Biostatistics...

(Med)

- Type of data:

1. Constant → عدد عيوني

2. Variable → Quantitive → continuous (fraction, decimals) BP / Hb%
   - Quantitive → Discrete (integer) heart rate
   - Qualitative → Ordinal (ranked) (logically) degree of succes
   - Qualitative → Nominal (مافي أفضلية) (Male / femal) (Blood groop)

- Measures of central tendency:

1. Mean = average = $\frac{\Sigma X}{n} = \bar{X}$

2. Median = middle value (order from lower to upper value)
   - → above 50% belowe 50% #
   - → # أفضل طريقة لعينات أرقام كبيرة أو صغيرة
   - → # ما بتأثر بقيم الكبيرة مثل Mean

3. Mode = most frequent (يمكن أكثر من رقم)

- Measures of dispersion:

1. Range → largest − smallest value.

2. Interquartile range → $Q_3 - Q_1$ ($25^{th} - 75^{th}$)

3. Variance → $\frac{\Sigma (X_i - \bar{X})^2}{n-1} = s^2$

4. standard deviation → $\sqrt{s^2} = s$

- nummber of interval to achieve the main advantage of tabular presentation (summarization) is (4−12)

- Width usually (5, 10, 15) calculate by Upper limit − lower Limit (interval contenous)

by Uppe limit − lower limit + 1 (interval desecrete)

- Real limit (upper limit +0.5 , Lower limit -0.5)
- Grapghical presentation → line graph → time (temperature, birth, death)
  - → Bar chart → for qualitative and quantitative discreet
  - → Histogram → continuous quantitative (no space between columns)
  - → Frequency polygon (mid point of interval on X axis)
  - → pie chart (four type of variable) (12) clock wise
- parameter → numerical information about population.
- statistics → numerical information about sample.
- percentiles or quantiles = np /100

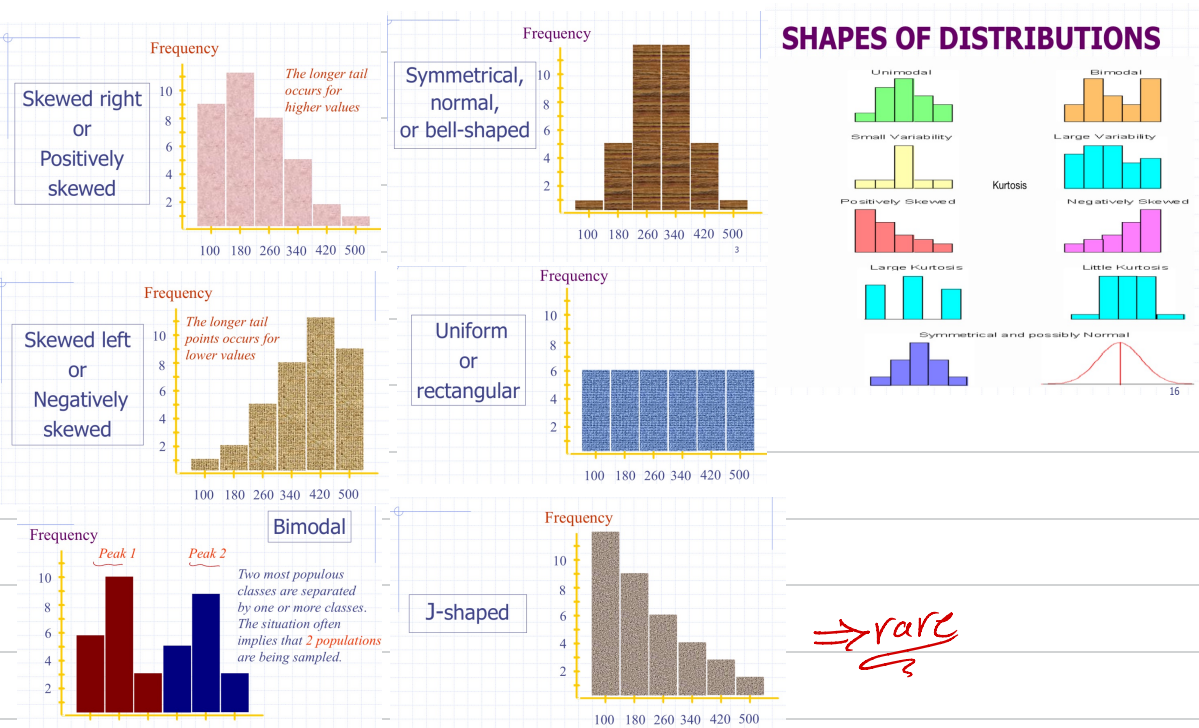
أول خطوة ترتب الأرقام في جدول ثم نحسب عالقانون (np/100) طلع رقم عشري أجبره وطلعه من الجدول وإذا طلع رقم صحيح بأخذ رقم اللي بعده كمان وبروح على الجدول وبجد المتوسط للقيمتين.

- Coefficient of variation → بستخدمه عشان أعرف مين أقل استقرار

$$CV = \frac{S}{\bar{X}} * 100\%$$

CV ↑ variance ↑ stable, uniform ↓

- SD ↑ → wide / flat
- SD ↓ → narrow / tall

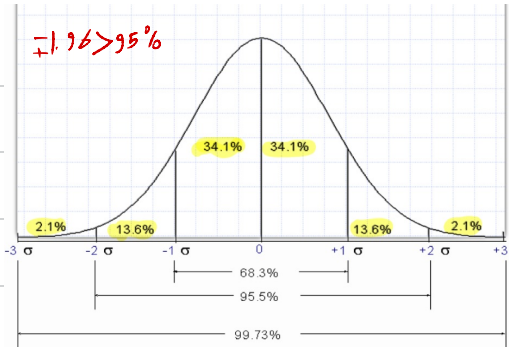


- symmetrical ⇄ normal / not normal But normal must be symmetrical
- frequency distribution may be represented in the form graph (general variability and symmetry) and equation (and mean, median, mode) same location
- symmetry of variation is indicated by skewness
- tail of distribution indicated the type of skewness (right or left)
- kurtosis → flatness or peakedness
  - large → high concentration on middle and out of tail
  - tilted → high concentration on middle and thin in tail

-standard normal distribution

$\bar{X} = 0 \qquad S/_S = 1$

$z = \frac{X - \bar{X}}{S}$

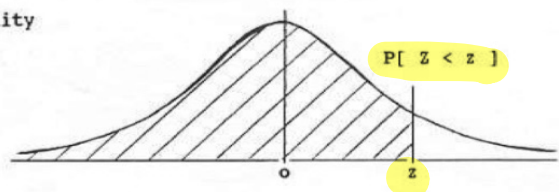


STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z i.e.

$$P[\,Z < z\,] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\, dZ$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

(Final)

- confidence intervals → range of likely values for the parameter.

$$\left(\bar{X} + \frac{\alpha\sigma}{\sqrt{n}}, \bar{X} - \frac{\alpha\sigma}{\sqrt{n}}\right)$$

- likely value (true) → critical values for standard normal distribution.

α

confidence coefficient ⇒ 95% $\alpha = 1.96$

⇒ 90% $\alpha = 1.65$

⇒ 99% $\alpha = 0.01$

- large sample ↓ standard error
- Interval coefficient narrow → more precise
- t-distribution → bell shape (continuous data)
  - similar to z curve but → degree of freedom (n-1)
  - → wider tail
- ↑ degree of freedom ↑ approaches the standard normal distribution (or infinity)
- p-values → chance
- statistical significance doesn't of necessity clinical significance
- Null hypothesis → no difference (H_0)
- Alternative hypothesis → have difference (H_1) (more or less) reject Null
- Type I error → rejecting H_0 when H_0 true (α)
- Type II error → accepting H_0 when H_1 true (β) (power of test)

- Hypothesis test ⇒ decide whether difference or no difference between sample

1. One-sample t-test

$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ إذا كان معطى متوسط للسكان population نستخدم

2. Independent / two-sample t-test

إذا كان معطى متوسط وانحرافين للعينتين نستخدم

$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2p}{n_1} + \frac{s^2p}{n_2}}}$ # $s^2p = \frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1 + n_2 - 2}$

3. Dependent t-test (paired)

$t = \frac{\bar{X}_D - \mu_0}{SD/\sqrt{n}}$ نستخدم (pre / post) (before, after) إذا كان معطى عينة قبل وبعد

D ⇒ difference

* بعد إيجاد t بالثلاثة أنواع نحدد degree of freedom (n-1) من الجدول إذا طلعت اللي حسبناها أقل من اللي طلعت من جدول ف لا يوجد اختلاف # accepted H_0 or fail to reject H_0

أما إذا طلعت اللي حسبناها أكبر من اللي طلعناها من الجدول ف يوجد اختلاف # accepted H_1 / reject H_0

4. Chi-square (X^2) test (relation, contingency table)

$E = \frac{\text{Total column} * \text{total rows of cell}}{\text{Grand total}}$

$X^2 = \sum \frac{(O - E)^2}{E}$

* إذا طلعت X^2 أكبر من α^2 إذا rejected H_0

* إذا طلعت X^2 أقل من α^2 fail to reject H_0