Q3 Statistics

x_i = Blood pressure Before
y_i = Blood pressure After
n = total observations in each x_i & y_i = 100 ⇒ $\boxed{n=100}$

(a) Measure Dispersion in each and interprete the result.

→ Measure of Dispersion for Blood pressure Before (x_i)

① Range

$(x_i)_{max} = 148$, $(x_i)_{min} = 120$

Range = maxval − minvalue
= 148 − 120 = 28

$\boxed{Range_{x_i} = 28}$

② variance

$$\boxed{\sigma^2 = \sum_{i=1}^{N} \frac{(x_i-\mu)^2}{N}}$$

using Excel

$\boxed{\sigma_x^2 = 43.53}$

③ Std. deviation

$\boxed{S.D._x = \sqrt{\sigma_x^2} = 6.59}$

– using Excel

→ Measure of Dispersion for Blood pressure After (y_i)

① Range

$(y_i)_{max} = 141$; $(y_i)_{min} = 118$

Range = 141 − 118
= 23

$\boxed{Range_y = 23}$

② variance

using Excel

$\boxed{\sigma_y^2 = 47.44}$

③ Std deviation

$\boxed{S.D._y = 6.88}$

note

- There are other Measures of Dispersions also, there are many but these are the common measures of Dispersions

- As both data have similar units that's why Absolute Measure of Dispersion & is considered. Relative measure of Dispersion will not provide any additional info in this case.

Interpretation :

"After Blood pressure" can be cause of anything like After some Activity, Medication, Meal etc. In our interpretation we will refer it as an "Event" because we have no clue about it.

① Before the Event the max^m value & min value observed among Individuals are 148 & 120 and After the Event the max^m & min values are 141 and 118. we can clearly see the Dec in max^m & min value After the event.

② Corresponding to the respective Data Before and After the Range. before the Event was 28 and After 23 (or less spread out) it become more narrower

③ The inc in values of std. deviation & variance clearly indicate increase increase in variability of Blood pressure after the event.

ⓑ Calculate mean and 5% CI and plot it in a graph.

Before

$n = 100$

$\bar{x} = 133.91$

CI = 5%

$S = 6.598 \simeq 6.60$

$\alpha = 95\%$

} using Excel

After

$n = 100$

$\bar{y} = 128.36$

CI = 5%

$S = 6.88 \simeq 6.8$

$\alpha = 95\%$

} using Excel

• we will use Z-Distribution in calculation of lower & Higher CI. Because as sample size increases T-Dist approaches Z-Dist. So std. dev. will be considered as std. dev. of population.

lower CI = Point Estimate − margin of Error

$= 133.91 - Z_{\frac{0.95}{2}} \frac{\sigma}{\sqrt{n}}$

$= 133.91 - 0.06 \times \frac{6.6}{10}$

$\boxed{= 133.87}$

Higher CI $= 133.91 + 0.06 \times \frac{6.6}{10}$

High CI = 133.94

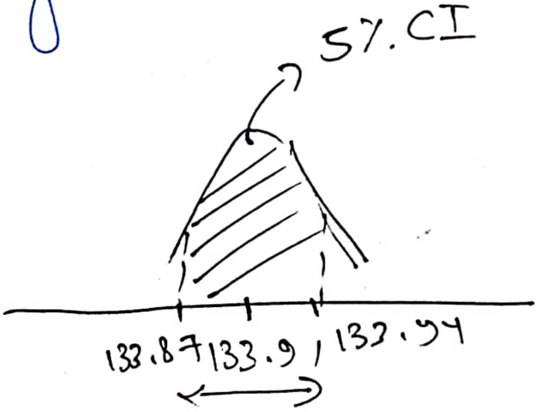
lower CI = Point Estimate − margin of Error

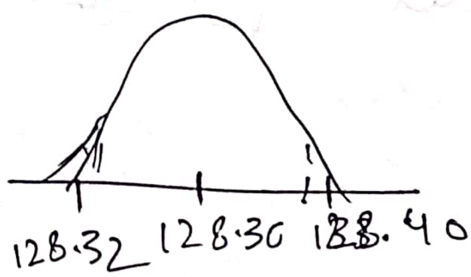
$= 128.36 - Z_{\frac{0.95}{2}} \times \frac{6.8}{10}$

$\boxed{\text{lower CI} = 128.32}$

High CI $= 128.36 + Z_{\frac{0.95}{2}} \times \frac{6.8}{10}$

$\boxed{\text{High CI} = 128.40}$





ⓒ Calculate the Mean absolute deviation and standard deviation and interpret the results.

Mean absolute Deviation Before = 0	Mean Absolute Deviation After = 0

– calculated using Excel

Standard deviation Before = 6.6

Standard deviation After = 6.8

ⓓ Calculate the ~~covariance~~ correlation coefficient and check the significance of it at 1% level of significance.

r = Sample correlation coefficient. (calculated)

n = Sample size = 100

ρ = Population correlation coefficient (unknown)

using Excel : $\boxed{\text{Pearson correlation coefficient } (r) = 0.97}$

Theory (Approach)

— Testing significance of correlation coefficient.

- Hypothesis test of the "significance of the correlation coefficient" is done to decide wether the linear relationship in the sample data is strong enough to use for the population data.

- we have no population data so we cannot calculate population correlation coefficient. But we have only sample data, The sample correlation coefficient r is our estimate (point estimate) of the unknown population correlation coefficient (as parameter)

Significantly different from zero

— means: correlation coefficient is "significant".

Conclusion: There is significant evidence to conclude that there is significant linear relationship b/w x and y because correlation coefficient is significantly different from zero.

- Linear regression line can be used to model the linear relation b/w x & y in the population

- or

Hypothesis test

Close to zero

— means: correlation coefficient is "not significant"

Conclusion: insufficient evidence to conclude that there ~~is sufficient~~ There is significant relation b/w x & y because the correlation is not significantly different from zero.

- or

- we cannot use regression line to model a linear relation b/w x & y in the population.

$r = 0.97$

ρ = Pop. correlation Coeffint

$\alpha = 1\%$ $\therefore$ CI = 99%.

Null Hypothesis $H_0: \rho = 0$ (there is no significant correlation b/w ... in pop)

Alternate Hypothesis $H_1: \rho \neq 0$ (there is significant correlation b/w ... in pop)

Calculating t-statistics (p-value is calculated using a t-Dist. with n-2 DOF)

$$\boxed{t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}}$$

$$= \frac{0.97\sqrt{100-2}}{\sqrt{1-0.97}}$$

$$\boxed{t = 55.44}$$

critical t-value for 2 tailed test at 1%. significance with n-2 DOF which is 100-2=98

Dof = n-2 = 98