الفصل الدراسي الثاني ٢٠١٤-٢٠١٥

zenon

Principles of statistics

Summery

إعداد : براءة مصطفى

/groups/zenooon
/zenooon
http://www.engzenon.com

Relative Frequency of continous Data.

التكرار النسبي للبيانات المتصلة.

الفئة Class	التكرار freq	التكرار النسبي Relative freq
1-5	1	1 / 10
6-10	4	4 / 10
11-15	3	3 / 10
16-20	2	2 / 10 → مجموع التكرارات الكلي. (total)
	tot = 10	تكرار الفئة (freq)

Commulitave frequency of continous Data.

التكرار التراكمي للبيانات المتصلة.

Class	freq	commulitave freq
1-5	1 →	1
6-10	4+1 →	5
11-15	3+5 →	8
16-20	2+8 →	10

(تكرار الفئة + مجموع تكرارات الفئات قبلها)

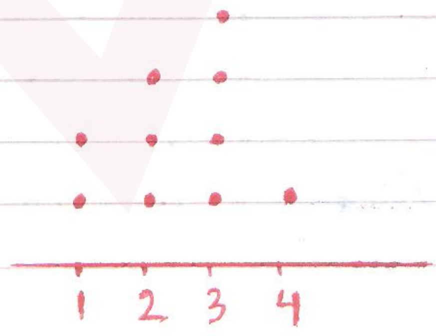
Graphical Presentation of Data

التمثيل البياني

1 Discrete Data → البيانات المنفصلة → « ٣ طرق »

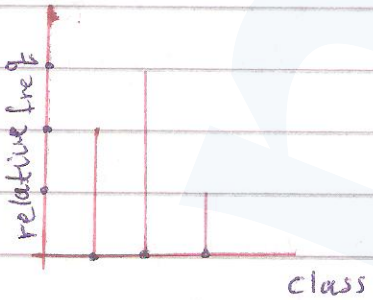
1. Dot plot of Data

Ex. 3, 4, 2, 1, 2, 3, 3, 2, 1, 3

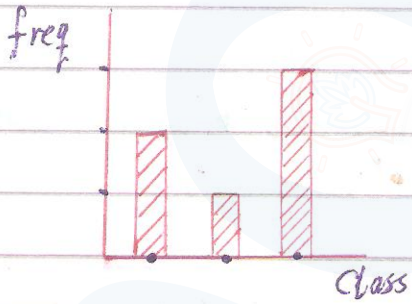


2. Line Diagram

التكرار النسبي ←

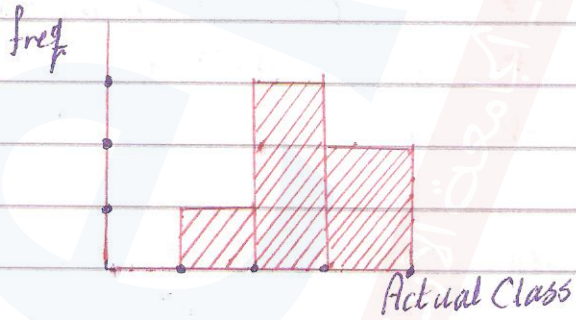


3. Bar Chart

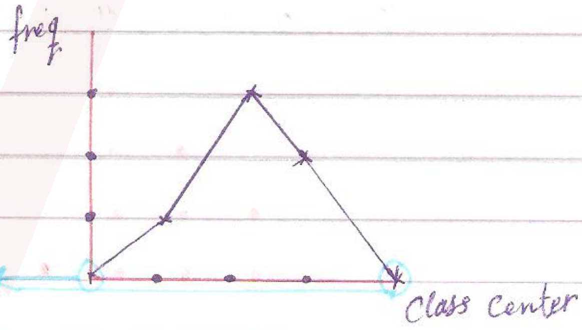


$\boxed{2}$ Continous Data: → البيانات المتصلة → (٣ طرق)

1. Histogram

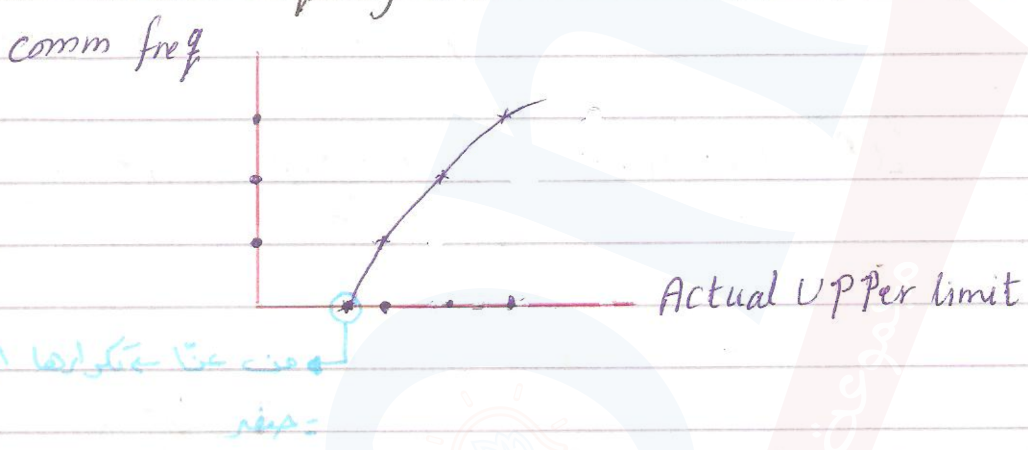


2. Frequency Polygon → مضلع



نختار أرقاماً من خارج الفترة
يكون تكرارها = صفر

3. Commulitave frequency curve   منحنى التكرار التراكمي



Measures of location

A. Raw Data → البيانات الخام → بدون جدول

1. Arithmatic mean « الوسط الحسابي »

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$\sum_{i=1}^{n} x_i$ → مجموع العلامات

n → عدد الطلاب

2. Median. → الوسيط   Ex. 1, 3, -2, 4, 5

* ترتيب → -2, 1, 3, 4, 5

Ex. 1, 3, -2, 0, 4, 5

* ترتيب → -2, 0, 1, 3, 4, 5 → $\frac{1+3}{2} = 2$

3. Mode → المنوال

• الفئة الأكثر تكراراً •

9. Percentiles → المئوية

(حسب السؤال)

عدد المشاهدات ← نسبة المشاهدات

$$100^{th} \text{ Percentile} = \begin{cases} x_{[np]+1}, & np \text{ is not integer (غير صحيح)} \\ \dfrac{x_{np} + x_{np+1}}{2}, & np \text{ is integer (صحيح)} \end{cases}$$

← القانون العام

%80 → هي العلامة التي تحتها 80% من المشاهدات .

- First Quartile → $Q_1 = P_{25}$ = العلامة التي تحتها ربع المشاهدات
- Second Quartile → $Q_2 = P_{50}$ = Median
- Third Quartile → $Q_3 = P_{75}$

B. Group Data → بيانات مجدولة (مع جدول)

1. Mean

$$\mu = \frac{\sum x_i \cdot f_i}{\sum f_i}$$

x_i → class center (مركز الفئة)

f_i → تكرار الفئة

$\sum f_i$ → مجموع التكرارات

Class	freq (f_i)	Class center (x_i)	$x_i \cdot f_i$
1 - 5	1	3 $\left(\frac{1+5}{2}\right)$	3
6 - 10	4	8 $\left(\frac{6+10}{2}\right)$	32
11 - 15	3	13 $\left(\frac{11+15}{2}\right)$	39
16 - 20	2	18 $\left(\frac{16+20}{2}\right)$	36
	$\sum f_i = 10$		$\sum x_i \cdot f_i = 110$

2. Mode

Class	freq
1 - 5	1
6 - 10	4
11 - 15	3
16 - 20	2

- Modian class → 6 - 10
  الفئة الأكثر تكراراً
- Mode $= \frac{6+10}{2} = 8$ Class center

3. Median + Percentiles.

لو طلب (Median) يعني ⇐ 50^{th} Percentile

Class	freq	Actual UPPer limit	comm freq
1 - 5	11	5.5	11
6 - 10	14	10.5	25
11 - 15	13	15.5	38
16 - 20	12	20.5	50

$x = P_{70} \leftarrow \cdots \rightarrow 35$

- Find the 70^{th} Percentile $= P_{70}$

ممكن ، يكون كسر

1. نجد الموقع → position of $P_{70} = np = 50(0.70) = 35$
   موقعها من التكرارات التراكمية.

2. نسبة وتناسب → $\frac{15.5 - 10.5}{x - 10.5} = \frac{38 - 25}{35 - 25}$

$$\frac{5}{x - 10.5} = \frac{13}{10} \rightarrow x = 14.3 = P_{70}$$

- Find the number of student with grade more than 13.

عدد الطلاب (من التكرارات التراكمية) الذين تزيد علاماتهم عن 13

$\frac{13 - 10.5}{15.5 - 10.5} = \frac{x - 5}{38 - 25} \rightarrow x = 11.5 \approx 11$ student

« 11 ⇐ عدد الطلاب الذين علاماتهم أقل من 13 »

$50 - 11.5 = 38.5 \approx 38$

لـ نأخذ المتممة.

Measures of variation مقاييس التشتت / التغير

1. Raw Data «مش جداول»

A. Range المدى

$$\boxed{\underset{\text{أكبر قيمة}}{\text{Range} = \text{maximum}} - \underset{\text{أقل قيمة}}{\text{minimum}}}$$

B. Interquartile Range

$$\boxed{IQR = Q_3 - Q_1 = P_{75} - P_{25}}$$

بنفس الطريقة السابقة (من غير جدول)

C. Standard Deviation الانحراف المعياري + variance التباين

Standard Deviation

- Population (بحكي بالسؤال) $\rightarrow \sigma = \sqrt{\dfrac{\sum (x_i - \mu)^2}{n}}$ (x_i → Class center, μ → mean (Population), n → العدد)
- Sample (بحكي بالسؤال) $\rightarrow S = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n-1}}$ ($\bar{x}$ → mean (sample))

* يمكن كتابة القوانين بطريقة أخرى

$$\sigma = \sqrt{\frac{\sum x_i^2 - n(\mu)^2}{n}}$$

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

variance = (Standard Deviation)^2

*Population $\rightarrow \sigma^2 = \dfrac{\sum (x_i - \mu)^2}{n}$

*Sample $\rightarrow S^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

حسب شو يكون كاتب بالسؤال pop. أو Sample

الوسط الوسيط الانحراف

d) Mean + Median deviation.

* Sample :

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

$$\text{mean deviation} = \frac{\sum |x_i - \text{mean}|}{n} \qquad \text{median deviation} = \frac{\sum |x_i - \text{median}|}{n}$$

2. Grouped Data « جداول. »

A) Standard Deviation

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2 \cdot f_i}{\sum f_i}} \qquad S = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i - 1}}$$

f_i → التكرار

Population مع تكرار   Sample مع تكرار

هش من عنا ، من السؤال .

حكيلنا s أو p.

Example: In the sample data, find standard deviation.

✓ ع طول نستخدم قانون s

✓ ومعطينا جدول مع تكرار ← نستخدم قوانين G.D

Class	freq (f_i)	class center (x_i)	$x_i \cdot f_i$	$(x_i - \bar{x})^2 \cdot f_i$
1 - 5	1	$\frac{1+5}{2} = 3$	3	$(3-11)^2 * 1 = 64$
6 - 10	4	8	32	$(8-11)^2 * 4 = 36$
11 - 15	3	13	39	$(13-11)^2 * 3 = 12$
16 - 20	2	18	36	$(18-11)^2 * 2 = 98$
			110	210

Sol: $\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{110}{10} = 11$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i - 1}} = \sqrt{\frac{210}{9}}$$

B) Mean + Median Deviation

$$\text{Mean deviation} = \frac{\sum |x_i - \text{mean}| \cdot f_i}{\sum f_i}$$

$$\text{Median deviation} = \frac{\sum |x_i - \text{median}| \cdot f_i}{\sum f_i}$$

D) Inter quartile Range

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

→ percentiles نفس الطريقة اللّي نطلّع منها من الجدول.

Comparing two collections.

1. Z. Score → للمقارنة بين علامتين موجودتين في مجموعتين مختلفتين.

$$Z = \frac{\overset{\text{العلامة}}{x} - \overset{\text{الوسط}}{\bar{x}}}{\underset{\text{الانحراف}}{S}}$$

كلهم بالسؤال :)

رح نفهم من المثال (٢)

- Example

		sec 1	sec 2
mean	$(\bar{x})$	60	70
standard dev	(S)	7	4
Grade	(x)	64	68

Which is better, 64 in sec 1 or 68 in sec 2 ??

Sol.

sec₁ : $Z_1 = \frac{x_1 - \bar{x}_1}{s_1} = \frac{64 - 60}{7} = 0.57$

sec₂ : $Z_2 = \frac{x_2 - \bar{x}_2}{s_2} = \frac{68 - 70}{4} = -0.50$

$Z_1 > Z_2 \rightarrow 0.57 > -0.50$

$\therefore$ 64 in sec_1 is better than 68 in sec_2 ^_^

2. Coefficient of variation معامل التغيُّر

$$\boxed{C.v = \frac{S}{\bar{x}} * 100\%}$$

Example:

	sec_1	sec_2
$\bar{x}$	30	100
S	4	6

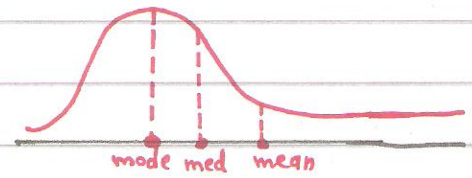
Sol.

$(C.v.)_1 = \frac{S_1}{\bar{x}_1} * 100\% = \frac{4}{30} * 100\% = 13.3\%$

$(C.v.)_2 = \frac{S_2}{\bar{x}_2} * 100\% = \frac{6}{100} * 100\% = 6\%$

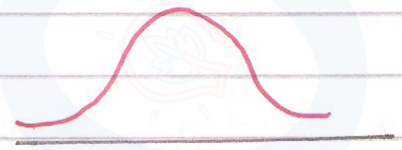
The variability in sec_1 is more than that in sec_2

3. Shape Measures قياسات الأشكال



$Mode < median < mean$ $\qquad$ $mode > median > mean$

حفظ $\qquad$ بيجي منهم $\qquad$ حفظ



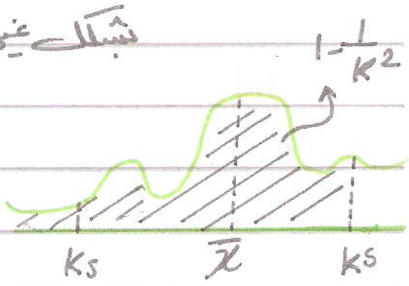
$Mean = Mode = median$

Applications

1. Chebyshev's Inequality

باختصار: نستخدم هذه الطريقة لحساب نسبة مشاهدات يحدد نترتها السؤال، ويكون توزيع المشاهدات بشكل غير منتظم

$$1-\frac{1}{K^2}$$



2. Bell-shaped Distribution

هاد أسهل ← لأنو النسبة تكون جاهزة بس لازم نطلع قيمة K ^_^

$(\bar{x} - s, \bar{x} + s) \rightsquigarrow k=1$, we have 68% of the observations
$(\bar{x} - 2s, \bar{x} + 2s) \rightsquigarrow k=2$, we have 95% of the observations
$(\bar{x} - 3s, \bar{x} + 3s) \rightsquigarrow k=3$, we have 99% of the observations.

حفظ

Updating Descriptive Measures

① The mean

A. combined samples «دمج عينتين»

$$\bar{x}_{compined} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

B. Adding, Deleting New Data

$$\bar{x}_{new} = \frac{n_{old}\ \bar{x}_{old} \mp \dots}{n_{new}}$$

C. Updating the variance

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

$$S^2_{pooled} = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2}$$

$$S^2_{combined} = S^2_{pooled} + \frac{n_1 n_2}{n_1 + n_2} \cdot \frac{(\bar{x}_1 - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

* تأثير العمليات الحسابية *

1

Mean
Median
Mode

} ← تتأثر بجميع العمليات الحسابية

percentiles → ✓ تتأثر بجميع العمليات الحسابية
✓ إذا ضربنا أو قسمنا بسالب نضرب ونقسم بالمعكوس

مثال :

$P_{35} = 59$ و $P_{65} = 80$

If each grades is divided by -3 and then we add 11

find $\underset{\text{بعد}}{(P_{35})_y}$ ، $\underset{\text{بعد}}{(P_{65})_y}$

Sol:

$$y = \frac{x}{-3} + 11$$

$$(P_{35})_y = \frac{\overset{\text{قبل}}{(P_{65})_x}}{-3} + 11 = \frac{80}{-3} + 11$$

$$(P_{65})_y = \frac{(P_{35})_x}{-3} + 11 = \frac{59}{-3} + 11$$

2

Range
IQR
std.dev

* لاتتأثر بالجمع والطرح
* في حالة * ، ÷ نأخذ القيمة المطلقة

variance →

* لاتتأثر بالجمع والطرح
* في حالة * ، ÷ نضرب ونقسم بالمربع

Example

The Observations in a sample have
Range = 60, IQR = 53, std.dev = 15
$x \rightsquigarrow$ before, $Y \rightsquigarrow$ After

If each observation devided by −3 then we subtract 5, find the new Range, IQR, st.dev, variance.

Sol: $y = \frac{x}{-3} - 15$

- $(Range)_y = (60) * \left|-\frac{1}{3}\right|$
- $(IQR)_y = (53) * \left|-\frac{1}{3}\right|$
- $(std.dev)_y = (15) * \left|-\frac{1}{3}\right|$
- $(variance)_y = \left(-\frac{1}{3}\right)^2 * \underbrace{(5)^2}_{S^2}$

CH# 2

Random Expirement $\Rightarrow$ التجربة العشوائية (رمي حجر نرد ، قطعة نقود)

Sample space $\Rightarrow$ الفضاء العيني (Ω)

$\cap$ « Intersection » $\Rightarrow$ تقاطع (OR)

$\cup$ « union » $\Rightarrow$ اتحاد (AND)

$\bar{A} \Rightarrow \Omega - A$

disjoint events $\Rightarrow$ حوادث منفصلة $(A \cap B = \emptyset)$

Independent events $\Rightarrow$ حوادث مستقلة $(P(A|B) = A$

$P(A \cap B) = P(A) \cdot P(B))$

Ex. Tossing a coin two times = Tossing two coins

T → Tail كتابة H → Head صورة

$\Omega = \{ HH, TH, HT, TT \}$

Ex. Rolling two dice (die) حجر نرد once

$\Omega = \{ (1,1), (1,2), (1,3), (1,4), (1,5), (1,6)$

$(2,1), (2,2), (2,3), \ldots\ldots (2,6)$

$(3,1) \ldots\ldots (3,6)$

$(4,1) \ldots\ldots (4,6)$

$(5,1) \ldots\ldots (5,6)$

$(6,1) \ldots\ldots (6,6) \}$

$$P(A) = \frac{n(A)}{n(\Omega)}$$

$n(A)$ → عدد عناصر الحادث

$n(\Omega)$ → عدد عناصر الفضاء العيني = (عدد الاحتمالات) عدد المحاولات.

$P(A)$ → احتمال أي حادث

• Permutations and Compinations (التباديل والتوافيق)

الاختلاف بين التوافيق والتباديل ← التباديل ← تهتم بالترتيب.
التوافيق ← لا تهتم بالترتيب

PAST PAPER

① If two balls are selected at random without replacement from a box contaning 4 red and 6 black balls, then the Probability that the two balls are of different colors is :

4	6
R	B

$$P(RB \cup BR) = P(RB) + P(BR)$$

$$= \frac{n(RB)}{n(\Omega)} + \frac{n(BR)}{n(\Omega)}$$

$n(BR) \rightarrow$ سحب كرتين مختلفتين

$n(\Omega) \rightarrow$ سحب كرتين بدون شروط

$$= \frac{4 * 6}{10 * 9} + \frac{4 * 6}{10 * 9} = \frac{48}{90}$$

لـ بدون ارجاع.

② A factory is manufactoring bags using three machines A, B and C. Machine A Produced **25%** machine B produces **35%** and machine C produces **40%** of the total products. suppose that 5% of the products from machine A, 4% from B, and 2% from C are defective. one bag is chosen at Random from a Products of this factory.

A) what is the probability that the bag is defective?

$E_a = 0.25$, $E_b = 0.35$, $E_c = 0.40$

$$P(D) = P(D|E_a)\,P(E_a) + P(D|E_b)\,P(E_b) + P(D|E_c)\,P(E_c)$$

$$= (0.05)(0.25) + (0.04)(0.35) + (0.02)(0.40)$$

$$= 0.012 + 0.014 + 0.008$$

$$= 0.034$$

B) If the bag is found defective, then what is the probability that it was manufactured by machine C ?

$$P(\ |D) = \frac{P(E_c)\,P(D|E_c)}{P(E_c)\,P(D|E_c) + P(E_b)\,P(D|E_b) + P(E_a)\,P(D|E_a)}$$

$$= \frac{(0.40)(0.02)}{(0.40)(0.02) + (0.04)(0.35) + (0.05)(0.25)}$$

$$= \frac{0.008}{0.345} = 0.023$$

- $P(A) = 0.7$, $P(B) = 0.6$

$P(A|B) = \frac{5}{6}$, find $P(A \cup B) = ?$

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

$$= 0.7 + 0.6 - 0.5$$

$$= 0.8$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = 0.5$$

- $P(A \cap \bar{B}) = 0.4$, $P(B \cap \bar{A}) = 0.1$

$P(\overline{A \cap B}) = 0.6$, $P(\bar{A}) = ??$

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

$$P(A) = 0.8$$

$$P(\bar{A}) = 1 - P(A) = 1 - 0.8$$

$$= 0.2$$

- $P(A) = 0.7$, $P(B) = 0.6$, $P(A \cap B) = 0.5$

find $P(A|\bar{B})$??

$$P(A|\bar{B}) = \frac{P(A \cap \bar{B})}{P(\bar{B})} = \frac{0.2}{0.4} = \frac{1}{2}$$

$$P(\bar{B}) = 1 - P(B) = 1 - 0.6 = 0.4$$

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

$$= 0.7 - 0.5 = 0.2$$

Ex. Let A and B be independent events

$P(A) = 0.3$, $P(A \cup B) = 0.8$, find ① $P(B)$

② $P(\bar{A} \cap \bar{B})$

③ $P(A \cap \bar{B})$

① $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$0.8 = 0.3 + P(B) - P(A).P(B)$

$0.8 = 0.3 + P(B) - 0.3\,P(B)$

$0.5 = P(B)\,[1 - 0.3]$

$0.5 = 0.7\,P(B) \rightarrow P(B) = \frac{5}{7}$

② $P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0.8 = 0.2$

③ $P(A \cap \bar{B}) = P(A) \cdot P(\bar{B}) = (0.3)(\frac{2}{7})$

• Univariate Random variable. (x)

متغير واحد

Ex. Rolling a die, let x be a number appears

Range $\leftarrow R_x = \{1, 2, 3, 4, 5, 6\}$

Discrete Random variable $\leftarrow x$ قيم

أرقام منفصلة وليست فترات.

• A Probability density function (Discrete)

اقترانه الكثافة الاحتمالية ← شروطه:

① $f(x_i) \geq 0 \rightarrow$ كل احتمال من الاحتمالات أكبر من صفر

② $\sum f(x_i) = 1 \rightarrow$ مجموع الاحتمالات جميعها يساوي 1

• A probability density function (continous)

شروطه ←:

① $f(x) \geq 0 \rightarrow$ كل احتمال أكبر من صفر لكل قيم x

② $\int f(x) \cdot dx = 1 \rightarrow$ تكامل الاحتمالات = صفر

مقارنة بين المتغير العشوائي المنفصل و المتصل

x is a discrete Random variable	x is a continous Random variable
① $P(1 < x \leq 3)$ $x = 2, 3$ لانأخذ فترات $= P(x=2) + P(x=3) = f(2) + f(3)$	① $P(1 < x \leq 3)$ $x = [1, 3]$ $= \int_1^3 f(x) \cdot dx$
② $P(1 \leq x \leq 3)$ $x = 1, 2, 3$ $= P(x=1) + P(x=2) + P(x=3)$ $= f(1) + f(2) + f(3)$ نهتم بالمساواة	② $P(1 \leq x \leq 3)$ $x = [1, 3]$ $= \int_1^3 f(x) \cdot dx$ X نهتم بالمساواة
③ $P(x=2) = f(2)$	③ $P(x=2) = \int_2^2 f(x) \cdot dx = 0$ $P(x=2) \neq f(2)$

- Expection of x
  - $E(x) = \sum x \cdot f(x) \rightarrow x$ is a discrete random variable.
  - $E(x) = \int x \cdot f(x) \rightarrow x$ is a continous random variable.

- $E(x) = \mu_x$ (Expection of x = mean of x)
- variance $\rightarrow v(x) = E(x^2) - (E(x))^2$

$$\boxed{E(x^2) \neq (E(x))^2} \quad = E(x^2) - (\mu_x)^2$$

$$v(ax+b) = a^2 \cdot v(x)$$

$$\begin{cases} E(x^2) = \sum x^2 f(x) \\ E(3x+1) = \sum (3x+1) \cdot f(x) \end{cases}$$

- **Bivariate Random variable** (x, Y)

متغيرين

شروط اقتران الكثافة الاحتمالية لـ x و Y :

① $0 \leq f(x,Y) \leq 1$

② $\sum_{Y} \sum_{x} f(x,Y) = 1$

Ex. Tossing a fair Coin 3 times.

let x be the number of heads in the 3 trials.

" Y " " " " the first and second trials.

Ω	x	Y	(x,Y)
TTT	0	0	$(0,0) \rightarrow f(0,0)$
TTH	1	0	$(1,0)$ احتمال أن تكون
THT	1	1	$(1,1)$ $\frac{1}{8}$ = مع بعض $Y=0, x=0$
HTT	1	1	$(1,1)$
THH	2	1	$(2,1)$
HTH	2	1	$(2,1) \rightarrow f(2,1) = \frac{2}{8}$
HHT	2	2	$(2,2)$
HHH	3	2	$(3,2)$

- The joint p.d.f of x and Y

$Y \backslash x$	0	1	2	3
0	$\frac{1}{8}$	$\frac{1}{8}$	0	0
1	0	2/8	2/8	0
2	0	0	1/8	1/8

- The Marginal Probability p.d.f. of x and Y

$Y \backslash x$	0	1	2	3	$f_Y(Y)$
0	1/8 +	1/8 +	0 +	0 =	2/8
1	+ 0	2/8	2/8	0	4/8
2	+ 0	0	1/8	1/8	2/8
$f_x(x)$	= $\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	

- $\text{cov}(x,Y) = E(xY) - \mu_x \mu_Y \rightarrow \mu_Y = E(Y) = \sum Y f_Y(Y)$

  $E(xY) = \sum xY\, f(x,Y)$
- $\text{cov}(ax+b, cY+d) = ac \ \text{cov}(x,Y)$
- x and Y are independent $\Rightarrow \text{cov}(x,Y) = 0$

  $f(x,Y) = f_x(x) \cdot f_Y(Y)$

- Correlation

$$\rho = \text{Corr}(x,Y) = \frac{\text{Cov}(x,Y)}{\sqrt{v(x)} \cdot \sqrt{v(Y)}}$$

$\text{Cov}(x,Y) \rightarrow E(xY) - \mu_x \mu_Y$

$v(Y) \rightarrow E(Y^2) - \mu_Y^2$

$E(Y^2) = \sum Y^2 \cdot f_Y(Y) \leftarrow E(Y^2) - \mu_Y^2$

$f_Y(Y) \rightarrow$ Marginal Prob. for Y

$$\text{Corr}(ax+b, cY+d) = \begin{cases} \text{Corr}(x,Y), & ac > 0 \\ -\text{Corr}(x,Y), & ac < 0 \end{cases}$$

$\boxed{1}$ $E(a_1x_1 + a_2x_2 + a_3x_3) = a_1E(x_1) + a_2E(x_2) + a_3(E(x_3))$

$\boxed{2}$ $v(a_1x_1 + a_2x_2 + a_3x_3)$

$= (a_1)^2 v(x_1) + a_2^2 \cdot v(x_2) + a_3^2 \, v(x_3)$

$+ 2\,[a_1a_2 \,cov(x_1,x_2) + a_1a_3\, cov(x_1,x_3)$

$+ a_2a_3\, cov(x_2,x_3)]$

if x_1, x_2, x_3 are independent

يصبح هذا الحد = صفر

27

CH#3 Discrete Probability Distribution.

التوزيعات الاحتمالية المنفصلة

① The Binomial Distribution.

الصيغة العامة ← $X \sim B(n,p)$

$q = 1 - p$

القانون ← $P(X=k) = \binom{n}{k} p^k q^{n-k}$

$k = 0, 1, \ldots, n$

يمكن الاستغناء عنه باستخدام الجدول

n: عدد المحاولات

p: احتمال النجاح في كل محاولة ← ثابت

$k(x)$: (عدد مرات النجاح المطلوبة في السؤال)

Example: suppose that the chance of heading the target in a single shot is 0.3. p

find the probability of heading the target three k times out of 10 trials.
n

Sol: $x \sim B(\overset{n}{10}, \overset{p}{0.3})$

$$P(x=k) = \binom{n}{k} p^{k} q^{n-k}$$

$$P(x=3) = \binom{10}{3} (0.3)^{3} (0.7)^{10-3} = \dots$$

ويمكن الحل باستخدام الجدول (أسهل) ::

↓

لكن: الجدول نظامه نظام تراكمي يعني لو طلب احتمال ظهور الحدث ٣ مرات مثلاً يعني بدنا نوخد الـ ٣ واللي قبلها

$(x \le ..)$ ::

أمثلة للتوضيح

① $P(x<7) = P(x \le 6)$

② $P(x \le 7) \longrightarrow$ مباشرة من الجدول

③ $P(x>7) = 1 - P(x \le 7)$

④ $P(1<x<7) = P(x \le 6) - P(x \le 1)$

⑤ $P(1 \le x \le 7) = P(x \le 7) - P(x \le 0)$

⑥ $P(1 < x \le 7) = P(x \le 7) - P(x \le 1)$

فهم أبداً مش حفظ :)

- Mean and variance

$x \sim B(n,p) \Rightarrow \mu = E(x) = np$

$v = \sigma^2 = npq$

② The Poisson distribution

$x \sim Poisson(\mu)$ ← الصيغة العامة

$$P(x=k) = \frac{e^{\mu} \cdot \mu^{k}}{k!}$$ → القانون

يمكن الاستغناء عنه باستخدام الجدول

μ: الوسط الحسابي (المعدل)

$e \simeq 2.7$

Example: suppose that x is the numper of patients admitted to the ICU at the university hospital Perday, If the average numper of patients admitted to the ICU perday is 3. what is the Prob. in agivenday there will be exactly two ddmissions ?

Sol: $x \sim \text{Poisson}\ (3)$

$$P(x=k) = \frac{e^{-\mu} \cdot \mu^{k}}{k!} = \frac{e^{-3} \cdot (-3)^2}{2!} = \ldots$$

وممكن الحل باستخدام الجدول («نفس فكرة ال Binomial «دراكي»)

- **Mean and variance.**

$$\mu = \sigma^2$$

mean = variance

③ The Geometric Distribution

$x \sim \text{Geom}\ (P)$ ← الصيغة العامة

$$P(x=k) = q^{k-1}\ P$$

القانون ، ولا يمكن استخدام الجدول :D

Example: the Prob. of hitting atarget for asingle trial is 0.6 find the Prob. of hitting the target for the first time is the 4^{th} trial

x = number of trial till we hit the target for the first time

Sol:: $P(x=4) = (0.6)\ (0.4)^3 = \ldots$

• Mean and variance.

$$\mu = E(x) = \frac{1}{p}$$

$$v = \sigma^2 = \frac{q}{p^2}$$

④ The hyper Geometric Distribution

الصيغة العامة $x \sim HyperGeom(n; M, N)$

$$P(x=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

القانون ، ولا يمكن استخدام الجدول :

N: عدد العناصر الكلي (بدون شروط)
n: عدد العناصر التي تم سحبها دون إرجاع
M: عدد عناصر المجموعة (الفئة) التي نسحب منها

Example: A box contains 20 bulbs , 6 of them are deffective , if 4 bulbs are drawn without RePlacement , find the Prob. of getting exactly (3) deffective bulbs.

Sol: $M = 6,\ N = 20,\ n = 4,\ k = 3$

$$p(x=3) = \frac{\binom{6}{3}\binom{16}{1}}{\binom{20}{4}} = \ldots$$

N = 20

M	N-M
6	14

draw 4 bulb (n)

• Mean and variance

$$\mu = n \frac{N}{M}$$

$$\sigma^2 = n\left(\frac{M}{N}\right)\left(1 - \frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$$ سهل

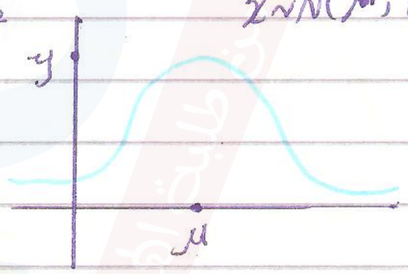
34

CH#4 Normal Probability Distribution

توزيع متصل

$x \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-(x-\mu)^2/2\sigma^2}$$



* شبيه بال Bell-Shaped
* يوجد تماثل

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

لا تهم المساواة

ما رح نحتاج نستخدم هاد القانون ☺

Example: The grades of students are normally distribute with mean (64) and standard deviation (5). If a student is choosen Randomly, find the Probability that this studend grade is:

① Less than 68 → $x \sim N(64, 25)$ → ليست في الحالة المعيارية

$P(x < 68)$ $Z \sim N(0,1)$

$$Z = \frac{x-\mu}{\sigma} = \frac{68-64}{5} = 0.8$$

حتى نحافظ على التماثل

نحول إلى الحالة المعيارية

$P(Z < 0.8) = 0.7881$ من الجدول

② Between 54 and 74

$$P(54 \leq x \leq 74)$$

الى الحالة المعيارية . الى الحالة المعيارية

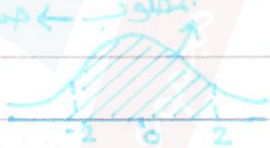
$$Z_{54} = \frac{x-\mu}{\sigma} = \frac{54-64}{5} = -2$$

$$\Rightarrow P(-2 \leq Z \leq 2)$$

$$Z_{74} = \frac{74-64}{5} = 2$$

المطلوب ← صعب نطلعه من الجدول

$$P(-2 \leq Z \leq 2)$$



$$= P(Z \leq 2) - P(Z \leq -2)$$

③ More than 56.3

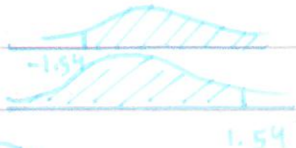
$$P(x > 56.3)$$

$$Z = \frac{56.3-64}{5} = -1.54$$

$$= P(Z > -1.54)$$

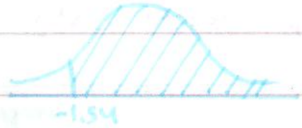
لم نتعامل مع إشارة الأكبر في جدول الـNormal !!

$$P(Z > -1.54)$$
$$= P(Z < 1.54)$$



OR

$$P(Z > -1.54)$$



$$= 1 - P(Z < -1.54) = 1 - 0.0618$$
$$= 0.9293$$

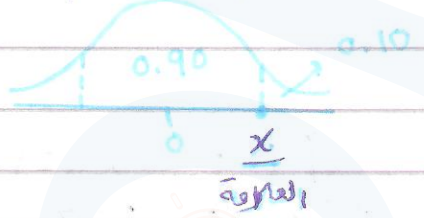
④ Find the grade above wich we have 10% of the ...

العلامة التي فوقها 10% من الطلاب

= " تحتها 90% " "

= ⊜: 90^{th} Percentile

∴ find the 90^{th} Percentile



بنشتغل بالعكس

$P(Z < Z_0) = \underline{0.90}$

Z_0 ← الحالة المعيارية للعلامة المطلوبة :   0.90 ← الاحتمال

	0.08
⋮	
1.2	0.8997
⋮	

0.8997 أقرب إلى 90

$Z_0 = 1.28$

$Z_0 = \frac{x-\mu}{\sigma} \Rightarrow 1.28 = \frac{x-64}{5}$

$x = 70.4 = 90^{th}\,P$

الفرق في هذا الفرع أنه أعطانا الاحتمال وطلب العلامة

أما في الأفرع السابقة فهو يعطينا العلامة ويطلب الاحتمال لها.

The Approximations

1. $x \sim B(n,p) \longrightarrow x \sim poisson(\mu)$

   when.. $n \geq 30$, $p < 0.1$

   $\mu = np$ ← μ لما نقرّب رح نحتاج

2. $x \sim HyperGeom(n;N,M) \longrightarrow x \sim B(n,p)$

   when.. N is Large , $\frac{n}{N} < 0.05$

   $p = \frac{M}{N}$ ← p رح نحتاج لـ

3. $x \sim B(n, p) \longrightarrow x \sim N(\mu, \sigma^2)$

when.. $n \geq 30$, $p > 0.1$

$\mu = np$, $\sigma^2 = npq$ ← σ^2, μ بنحتاج لـ

4. $x \sim Poisson(\mu) \longrightarrow x \sim N(\mu, \sigma^2)$

when.. μ is large

$\sigma^2 = \mu$ ← σ^2 بنحتاج لـ

CH # 5 Sampling Distribution

باختصار: بيتكلم هذا الفصل عن العينة التي تسحب من توزيع الـ NORMAL .

	Population	Sample	
mean	μ	$\bar{x}$	$\bar{x}_1 - \bar{x}_2$
var.	σ^2	S^2	S_1^2 / S_2^2
prob.	P	$\hat{P}$	$\hat{P}_1 - \hat{P}_2$

1. The distribution of the sample Mean.

$$x \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$\bar{x}$

- $n \geq 30$ → Z.dist
- $n < 30$
  - σ^2 (known) → Z.dist
  - σ^2 (unknown) ← for population → t.dist

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

* الاختلاف بينهم بالجدول *

Example: In a certain population, the weights are normally distributed with mean 76 (μ) and st. dev 5 (σ). Choosing a sample of size 16 (n), find the probability that:

① The sample average ($\bar{x}$) is more than 80

$$\left.\begin{array}{ll} \text{Population}: & x \sim N(\overset{\mu}{76}, \overset{\sigma^2}{25}) \\ \text{Sample}: & \bar{x} \sim N(\mu, \frac{\sigma^2}{n}) \quad , n = 16 \end{array}\right\}$$

$* P(\bar{x} > 80) \rightarrow n < 30$, σ^2 is known $\rightarrow$ Z - dist

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{80 - 76}{\frac{5}{\sqrt{16}}} = 3.2$$

$$P(Z > 3.2) = P(Z < -3.2) = 0.0007$$

② Their sum (Σ) of their weights is less than 1200 kg.

Remember

$$\bar{x} = \frac{\Sigma x_i}{n}$$

$$P\left(\sum_{i=1}^{16} x_i < 1200\right)$$

حتى نحصل على توزيع معروف

نقسم على n

$$= P\left(\frac{\Sigma x_i}{n} < \frac{1200}{n}\right) = P(\bar{x} < 75)$$

$$Z = \frac{75 - 76}{\frac{5}{\sqrt{16}}} = -0.86$$

$$P(Z < -0.86) = 0.2119$$

② The Distribution of the different between two samples mean $(\bar{x}_1 - \bar{x}_2)$

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \bar{x}_1 - \bar{x}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}\right)$$

المفروض (-) بس حطينا (+) لأنه مابصير الـ var يكون سالب :

Example: The grades of females and males students in Calculus are normally distributed with means 70 and 65 respectively and standard deviation 8 and 10 respectively.
In samples of 15 females and 20 males, find the prob. that the female students will have an average more than the male student average by 2?

- Population 1 (females): $N(\underset{\mu}{70}, \underset{\sigma^2}{64})$
  Population 2 (Males): $N(65, 100)$
- sample 1 (females): $n_f = 15$
  sample 2 (Males) $n_m = 20$

$P(\bar{x}_f > \bar{x}_m + 2)$ المطلوب
$= P(\bar{x}_f - \bar{x}_m > 2)$

$$\bar{x}_f - \bar{x}_m \sim N\left(\mu_f - \mu_m, \frac{\sigma_f^2}{n_f} + \frac{\sigma_m^2}{n_m}\right)$$

variance is known in two pop.
↳ Z - dist

$$Z = \frac{\bar{x}_f - \bar{x}_m - (\mu_f - \mu_m)}{\sqrt{\frac{\sigma_f^2}{n_f} + \frac{\sigma_m^2}{n_m}}} = \frac{2 - (70 - 65)}{\sqrt{\frac{64}{15} + \frac{100}{20}}} = -0.985$$

$= P(Z > -0.985) = P(Z < 0.99)$ مع تجبير

③ The Distribution of the sample variance : S^2

له توزيع خاص (جدول خاص)

للتحويل إلى الحالة المعيارية ← $\chi^2 = \dfrac{(n-1)\,S^2}{\sigma^2}$ (d.f) χ^2-dist

Example: If a sample of size $n=6$ is drawn from a population with variance $\sigma^2 = 10$, find:

$P(S^2 > 18.4727)$

نحول إلى الحالة المعيارية

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(6-1)(18.4727)}{10} = 9.2364$$

$$= P(\chi^2 > 9.2364) = \%.10$$

> نتعامل مع إشارة الأكبر

من الجدول.

④ The Distribution of the Ratio of the sample variances $\left(\dfrac{S_1^2}{S_2^2}\right)$

له توزيع خاص (جدول خاص)

عند التحويل للحالة المعيارية ← $F = \dfrac{S_1^2}{S_2^2} \cdot \dfrac{\sigma_2^2}{\sigma_1^2}$ F-dist

Example: Consider two independent Random samples from 2 Normal Population:

$n_1 = 5$, $n_2 = 8$ و $\sigma_1^2 = 50$, $\sigma_2^2 = 20$

Find $P\left(\dfrac{S_1^2}{S_2^2} > 7.4\right)$ ⟹ تحول إلى المعياري

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}$$

$$= 7.4 \cdot \frac{20}{50} = 2.96 \Rightarrow P(F > 2.96) = 0.10$$

من الجدول.

⑤ The Distribution of the sample proportion : $\hat{p}$

توزيع الـ NORMAL
Z·dis

$$\hat{p} \sim N\left(p,\ \frac{pq}{n}\right)$$

mean variance

نحول إلى الحالة المعيارية عن طريق

$$Z = \frac{(\hat{p} - p)}{\sqrt{\frac{pq}{n}}}$$

Example: suppose that 10% of a certain population are are difectives.

If 400 Items are drawn from the Population, what is the Probability that the sample proportion will be more than %12

$\underline{n = 400}$ Sample , $\underline{p = 0.10}$ Population

المطلوب ⇐ $P(\hat{p} > 0.12)$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.12 - 0.10}{\sqrt{\frac{(0.10)(0.90)}{400}}} = 1.33$$

$$P(Z > 1.33) = P(Z < -1.33) = 0.0918$$

(6) The Distribution of the difference between two samples proportion $(\hat{p}_1 - \hat{p}_2)$

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2 \, , \, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$$

Example: suppose that 50% of population A own cars.
35% " " B " "

If a sample of size 100 is drawn from A
and " " " 80 " " " B

Find the Prob. that the difference between the samples proportions $(\hat{p}_A - \hat{p}_B)$ will be between 0.1 and 0.2

- Pop A : $p_A = 0.50$ • Sample A : $n_A = 100$
- Pop B : $p_B = 0.35$ Sample B : $n_B = 80$

$$P(0.1 < \hat{p}_A - \hat{p}_B < 0.2)$$ المطلوب

نحول إلى المعياري

$$Z_{0.1} = \frac{(\hat{p}_A - \hat{p}_B) - (p_A - p_B)}{\sqrt{\frac{p_A q_A}{n_A} + \frac{p_B q_B}{n_B}}} = \frac{0.1 - (0.5 - 0.35)}{\sqrt{\frac{(0.5 * 0.5)}{100} + \frac{(0.35 * 0.65)}{80}}}$$

$$= -0.68$$

$$Z_{0.2} = 0.68$$

$$P(-0.68 < Z < 0.68)$$
$$= P(Z < 0.68) - P(Z < -0.68)$$
$$= 0.7517 - 0.2483$$
$$= 0.5034$$

CH #6 + 7

Example: The salaries of teachers in Jordan for 1990 – 2000 are normally distributed with std. dev 50 J.D. The average salary based on a sample of 400 teachers for 1990 – 2000 was 215 J.D per month

Population: $x \sim N(\overset{?}{\mu}, \overset{\sigma^2}{50^2})$

Sample: $n = 400$, $\bar{x} = 215$

(a) What is the Point estimate of the mean 1990 – 2000 salaries and its S.E (Standard Error).

① Point Estimat of (μ) is $(\bar{x})$

② $S.E.(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{400}} = 2.5 \ll \bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

(Standard Error)

انحراف معياري

(b) Give 90% C.I for the mean 1990 - 2000 salaries.

$(1-\alpha)100\%$ C.I for μ

$\bar{x} \mp Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

90% C.I for μ

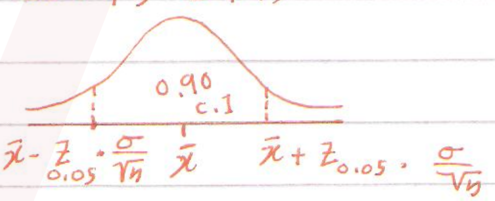
$\bar{x} \mp Z_{0.05} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow \alpha = 0.10, \frac{\alpha}{2} = 0.05$

0.90 متممة

$215 \mp (1.64) \cdot (\frac{50}{\sqrt{400}}) = 215 \mp 4.1$

$(215 - 4.1, 215 + 4.1)$

$(210.9, 219.1)$



Eror Of Estimation $= Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

ممكن تكون n مجهولة بنطلعها من نفس القانون

Consepts of Hypothesis Testing.

- Null Hypothesis (H_0): True unless proved to be false

صحيح حتى يثبت الخطأ.

- Alternative Hypothesis (H_1): الهدف الـ Test

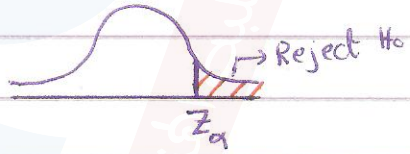
Testing for μ

* خطوات الحل:

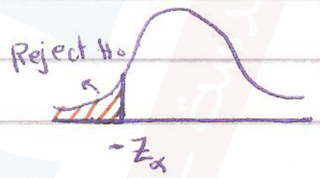
① نحدد H_1 من السؤال.

② نحدد Critical / cut Point بناءً على α

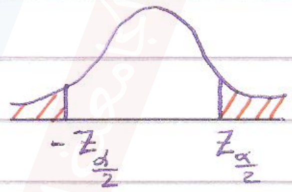
$H_1: \mu > \mu_0$



$H_1: \mu < \mu_0$



$H_1: \mu \neq \mu_0$



③ نتأكد من موقع H_0 عن طريق

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

نستخدم Z أو t حسب:

$\bar{x}$
- $n \geq 30$ → Z
- $n < 30$
  - σ^2 known → Z
  - σ^2 unknown → t

Testing for $\hat{p}$

خطوات الحل :

① نفس أول خطوتين من test μ

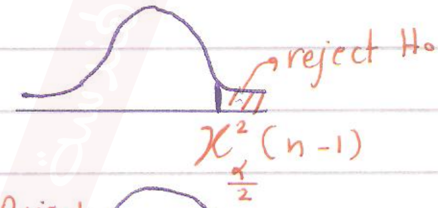
② $$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Testing for variance

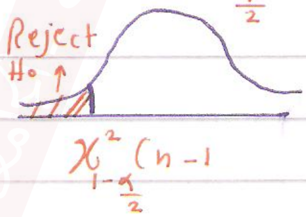
① نحدد H_1 من السؤال

② نحدد critical / cut point بناءً على α

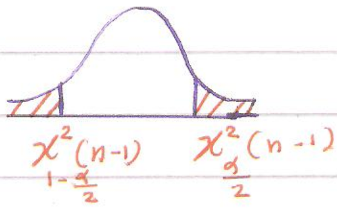
$H_1 : \sigma^2 > \sigma_0^2$



$H_1 : \sigma^2 < \sigma_0^2$



$H_1 : \sigma^2 \neq \sigma_0^2$



③ نحدد الموقع

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$