Editing file

Feedback

Statistical tests for qualitative variables

Objectives:

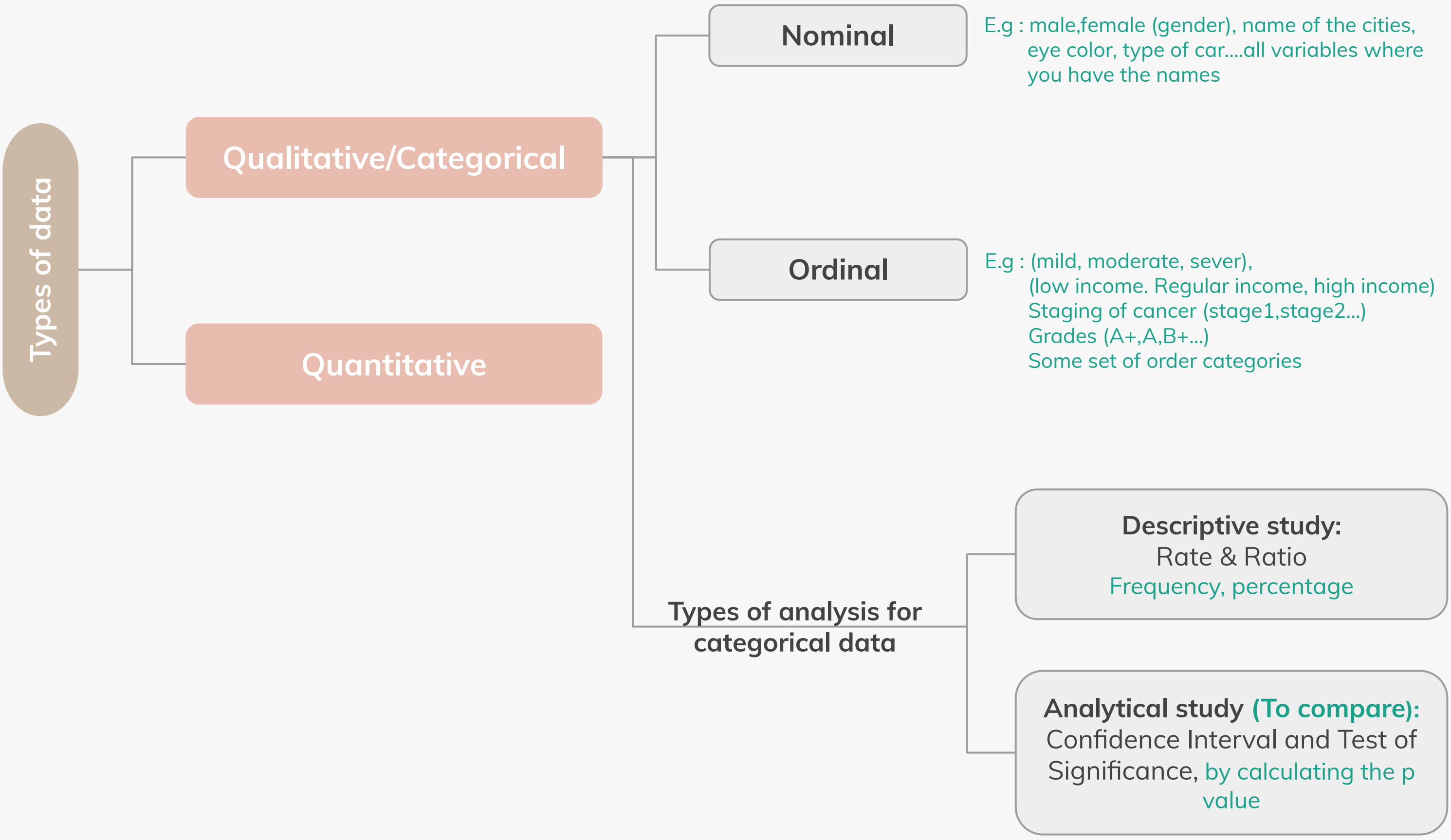
1. Able to understand the factors to apply for the choice of statistical tests in analysing the data.
2. Able to apply appropriately Z-test, Chi-square test, Fisher's exact test & Macnemar's Chi-square test.
3. Able to interpret the findings of the analysis using these four tests.

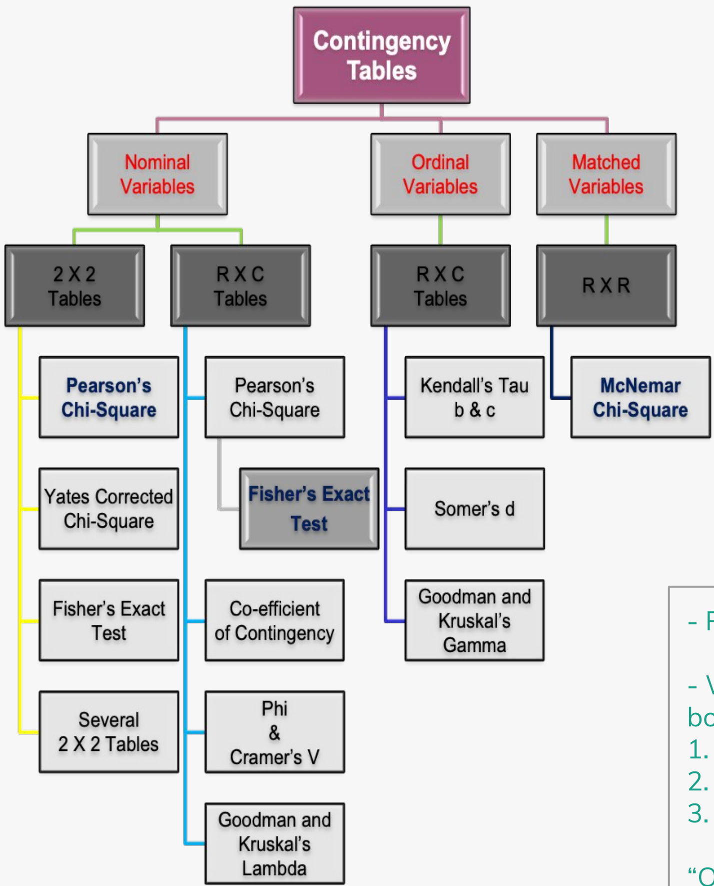
Click here for the practical

Color Index:

- Boys' Slides
- Girls' Slides
- Doctors Notes
- Golden Notes
- Important
- Extra

Statistical Test





- R x C table = rows and columns

- We only will cover the blue boxes in this lectures:
1. Pearson's Chi-square
2. Fisher's Exact Test
3. McNemar Chi-square

"Other things are not necessary for now"

Statistical Test

In the exam everything will be given, there will be no need for the calculations, you just have to pick which one is the appropriate statistical test.

The appropriate statistical test :

Choosing appropriate statistical test is based on three aspects of the data:

- ✓ Types of variables.
- ✓ Number of groups being compared.
- ✓ Sample size.

Remember all statistical tests has to satisfy some of the requirements or assumptions from the data, otherwise it is a misuse of the statistical test

<table>
<tr><th>Test</th><th>Study variable</th><th>Outcome variable</th><th>Comparison</th><th>Sample size
Requirement</th><th>Expected frequency
Requirement</th></tr>
<tr><td>Chi-square Or
Pearson's Chi square</td><td rowspan="4">Qualitative
(categorical)</td><td rowspan="4">Qualitative
(categorical)</td><td>Two or more proportions
E.g : (Two proportions): Prevalence of exercise among female and male.
(More than two proportions):
1- Prevalence of exercise among gp A, gp B, female gp.
2- Prevalence of hypertension among 4 age gps.</td><td>> 20</td><td>>5</td></tr>
<tr><td>Fisher's exact test</td><td>Two proportions
For Categorical nominal data, E.g:
When you put a 2x2 table, such as case-control, cohort, cross-sectional or anything...and you want to study the association between smoking & developing cancer
The columns are the outcome variable
The rows are the exposure variable,
You'll get two groups for each the outcome & exposure
Exposure (smoking)→ smoker or non-smoker
Outcome (cancer)→ present (cases) or absent (contr</td><td>< 20</td><td></td></tr>
<tr><td>Mcnemar's test (for paired samples)</td><td>Two proportions
Two proportion means → 2x2 table
Two study designs will become relevant→ matched case control to remove the confounding effect (ex, most commonly age & gender) & cross over trial
Matching in case control is most likely done during the study design if not then it's done in the analysis</td><td>Any</td><td></td></tr>
<tr><td>Z-test</td><td>- Sample proportion with population proportion
- Two sample proportions</td><td>Larger in each group (>30)</td><td></td></tr>
</table>

Chi-square Test

Or Pearson's Chi square

Purpose

To find out whether the association between two categorical variables are statistically significant.

Odds ratio quantifies the association (measure), chi square tests the association (is it there or not)

Null hypothesis

There is no association between two variables.

Equation:

"No need to worry, the formula is just for explanation, not for the exam"

$$\chi^2 = \sum \left[\frac{(O - E)^2}{E} \right]$$

Figure for each cell

- The summation is over all cells of the contingency table consisting of **r** rows and **c** columns.
- **O** is the <u>observed</u> frequency. (What we observe from the data; how many smoker? Non smokers? Out of that how many have the outcome?)
- $\hat{E}$ is the <u>expected</u> frequency. (We get the expected data from this formula)

$$\hat{E} = \frac{\left(\text{total of row in which the cell lies}\right) \cdot \left(\text{total of column in which the cell lies}\right)}{\text{(total of all cells)}}$$

- The degrees of freedom are **df** = **(r-1)(c-1)**.

r = number of rows
c - number of columns

- Reject H_o if $x^2 > x^2_{.\alpha,df}$

Degrees of freedom :
Chi-square test→ related to number of r & c → (r-1)(c-1)
Student t-test → related to number of sample size → n-1 or n1+n2-1

Click here to check out the degree of freedom practical lecture

Requirements:

Prior to using the chi square test, there are certain requirements that must be met:

- The data must be in the form of <u>frequencies</u> counted in each of a set of categories. Percentages cannot be used. Most commonly used test & most commonly misused. In some literature they use percentages instead of frequencies conveniently to escape from the low sample size of the data and percentages might lead to misreadings. Example: if we have in one category 10 cases and 3 of them got exposed which is calculated as 30% some will write 30 in the calculation instead of 3. 2 responds out of 5 this is a small frequency but the percentage will be nearly 40%
- The total number observed must exceed 20. The higher the better
- The expected frequency under the H_0 hypothesis in any one fraction must not normally be less than 5. It can be flexible if one cell out of 4 is less than 5 but not more than one cell.
- All the observations must be independent of each other. In other words, one observation must not have an influence upon another observation. Example: looking at smokers and non-smokers, female and male or people who exercise and people who don't which is mutually exclusive and won't overlap.

Application:

- Testing for independence (or association). Know the difference between "<u>tests</u> of association" such as statistical tests (Chi-square, Fisher's and McNemar test) & "<u>measure</u> of association" such as Odds Ratio (for prospective study, case control and cross sectional) and Relative Risk (for retrospective study and RCT), the first will only tell you if there is an association or not, while the second one will tell how much is the association.
- Testing for homogeneity.
- Testing of goodness-of-fit. (not required)

Chi-square Test

Problem:

- **Objectives**: smoking is a risk factor for MI.
- **Null hypothesis**: smoking does not cause MI.

Remember the causation! Even if there is a significant association between smoking and MI, it DOESN'T mean smoking ALONE will cause MI‹ there is a big criteria to rule out causation.

	Disease (MI)	No disease (MI)	Total
Smoker	29	21	50
Non-smoker	16	34	50
Total	45	55	100

	MI	Non-MI
Smoker	29 O / E 22.5	21 O / E 27.5
Non smoker	16 O / E 22.5	34 O / E 27.5

$$\hat{E} = \frac{\left(\text{total of row in which the cell lies}\right)\cdot\left(\text{total of column in which the cell lies}\right)}{\text{(total of all cells)}}$$

- **Expected frequency**:

1. For cell a: 50 * 45 / 100 = 22.5
2. For cell b: 50 * 55 / 100 = 27.5
3. For cell c: 50 * 45 / 100 = 22.5
4. For cell d: 50 * 55 / 100 = 27.5

- Remember if you take the summation of the expected frequency of cells from the same row or column you will get a value similar to the total of that row or column for example If you add expected frequency of a and b you'll get 50

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

- You will subtract each expected frequency from its cell, square this value then divide by the expected frequency. Lastly summation of all the values.
- If difference between Observed & Expected frequency:
  large→ significant association
  small→ no significant association
- Degree of freedom: **df**= **(r-1) (c-1)** = (2-1) (2-1) = 1

So :

- Critical value **(look the table at page 13)** at 0.05 level of significance and with 1 **df** = 3.84
- Calculated value = 6.84 (when you get a larger chi square value → you get a smaller p-value, and smaller p-value means = statistically significant) (Smaller chi square value → you get a larger p-value, and larger p-value means = statistically insignificant)
- Calculated value 6.84 is greater than critical (table) value 3.84 at 0.05 level with 1 d.f.f.(degrees of freedom) Hence we reject our H_0 and conclude that there is highly statistically significance association between smoking and MI.

Problem:

- Find out whether the gender is equally distributed among each age group

Gender (Nominal)	Age (Ordinal)			Total
	<30	30-45	>45	
Male	60 (60)	20 (30)	40 (30)	120
Female	40 (40)	30 (20)	10 (20)	80
Total	100	50	50	200

Chi-square Test

Problem: Test for homogeneity (similarity)

- To test similarity between frequency distribution or group. It is used in assessing the similarity between non-responders and responders in any survey

Age (yrs)	Responders	Non-responders	Total
<20	76 (82)	20 (14)	96
20-29	288 (289)	50 (49)	338
30-39	312 (310)	51 (53)	363
40-49	187 (185)	30 (32)	217
>50	77 (73)	9 (13)	86
Total	940	160	1100

Problem: Association between DM and heart disease

Contradictory opinions:

- A diabetic's risk of dying after a first heart attack is the same as that of someone without diabetes. There is no link between diabetes and heart disease.

Vs :

- Diabetes takes a heavy toll on the body and diabetes patients often suffer heart attacks and strokes or die from cardiovascular complications at a much younger age.

- So we use hypothesis test based on the latest data to see what's the right conclusion.
- There are a total of 5167 managed-care patients, among which 1131 patients are non-diabetics and 4036 are diabetics (Nominal data). Among the non-diabetic patients, 42% of them had their blood pressure properly controlled (therefore it's 475 of 1131). While among the diabetic patients only 20% of them had the blood pressure controlled (therefore it's 807 of 4036). The frequency is 807 of 4036 and 475 of 1131

	Controlled	Uncontrolled	Total
Diabetes	807	3229	4036
Non-diabetes	475	656	1131
Total	1282	3885	5167

Chi-square Test

Problem: Association between DM and heart disease

Data:

- Diabetes: 1 = Not have diabetes, 2= Have diabetes
- Control: 1 = Controlled, 2- Uncontrolled

Make sure to label the rows & columns of the 2x2 table correctly:
outcome→ columns, exposure→ rows

1st column→ first category (controlled)
2nd category (uncontrolled)

1st row→ exposed (diabetes)
2nd row→ non-exposed (non-diabetes)

Diabetes * Control crosstabulation

		Control		Total
		1.00	2.00	
Diabetes	1.00	807	3229	4036
	2.00	475	656	1131
Total		1282	3885	5167

Diabetes * Control crosstabulation (SPSS Results)

			Control		Total
			1.00	2.00	
Diabetes	1.00	Count % within DIABETES % within CONTROL % of Total	807 20.0% 62.9% 15.6%	3229 80.0% 83.1% 62.5%	4036 100.0% 78.1% 78.1%
	2.00	Count % within DIABETES % within CONTROL % of Total	475 42.0% 37.1% 9.2%	656 58.0% 16.9% 12.7%	1131 100.0% 21.9% 21.9%
Total		Count % within DIABETES % within CONTROL % of Total	1282 24.8% 100.0% 24.8%	3885 75.2% 100.0% 75.2%	5167 100.0% 100.0% 100.0%

Hypothesis Test:

- H_0: There is no association between diabetes and heart disease.
  (There is no association between diabetes and heart disease (or) Diabetes and heart disease are independent.).
- H_A: There is an association between diabetes and heart disease.
  (There is an association between diabetes and heart disease (or) Diabetes and heart disease are dependent.).
- Assume a significance level (false positive, alpha) of 0.05

Chi-square Test

Problem: Association between DM and heart disease

SPSS Output:

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided) P-Value	Exact. Sig. (2-sided)	Exact. Sig. (1-sided)
Pearson Chi-Square Most relevant	229.268 [b]	1	.000		
Continuity Correction [a]	228.091	1	.000		
Likelihood Ratio	212.149	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	229.224	1	.000		
N of Valid Cases	5167				

a. Computed only for a 2x2 table
b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 280.62.

In this data if you calculate the expected values, it will be greater than 5, since it is a huge data., That is what the SPSS has reported here. If the statement isn't appropriate you have to use fisher's exact test→ SPSS by default will do it

- The computer gives us a Chi-Square Statistic of 229.268.
- The computer gives us a p-value of .000, DOESN'T mean it is zero, it means that there is a number but it is smaller than 4 digits i.e., (<0.0001). Put symbol less than "<", don't use equal "=" and **don't** write .000
- Because our p-value is less than alpha (0.05), we would reject the null hypothesis.
- There is sufficient evidence to conclude that there is an association between diabetes and heart disease.

Fisher’s Exact Test

The method of Yates's correction was useful when manual calculations were done. Now different types of statistical packages are available. Therefore, it is better to use Fisher's exact test rather than Yates's correction as it gives exact result.

$$Fisher's\ Exact\ Test = \frac{R_1!R_2!C_1!C_2!}{n!a!b!c!d!}$$

!: Factorial
R: Row
C: Column
N: Total frequency

a: Cell A
b: Cell B
c: Cell C
d: Cell D

Values in a contingency table

Examples:

Here are examples of where the data’s sample size is small and if we calculate the expected frequency for this table it will be less than 5, therefore we can’t apply Chi-squares test → use fisher’s exact test instead

- The following data relate to suicidal feelings in samples of psychotic and neurotic patients:

	Psychotics	Neurotics	Total
Suicidal feelings	2	6	8
Non-suicidal feelings	18	14	32
Total	20	20	40

- The following data compare malocclusion of teeth with method of feeding infants.

	Normal teeth	Malocclusion	Total
Breast fed	4	16	20
Bottle fed	1	21	22
Total	5	37	42

McNemar’s Test

Also called McNemer’s Chi-square test

When to use

When we have a paired (dependent) samples and both the exposure and outcome variables are qualitative variables (Binary).

Situation

Two paired binary variables that forma particular type of 2 x 2 table

E.g. matched case-control study or cross-over trial

Cross-over trial (a type of randomized control trial) that is used when there is a limited number of subjects which will also be the comparison group. For example: when we have 20 patients and we give them a certain intervention after we get an outcome there will be a wash-out period then the same patients will take another intervention then we compare (when you give treatment A to the group and after 3 months (wash-out period) you give the same group treatment B so the group becomes a comparison group) another situation is when you give first group treatment A and second group treatment B and after a while you cross the treatments between the groups

In statistics we call the cross-over groups dependent, meaning that they are related. Also the same goes for the groups of matched case-control study.

Problem

- A researcher has done a matched case-control study of endometrial cancer (cases) and exposure to conjugated estrogens (exposed).
- In the study cases were individually matched 1:1 to a non-cancer hospital-based control, based on age, race, date of admission, and hospital.

Why we match the age, gender? To remove confounding factors

Here, there are 5 variables to match, which is difficult. We usually do matching for basic characteristics such as age & gender

McNemar's Test

Example:

1. Data :

	Cases	Controls	Total
Exposed	55	19	74
Non-exposed	128	164	292
Total	183	183	366

- We can't use a Chi-square because:
    - Observations are not independent - they are paired (dependent)
- The information in the standard 2 x 2 table used for unmatched studies is insufficient because it doesn't say who is in which pair.
    - Ignoring the matching.
- We must present the 2 x 2 table differently:
    - Each cell should contain a count of the number of pairs with certain criteria, with the columns and rows respectively referring to each of the subjects in the matched pair.
- So we will be constructing a matched 2 x 2 table:

Cases	Controls		Total
	Exposed	Non-exposed	
Exposed	e	f	e+f
Non-exposed	g	h	g+h
Total	e+g	f+h	n

The problem is in f and g cells they are mismatched which is the false positive and false negative. McNemar test will only take the difference between f and g unlike Chi-square test which takes the 4 cells. If the difference is large→ significant association and if it is small→ no significant association

2. Data will be:

The values below are pairs NOT individual values

Cases	Controls		Total
	Exposed	Non-exposed	
Exposed	12	43	55
Non-exposed	7	121	128
Total	19	164	183

McNemar's Test

Formula

- The odds ratio is **f/g**.
- The test is :

$$X^2 = \frac{(|f - g| - 1)^2}{f + g}$$

- Compare this to the X^2 distributed in 1 **df.** Why 1df? Because it's 2 rows→ it is meant only for 2x2 table not a bigger dimension table (2-1)(2-1)=1

Example:

$$X^2 = \frac{(|43 - 7| - 1)^2}{43 + 7} = \frac{1225}{50} = 24.5$$

- **P** <0.001, Odds Ratio = 43/7 = 6.1 Previously the routine odds ratio formula = ad/bc, BUT here it is = f/g
  Interpretation of odds ratio: The odds of exposure is 6.1 times more in cancer patients compared to non-cancer patients. (the outcome has already occurred so the statement should be related to the exposure not to the outcome)
- $p_1 - p_2$ = (55/183) – (19/183) = 0.197 (20%)
- **s.e.**($p_1 - p_2$) = 0.036
- **95% CI** : 0.12 to 0.27 (or 12% to 27%)
- Degrees of Freedom
  - **df** = **(r-1) (c-1)**
  - = (2-1) (2-1) = 1
- Critical Value at 0,05 level of significance and 1 df **(look the table at page 13)** = 3.84
- Calculated value (X^2) = 25.92

- Calculated value 25.92 is greater than critical (table) value at 0.05 level with 1 d.f.f (degrees of freedom)
- Hence we reject our H_o and conclude that there is highly statistically significant association between Endometrial cancer and Estrogens.

Stata output:

Stat is another software

Cases	Controls: Exposed	Controls: Non-exposed	Total
Exposed	12	43	55
Non-exposed	7	121	128
Total	19	164	183

- McNemar's chi2(1) = 25.92
  Prob > chi2 = 0.0000
- Exact McNemar significance probability = 0.0000

		[95% conf. interval]	
Cases	.3005464		
Controls	.1038251		
Difference	.1967213	.1210924	.2723502
Ratio	2.894737	1.885462	4.444269
Rel.diff.	.2195122	.1448549	.2941695
Odds ratio	6.142857	2.739772	16.18458

(exact)

Two-tailed critical ratios of X^2

Two-tailed critical ratios of χ^2

Degrees of freedom df	.10	.05	.02	.01
1	2.706	3.841	5.412	6.635
2	4.605	5.991	7.824	9.210
3	6.251	7.815	9.837	11.341
4	7.779	9.488	11.668	13.277
5	9.236	11.070	13.388	15.086
6	10.645	12.592	15.033	16.812
7	12.017	14.067	16.622	18.475
8	13.362	15.507	18.168	20.090
9	14.684	16.919	19.679	21.666
10	15.987	18.307	21.161	23.209
11	17.275	19.675	22.618	24.725
12	18.549	21.026	24.054	26.217
13	19.812	22.362	25.472	27.688
14	21.064	23.685	26.873	29.141
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

TABLE 9 Critical Values of the Chi-Square Distribution

Note: Column headings are non-directional (omni-directional) P-values. If H_A is directional (which is only possible when df = 1), the directional P-values are found by dividing the column headings in half.

	TAIL PROBABILITY						
df	0.20	0.10	0.05	0.02	0.01	0.001	0.0001
1	1.64	2.71	3.84	5.41	6.63	10.83	15.14
2	3.22	4.61	5.99	7.82	9.21	13.82	18.42
3	4.64	6.25	7.81	9.84	11.34	16.27	21.11
4	5.99	7.78	9.49	11.67	13.28	18.47	23.51
5	7.29	9.24	11.07	13.39	15.09	20.51	25.74
6	8.56	10.64	12.59	15.03	16.81	22.46	27.86
7	9.80	12.02	14.07	16.62	18.48	24.32	29.88
8	11.03	13.36	15.51	18.17	20.09	26.12	31.83
9	12.24	14.68	16.92	19.68	21.67	27.88	33.72
10	13.44	15.99	18.31	21.16	23.21	29.59	35.56
11	14.63	17.28	19.68	22.62	24.72	31.26	37.37
12	15.81	18.55	21.03	24.05	26.22	32.91	39.13
13	16.98	19.81	22.36	25.47	27.69	34.53	40.87
14	18.15	21.06	23.68	26.87	29.14	36.12	42.58
15	19.31	22.31	25.00	28.26	30.58	37.70	44.26
16	20.47	23.54	26.30	29.63	32.00	39.25	45.92
17	21.61	24.77	27.59	31.00	33.41	40.79	47.57
18	22.76	25.99	28.87	32.35	34.81	42.31	49.19
19	23.90	27.20	30.14	33.69	36.19	43.82	50.80
20	25.04	28.41	31.41	35.02	37.57	45.31	52.39
21	26.17	29.62	32.67	36.34	38.93	46.80	53.96
22	27.30	30.81	33.92	37.66	40.29	48.27	55.52
23	28.43	32.01	35.17	38.97	41.64	49.73	57.08
24	29.55	33.20	36.42	40.27	42.98	51.18	58.61
25	30.68	34.38	37.65	41.57	44.31	52.62	60.14
26	31.79	35.56	38.89	42.86	45.64	54.05	61.66
27	32.91	36.74	40.11	44.14	46.96	55.48	63.16
28	34.03	37.92	41.34	45.42	48.28	56.89	64.66
29	35.14	39.09	42.56	46.69	49.59	58.30	66.15
30	36.25	40.26	43.77	47.96	50.89	59.70	67.63

Z-test

Normal distribution

For sample proportion with population proportion:

Problem:

In an otological examination of school children, out of 146 children examined 21 were found to have some type of otological abnormalities. Does it confirm with the statement that 20% of the school children have otological abnormalities?

- **Question to be answered:**

  Is the sample taken from a population of children with 20% otological abnormality?

  The prevalence in the sample taken is 14.38% (21/146), and the prevalence in the population is 20% so the point of the test here is to test whether the 14% (taken from the sample) is similar to the 20% (prevalence of what is reported from the population)

- **Null hypothesis:**

  The sample has come from a population with 20% otological abnormal children. (there is no difference)

- **Test statistics:**

$$z = \frac{p-P}{\sqrt{\frac{pq}{n}}} = \frac{14.4-20.0}{\sqrt{\frac{14.4 * 85.6}{146}}} = 1.69$$

P – Population. Prop.
p- sample prop.
n- number of samples

- **Comparison with theoretical value:**

  $Z \sim N(0,1)$; $Z_{0.05} = 1.96$

  The prob. of observing a value equal to or greater than 1.69 by chance is more than 5%.

  We therefore do not reject the Null Hypothesis.

  It falls in the bigger area, the centre because it's less than 1.96

- **Inference:**

  There is an evidence to show that the sample is taken from a population of children with 20% abnormalities.

Z-test

For two independent sample proportion:

Example:

Researchers wished to know if urban and rural adult residents of a developing country differ with respect to prevalence of a certain eye disease. A survey revealed the following information.
Test at 5% level of significance, the difference in the prevalence of eye disease in the 2 groups

Residence	Eye disease		Total
	Yes	No	
Rural (Small village)	24	276	300
Urban (Bigger city)	15	485	500

- **Test statistics:**

Absolute difference

Standard error

$$Z = \frac{\boxed{P_1 - P_2}}{\boxed{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}}}$$

P1= proportion in the first group
P2= proportion in the second group
n1= first sample size
n2= second sample size

- **Critical Z:**
  - ✓ 1.96 at 5% level of significance. (alpha)
  - ✓ 2.58 at 1% level of significance.

So:

- **P1** = 24/ 300 = 0.08
- **P2** = 15 / 500 = 0.03

$$Z = \frac{0.08 - 0.03}{\sqrt{\frac{0.08(1-0.08)}{300} + \frac{0.03(1-0.03)}{500}}} = 2.87$$

2.87 > 1.96 (from Z-table at α=0.05)

It falls in the smaller area, the sides because it's more than 1.96

- Hence we can conclude that, the difference of prevalence of eye disease between the two groups is statistically significant.

- Even though the absolute difference is small but it resulted in significant association why? Because the sample size is small thus we got a significant result (large number > 1.96)

Summary

In conclusion

When both the study variables and outcome variables are categorical (Qualitative):
Apply :

- Chi square test (for two and more than two groups). Used more than Z-test for the association
- Fisher's exact test (Small samples).
- McNemar's test (for paired samples).
- Z-test for single sample (comparing sample proportion with population proportion) and two samples (two sample proportions).

<table>
<tr><th>Test</th><th>Study variable</th><th>Outcome variable</th><th>Comparison</th><th>Sample size</th><th>Expected frequency</th></tr>
<tr><td>Chi-square</td><td rowspan="4">Qualitative</td><td rowspan="4">Qualitative</td><td>Two or more proportions</td><td>> 20</td><td>>5</td></tr>
<tr><td>Fisher's exact</td><td>Two proportions</td><td>< 20</td><td></td></tr>
<tr><td>Mcnemar's</td><td>Two proportions</td><td>any</td><td></td></tr>
<tr><td>Z-test</td><td>- Sample proportion with population proportion
- Two sample proportions</td><td>Larger in each group (>30)</td><td></td></tr>
</table>

Test	Equation
Ch-square	$X^2=\sum\left[\frac{(o-e)^2}{e}\right]$
Fisher's exact	$=\frac{R_1!R_2!C_1!C_2!}{n!a!b!c!d!}$
McNemar's test	$X^2=\frac{(\lvert f-g\rvert-1)^2}{f+g}$
Z-test	$z=\frac{p-P}{\sqrt{\frac{pq}{n}}}$ $Z=\frac{P_1-P_2}{\sqrt{\frac{P_1(1-P_1)}{n_1}+\frac{P_2(1-P_2)}{n_2}}}$

Questions

(1) What is the best test for matched case-control study?

A) Chi-square
B) McNemar's
C) Fisher's exact
D) Z-test

(2) In Z-test we compare...

A) Two population proportions
B) Three proportions
C) No comparisons (single sample)
D) Sample proportion with population proportion

(3) Which one of the following is not considered one of the requirements of chi-square test?

A) Sample size > 20
B) Frequency > 5
C) Two or more proportions
D) Sample size < 20

(4) What is the best test to use when you have a small sample size (<20)?

A) Chi-square
B) Fisher's exact
C) McNemar's
D) Z-test

(5) What is the purpose of chi-square test?

A) To test for differences
B) To test for association
C) To calculate the sample size
D) To measure the association

(6) Which test is used for large sample size (>30)?

A) Chi-square
B) Fisher's exact
C) McNemar's
D) Z-test

Answers:
1: B | 2: D | 3: D | 4: B | 5: B | 6: D

Research's Dream Team Board

We want to start off by thanking 438 for all the help they've provided us with and for allowing us to use their teamwork.

Special and **HUGE** thanks to:

438's amazing leaders:

Alanoud Salman, Aued Alanazi, Lama Alassiri, Taif Alshammari & Mohsen Almutairi

439's marvelous leaders:

Rania Almutiri & Homoud Algadib

439's supercalifragilisticexpialidocious academic leaders:

Muneerah Alsadhan & Mishal Althunayn

And lastly but not least to the most incredible members, note takers and revisers:

438:

- Aued Alanazi
- Mashal Abaalkhail
- Shahad BinSelaym
- Sarah AlHelal
- Yasmeen Almousa
- Abdulrhman Al-Mezaini
- Omar AlOtaibi
- Ateen Almutairi
- AlHanouf Alhaluli
- Deana Awartani
- Meshaal Alghanim
- Noura AlTurki
- Nawaf Albhijan
- Abdulaziz Alghamidi
- Musab Alsaadan
- Nouf Albrikan
- Taif AlOtaibi
- Amirah AlZahrani
- Jehad Alorainy
- Nouf Alhumaidhi
- Tariq Alanezi
- Ali Abdulaziz
- Lama Alzamil
- Abdullah Alnuwaybit
- Leena AlNasser
- Reham AlTurki
- Abdulaziz Redwan
- Ajeed Al-Rashoud
- Shahad Alsalamh
- Bassam Khwaitter
- Bader Alwhaibi
- Fahad Alsultan

439:

- Banan Alqady
- Sadem Al Zayed
- Bushra Alotaibi
- Osama alharbi
- Feras Alqaidi
- Mona Alomiriny
- Albandari Alanazi
- Haya Alanzai
- Rayan jabaan
- Shaden alobaid
- Muneerah Alsadhan
- Rand AlRefaei
- Nourah Alklaib
- Hessah Alalyan
- Sara Alharbi
- Abdulaziz Alomar
- Noura Alkathiri
- Dana Naibulharam
- Hessah Fahad
- Mohamed Albabtain
- Mishal Althunayan
- Aseel Alshehri
- Mohamed Alquhidan
- Fatimah Alhelal
- Hamad Almousa
- Leen Almadhyani
- Norah Almasaad
- Rakan Aldohan
- Shatha Aldhohair
- Samar Almohammedi
- Abdulaziz Alamri
- Noura Alshathri
- Shahd Almezel
- Abdulaziz Alderaywsh
- Lama Alahmadi
- Fahad Alajmi
- Shaden Alsaiedan
- Renad Alhomaidi
- Fatimah binmeather
- Ghada aljedaie

We hope we didn't forget anyone...

Sincerely,
Abdulrhman Alsuhaibany, Shuaa Khdary, Sarah AlQuwayz

Research Team

RESEARCH TEAM 438

Keep it simple, Keep it focused

Thank you for checking our work!

Leaders:

Shuaa Khdary   Sarah AlQuwayz

Abdulrhman Alsuhaibany

Member:

Albandari Alanazi

Note Taker:

Leen Almadhyani   Hamad Almousa

Contact us:

Research4390@gmail.com