Chapter 10

Ch7 test hypothesis 1 sample
ch8 2 Sample $\mu_1 - \mu_2$
ch10 2 Sample $(\hat{p}_1 - \hat{p}_2)$

DATA

- qulitative
  نوعية
  "Catagorical"
- quantitave
  كمية

Examples:
1. Blood group (A, B, O, AB)
2. Sick / not sick
3. Male / female

μ, σ^2, p → $\bar{X}$, S^2, $\hat{p}$

population 3 parameters

* Test of hypothesis between 2 Sample proportions $(\hat{p}_1 - \hat{p}_2)$

يوجد طريقتين

- normal method z $\begin{pmatrix} n p_1 q_1 > 5 \\ n p_2 q_2 > 5 \end{pmatrix}$
- Contigency table

Note that
These two approaches are *equivalent* in that they always yield the same *p*-values, so which one is used is a matter of convenience.

$H_0: p_1 = p_2$ Vs $H_1: p_1 \neq p_2$

Test Stat

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - \cancel{(p_1 - p_2)}^{\,0}}{\sqrt{p^* q^* \left(\frac{1}{n} + \frac{1}{n}\right)}}$$

$$p^* : \text{pooled proportion} = \frac{x_1 + x_2}{n_1 + n_2}$$

$\boxed{X : n_1 * \hat{p}_1}$ $\boxed{y = n_2 * \hat{p}_2}$ →

في حال عدم اعطاء قيمة
x_1 و x_2

$$Z_{corr} = \frac{|\hat{p}_1 - \hat{p}_2| - \left(\frac{1}{2n_1} + \frac{1}{2n_2}\right)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

For a two-sided level α test,
if $z > z_{1-\alpha/2}$ ← اي بنحسبها
then reject H_0;
if $z \le z_{1-\alpha/2}$
then accept H_0.

p value > α
accept H_0 and
rject H_1

p value < α
reject H_0 and
accept H_1

Example : 2 types of medication for hives are being tested to determine if there is a difference . Twenty out of a random sample of 200 adults given medication " A" still had lives 30 mintues after taking the medication . Twelve out of another Random sample of 200 adults given medication "B" still had hives 30 mins after taking the medication . Test using 1% significance level when :

$\alpha = 0.01$

1.no continuity correction applied

$H_0 : p_1 = p_2$ vs. $H_1 : p_1 \neq p_2$

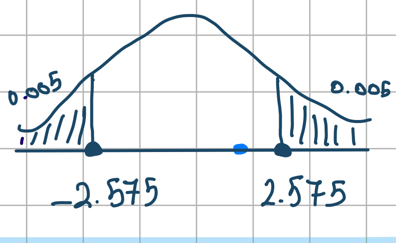
$x = 20$	$x_2 = 12$
$n = 200$	$n = 200$
$\hat{p}_1 = \frac{1}{10} = 0.1$	$\hat{p}_2 = \frac{6}{100} = 0.06$

بناخذ اول منزلتين

$$Z = \frac{(0.1 - 0.06)}{\sqrt{0.08 * 0.92 \left(0.01\right)}} = 1.47$$

$\alpha = 0.01$

$\frac{\alpha}{2} = 0.005$



So we accept H_0 and Reject H_1 $\Leftarrow$

2.54	.9945	.0055
2.55	.9946	.0054
2.56	.9948	.0052
2.57	.9949	.0051
2.58	.9951	.0049
2.59	.9952	.0048
2.60	.9953	.0047
2.61	.9955	.0045

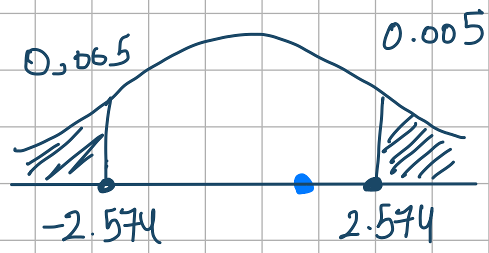
$+ \div 2$

2.Applied Z corr on your answer

$H_0: p_1 = p_2$ Vs $H_1: p_1 \neq p_2$

$$Z = \frac{|0.1 - 0.06| - \left(\frac{1}{400} + \frac{1}{400}\right)}{\sqrt{0.08 * 0.92 * \left(\frac{2}{400}\right)}} = 1,29$$

∴ we accept Ho and Reject H_1 ⟸

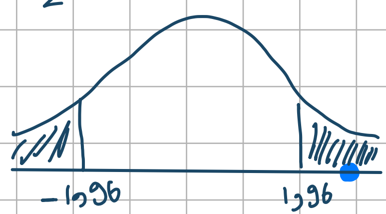


كتاب

Example : A study looked of the effect of OC we on heart disease in women (40-44) y/o . The Research found that among (5000) current OC users at baseline 13 women develops myocardial infarction (MI) over 3 years period wherease among (10 000) non OC users 7 asses the statistcal significance of the results

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$



$x = 13$ $\quad$ $x = 7$

$n = 5000$ $\quad$ $n = 10000$

$\hat{p} = 0,0026$ $\quad$ $\hat{p} = 0,0007$

$$Z = \frac{|0,0026 - 0,0007| - \left(\frac{1}{10000} + \frac{1}{20000}\right)}{\sqrt{0,0013 * 0.9987 * \left(\frac{1}{5000} + \frac{1}{10000}\right)}}$$

$$\frac{1,9 * 10^{-3} - 1,05 * 10^{-3}}{6.24093743 * 10^{-4}} = 2.77$$

we reject Ho and accept H_1

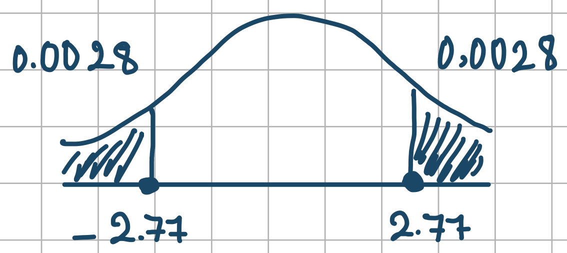
p-value = 2 * 0.0028

$\boxed{= 0.0056}$

p value < α

reject H_0 and accept H_1

highly significant



Guidelines for Judging the Significance of a *p*-Value

If $.01 \le p < .05$, then the results are *significant*.
If $.001 \le p < .01$, then the results are *highly significant*.
If $p < .001$, then the results are *very highly significant*.
If $p > .05$, then the results are considered *not statistically significant* (sometimes denoted by NS).
However, if $.05 < p < .10$, then a trend toward statistical significance is sometimes noted.

VH ← 0.001 —hs— 0.01 —s— 0.05 —— 0.10

→ not statistically significant

↓ trend toward statistical significant

Example : police officers in new york city can stop a driver Who is not wearing their seat belt . In Boston , police officers can issue citafions to driver not wearing their seat belts only if the driver has been stopped for another violation Data from random samples of femal in 2002 is summarized as the following :

City	Drivers	wearing seatbelts
Boston	117	68
New york	220	183

$H_0: p_1 = p_2$ vs $H_1: p_1 \neq p_2$

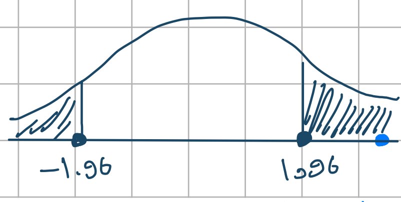
Is there compelling evidence to conclude a difference is Rate of drivers wear their seatbelts in Boston as compared to newyork?
(Assume cotinunity corredction is applied , use a= 0.05)

$$\hat{p}_1 = \frac{68}{117} = 0.58 \qquad \hat{p}_2 = \frac{183}{220} = 0.83$$

$$Z = \frac{|0.58 - 0.83| - \left(\frac{1}{2*117} + \frac{1}{2*220}\right)}{\sqrt{0.74 * 0.26 * \left(\frac{1}{117} + \frac{1}{220}\right)}} = 4.85$$

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$



$\therefore$ we reject H_0 and accept H_1

② The contingency table

Example: observed table (2x 2 table)

	Rt - hand	Lt - hand	Total
Males	43 O_{11}	9 O_{12}	52
Females	44 O_{21}	4 O_{22}	48
Total	87	13	100

52: Row margin

48: Row margin

87: Column margin

13: Column margin

100: Grand total

Row صف / Column عامود

Expected table

	Rt - hand	Lt - hand	Total
Males	$\frac{87*52}{100} = 45.24$ E_{11}	$\frac{13*52}{100} = 6.76$ E_{12}	52
Females	$\frac{87*48}{100} = 41.76$ E_{21}	$\frac{13*48}{100} = 6.24$ E_{22}	48
Total	87	13	100

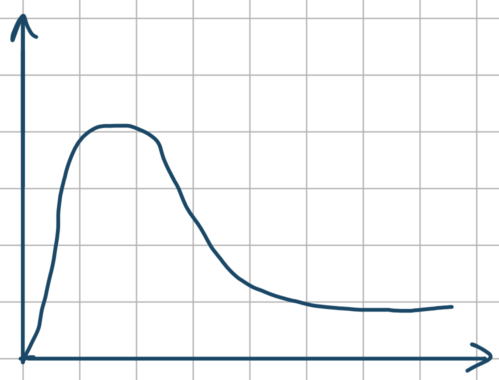
NOTE: $E = \frac{R * C}{\text{grand total}}$

بعد ما طلعنا ال Fexpected table كيف نعمل test hypothesis

$H_0: p_1 = p_2$ Vs $H_1: p_1 \neq p_2$

test stat
Contigency table
chi squared دائما

* Chi squared



⇒ skewed to the right

⇒ All values are positive

⇒ $d.f = (R-1)(C-1)$

<u>NOTE:</u> Always d.f in 2x 2 Contingency table is equale to 1

المساحة على يسار الرقم

TABLE 6 Percentage points of the chi-square distribution ($\chi^2_{d,u}$)[a]

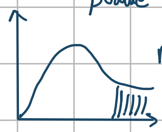
	u													
d	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995	.999
1	0.0^4393[b]	0.0^3157[c]	0.0^3982[d]	0.00393	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88	10.83
2	0.0100	0.0201	0.0506	0.103	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60	13.81
3	0.0717	0.115	0.216	0.352	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86	18.47
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75	20.52
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55	22.46
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28	24.32
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.95	26.12
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	4.60	5.23	6.27	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00	45.32

* General notes

① Always the test is Right tailed test

• لما بدي احسب ال Rejection region
دائماً Right tailed .

• لما بدي احسب ال pvalue
دائماً Right tailed
ما بضرب ب 2



② test statistic

$$\chi^2 = \frac{\Sigma (O-E)^2}{E}$$

$$= \frac{(O_{11}-E_{11})^2}{E_{11}} + \frac{(O_{12}-E_{12})^2}{E_{12}} + \frac{(O_{21}-E_{21})^2}{E_{21}} + \frac{(O_{22}-E_{22})^2}{E_{22}}$$

$$\chi^2_{Corr} = \frac{\Sigma(|O-E|-\frac{1}{2})^2}{E}$$

③ Always the Expected values are more than 5

Example : ***the following table list Results from an experiment designed to to test the ability of dogs to use their extraordinary sense of smell to detect malaria in samples of children's socks . The accompaying information shows the following :***

	Malaria was present	Malaria wasn't present	total
Dog was Correct	123	131	254
Dog was wrong	52	14	66
total	175	145	320

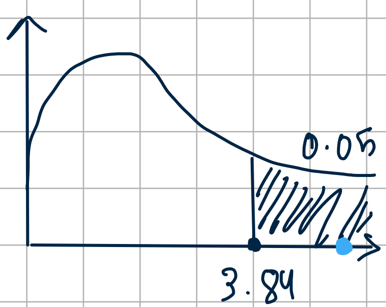
The p- value ,and then state the conclusion about the null hypothesis .

$$= \frac{(123-138.91)^2}{138.91} + \frac{(131-115.09)^2}{115.09} + \frac{(52-36.09)^2}{36.09} + \frac{(14-29.91)^2}{29.91} = \boxed{19.49}$$

	malaria	malria wasn't present
Dog was ✓	138.91	115.09
Dog was ✗	36.09	29.91

$\alpha = 0.05$

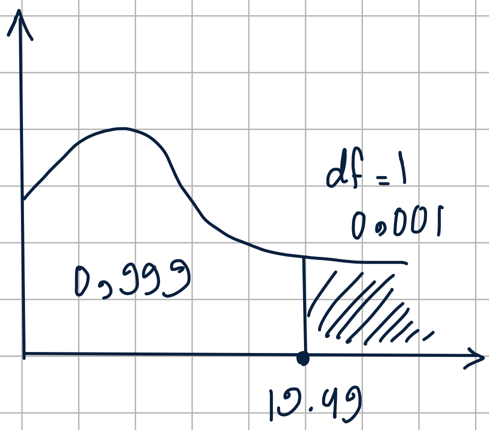
$d.f = 1$



So we reject H_0 and accept H_1

p- value = 0.001

highly significant



$$\chi^2_{corr} = \frac{\left(|123 - 138.91| - \frac{1}{2}\right)^2}{138.91} + \frac{\left(|131 - 115.09| - \frac{1}{2}\right)^2}{115.09} + \frac{\left(|52 - 36.09| - \frac{1}{2}\right)^2}{36.09} + \frac{\left(|14 - 29.91| - \frac{1}{2}\right)^2}{29.91}$$

Example -: suppose we want to know if the Rate of smoking in males is different from Females in a sample of 203 Jordanian the observed values set as the following :

Use a= 0.001

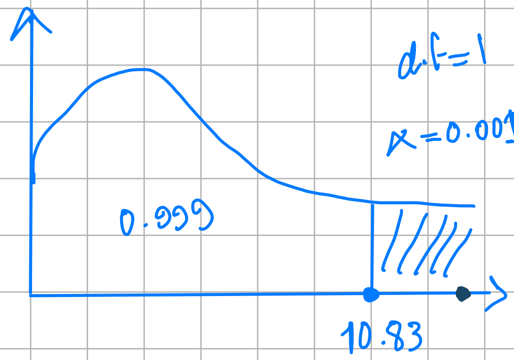
	Smoker	non smoker	total
Males	72	44	116
females	34	53	87
total	106	97	203

$$Z_{corr} = \frac{(|72-60.57|-\frac{1}{2})^2}{60.57} + \frac{(|44-55.43|-\frac{1}{2})^2}{55.43} + \frac{(|34-45.43|-\frac{1}{2})^2}{45.43} + \frac{(|53-41.57|-\frac{1}{2})^2}{41.57} = \boxed{12,74}$$

	smoker	not smoker
Males	60,57	55,43
female	45,43	41,57

$d.f = 1$

$\alpha = 0.001$



∴ we reject H_0 and accept H_1

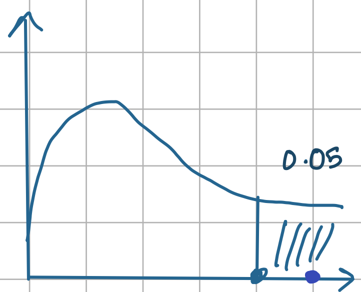
كتاب

Example : compute the expected table for the breast cancer data shown in the folllowing table :

	≥ 30	< 29	total
Case	683	2537	3220
Control	1498	8747	10245
total	2181	11284	13465

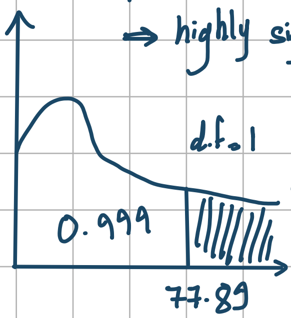
	≥ 30	< 29	total
Case	521.6	2698.4	3220
Control	1659.4	8585.6	10245
total	2181	11284	13465

$\chi^2_{corr} = \boxed{77.89}$



$\therefore$ we reject H_0 and accept H_1

p value = 0.001
→ highly significant



كتاب

Example : Assess the OC- MI data for statistical significane , using confingency table approach ?

2 × 2 contingency table for the OC–MI data in Example 10.6

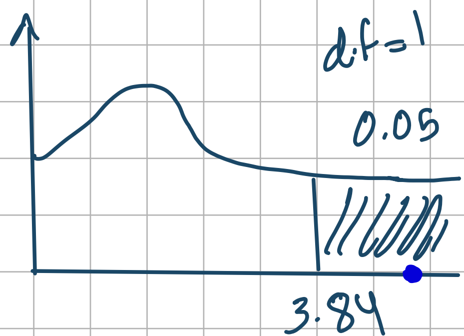
	MI incidence over 3 years		
OC-use group	Yes	No	Total
Current OC users	13	4987	5000
Never-OC users	7	9993	10,000
Total	20	14,980	15,000

الحل:

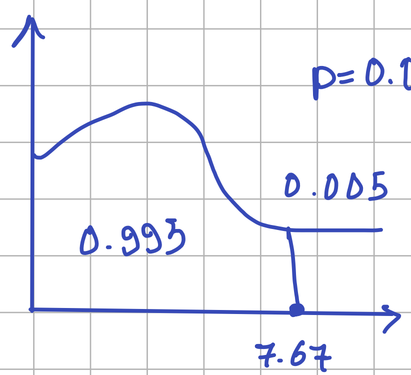
2 × 2 contingency table for the OC–MI data in Example 10.6

	MI incidence over 3 years		
OC-use group	Yes	No	Total
Current OC users	6.7	4993.3	5000
Never-OC users	13.3	9986.7	10,000
Total	20	14,980	15,000

$$x^2_{corr} = \frac{(|13-6.7|-\frac{1}{2})^2}{6.7} + \frac{(|4987-4993.3|-\frac{1}{2})^2}{4993.3} + \frac{(|7-13.3|-\frac{1}{2})^2}{13.3}$$
$$+ \frac{(|9993-9986.7|-\frac{1}{2})^2}{9986.7} = \boxed{7.67}$$



we reject Ho and accept H₁



highly significant

NOTES:

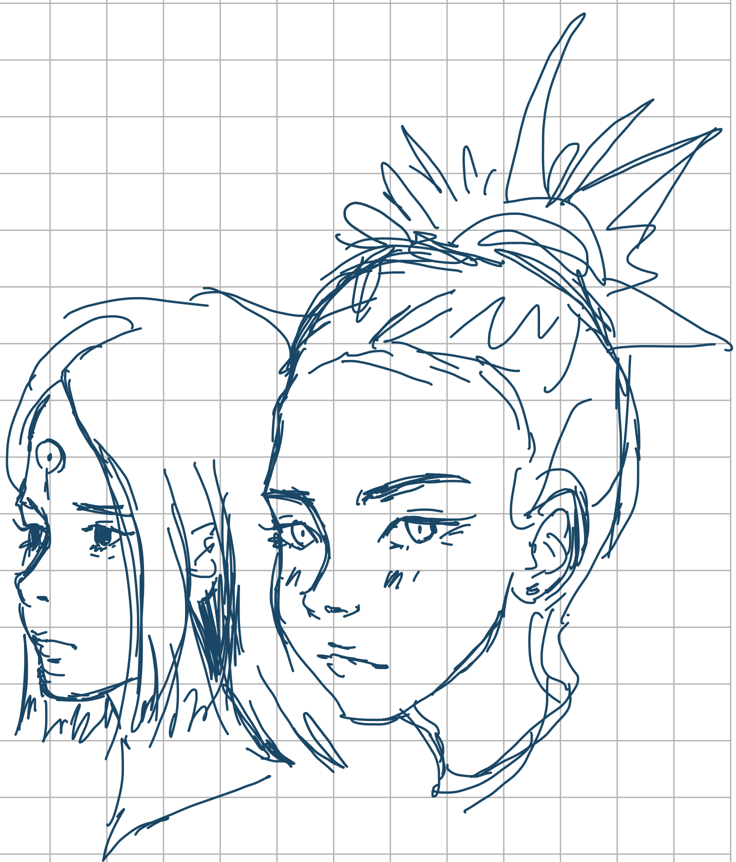
① The purpose of Contingency table is to Summarize a large set of data.

② x^2_{corr} is Called yates - Corrected chi squared.

③ Always the expected values are more than 5

④ The Contingency table is often used to determine If the two variable have an association

⑤ H_0 : If they are independent
H_1 : If they are dependent



* R x C Contingency table

	x	y	z	K
a				
B				
C				

من ناحية حساب ال expected
نفس جدول ال (2x2)

1. $E = \frac{R * C}{\text{Total}}$

2. test stat : Always χ^2 العادية

$$\chi^2 = \frac{\Sigma (O-E)^2}{E}$$

3. degree of freedom in R x C table is Calculated as the following : $(R-1) \times (C-1)$

4. The conditions of the table :

Ⓐ No cells has an expected <1

Ⓑ No more than 1/5 of the cells have expected value less than 5 ⟹

شرح للنقطة هاي "بتحكي هاي النقطة
أنو هو ما بجبرك بال table (RxC)
تكون كل القيم أكبر من 5 بس في
شرط أنو اي اشي بين ال (1-5)
مايزيد عن خمس البيانات يعني
بشكل أوضح في جدول الـ (2x2)
كان شرط أنو ال expected value
كلهم يكونو أكبر من 5 ولكن في
جدول ال (RxC) عادي لو كانو
مو أكبر من 5 بس لازم يكونو
فوق ال 1 وأي قيمة بين ال (1-5)
مسموح تكون موجودة ولكن مايزيد عددهم
عن خمس القيم"

• نفس طريقة حل جدول (2 x 2)

الهم انه هون ما بزبط نستخدم test stat Corr

Example : ***Assess the statistical significance in 300 persons giving the following :***

Table of Observed Values

Qualification / Martial status	Middle school	High school	Bachelor's	Master's	ph.D	Total
Never married	18	36	21	9	6	90
married	12	36	45	36	21	150
Divorced	6	9	9	3	3	30
Widowed	3	9	9	6	3	30
Total	39	90	84	54	33	300

Expected

Qualification / Martial status	Middle school	High school	Bachelor's	Master's	ph.D	Total
Never married	11,7	27	25.2	16.2	9.9	90
married	19,5	45	42	27	16.5	150
Divorced	3,9	9	8.4	5.4	3.3	30
Widowed	3,9	9	8.4	5.4	3.3	30
Total	39	90	84	54	33	300

Test stat

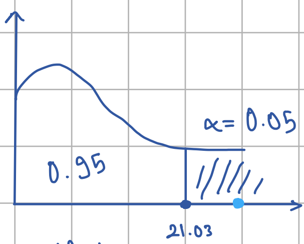
$$x^2 = \sum \frac{(O-E)^2}{E} =$$

$$3.39\ \overset{2.88}{\frac{(18-11.7)^2}{11.7}} + \overset{1.30}{\frac{(12-19.5)^2}{19.5}} + \overset{0.20}{\frac{(6-3.9)^2}{3.9}} + \overset{3}{\frac{(3-3.9)^2}{3.9}} + \overset{1.8}{\frac{(36-27)^2}{27}} + \overset{0}{\frac{(36-45)^2}{45}} + \frac{(9-9)^2}{9} +$$

$$0\ \frac{(9-9)^2}{9} + \overset{0.7}{\frac{(21-25.2)^2}{25.2}} + \overset{0.21}{\frac{(45-42)^2}{42}} + \overset{0.042}{\frac{(9-8.4)^2}{8.4}} + \overset{0.042}{\frac{(9-8.4)^2}{8.4}} + \overset{3.2}{\frac{(9-16.2)^2}{16.2}} + \overset{3}{\frac{(36-27)^2}{27}} +$$

$$1.06\ \frac{(3-5.4)^2}{5.4} + \underset{0.06}{\frac{(6-5.4)^2}{5.4}} + \underset{1.53}{\frac{(6-9.9)^2}{9.9}} + \underset{1.22}{\frac{(21-16.5)^2}{16.5}} + \underset{0.03}{\frac{(3-3.3)^2}{3}} + \underset{0.03}{\frac{(3-3.3)^2}{3}} = \boxed{23.57}$$

we reject H_0 and accept H_1



H_0: Martial status independent from qualification

H_1: Martial status dependent from qualification

Example : Assess the statistical significance of the data between 2 variables , the age of first birth and the prevefance of breast cancer

Data from the international study in Example 10.4 investigating the possible association between age at first birth and case–control status

Case–control status	Age at first birth <20	20–24	25–29	30–34	≥35	Total
Case	320	1206	1011	463	220	3220
Control	1422	4432	2893	1092	406	10,245
Total	1742	5638	3904	1555	626	13,465
% cases	.184	.214	.259	.298	.351	.239

Case–Control	<20	20–24	25–29	30–34	≥ 35	Total
Case	416.58	1348.26	933.60	371.86	149.70	3220
Control	1325.42	4289.74	2970.40	1183.14	476.30	10245
Total	1742	5638	3904	1555	626	13465

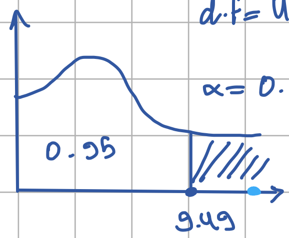
$$\chi^2 = \frac{\overset{22.40}{(320-416.58)^2}}{416.58} + \frac{\overset{15.01}{(1206-1348.26)^2}}{1348.26} + \frac{\overset{6.42}{(1011-933.60)^2}}{933.60} + \frac{\overset{22.34}{(463-371.86)^2}}{371.86}$$

$$+ \frac{\overset{33.01}{(220-149.70)^2}}{149.70} + \frac{\overset{7.04}{(1422-1325.42)^2}}{1325.42} + \frac{\overset{4.72}{(4432-4289.74)^2}}{4289.74} + \frac{\overset{2.02}{(2893-2970.40)^2}}{2970.40} +$$

$$\frac{\overset{7.02}{(1092-1183.14)^2}}{1183.14} + \frac{\overset{10.34}{(466-476.30)^2}}{476.30} = \boxed{130.32}$$

d.f = 4

$\alpha = 0.05$

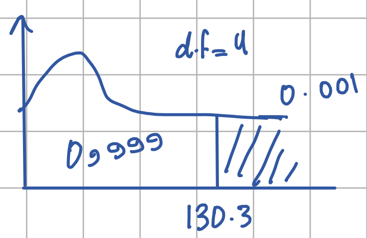


∴ we reject H_0 and accept H_1 ⇐

(There is an association Between the first birth age and breast Cancer)

p-value = 0.001

→ highly significant



Example: Determine to the 5% significane level whether school and are dependent

$\alpha = 0{,}05$

school \ Grade	A	B	C	Total
x	18	12	20	50
y	26	12	32	70
Total	44	24	52	120

Answer:

Expected

school \ Grade	A	B	C	Total
x	18.33	10	21.67	50
y	25.67	14	30,33	70
Total	44	24	52	120

$$x^2 = \frac{(18-18.33)^2}{18.33} + \frac{(12-10)^2}{10} + \frac{(20-21.67)^2}{21.67} +$$

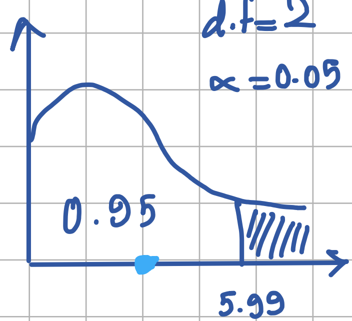
$$\frac{(26-25.67)^2}{25.67} + \frac{(12-14)^2}{14} + \frac{(32-52)^2}{52} = \boxed{0.916}$$

H_0: school is independent from the Grade

H_1: school is dependent on grade

d.f = 2

$\alpha = 0.05$



$\therefore$ we accept H_0 and reject H_1

(There is no association Between the school and Grade)

* Goodness of fit test (chi-squared)

اختبار حسن المطابقة ⟵

⟹ Approximation of discrete Random variable to Continous Random variable.

D ⟶ n

discrete بعطيك بيانات ⟶ ويطلب منك تحولها

لـ normal

لازم يكون في مساواة

$\Longrightarrow p(x < 16)$ [Discrete]

$\Longrightarrow p(x \leq 15)$ 15   16   (يمين)

١) أول خطوة نضع مساواة بحيث اخذ الرقم الي اصغر من 16 مباشرة

$p(x \leq 15)$ x   15

x أكبر من 15 اذا تقع على يمين

٢) نحدد اذا الرقم يقع على يمين او يسار

٣) اذا يقع على يمين بزيد $\frac{1}{2}$ اذا يسار بطرح $\frac{1}{2}$

$\Longrightarrow p(x \leq 15.5)$ [Continuity Correction]

NOTE

① $p(x \leq a) \Longrightarrow p(x \leq a+0.5)$

② $p(x \geq a) \Longrightarrow p(x \geq a-0.5)$

③ $p(a \leq x \leq b) \Longrightarrow p(a-0.5 \leq x \leq b+0.5)$

Example :

① $p(x > 18)$ [Discrete]

$= p(x \geq 19)$

$= p(x > 18.5)$

② $p(18 < x < 26)$

$= p(19 \leq x \leq 25)$

$= p(19-0.5 \leq x \leq 25+0.5)$

③ $p(18 \leq x < 26)$

$p(18 \leq x \leq 27)$

$= p(18-0.5 \leq x \leq 27+0.5)$

④ $p(18 < x \leq 25)$

$p(19 \leq x \leq 25)$

$p(19-0.5 \leq x \leq 25+0.5)$

Example : If the $\mu = 20$, $\sigma^2 = 16$

find :

① $p(x < 26)$

$p(x \leq 25) \Rightarrow p(x \leq 25.5)$

$\Rightarrow p\left(z \leq \frac{25.5 - 20}{4}\right)$

$\Rightarrow p(z \leq 1.25) = 0.8944$

1.23	.8907	.1093
1.24	.8925	.1075
1.25	.8944	.1056
1.26	.8962	.1038

② $p(18 < x \leq 26)$

$= p(19 \leq x \leq 26)$

$p(18.5 \leq x \leq 26.5)$

$p(x \leq 26.5) - p(x \leq 18.5)$

$p\left(z \leq \frac{26.5 - 20}{4}\right) - p\left(x \leq \frac{18.5 - 20}{4}\right)$

⋮ ⋮

ضغط الدم الأنبساطي

EXAMPLE 10.46

Hypertension Diastolic blood-pressure measurements were collected at home in a community-wide screening program of 14,736 adults ages 30–69 in East Boston, Massachusetts, as part of a nationwide study to detect and treat hypertensive people [6]. The people in the study were each screened in the home, with two measurements taken during one visit. A frequency distribution of the mean diastolic blood pressure is given in Table 10.20 in 10-mm Hg intervals.

We would like to assume these measurements came from an underlying normal distribution because standard methods of statistical inference could then be applied on these data as presented in this text. How can the validity of this assumption be tested?

TABLE 10.20 Frequency distribution of mean diastolic blood pressure for adults 30–69 years old in a community-wide screening program in East Boston, Massachusetts

Group (mm Hg)	Observed frequency	Expected frequency	Group	Observed frequency	Expected frequency
<50	57	69.0	≥80, <90	4604	4538.6
≥50, <60	330	502.5	≥90, <100	2119	2545.9
≥60, <70	2132	2018.4	≥100, <110	659	740.4
≥70, <80	4584	4200.9	≥110	251	120.2
			Total	14,736	14,736

Hypertension Compute the expected frequencies for the data in Table 10.20, assuming an underlying normal distribution.

Solution: Assume the mean and standard deviation of this hypothetical normal distribution are given by the sample mean and standard deviation, respectively ($\bar{x} = 80.68$, $s = 12.00$). The expected frequency within a group interval from a to b would then be given by

$P(X < 50)$

$= P(X \leq 49)$

$P(X \leq 49.5) \Rightarrow P\left(Z \leq \frac{49.5 - 80.68}{12}\right)$

$\Rightarrow P(Z \leq -2.598)$

قرب الرقم = 2.60

$= 0.0047$

2.59	.9952	.0048
2.60	.9953	.0047
2.61	.9955	.0045
2.62	.9956	.0044

هيك بنكون طلعنا ال expected $\Rightarrow$ $0.0047 * \underbrace{14736}_{\text{Total}} = 69$

$\Rightarrow p(50 \leq x < 60)$

$p(50 \leq x \leq 59)$

$p(x \leq 59.5) - p(x \leq 49.5)$

$p\left(z \leq \frac{59.5 - 80.68}{12}\right) - p\left(z \leq \frac{49.5 - 80.68}{12}\right)$

$p(z \leq -1.77) - p(z \leq -2.60)$

$0.0384 - 0.0047$

$= 0.0337$

1.76	.9608	.0392
1.77	.9616	.0384
1.78	.9625	.0375

2.58	.9951	.0049
2.59	.9952	.0048
2.60	.9953	.0047

$0.0337 * 14736$

$= 502.5$

Test stat

بعد ماتطلع ال expeted و ال square بتطلع ال ←

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

$$\frac{(57 - 69)^2}{69} + \dots\dots\dots\dots + \frac{(251 - 120.2)^2}{120}$$

$= \boxed{326.2}$

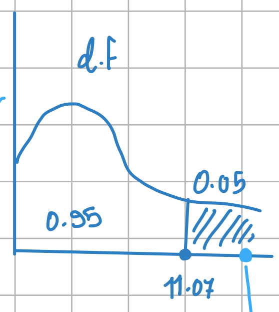
(μ, σ^2) estimater

$\alpha = 0.05$

$D.f = g - k - 1$ group

$8 - 2 - 1$

$= 5$



we reject H_0 and accept H_1

$\Rightarrow$ normal method doesn't provide an adequate fit to the data.

NOTES

1. we study the fit of the test to a data

2. Expected
   - → equality
   - → Continuity correction
   - → probability
   - → probability * grand total

لإيجاد

expected ال

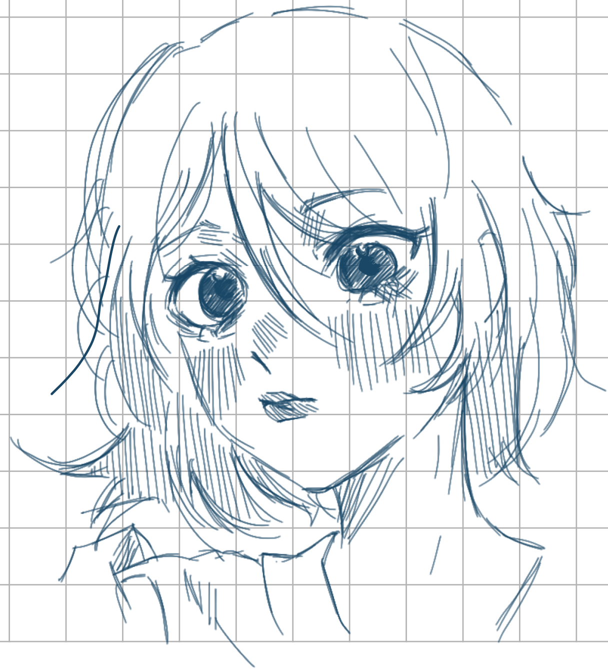
3. $\chi^2 = \sum \frac{(O-E)^2}{E}$

test stat

4. $Df = g - K - 1$

group عدد ال

عدد المقاييس



$\bar{x}$ / S

Example : *The mean weights of a sample of 200 patients is 52 kGs and the standard deviation is 3 kGs*

$\bar{X} = 52$

$S = 3$

weight	$w < 45$	$45 \leq w < 50$	$52 \leq w < 55$	$55 \leq w < 60$	$w \geq 60$
frequency	12	44	82	53	9
Expected	1.24	39.42	118.7	39.42	1.24

We would like to assume that these measurments came from the normal distibution how can the validity of this assumption be tested?

$P(w < 45)$

$= P(w \leq 44)$

$P\left(z \leq \frac{44.5 - 52}{12}\right) \Rightarrow P(z \leq -2.5)$

$= 0.0062$

$= 0.0062 \times 200 = 1.24$

$\Rightarrow P(45 \leq x < 50)$

$P(45 \leq x \leq 49)$

$P(44.5 \leq x \leq 49.5)$

$P(x \leq 49.5) - P(X \leq 44.5)$

$P\left(x \leq \frac{49.5 - 52}{3}\right) - P\left(x \leq \frac{44.5 - 52}{3}\right)$

$P(z \leq -0.83) - P(z \leq -2.5)$

$= 0.2033 - 0.0062 = 0.1971 * 200 = 39.42$

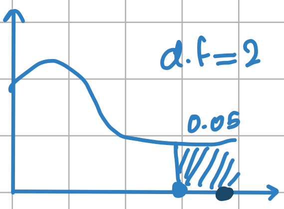
$$x^2 = \frac{(12-1.24)^2}{1.24} + \cdots\cdots\cdots + \frac{(9-1.24)^2}{1.24}$$

$= \boxed{158.49}$

$df = g - k - 1$

$= 5 - 2 - 1$

$= 2$



$\therefore$ we reject H_0 and accept H_1