of Hypothesis

1. Determine the appropriate test
2. Establish the level of significance
3. Formulate the statistical null hypothesis
4. Calculate the test statistic
5. Compare computed test statistic against a tabled/critical value

<b>Confidence</b>	<b>Z</b>
90%	1.645
95%	1.960
99%	2.576
99.9%	3.291

# Determine The Hypothesis: Whether There is an Association or Not

- Write down the **NULL HYPOTHESIS** and **ALTERNATIVE HYPOTHESIS** and set the **LEVEL OF SIGNIFICANCE**.
- $H_0$  : The two variables are independent
- $H_a$  : The two variables are associated
- We will set the **level of significance at 0.05.**

# For Example

- *Some null hypotheses may be:*
  - ‘there is no relationship between the smoking of and the common cold’.
  - ‘there is no difference in the location of superstores and small grocers shops’
  - ‘there is no connection between the size of farm and the type of farm’

# EXAMPLE

In a nutritional study, 100 children were given usual diet and 100 were given vitamin A and D tablets. After 6 months, average weight of group A was 29kg with SD of 1.8kg and average weight of group B was 30kg with SD of 2kg. Is the difference is significant?

$$SD1 = 1.8$$

$$n1 = 100$$

$$SD2 = 2$$

$$n2 = 100$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{SE(\bar{X}_1 - \bar{X}_2)}$$

# STANDARD ERROR OF DIFFERENCE BETWEEN TWO MEANS

$$SE (X1 - X2) = \sqrt{\frac{SD1^2}{n1} + \frac{SD2^2}{n2}}$$

# STANDARD ERROR OF DIFFERENCE BETWEEN TWO MEANS

$$SE (X_1 - X_2) = \sqrt{\frac{(1.8)^2}{100} + \frac{(2)^2}{100}}$$

# STANDARD ERROR OF DIFFERENCE BETWEEN TWO MEANS

$$SE (X1 - X2) = \sqrt{\frac{3.24+4}{100}}$$

# STANDARD ERROR OF DIFFERENCE BETWEEN TWO MEANS

$$SE (X_1 - X_2) = \sqrt{0.0724} = 0.27$$

$$Z = \frac{30 - 29}{0.27} = 3.7$$

- As obtained value of  $z$  (3.7) is higher than critical value (-1.96 or -2.58), the observed difference is highly significant, vitamins played a role in weight gain .

$z$	1.96	2.58
$p$	0.05	0.01

- Eg.2 In a study on growth of children, one group of 100 children had a mean ht of 60 cm and SD of 2.5 cm while another group of 150 had a mean ht of 62 cm and SD of 3 cm.
- Is the difference between the two groups statistically significant?

# STANDARD ERROR OF DIFFERENCE BETWEEN TWO PROPORTIONS

**Data is Qualitative**

$$Z = \frac{P_1 - P_2}{SE (P_1 - P_2)}$$

# STANDARD ERROR OF DIFFERENCE BETWEEN TWO PROPORTIONS

$$SE (P1 - P2) = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$$

# EXAMPLE

If swine flu mortality in one sample of 100 is 20% and in another sample of 100 it is 30%. Is the difference in mortality rate is significant?

$P_1 = 20$	$q_1 = 80$	$n_1 = 100$
$P_2 = 30$	$q_2 = 70$	$n_2 = 100$

$$Z = \frac{P_1 - P_2}{SE (P_1 - P_2)}$$

# STANDARD ERROR OF DIFFERENCE BETWEEN TWO PROPORTIONS

$$SE (\underline{P1 - P2}) = \sqrt{\frac{20 \times 80}{100} + \frac{30 \times 70}{100}}$$

# STANDARD ERROR OF DIFFERENCE BETWEEN TWO PROPORTIONS

$$\begin{aligned} \text{SE } (\underline{P1 - P2}) &= \sqrt{37} \\ &= 6.08 \end{aligned}$$

# EXAMPLE

If swine flu mortality in one sample of 100 is 20% and in another sample of 100 it is 30%. Is the difference in mortality rate is significant?

$$\begin{array}{lll} P_1 = 20 & q_1 = 80 & n_1 = 100 \\ P_2 = 30 & q_2 = 70 & n_2 = 100 \end{array}$$

$$Z = \frac{20-30}{6.08} = -1.64$$

- Obtained z value (1.64) is less than critical z value (1.96) , so  $P > 0.05$ , hence difference is insignificant at 95 % confidence limits.

z	1.96	2.58
p	0.05	0.01

- E.g in school A , tonsillectomy had been done in 23 students out of 50

In school B it was done in 77 out of 350.

Find if the difference observed in the two schools is by chance or due to some influence?

- Whooping cough in school A 25 cases  
Whooping cough in school B 215 cases