Biostatistics

تجميعة القوانين

* Mid :

1- Relative frequancy = $\frac{\text{No. of frequancies of each centa...}}{\text{total frequancy}}$

2- Percentage frequancy = R.f * 100%

غير مطابق فراحسب كلام الدكتورة

3- Sturges rule (K) = $1 + 3322\ (\log N)$

No. of classes تستخدم لإيجاد

4- Width of class interval = $\frac{\text{Range}}{K}$

5- Mide point = $\frac{X_1 + X_2}{2}$

* Measurments of Central tendency :

disadvantage = limitation = negativity

1- Mode (Mo)

2- Median (Md) ⇒ $\frac{n+1}{2}$

3- Mean ($\bar{x}$) = $\frac{\sum_{i=1}^{n} x_i}{N}$

4- Weighted mean = $\frac{w_1 x_1 + w_2 x_2 \dots + w_k x_k}{w_1 + w_2 \dots + w_k}$

* Measures of dispersion

1- Range (أكبر قيمة - أقل قيمة) or (أكبر قيمة، أقل قيمة)

2- Variance (S^2) = $\frac{\Sigma(x-\bar{x})^2}{N-1}$

3- Standard deviation = $\pm \sqrt{S^2}$ or $S.D = \sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{N} \Big/ (N-1)}$

4- $C.V = \frac{S.D}{\bar{X}} * 100\ \%$

* Standard error $= \frac{S.D}{\sqrt{N}}$

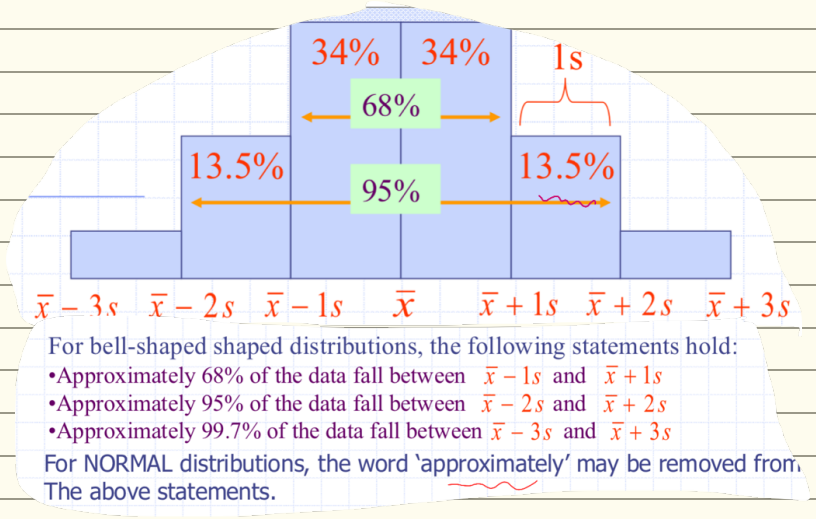
* حيث N هي حجم العينة في اي group العام.

* % within 95% $= \bar{x} \pm 1.96\ S.E$

* 95% of Conf. interval $= \bar{x} \pm 1.96\ S.E$

* % within 99% $= \bar{x} \pm 2.58 S.E$

*



** الأفضل خلال الامتحان رسم الرسمة ... .

* $Z = \frac{X_i - \bar{X}}{S.D}$

to transform the X value to its corresponding Z value.

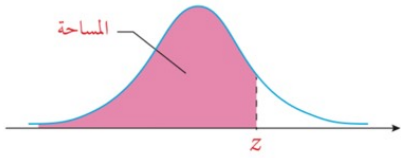
* x-scale ⇦ abscissa

* y-scale ⇦ ordinate

* If we want to represent the distribution of averages as a standard normal we use the :

$$Z = \frac{\bar{x}_i - \bar{\bar{x}}}{S.D/\sqrt{n}} = \frac{\bar{x}_i - \bar{\bar{x}}}{S.E}$$

* للاختبار



جدول التوزيع الطبيعي المعياري

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	0.5160	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	0.7852
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	0.8810	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	0.9772	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	0.9864	0.9868	0.9871	0.9875	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	0.9925	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	0.9979	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
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

* $P(z > a) = 1 - P(z < a)$

* $P(z < -a) = P(z > a) = 1 - P(z < a)$

* $P(a < z < b) = P(z < b) - P(z < a)$

* Final :

1- Degree of freedom (df) = n-1

2- Calculated $t = \frac{\bar{X} - \mu}{S.E}$ $\left(S.E = \frac{SD}{\sqrt{N}}\right)$

- To test a sample of normal continuous data, we need:
  1. **An expected value** = the **population or true** mean (mean of population) μ
  2. **An observed mean** = the average of your sample (mean of sample) $\bar{X}$
  3. **A measure of spread: standard error**
  4. **Degrees of freedom (df)** = n-1 (number of values used to calculate SD or SE)

  Then, we can calculate a **test statistic** to be compared to a known distribution

→ keep these concepts in your mind.

3- ↳ if p-value ≥ α → accept the Null. ((fail to reject))
↳ if p-value < α → reject the Null.

4. ↳ if calculated t > critical t → reject the Null.
↳ if calculated t < critical t → accept the Null.

* p-value : مقياس - تخبرك بمدى احتمال أن تكون النتائج التي حصلت عليها في تجربة أو دراسة ما عن طريق الصدفة فقط.

* α → probability of type I error (reject the Ho while Ho is true)

يعني شيء بأنه صحيح ولكن في الواقع هو خاطئ

or p-value represent the area that corresponds to the Z or the t-test statistic.

* يعني بتخبرني هل النتائج أجت بالصدفة ولا لأ؟

مثال فقط ← مثلاً لو P=0.05 هذا يعني أنه هنا احتمال 5 أشخاص من أصل 100 احتمال تكون نتائجهم طلعت معي هيك، بمحض الصدفة.

5- The test considered statistically significant when we reject the Null
↳ (p-value < α)

6- " " " " insignificant when we accept the Null Hyp.
↳ (p-value > α)

2 sampels → * For independent t-test: (unpaired)

* $df = ((n_1 + n_2) - 2)$

* pooled variance (S_p^2) = $\dfrac{S_1^2 (n_1 - 1) + S_2^2 (n_2 - 1)}{((n_1 + n_2) - 2)}$

* calculated $t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S^2p}{n_1} + \dfrac{S^2p}{n_2}}}$

one sample // (Befor + After) (نحنا نجري لنفس الشخص) → * For dependent t-test (paired):

* $t = \dfrac{\bar{X}_D - \mu_0}{S.E_D}$ (μ_0 of population → ميو) (S.E of the difference → D)

* $df = n - 1$

* Chi-square: 2x2 (الأفضل طبعًا) or for a x b
  1- independent.
  2- qualitative variable.
  3- Diechotomous (Yes/No) (أبيض/أسود) variable.

↳ the assumption of it:
  1- mutually exclusive (متعاكستان متعارضان) → 50% yes or 50% No
  2- the count of each cell must not be less than 5. (if less use fisher test.)

* $df = (r-1)(c-1)$

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

* $E = \dfrac{\text{total } c \times \text{total } r}{\text{Grand total}}$

* Continuity Correction (Yates) :

$$X^2 = \sum \frac{((O-E)-0.5)^2}{E}$$

* Used when df = 1
↳ means just for 2x2 table.

↳ resulting in small value for chi square.
↳ يعني بخليك متأكد أكثر في النتيجة.

* القانون الاحتياط ....

* (P_0) → true proportion (Rate) for total

* P_1 → Rate (proportation) succeded

	♂		♀		total	
succeeded	70	87.5%	90	75%	160	80%
not succeeded	10	12.5%	30	25%	40	
Total		80		120	200	

If the true population proportion of condition is
160/200 =0.8 40/200 = 0.2
Po =0.8 and
Rate (proportion) of succeeded ♂ (p_1)=70/80= 87.5%
Rate(proportion) of succeeded ♀ (p_2)= =90/120= 75%

$$Ho = P_1 = P_2 = P_0$$
$$H_A = P_1 \neq P_2 \neq P_0$$

????

3. SUMMARY OF STATISTICAL TESTS:

CATEGORICAL DATA	Enough data	Too little data (<5 in a cell)
Any r x c table	Chi-square	Fisher's Exact
CONTINUOUS DATA	**Normal** (even if transformed to normal) or large n	**Not normal**: (non-parametric tests)
One (group) sample	1-sample t-test	Kolmogorov-Smirnov
Two samples	2-sample t-test	Mann-Whitney U or Rank Sum
Paired data	1-sample t-test on paired differences (paired t-test)	Wilcoxon Signed-Rank
Three or more samples	Analysis of variance (ANOVA)	Kruskal-Wallis

* **Incidence rate** = $\frac{\text{No. of persons developing (new cases) a disease in a specific time and locality}}{\text{Total No. of population at risk}} * 1000$

* **Prevelance rate**: $\frac{\text{No. of persons who has the disease (existing) in a specific time and locality}}{\text{Total No. of population at risk}} * 1000$

Analysis of Case Control

	Cases (مصاب)	Control (غير مصاب)	total
Exposed (معرض)	A	B	A+B
Non (غير معرض)	C	D	C+D

Odds Ratio = $\frac{A}{B} \div \frac{C}{D} = \frac{AD}{BC}$ (البعض المعرض / البعض غير المعرض)

* Odd Ratio = $\frac{\text{Exposed}}{\text{Non Exposed}} = \frac{\frac{A}{B}}{\frac{C}{D}} = \frac{AD}{BC}$

تفسير النسبة الاحتمالية:

- إذا كانت النسبة الاحتمالية = 1: هذا يعني أنه لا يوجد فرق في الاحتمالات بين المجموعتين.
- إذا كانت النسبة الاحتمالية > 1: هذا يعني أن التعرض للعامل يزيد من احتمال حدوث النتيجة.
- إذا كانت النسبة الاحتمالية < 1: هذا يعني أن التعرض للعامل يقلل من احتمال حدوث النتيجة.

Analysis of Cohort:

* Estimation of risk :

Relative Risk (RR) = $\frac{\text{Incidence of disease among exposed}}{\text{Incidence of disease among non-exposed}} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$