ثلخيص قوانين مبادئ الإحصاء :-

Relative frequncy —→ Frequncy / Sample size.

If all the classes have the same class lengbth then:-

$$\text{class lenght} = \frac{\text{maximum observation} - \text{minimum observation}}{\text{number of classes}}$$

How to construct class:-

x ———— $(x + y - a)$ ⇒ x: minimum.
y: class lenght.
a: acuricy Unit.

Mid point = (Maximum + Minimum)/2 , If the class lenght of all classes are Identical then:-

Mid point for these class = (Mid point for previous class) + class lenght. ~~first class~~

Compulitive frequncy for first class = frequncy for these (first) class.

Compulitive frequncy for these class = (compulitve frequncy for previous class) + ~~compulitive~~ **frequncy for these class

** The compulitive frequancy for last class ≡ Sample Size.

Actual limite

(minimum − $\frac{1}{2}$(acuricy Unit) —— Maximum + $\frac{1}{2}$(acuricy Unit))

$\sum$ Relative frequancy = 1

The proportion of observation(Z) = frequancy of Z / number of obs.

← نفس التكرار النسبي.

Arithmatic mean $\bar{X}$:

1 for Raw data ⇒ $\frac{\text{Sum of the observation}}{\text{num of the observation}}$.

2 frequancy table without classes: → $\frac{\sum x f}{\sum f \text{(frequancy)}}$

observation.

3 $\sum \underline{\underline{x}}\, f_r$ → Relative frequancy. ← for frequancy table with classes.

mid point

Median:-

If n is even

$$\frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$$

If n is odd.

$$x_{(\frac{n+1}{2})}$$

حيث أن (n) هي عدد المشاهدات.

ولكن في حالة الجدول التكراري بدون فئات هي مجموع التكرار.

Percentiles:-

في حالة الـ Raw data أو Frequncy table without classes

$$P_{100} \rightarrow \frac{x_{np} + x_{np+1}}{2} \longrightarrow \text{If np is Integer.}$$

$$x_{\lceil np \rceil} \longrightarrow \text{If np is NOT Integer}$$

ملاحظة: يجب ترتيب البيانات من الأقل إلى الأعلى أولاً.

في حالة الـ frequncy table with classes

نُحدد قيمة الـ np (دون تقريب) ثم نُحدد موقعها في الـ compulitive freq وهي قيمة الـ حينئذٍ P100p هي قيمة المطلوبة لها في عمود الـ (lower actual limit.)

ولكن إن كان الموقع غير ظاهر أي أنه بين قيمتين

هذا أسهل للحفظ

$$\boxed{\frac{x - x_{\text{الحد السابق لـ}}}{x_{\text{الحد الذي بعد}} - x_{\text{الحد السابق لـ}}} = \frac{np - np_{\text{الحد السابق}}}{np_{\text{الحد الذي بعد}} - np_{\text{الحد السابق لـ}}}}$$

ملاحظة ← الـ Median في الجداول التكرارية لفئات هو الـ (P_{50}) .

→ Varince:-

for Raw data:-

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

for Frequency table:-

التكرار العادي .

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 f_i^2 - \frac{(\sum x_i f_i)^2}{\sum f \rightarrow n}}{n \leftarrow \sum f - 1}$$

At least for any $k>1$, $1-\frac{1}{k^2}$ of the observation between $\boxed{\bar{x}-ks \text{ and } \bar{x}+ks}$ → chybysh's Interval .

← لو طلب في السؤال كم عدد الـ مشاهدات بين X و Y بالضبط بدك أضرب الـ $(1-\frac{1}{k^2})$ في الـ Sample Size.

← لو طلب عدد المشاهدات اللي أقل من Z وأكبر من Y من كـ بين k ثم $\frac{1}{k^2}$ وهي العينة المطلوبة وبعد نضرب في الـ Sample Size .

← في حالة طلب الـ Interval يلي تحوي مثلاً على 50% من المشاهدات

Solution

$\frac{50}{100} = 1-\frac{1}{k^2}$ ⤳ (k) then by $\bar{x}-ks$ and $\bar{x}+ks$ you can get the Interval.

Updating statical Measures:-

$$\bar{X} = \frac{\sum x}{n}$$

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

$$\updownarrow$$

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

$$\updownarrow$$

$$\sum x^2 = (n-1)S^2 + n\bar{X}^2$$

Note:

Range = Maximum − Minimum

$IQR = Q_3(P_{75}) - Q_1(P_{25})$

Linear transformation of data:-

1. $\bar{Y} = a\bar{X} + b$. (for position)
2. $IQR(Y) = IQR(X) * |a|$
3. $S_y = S_x |a|$
4. $S^2_y = a^2 S_x^2$.
5. Range of y = Range of $x * |a|$.
6. P_{100p} → P_{100p} of $X+b$, $a > 0$; → $P_{100-p100}$ of $X+b$, $a < 0$.

Probility Rules:-

1. $P(\text{not } A) = P(\bar{A}) = 1 - P(A)$.

2. if $A \leqslant B$, then $P(A) \leqslant P(B)$

3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

4. $P(A \text{ and not } B) = P(A \cap \bar{B}) = P(A) - P(A \cap B)$.

5. $P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B})$

6. $P(\bar{A} \cap \bar{B}) = P(\bar{A}) + P(\bar{B}) - P(\bar{A} \cap \bar{B})$.

7. $P(A|B) = P(A \cap B) / P(B)$

الأهم كتير $\Leftarrow$ $\rightarrow$ $P(A \cap B) = P(B)\,P(A|B)$ OR $= P(A)\,P(B|A)$.

If A, B are disjoint, then $\leftarrow$ الأحداث المتباعدة :

$$P(A \cup B) = P(A) + P(B)$$

If A, B are Independent events $\leftarrow$ الأحداث المستقلة

then:-

$P(A \cap B) = P(A) * P(B)$ $\quad$ $P(A|B) = P(A)$ $\quad$ $P(B|A) = P(B)$.

Counting Techniques:-

$\boxed{m * n}$

← في حالة طلب عدد طرق فالجواب هو:

$m_1 * m_2 * m_3 * \cdots * m_n$

← في حال طلب الإحتمال:-

احتمال الخطوة الأولى * احتمال الخطوة الثانية * وهكذا.

(في الأمثلة تتوضح الفكرة).

Permutation

التبادل.

الترتيب مهم.

← في حالة عدد الأشياء التي معي يساوي عدد الخانات

مثلاً: 5 حروف و 5 خانات فعدد الطرق لعمل كلمات السر هو:-

$\underline{\underline{n!}} \rightarrow 5!$

← غير متساوي كأن يكون معي 8 حروف و 5 خانات.

عدد الطرق هو:-

$$P^{k}_{n} = \frac{n!}{(n-k)!}$$

Combination

التوافيق (الترتيب غير مهم).

$$C^{n}_{k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

← عدد الطرق هو:

↓

يمكن حسابها مباشرة من خلال الآلة الحاسبة

$\boxed{nCr}$

Baye's Rule :-

→ $P(B|A) = \dfrac{P(A|B) * P(B)}{P(A)}$.

→ Find $\underline{P(A)} = P(A|1) * P(1) + P(A|2) * P(2) + \cdots P(A|n) * P(n)$

. Total
Probility.

Expected Value

→ For discrete x, $E(x) = \sum x\, p(x)$

→ For continouse x, $E(x) = \int_{\text{over all } x} x\, p(x)\, dx$.

Variance:-

$Var(x) = E(x^2) - (E(x))^2$

→ For descrete x, $E(x^2) = \sum x^2 p(x)$.

→ $(E(x))^2 = (\sum x\, p(x))^2$

For continous x, $E(x^2) = \int_{\text{over all } x} x^2 p(x)\, dx$.

$(E(x))^2 = (\int x\, p(x)\, dx)^2$.

قوانين المتغيرين (x,y)

$\rightarrow E(XY) = \sum (xy\, P(X=x \cap Y=y)$

$\rightarrow cov(x,y) = E(XY) - E(X)E(Y)$

$\rightarrow corr(x,y) = \dfrac{cov(x,y)}{\sqrt{Var(x)} * \sqrt{Var(y)}}$

$\Rightarrow$ قوانين العلاقات :--

$\rightarrow E(ax + by + c) = a\,E(x) + b\,E(y) + c$

$\rightarrow Var(ax + by + c) = a^2 Var(x) + b^2 Var(y) + 2ab\, cov(x,y)$

$a=1 \quad b=1$

$Var(x+y) = 1\,Var(x) + 1\,Var(y) + 2\,cov(x,y)$

وهكذا

$\rightarrow cov(ax+b, cy+d) = ac\, cov(x,y)$

$\rightarrow corr(ax+b, cy+d) =$

$corr(x,y)$ if $ac > 0$

$-corr(x,y)$ if $ac < 0$.

Independent ⊙

1. $P(X \cap Y) = P(X) * P(Y)$

2. $\text{corr}(X,Y) = \text{cov}(X,Y) = 0$

3. $\text{Var}(aX + bY + c) = a^2 \text{Var}(X) + b^2 \text{Var}(Y)$

4. $E(XY) = E(X) * E(Y)$

Binomial Distribution :-

number of trials

$X \sim \text{Binomial}(n, p)$

has a distribution

probapility of Succes

→ PDF of X → $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

→ $E(X) = np$

→ $\text{Var}(X) = np(1-p)$

Poisson Distribution.

$X \sim \text{poisson}(\mu)$

0, 1, 2, ...

→ PDF of X → $P(X = k) = \dfrac{e^{-\mu} \mu^k}{\boxed{k}!}$ | → $E(X) = \text{Var}(X) = \mu$.

... 2, 1, 0 ←

$np < 5 \longrightarrow$ poisson.

Geometric Distribution

→ P.d.f $P(X=k) = (1-p)^{k-1} p$

→ $E(x) = \frac{1}{p}$

→ $Var(x) = \frac{1-p}{p^2}$.

Hypergeometric Distribution

→ P.D.f of X: $P(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$

→ $E(x) = n * \frac{M}{N}$

→ $Var(x) = n * \frac{M}{N}\left(1 - \frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$

(If selected with replacment then

$\boxed{X \sim Binomial\left(n, \frac{M}{N}\right)}$

without replacment —

$\boxed{X \sim Hypergeom(n, M, N)}$.

Scanned by CamScanner

Normal Distribution

→ $F(x) \to Pdf = \frac{1}{\sqrt{2\pi}\sigma} * e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

→ $X \sim N(\mu, \sigma^2)$, if standard then .

$\frac{X-\mu}{\sigma} \sim N(0,1)$ $\quad Z \sim N(0,1)$.

, If x is a contenouse then :
and $a<b$

① $P(x=a)=0$

② $P(a \leq x \leq b) = \int_a^b F(x)\,dx$

$= P(a \leq x \leq b) = P(a<x<b)$

$$\boxed{\text{Ploop of } N(\mu, \sigma^2) = \sigma * (\text{Ploop of } N(0,1)) + \mu}$$

When $x \sim \text{Binomial}(n,p)$, $np \geq 5$ and $n(1-p) \geq 5$

you can approximated by $N(np, np(1-p))$

$P(x=a)$ in discrete corrected to → $P(a-\frac{1}{2}<x<a+\frac{1}{2})$

$P(a \leq x \leq b)$ corrected to → $P(a-\frac{1}{2}<x<b+\frac{1}{2})$

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

if n is large ($n \geq 30$) then the $\bar{X} \sim N(\mu_x, \frac{\sigma_x^2}{n})$ regardless the distribution of x.

if $X \sim N(\mu_x, \sigma_x^2)$ then $\bar{X} \sim N(\mu_x, \frac{\sigma_x^2}{n})$ ← Fact - 1.

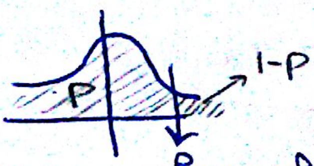
$P(x_1 + x_2 + \dots x_n > b) = P\left(\frac{x_1 + x_2 + \dots + x_n}{n} > \frac{b}{n}\right)$

$= P(\bar{x} > \frac{b}{n})$.

T-distribution:

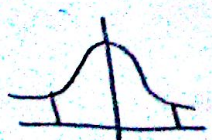
↳ the distribution of $\boxed{\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}}$ is t-distribution at 1-n decrease of freedom.

→ when $n \geq 30$ we approximate t-distribution ~~Gam~~ by $N(0,1)$.



Proop of t distribution = $t_{(1-p)}$ at d.f (1-n).

→ $\hat{p} > 30$ use N. distribution but $\hat{p}^n > 30$ OR $30 > \hat{p}^n$ use t-distribution.



[illegible]

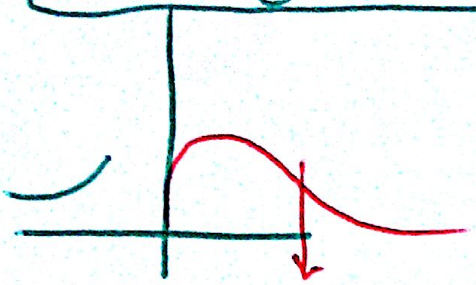
→ $\boxed{\text{Proop of } \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \text{Proop of t-distribution.}}$

χ^2-distribution (distribution of S^2).

diffrent.

$$N(\mu, \sigma^2) \rightarrow \frac{(n-1)S^2}{\sigma^2}.$$

$$\boxed{\frac{\text{Pioop of } S^2 * (n-1)}{\sigma^2} = \text{Pioop of } \chi^2\text{-distribution at d.f} = n-1}$$



χ^2 at $(1-p)$ at d.f $n-1$.

Distribution of $\hat{p}$:-.

→ $\hat{p} = \frac{x}{n}$

$$E(\hat{p}) = E\left(\frac{x}{n}\right) = \frac{1}{n} * E(x) = \frac{1}{\cancel{n}} * \cancel{n} * p = p$$

$$Var(\hat{p}) = Var\left(\frac{x}{n}\right) = \frac{1}{n^2} Var(x) = \frac{1}{n^2} * n * p(1-p)$$

$$= \frac{p(1-p)}{n}$$

→ $np \geq 5$ and $n(1-p) \geq 5$ then $\hat{p} \sim N(p, (p(1-p)/n))$.

Estimating μ.

$(1-\alpha)100\%$ C.I for μ is $\rightarrow \left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) \leftarrow n \geq 30$.

If $n < 30$ use t-distribution (if the previous dist is Normal.)

$$\left(\bar{X} - t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right).$$

For the Interval of $(1-\alpha)100\%$ C.I for μ $(f, \tilde{f})$ the $\leftarrow$

$\bar{X} = (f + \tilde{f})/2$.

Estimating σ^2.

a $(1-\alpha)100\%$ for σ is $\left(\frac{(n-1)s^2}{Z^2_{\frac{\alpha}{2}}}, \frac{(n-1)s^2}{Z^2_{1-\frac{\alpha}{2}}}\right)$

but For Just σ not σ^2 $\rightarrow$ جذر لطرفي الفترة.

A $(1-\alpha)100\%$ C.I for p.

Recall : $\dfrac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \sim N(0,1)$ We cant use it to give a $(1-\alpha)100\%$ C.I for p because $p(1-p)$ is unknown.

Fact:-

$\dfrac{\hat{p}-p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}}$ almost has $N(0,1)$

Estimating $\hat{p}$

↓

Thus a $(1-\alpha)100\%$ C.I for p is given by $\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}\right)$

Recall:

When $\theta = \mu$ or P

$(1-\alpha)100\%$ C.I estimating error $= \frac{1}{2} *$ length of the interval.

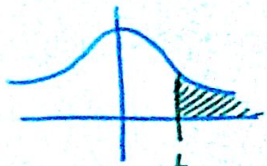
$$= \begin{cases} Z_{\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}} & \text{when } \theta = \mu \\[2ex] Z_{\frac{\alpha}{2}} \sqrt{\dfrac{P(1-P)}{n}} & \text{when } \theta = P \end{cases}$$

Hypothesis testing:-

1. determine the Value of H_0 and H_1.

$H_0 : \theta = \theta_i$

$H_1 : \theta > \theta_i \longrightarrow$ Right Side test

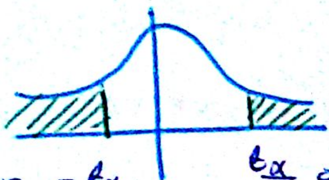


t_α or Z_α or χ^2_α.

$H_1 : \theta < \theta_i \longrightarrow$ left side test

$\chi^2_{(1-\alpha)}$ or $-Z_\alpha$ or $-t_\alpha$

$H_1 : \theta \neq \theta_i \longrightarrow$ double test



$\chi^2_{(1-\frac{\alpha}{2})}$ or $-Z_{\frac{\alpha}{2}}$ or $-t_{\frac{\alpha}{2}}$ $\qquad$ $t_{\frac{\alpha}{2}}$ or $Z_{\frac{\alpha}{2}}$ or $\chi^2_{\frac{\alpha}{2}}$

2. determine the type of carve.

Normal or t-distribution or ch-distribution

3. Test static: the low different from parameter to another.

$\boxed{1}$ $\theta = \mu$ If $n > 30 \longrightarrow$ Normal

$n < 30 \longrightarrow$ t-distribution.

$$\text{test static} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$\sigma \rightarrow$ use s if σ unknown.

$\boxed{2}$ $\theta = p$ $n > 30 \longrightarrow$ Normal

$n \cancel{<} 30 \longrightarrow$ t-distribution.

$$\text{test static} = \frac{\hat{p} - p}{\sqrt{\frac{p_i(1-p_i)}{n}}}$$

$\boxed{3}$ $\theta = \sigma^2$

chi^2-distribution.

$$\text{test static} = \frac{(n-1)s^2}{\sigma_i^2}.$$

$\boxed{4}$ determine the position of $\boxed{\text{test static}}$ value on the curve:-

① If ~~\#~~ if the shadow position (rejection region)

Then :-

① we can reject H_0

② evedence H_1.

③ type error $\longrightarrow$ type 1 $\longrightarrow$ $p(\text{error}1) = \alpha$.

② If if not the shadow position (accept region).

Then :-

① we cant reject H_0.

② we cant evedence H_1

③ type error $\longrightarrow$ type 2 $\longrightarrow$ $p(\text{error }2) = \beta$ $\rightarrow$ ده نفس اجابة