AP Ashrafi NIGAR Sir

# Statistics

29 AUG 23

The main purposes of statistics are classified under three headings.

• Description • Analysis & • Prediction.

Individual data are not important but they are considered as a means to measure a certain physical property of interest, the test of hypothesis on the prediction of future occurances under given conditions.

Whatever the final objective of the experiment statistical methods of "inductive inference" in which a particular set (or sets) of data — "the sacalled realization of the samples" is used to drao inferences of general nature on a "population" under study.

Inductive inference drawn from incomplete information may be wrong even if the original information is not. In the field of statistics, the positivity is after related to the process of data collection on the other hand to the fact that we can only make probabilistic statement predictions. It is evident that insufficient on biased data or/and the failure to consider an important influencing factor in the experiment may lead to incorrect conclusion.

# 5 essential steps to solve problems: (in industry)

1. State the problem on question.
2. Collect and Analyze data
3. Interpret data and make decision.
4. Impliment and verify decision.
5. Plan next action.

(Scheaffer) পড়বে।

Data ⟶ Qualitative (categorical)
└⟶ Quantitative (numerical)

(*) Data in the form of numerical measurements or counts are refered to as quantitative.

(*) Data in the form of classification into different groups or categories are referred to as quantitative data.

# Quantitative কে Qualitave এ convert করা যায়, Lets 70-80% হচ্ছে Good weather. এখানে Qunatitative কে Qualitative এ convert করলাম

· Represulation: ① Graphical
② 

Measure of centre:
Mean / Averave / ~~median~~:

Median: মধ্যক: ascending order এ সাজালে middle point হচ্ছে median

$$l = \frac{(n+1)}{2}$$ (the dataset should be in ordered)

$$\bar{x}_m = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2},$$

mean: গড়
median: মধ্যমা
Mode: সর্বাধিক

Relative humidity of 30 days of a (05 Sep 23)

(b)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 60 | 63 | 64 | 71 | 67 | 73 | 79 | 80 | 83 | 81 |
| 86 | 90 | 96 | 98 | 98 | 99 | 89 | 80 | 77 | 78 |
| 71 | 79 | 76 | 84 | 85 | 82 | 90 | 78 | 79 | 79 |

Taking 2nd Table:

| x |
|---|
| 77 |
| 78 |
| 80 |
| 86 |
| 89 |
| 90 |
| 96 |
| 98 |
| 98 |
| 99 |

∴ range = (99 − 77) = 22

$$\text{Mean} = \frac{\sum_{i=1}^{10} x_i}{n}$$

$= n = 10.$

$$\text{Median} = \frac{89+90}{2} =$$

Mode = 98.

| x | fx |
|---|---|
| 77 → | 1 |
| 78 → | 1 |
| 80 → | 1 |
| 86 → | 1 |
| 89 → | 1 |
| 90 → | 1 |
| 96 → | 1 |
| 98 → | 2 |
| 99 → | 2 |

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i}$$

Range = Highest − lowest class interval → ✓

$$\text{Variance} = \frac{\sum (x_i - \bar{x})^2}{n-1}, \text{ or, } \sigma^2 \text{ or } s^2.$$

$$\therefore \sigma = \sqrt{\sigma^2} \text{ or } \sqrt{s^2}. \rightarrow \text{Static Deviation}$$

Grouped Data #

| Class | fi frequency | (xi) midpoint | fixi | Cumulative frequency | $(x-\bar{x})^2$ | $f_i(x-\bar{x})^2$ |
|---|---|---|---|---|---|---|
| 60-64 | 3 | 62 | 186 | 3 | 330.15 | 990.45 |
| 65-69 | 1 | 67 | 67 | 4 | 173.45 | 173.45 |
| 70-74 | 4 | 72 | 288 | 8 | 66.42 | 256.68 |
| 75-79 | 7 | 77 | 539 | 15 | 9.92 | 69.44 |
| 80-84 | 6 | 82 | 492 | 21 | 3.35 | 20.1 |
| 85-89 | 3 | 87 | 261 | 24 | 46.65 | 139.95 |
| 90-94 | 2 | 92 | 184 | 26 | 139.95 | 279.9 |
| 95-99 | 4 | 97 | 388 | 30 | 283.25 | 11.33 |
| | $n=30$. | | $\Sigma = 2405$. | | | 3062.97 |

Mean. $\bar{x} = 8.17$, Mode 79, Median 80,

$$(\bar{x} \text{ or } \mu) \text{ Mean} = \frac{\sum_{i=1}^{n} f x_i}{n}$$

$$\therefore \bar{x} \text{ or } \mu = \frac{2405}{30} = 80.167 \text{ or } 80.17$$

Modal class → (75–79) যেখানে সবচেয়ে class এর max frequency থাকে।

∴ Mode : 79 (frequency 4).

Median class → $\frac{n}{2}$ তম Class এর cumulative freq এ থাকে।

$$\text{Estimated Median} = L + \left[\frac{\frac{n}{2} - cf}{f}\right] w.$$

(cf → of pre median class)

L = lower class boundary of the group
cf = cumulative frequency
f = frequency
w = width
n = total no of values.

$$= 75 + \left[\frac{\frac{30}{2} - 8}{7}\right] \times 5$$

$$= 80$$

$$\text{Estimated Mode} = L + \frac{(f_m - f_{m-1})}{(f_m - f_{m+1}) + (f_m - f_{m-1})} \times w$$

$$= \left(75 + \frac{7-4}{(7-4)+(7-6)} \times 5\right)$$

w = group width
L = lower boundary of the modal class.
$f_m$ = frequency of the group
$f_{m+1}$ = frequency of the group after modal group.
$f_{m-1}$ = " " " " before " "

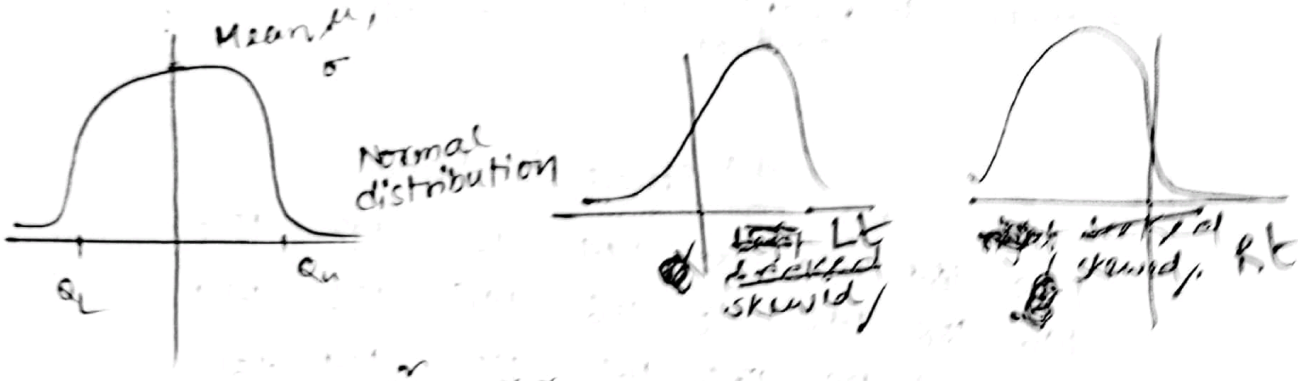
L = 75
$f_m$ = 7
$f_{m+1}$ = 6
$f_{m-1}$ = 4

mean square deviation.

(12 Sep 23)

Variance, $\sigma^2$ or $s^2$ $= \frac{\sum_{i=1}^{n} f_i (x - \bar{x})^2}{(n-1)} = \frac{3062.97}{30-1} \approx 105.62$

Standard Deviation: $\sqrt{s^2} = \sigma = 10.2$




$\bar{y} = a\bar{x} + b$ and $s_y^2 = a^2 s_x^2$.

(13 SEP 23)

Empirical Rule:

For any mounted shaped, nearly symmetric distribution of data

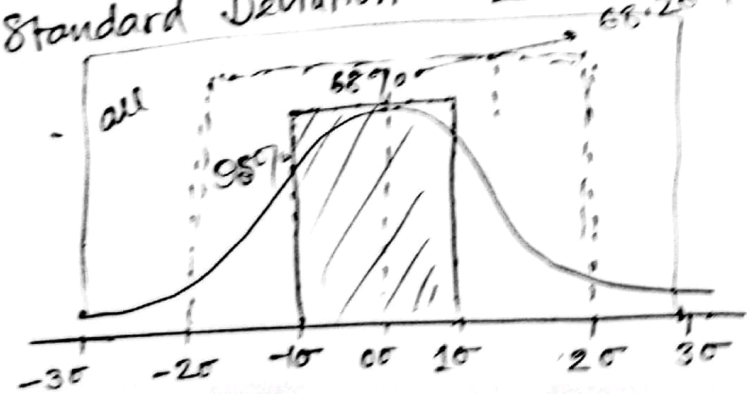
interval

- The $\bar{x} \pm s$ contains 68% data points. → 68.26%
- The interval $\bar{x} \pm 2s$ contains 95% data → 95.44%
- → → $\bar{x} \pm 3s$ → all data → → 99.72%

## Standard Normal Distribution:

Mean → $\mu = 0$

Standard Deviation = 1 (σ)



(19 Sep 23)

## Standard Values and z-scores:

$$z\text{-score} = \frac{\text{measurement} - \text{mean}}{\text{standard Deviation}}$$

$$z\text{ score} = \frac{189,966.99 - 37298.013}{84.8369} = 1.81$$

### 40 students Data

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 138 | 164 | 150 | 132 | 144 | 125 | 149 | 157 |
| 146 | 158 | 140 | 147 | 136 | 148 | 152 | 144 |
| 168 | 126 | 138 | 176 | 163 | 119 | 154 | 165 |
| 146 | 173 | 142 | 147 | 135 | 153 | 140 | 135 |
| 161 | 145 | 135 | 142 | 150 | 156 | 145 | 128 |

Largest weight = 176 lb

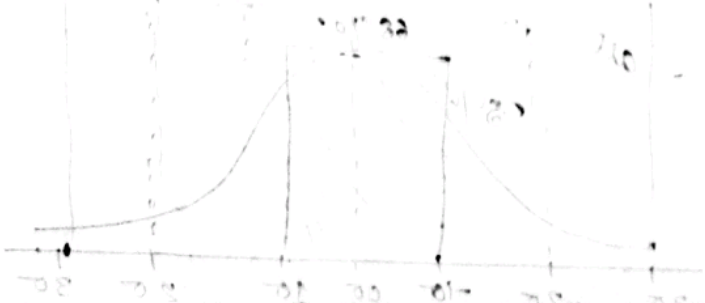
Smallest " = 119 lb

∴ Range = (176 − 119) = 57 lb.

$\frac{57}{5} \simeq 11$ → (total dividation)

interval →

$\frac{57}{9} \simeq 7$



if 9

| | $f_i$ |
|---|---|
| 118 – 126 → | 3 |
| 127 – 135 → | 5 |
| 136 – 144 → | 9 |
| 145 – 153 → | 12 |
| 154 – 162 → | 5 |
| 163 – 171 → | 1 |
| 172 – 180 → | 2 |
| | 70 |

| | frequency ($f_i$) | $x_i$ Midpoint |
|---|---|---|
| 118 – 122 | 1 | 120 |
| 123 – 127 | 2 | 125 |
| 128 – 132 | 2 | 130 |
| 133 – 137 | 1 | 135 |
| 138 – 142 | 6 | 140 |
| 143 – 147 | 8 | 145 |
| 148 – 152 | 9 | 150 |
| 153 – 157 | 1 | 155 |
| 158 – 162 | 2 | 160 |
| 163 – 167 | 3 | 165 |
| 168 – 172 | 1 | 170 |
| 173 – 178 | 2 | 175 |

Given $\bar{x} = 500\$$, $s = 125\$$

(a) increase price of each by 50$
(b) increase by 10%

Std variance: $\sqrt{s^2} = \sigma$

variance $s^2 = \sigma^2$

$$\bar{y} = a\bar{x} + b$$
$$s_y^2 = a^2 s_x^2$$

Ex: 1.1 (Pg. 51 of 839)

#

(a) $y_i = x_i + 50$; $\bar{y} = \bar{x} + 50 = 550\$$

standard deviation unchanged.

(b) $y_i = 1.10 x_i$; $\bar{y} = 1.10\bar{x} = 1.10(500) = 550\$$

standard deviation, $s_y = \sqrt{1.10^2 \cdot s_x^2}$

$= \sqrt{(1.10)^2 \times (125)^2}$

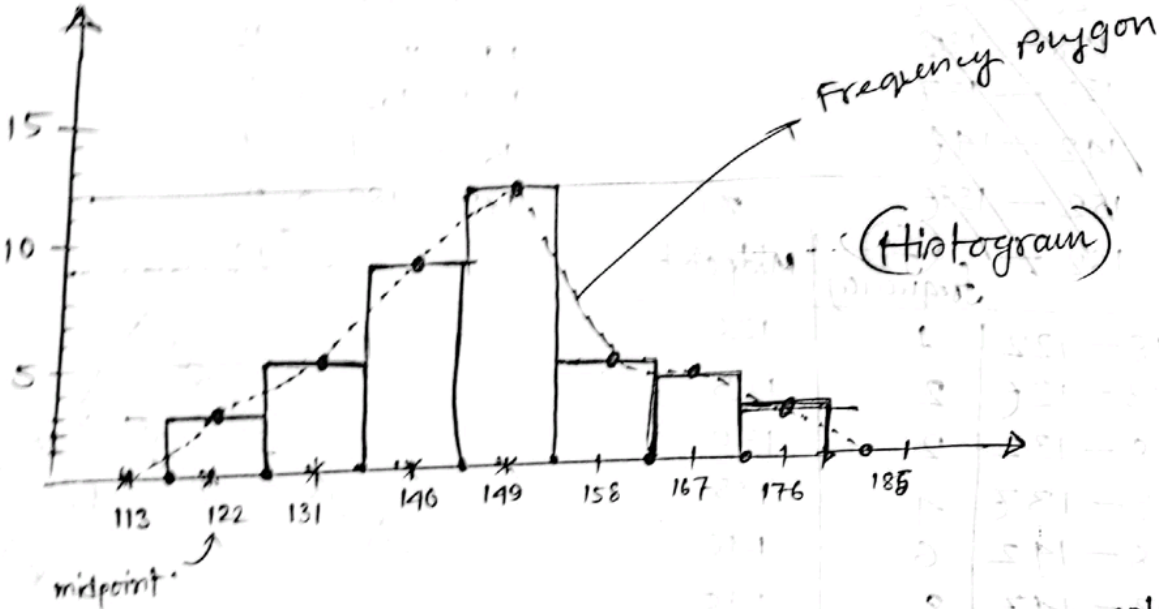
$= 137.5 > 125$

$s_{x^*}^2, a^2$

Histogram: first & Last interval

(26 Sep 23)

Frequency Polygon: Midpoint Required
Histogram : First & Last point.




Chebyshev's Inequality

Let $\bar{x}$ and $s$ be the sample mean and sample standard deviation of data set assuming that $s > 0$. Chebyshev's inequality states that for any value of $K \geq 1$, greater than $100\left(1-\frac{1}{K^2}\right)$ percent of the data lie within the interval from $\bar{x} - Ks$ to $\bar{x} + Ks$.

$100\left(1-\frac{1}{K^2}\right)$

$(\bar{x} - 1.5s, \bar{x} + 1.5s)$ interval এ কত % data থাকে,

Suppose, $\frac{3}{2} = K$,

$100 - \left(1 - \frac{1}{9}\right)$

$= 100\left(\frac{5}{9}\right)$

$\approx 55.96$ percent of data. We within $(\bar{x} - 1.5s \to \bar{x} + 1.5s)$

Let $S_K := \{i : \leq i \leq$

$$\frac{N(S_K)}{n} \geq 1 - \frac{n-1}{nK^2} > 1 - \frac{1}{K^2}.$$

$N(K)$ = Number of $i$ : $|x_i - \bar{x}| \geq Ks$

$|x_i - \bar{x}| < Ks$

$\bar{x} - Ks < x_i < \bar{x} + Ks$.

$$\frac{N_K}{n} < \frac{1}{K^2}$$

03 OCT 23

## Variance

$$S^2 = \frac{\sum f_i (x_i - \bar{x})^2}{n-1}$$

যখন প্রত্যেকটা data এর frequency 1 ∴ সব data আলাদা তখন $= \frac{\sum (x_i - \bar{x})^2}{n-1}$

CT Exam Question Review

Estimated

$n = 64$

$\therefore \frac{n}{2} = \frac{64}{2}$

$= 32$

∴ Median

$= \frac{32^{nd} + 33^{rd} \text{ entry/data}}{2}$

$= \frac{5+5}{2}$

$= 5.$

Class width = 2000

$2 \to 3 = 2$ (w)

$2 \to 4 = 4$ (w)

The following data represents the lifetimes (in hrs) of a sample of 40 transistors.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 112 | 121 | 126 | 108 | 141 | 104 | 136 | 134 |
| 121 | 118 | 143 | 116 | 108 | 122 | 127 | 140 |
| 113 | 117 | 126 | 130 | 134 | 120 | 131 | 133 |
| 118 | 125 | 151 | 147 | 137 | 140 | 132 | 119 |
| 110 | 124 | 132 | 152 | 135 | 130 | 136 | 128 |

(a) Determine the sample mean, median and mode.
(b) Draw cumiliative relative frequency plot.
(c) Are the data appearently normal?
(d) What parcentage of data fall within $\bar{x} \pm 1.5s$?
(e) Compare the result in (d) to that given by empirical rule
(f) " " " " bound by Chebyshev's inequality.

Soln:
Mean = 127.425
Median = 127.5
Mode = 108
Standard Deviation = 11.873
Range = 152 − 104 = 48


formula = $100\left(1-\frac{1}{K^2}\right)$

$(\bar{x}-Ks, \bar{x}+Ks)$

| | freq | relative freq | cumulative freq |
|---|---|---|---|
| 104 – 116 | 7 | $\frac{7}{40}$ | 7 |
| 117 – 128 | 14 | $\frac{14}{40}$ | 21 |
| 129 – 140 | 14 | $\frac{14}{40}$ | 35 |
| 141 – 152 | 5 | $\frac{5}{40}$ | 40 |
| | $n=40$ | | |

chebyshev's inequality yields that $100\left(1-\frac{1}{(3/2)^2}\right)$

$\Rightarrow 100\times\frac{5}{9} = 55.55$

percent of data lies in the interval

$(\bar{x}-1.5s, \bar{x}+1.5s) = \{(127.425-1.5\times11.873); (127.425+1.5\times11.873)$

$= (109.615, 145.235)$

standard deviation → s

Chebyshev's band:

$$\frac{N(s_{1.5})}{40} \geq 1-\frac{n-1}{nK^2} > 1-\frac{1}{K^2} = \frac{5}{9}$$

$$\boxed{\frac{N(sK)}{n} \geq 1-\frac{n-1}{nK^2} > 1-\frac{1}{K^2}}$$

$$1-\frac{40-1}{40\cdot\left(\frac{3}{2}\right)^2} > \frac{5}{9}$$

$$\therefore \frac{51}{90} > \frac{5}{9}.$$