HPH 506

Analysis of Contingency Tables

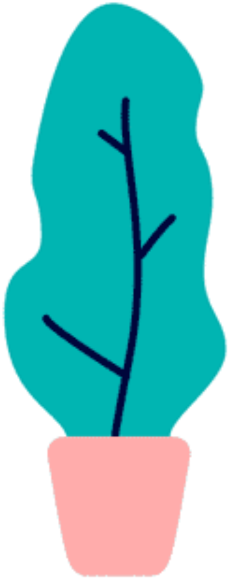
Study Objectives

Apply

- Chi-square goodness-of-fit test
- Chi-square test for independence

Master

- McNemar’s test for paired data
- Mantel – Haenszel Test for stratified data



Review Quiz

What we learned

Outcome	Explanatory Variable	Parametric Method	Nonparametric Method
Continuous	Two independent groups	Two sample t-test	Wilcoxon Rank Sum test
Continuous	Paired groups	Paired t-test	Wilcoxon Sign Test Wilcoxon Signed Rank test
Continuous	More than two groups	ANOVA	Kruskal Wallis test

P value

The **P-value** (also called the **observed significance level**) is a measure of consistency between the null hypothesis and the observed sample.

In other words, the **P-value** is the likelihood of whether the sample data are consistent with H_0.

Outline

One categorical variable	Chi-Square Goodness of Fit Test
Two non-paired categorical variables	Chi-Square Test of Independence
Two paired categorical variables	McNemar's test
Two categorical variables stratified by a third categorical variables	Cochran-Mantel-Haenszel test

One Sample Test

for Population Proportion π	for Population Mean μ
p is the sample proportion from a random sample The sample size n is large (np $\geq$ 5 and n(1- p) $\geq$ 5)	The sample size n is large (generally n $\geq$ 30) If the population distribution is normal, any sample size is OK.
One sample **Z -test**	One sample **t-test** (df=n-1)
$$Z = \frac{p - \pi_0}{\sqrt{\pi_0 (1 - \pi_0) / n}}$$	$$t = \frac{\bar{x} - Hypothesized\ value}{s/\sqrt{n}}$$
Confidence Interval: $$p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$	Confidence Interval: $$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Chi-square test statistics

$$\text{STATISTICAL TEST} : \frac{\text{OBSERVED DIFFERENCE} - \text{WHAT WE EXPECT IF THE NULL IS TRUE}}{\text{AVERAGE VARIATION}}$$

$$t = \frac{\bar{x} - Hypothesized\ value}{s/\sqrt{n}}$$

$$X^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

8.1 Goodness of Fit Test

Chi-square Test -One categorical variable Test

Goodness of Fit Test

Chi-Square Goodness of Fit Test

$$H_0 : P_{HEALER} = 15\%, P_{TANK} = 20\%, P_{ASSASSIN} = 20\%, P_{FIGHTER} = 45\%$$

OBSERVED DATA

HEALER	TANK	ASSASSIN	FIGHTER
25	35	50	90

EXPECTED DATA

HEALER	TANK	ASSASSIN	FIGHTER
15% OF 200 = 30	20% OF 200 = 40	20% OF 200 = 40	45% OF 200 = 90

OBSERVED LESS EXPECTED

HEALER	TANK	ASSASSIN	FIGHTER
25 – 30 = -5	35 – 40 = -5	50 – 40 = 10	90 – 90 = 0

$= 0$

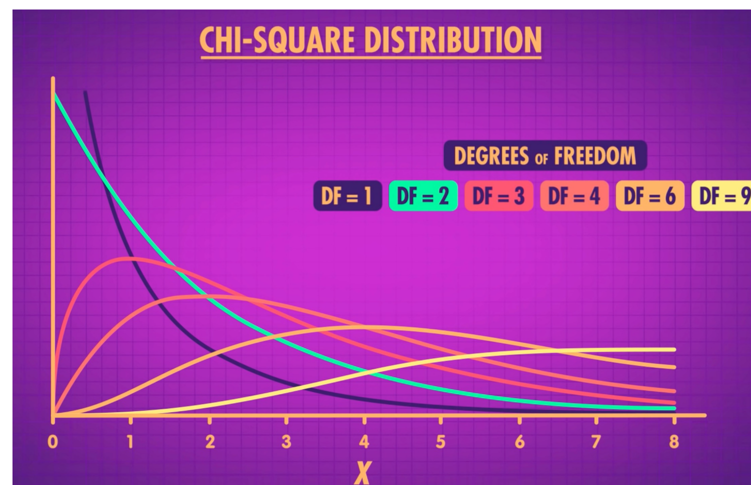
CHI-SQUARE

HEALER	TANK	ASSASSIN	FIGHTER
$-5^2 = 25$	$-5^2 = 25$	$10^2 = 100$	$0^2 = 0$

CHI-SQUARE STATISTIC

HEALER	TANK	ASSASSIN	FIGHTER
25/30	25/40	100/40	0/90

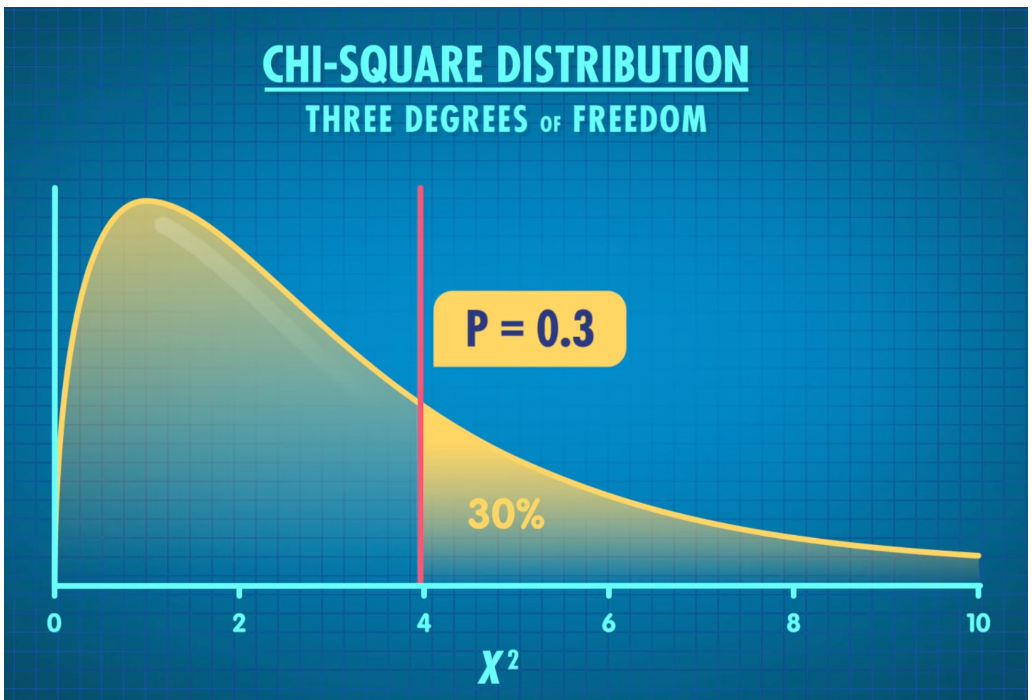
$$0.83 + 0.625 + 2.5 + 0 = 3.958$$

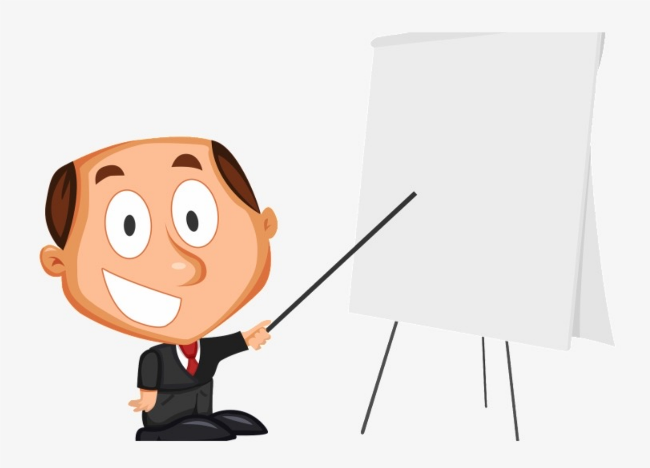


Degree of Freedom

OBSERVED STATE

HEALER	TANK	ASSASSIN	FIGHTER
25	35	50	90





8.2 Two categorical variables test
Chi-square $\boldsymbol{\chi}^2$ test of independence

8.2.1 Chi-Square Test of Independence

- Variable 1: What Hogwarts houses nerdfighters were in
- Variable 2: if they liked pineapple on pizza
- Whether Pineapple on Pizza preference is independent of Hogwarts House
- Does liking pineapple on pizza affect the probabilities of you identifying with each of the houses?

Test of Independence

	GRYFFINDOR	HUFFLEPUFF	RAVENCLAW	SLYTHERIN
NO	79	122	204	74
YES	82	130	240	69

Test of Independence: degree of freedom

CALCULATING THE CHI-SQUARE STATISTIC

	GRYFFINDOR	HUFFLEPUFF	RAVENCLAW	SLYTHERIN	
NO	79	122	204	74	479
YES	82	130	240	69	521
	161	252	444	143	1000

$$(r-1)(c-1)$$

Test of Independence: expected value

$$\frac{161}{1000} = .161 \text{ OR } 16.1\%$$

CALCULATING THE CHI-SQUARE STATISTIC

	GRYFFINDOR	HUFFLEPUFF	RAVENCLAW	SLYTHERIN	
NO	479 × 16.1%	479 × 25.2%	479 × 44.4%	479 × 14.3%	479
YES	521 × 16.1%	521 × 25.2%	521 × 44.4%	521 × 14.3%	521
	161/1000 = 16.1%	252/1000 = 25.2%	444/1000 = 44.4%	143/1000 = 14.3%	1000

Test of Independence: χ statistic

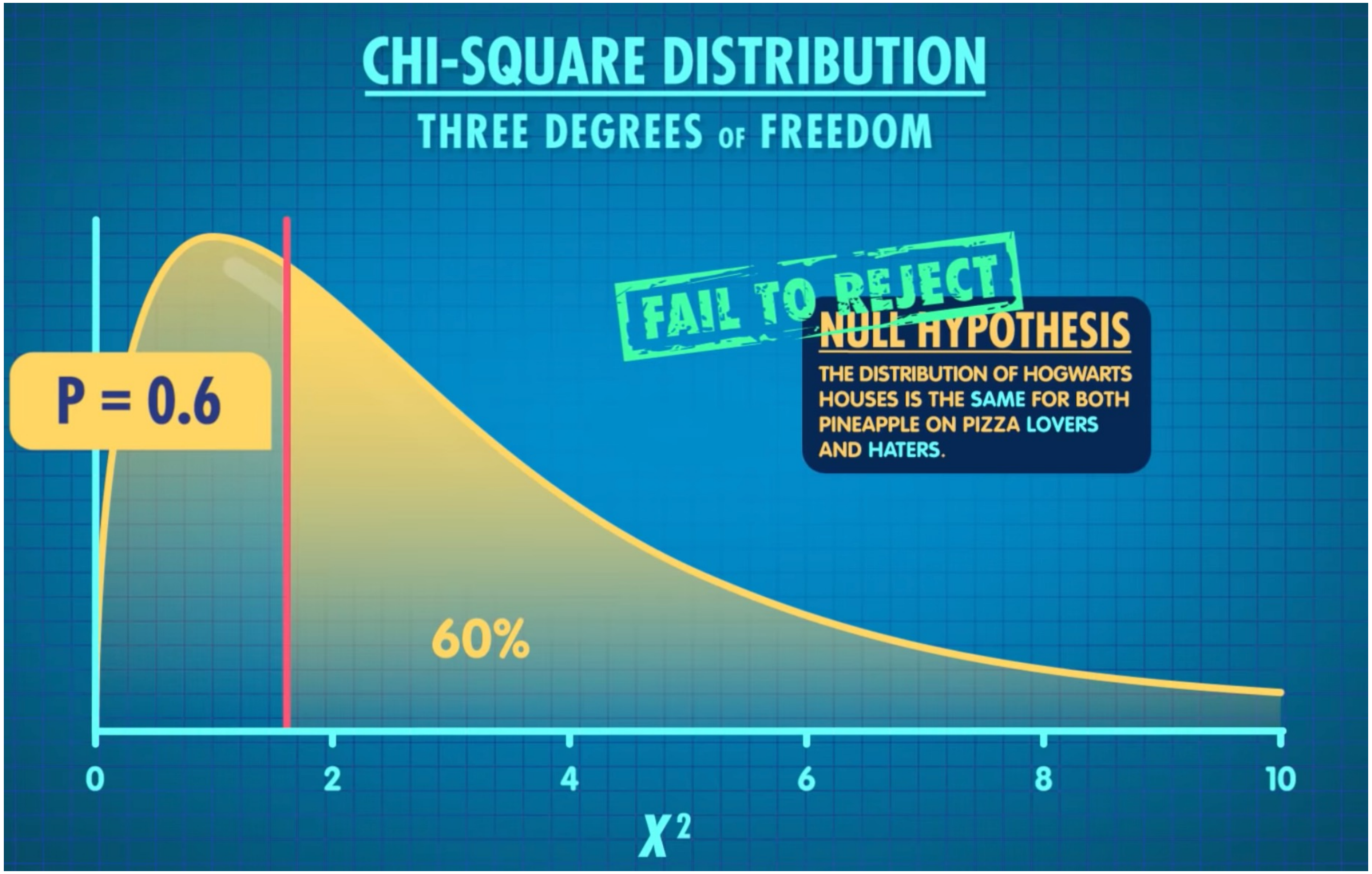
CALCULATING THE CHI-SQUARE STATISTIC

	GRYFFINDOR	HUFFLEPUFF	RAVENCLAW	SLYTHERIN	
NO	77.12	120.71	212.68	68.5	479
YES	83.88	131.29	231.32	74.5	521
	161	252	444	143	1000

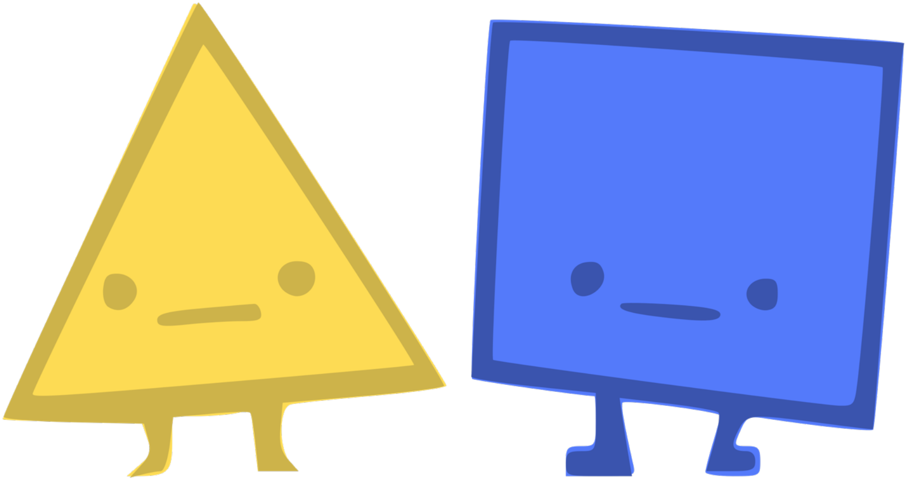
CHI-SQUARE STATISTIC

$$X^2 = 1.6$$

Test of Independence

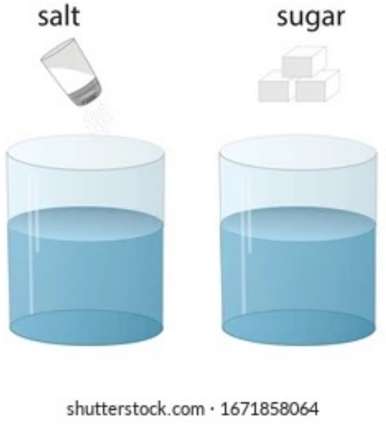


8.2.2 Test of homogeneity



Test of homogeneity

- Looking at whether it's likely that different samples come from the same population.
- In the test of homogeneity, we **select random samples from each subgroup or population separately and collect data on a single categorical variable**.
- The null hypothesis says that the distribution of the categorical variable is the same for each subgroup or population. Both tests use the same chi-square test statistic.



Homogeneity vs independence

- **Calculation is identical**
- The difference is **a matter of design**. In the test of independence, observational units are collected at random from a population and two categorical variables are observed for each unit. In the test of homogeneity, the data are collected by randomly sampling from each sub-group separately.

Test of homogeneity

	Living alone: Sample size: 32 %	Living with others Sample size: 295 %
On a diet	6	8
Watch what I eat and drink	72	49
Eat and drink whatever I feel like	22	42
Total	100	100

Observed counts	Living alone Counts	Living with others Counts	Total Counts
On a diet	2	25	27
Watch what I eat and drink	23	146	169
Eat and drink whatever I feel like	7	124	131
Total	32	295	327

Expected counts	Living alone Expected Counts	Living with others Expected Counts
On a diet	2.6	24.4
Watch what I eat and drink	16.5	152.5
Eat and drink whatever I feel like	12.8	118.2

Calculation of Expected Value Formula

Probability: *(# events that occurred in a period)/(number of entities eligible)*.

	Variable A		
Variable B	Yes	No	
Yes	a	b	N_{row1}=a+b
No	c	d	N_{row2}=c+d
	N_{col1}=a+c	N_{col2}=b+d	N_{Total}

If B is independent with A, $P_{AB}=P_A*P_B$

Null hypothesis always assume no difference, same value.

Here, variable A's proportion has no difference in variable B's group means variable A and Variable B are independent.

P_A = Count(A)/N_{total} = N_{col1}/N_{total}
P_B = Count(B)/N_{total} = N_{row1}/N_{total}

$P_{AB} = P_A * P_B = N_{col1} * N_{row1} / N^2_{total}$

Expected Count (AB) = $P_{AB} * N_{total} = N_{col1} * N_{row1} / N_{total}$
Expected Count ($A\bar{B}$)= $P_{A\bar{B}} * N_{total} = N_{col1} * N_{row2} / N_{total}$
Expected Count ($\bar{A}$B)= $P_{\bar{A}B} * N_{total} = N_{col2} * N_{row1} / N_{total}$
Expected Count ($\bar{A}\bar{B}$)= $P_{\bar{A}\bar{B}} * N_{total} = N_{col2} * N_{row2} / N_{total}$

Test of Independence: Calculation

If B is independent with A, $P_{AB}=P_A*P_B$

$P_{smoking}=50/100$, $P_{cancer}=50/100$, $P_{smoking\ \&\ cancer}=50*50/100^2$,

$N_{smoking\ \&\ cancer}=P_{smoking\ \&\ cancer}*100=50*50/100$

Observed

		Cancer	
Smoking		Yes	No
	Yes	49	1
	No	1	49

Expected under H0

		Cancer	
Smoking		Yes	No
	Yes	25	25
	No	25	25

Probability: *(# events that occurred in a period)/(number of entities eligible)*.

Independence Test, Smoking Example

Observed

		Cancer	
Smoking		Yes	No
	Yes	49	1
	No	1	49

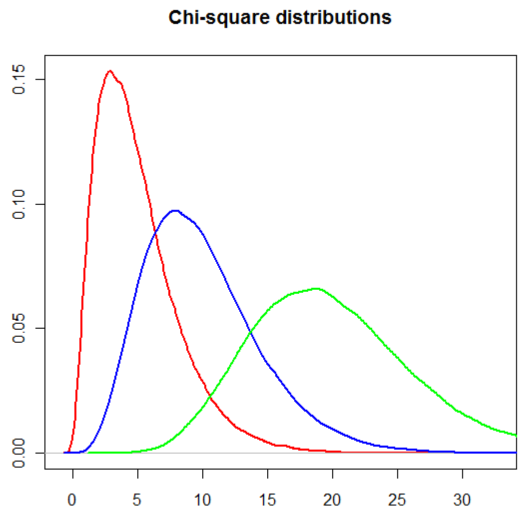
Expected under H0

		Cancer	
Smoking		Yes	No
	Yes	25	25
	No	25	25

the Chi-square statistic:

$$TS = \frac{(49-25)^2}{25} + \frac{(1-25)^2}{25} + \frac{(1-25)^2}{25} + \frac{(49-25)^2}{25} = 355$$

P value <0.001, we reject H0 and claim that smoking is associated with Cancer.



Bicycle Example

Is wearing helmet associated with (independent of) head injuries?

In other words, the proportions of persons suffering head injuries are different between the two populations (wearing helmet or not)

Head	**Wearing Helmet**		
Injury	Yes	No	Total
Yes	17	218	235
No	130	428	558
Total	147	646	793

Bicycle Example

Head	Wearing Helmet		
Injury	Yes	No	Total
Yes	17	218	235
No	130	428	558
Total	147	646	793

H0: Whether the proportions of persons suffering head injuries are identical in the two populations (wearing helmet or not)

Head	Wearing Helmet		
Injury	Yes	No	Total
Yes	$\frac{235*147}{793}$	$\frac{235*646}{793}$	235
No	$\frac{558*147}{793}$	$\frac{646*558}{793}$	558
Total	147	646	793

Null hypothesis: H_0: $\pi_1 - \pi_2 = 0$

Alternative hypothesis: H_a: $\pi_1 - \pi_2 \neq 0$

Sample proportions: $p_1 = \frac{17}{147} = 12\%$, $p_2 = \frac{218}{646} = 34\%$

We want to know whether the two population proportions are significantly different.

Bicycle Example

- ❖ H0: wearing helmet is not associated with (independent of) head injuries.
- ❖The idea

  1. calculate the expected number in each cell under H0
  2. Compare the expected numbers to the observed numbers and test the significance of the deviation

the <u>Chi-square</u> statistic:

TS = $\mathbf{27.27}$

For the test of a 2x2 table,

Degrees of freedom = (2-1)*(2-1) = 1.

P value <0.0001

❖ H0: π_1 - π_2 = 0 -- the proportions of persons suffering head injuries are identical in the two populations (wearing helmet or not)

❖H0: wearing helmet is not associated with (in other words independent of) head injuries.

Goodness of Fit Test vs Test of Independence

	Chi-Square Goodness of Fit Test	Chi-Square Test of Independence
Number of variables	One	Two
Purpose of test	Decide if sample's categorical variable has same proportion distribution as is in H_0	Decide if two categorical variables might be related or not
Example	Whether the top players have the same distribution as the average of all players	Whether Pineapple on Pizza preference is independent of Hogwarts House
Hypotheses in example	H_o: Top rank players have the same proportions on choosing roles as to overall players. H_a: Top rank players have different proportions on choosing roles compared to overall players	H_o: Pineapple on Pizza preference is independent of Hogwarts House H_a: Pineapple on Pizza preference is dependent of Hogwarts House
Conclusion	we fail to reject the null. The sample we took didn't give us any statistically significant evidence that the top rank players have different distributions.	we fail to reject the null hypothesis. So we don't have evidence that Hogwarts House is dependent on Pineapples on pizza preference

Goodness of Fit Test

```
proc freq data = sashelp.cars;
tables origin /chisq testp=(0.35 0.40 0.25);
run;
```

The FREQ Procedure

Origin	Frequency	Percent	Test Percent	Cumulative Frequency	Cumulative Percent
Asia	158	36.92	35.00	158	36.92
Europe	123	28.74	40.00	281	65.65
USA	147	34.35	25.00	428	100.00

Chi-Square Test for Specified Proportions	
Chi-Square	28.9725
DF	2
Pr > ChiSq	<.0001

Based on the p value, the distribution of car origin is significantly different from the proportion 0.35, 0.40, 0.25.

Test of Independence

```
Proc freq data=a;
Tables INJ * HEL/chisq;
Run;
```

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of INJ by HEL		
	HEL		
INJ	0	1	Total
0	428 53.97 76.70 66.25	130 16.39 23.30 88.44	558 70.37
1	218 27.49 92.77 33.75	17 2.14 7.23 11.56	235 29.63
Total	646 81.46	147 18.54	793 100.00

Statistics for Table of INJ by HEL

Statistic	DF	Value	Prob
Chi-Square	1	28.2555	<.0001

Based on the p value, the % of head injury in Helmet wearing group is significantly different from the % of head injury in Not wearing group.

small sample size

If the expected numbers are less than 5 for some cells, use Fisher's exact test. SAS will automatically provide Fisher's exact test if the sample size is small.

2x3 table

A Genome-Wide Association Study (GWAS) example:
Data are collected for genotypes and phenotypes for a number of unrelated individuals

- Genotypes : often SNPs, minor allele homozygous, heterozygous, and major allel homozygous sites are coded as 0, 1, and 2
- Phenotypes: categorical data (e.g. case/control)

For each SNP, create a contingency table

	AA	Aa	aa	Total
Case				
Control				
Total				

A Chi-square test with (2-1)*(3-1)=2 degrees of freedom can be used.

Original Article | Published: 20 February 2020

Genetic polymorphisms of HbE/beta thalassemia related to clinical presentation: implications for clinical diversity

Nurul Fatihah Azman, Wan Zaidah Abdullah, Sarifah Hanafi, R. Diana, Rosnah Bahar, Muhammad Farid Johan, Bin Alwi Zilfalil & Rosline Hassan

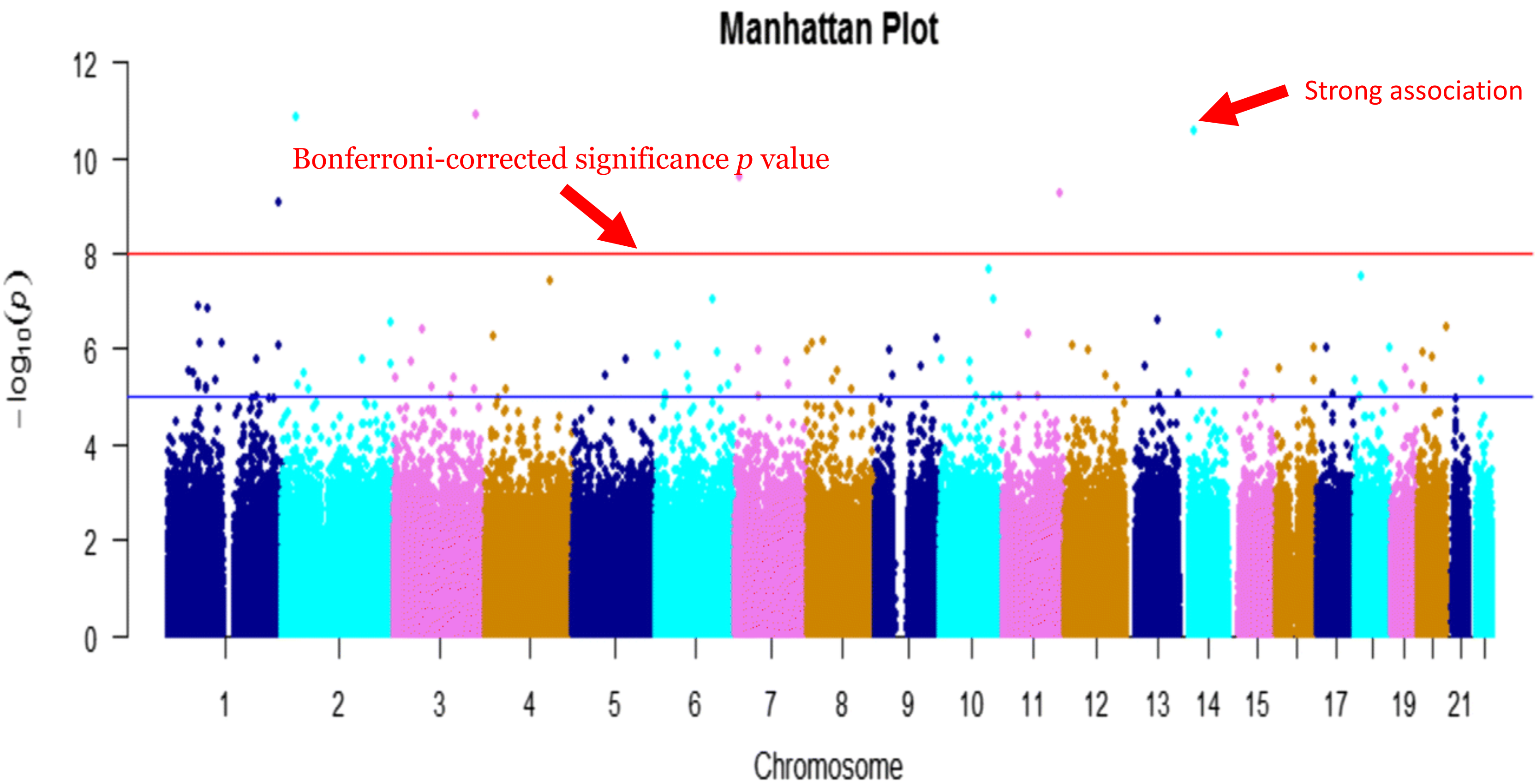
Annals of Hematology (2020) | Cite this article

A case-control study was conducted among Malay transfusion-dependent HbE/β-thalassemia patients. Patients who were confirmed HbE/β-thalassemia were recruited and genotyping study was performed on these subjects.

A total of 98 transfusion-dependent HbE/β-thalassemia patients were successfully recruited and genotyped. Out of these 98 patients, 43 (43.9%) were classified as moderate and 55 (56.1%) were severe cases.

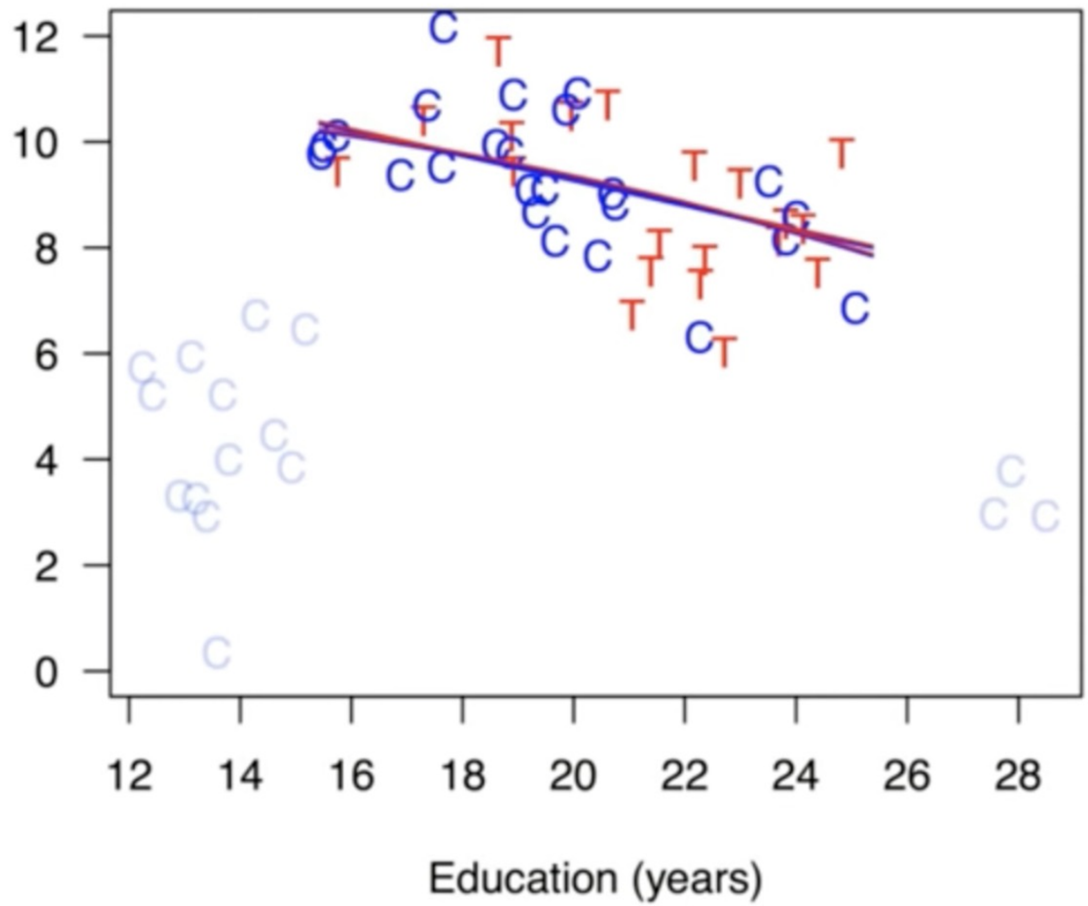
During data analysis, 645,831 SNPs were passed and included in this association study.

The severity of the patients was classified using Pearson chi-square test for categorical variables or Fisher exact test.



Example of a Manhattan plot of the association between severe and moderate severity type of HbE/beta thalassemia. The plot shows $-\log_{10}p$ value for each SNP against the chromosomal location. The *x-axis* represents chromosomal position and the *y-axis* shows $-\log_{10}p$-values. The blue line indicates suggestive p value 10^{-5} while the red line indicates Bonferroni-corrected significance p value. The highest dot with significance p value ($< 10^{-8}$) was at chromosome 3, 2, 14, 7, 11, and 1 with rs7372408, rs13398071, rs10483436, rs2355081, rs10750301, and rs6690359, respectively

Paired Data



Paired Data

	Treatment	Control
Concordant	Yes	Yes
Concordant	No	No
Discordant	Yes	No
Discordant	No	Yes

- **Concordant pairs:** Outcome is the same for pairs.
  - Note that concordant pairs give us no information
- **Discordant pairs:** Different outcomes for each pair.
  - All of the information about group differences is given by the discordant pairs.

- **McNemar's test** focuses solely on the discordant pairs.
- Under H0, we would expect the discordant pairs to be approximately equally distributed between cell (Yes,No) and cell (No,Yes); that is H0: $p_{Y,N}=p_{N,Y}=50\%$

Two discrete Paired data

Suppose we want to survey the same group of participants in 2016 and in 2020 again and test whether the proportion of voting Trump changes.
A fake data:

Paired variable

Outcome variable

	2016 (before)	2020 (after)	N (%)	
Voting	Yes	Yes	30	No change, provide no information
	No	No	60	No change, provide no information
	Yes	No	50%??	Change
	No	Yes	50%??	Change

Under the null hypothesis (there is no change in voting % between 2016 and 2020), the # of Yes-> No and No ->Yes groups will be equal.

144 pairs for MI(Yes,No)

McNemar's Test

Paired Dichotomies

e.g. Pairs matched on age & sex:

Diabetes	M.I. Yes	M.I. No	Total
Yes	46	25	71
No	98	119	217
Total	144	144	288

McNemar’s test for paired data

❖**Example:** MI & Diabetes. Are they independent/associated?

M.I.	No M.I.		Total
	Diabetes	No Diabetes	
Diabetes	9	37	46
No Diabetes	16	82	98
Total	25	119	144

In this case, we can use McNemar’s test for the paired data.

Calculation

Discordant entries: 37 & 16

$$X^2 = \frac{[|37 - 16| - 1]^2}{37 + 16}$$

$$= 7.55$$

$$X^2_{1,.010} = 6.63$$

$$X^2_{1,.001} = 10.83$$

$$.001 < p < .010$$

M.I.	No M.I. Diabetes	No M.I. No Diabetes	Total
Diabetes	9	37	46
No Diabetes	16	82	98
Total	25	119	144

McNemar's test for paired data

McNemar's test
H0: diabetes is not associated with MI

```
/*
McNemar's test
using already calcuated frequency data*/

data a;
input MI $ NO_MI $ wght;
datalines;
D D 9
D N 37
N D 16
N N 82
;
run;
proc freq; tables MI*NO_MI/agree;
weight wght;
run;
```

The FREQ Procedure

Frequency / Percent / Row Pct / Col Pct

Table of MI by NO_MI

MI	NO_MI: D	NO_MI: N	Total
D	9 6.25 19.57 36.00	37 25.69 80.43 31.09	46 31.94
N	16 11.11 16.33 64.00	82 56.94 83.67 68.91	98 68.06
Total	25 17.36	119 82.64	144 100.00

Statistics for Table of MI by NO_MI

McNemar's Test	
Statistic (S)	8.3208
DF	1
Pr > S	0.0039

Research Letter

FRE

January 23/30, 2018

Characteristics of Interim Publications of Randomized Clinical Trials and Comparison With Final Publications

Steven Woloshin, MD, MS[1]; Lisa M. Schwartz, MD, MS[1]; Pamela J. Bagley, PhD, MSLS[2]; et al

» Author Affiliations | Article Information

JAMA. 2018;319(4):404-406. doi:10.1001/jama.2017.20653

because interim results are new and often promising, they may generate substantial interest, which can be misleading if results change. We describe the characteristics of interim publications from ongoing randomized trials and compare their consistency and prominence with those of final publications.

Table 2

Characteristics	Publication		P Value[b]
	Interim	Final	
Journal prominence			
High impact factor (≥20), No. (%)	16 (22)	17 (23)	.70
Top-5 impact factor for general medical journal, No. (%)	10 (14)	8 (11)	.60
Both interim and final publications in top-5 journal, No. (%)		3 (4)	

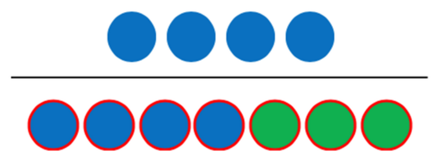
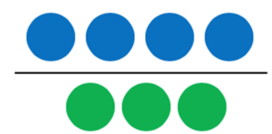
[b] *P* values for paired differences (McNemar test for dichotomous variables and signed rank test for continuous variables).

Treatn

The % of interim publication on high journal is not significantly different for the % of final publication on high journal (p=0.70).

Recall Odds Ratio

Probability vs Odds

	Risk	Odds
Mathematically	$P(p) = \frac{p}{p+q}$	$O(p) = \frac{\frac{p}{p+q}}{\frac{q}{p+q}} = \frac{p(p+q)}{q(p+q)} = \frac{p}{q}$
Graphically		

Odds Ratios in Case-Control and Cohort Studies

Cohort	Develop disease	Do not develop disease
Exposed	a	b
Not exposed	c	d

Odds ratio=

$$\frac{\text{Odds that an exposed person Develops disease}}{\text{Odds that a non-exposed Person develops disease}}$$

$$= \frac{a/b}{c/d}$$

$$= \frac{ad}{bc}$$

Case-control	Cases	Controls
History of exposure	a	b
No history of exposure	c	d

Odds ratio =

$$\frac{\text{Odds that a case was exposed}}{\text{Odds that a control was exposed}}$$

$$= \frac{a/c}{b/d}$$

$$= \frac{ad}{bc}$$

Odds Ratio

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of C_sec by EFM		
	EFM		
C_sec	0	1	Total
0	2745 47.12 52.41 92.30	2493 42.80 47.59 87.44	5238 89.92
1	229 3.93 39.01 7.70	358 6.15 60.99 12.56	587 10.08
Total	2974 51.06	2851 48.94	5825 100.00

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.7213	1.4457	2.0495
Relative Risk (Column 1)	1.3433	1.2102	1.4911
Relative Risk (Column 2)	0.7804	0.7271	0.8375

```
proc freq;
tables C_sec*EFM/measures;
run;
```

We can test whether the odds ratio is significantly different from 1
H0: Odds Ratio=1 using the confidence interval

Since the confidence interval does NOT contain 1, we can conclude that the odds ratio is significant and C section is associated with EFM exposure.

Use Chi-square test to obtain exact p value.

How to read SAS summary table

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of C_sec by EFM

C_sec	EFM: 0	EFM: 1	Total
0	2745	2493	5238
	47.12	42.80	89.92
	52.41	47.59	
	92.30	87.44	
1	229	358	587
	3.93	6.15	10.08
	39.01	60.99	
	7.70	12.56	
Total	2974	2851	5825
	51.06	48.94	100.00

"Row Pct" and "Col Pct" refers to "row total" and "col total" respectively in the denominator

Row Pct= $\frac{cell\ Count}{Row\ Total}$=
2493/5238=47.59
Means in row variable's value group, how many proportion of column variable's value.
In this case, 47.6% is the proportion of C_sec=0 in group where EFM=1

Col Pct= $\frac{Cell\ Count}{Column\ Total}$=
358/2851=12.6
Means in column variable's value group, how many proportion of row variable's value.
In this case, 12.6% is the proportion of C_sec=1 in group where EFM=1

The Mantel – Haenszel Test

Test procedure:

1. Test H0: OR(1)=OR(2) – whether the odds ratios are identical across strata or repeated groups (smokers vs non-smokers).

2. If significant, stop.

3. If not significant, compute a summary odds ratio combining all groups.

4. Test H0: summary OR=1 -whether the summary odds ratio is significant.

The Mantel – Haenszel Test, Coffee Example

Smokers

Myocardial	Coffee		
Infarction	Yes	No	Total
Yes	1011	81	1092
No	390	77	467
Total	1401	158	1559

OR(s)=2.46

Nonsmokers

Myocardial	Coffee		
Infarction	Yes	No	Total
Yes	383	66	449
No	365	123	488
Total	748	189	937

OR(non)=1.96

1. H0: OR(1)=OR(2), $p>0.10$, fail to reject H0.
2. Since the two Odds ratios are not significantly different,
The two data sets can be combined and a summary odds ratio was computed as summary OR=2.18
3. Test H0: summary OR=1. Based on the data and $p<0.0001$, we conclude that coffee is associated with MI overall.

Cochran-Mantel-Haenszel Equations

	Outcome Present	Outcome Absent	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
	a+c	b+d	n

$$\hat{RR}_{CMH} = \frac{\sum \frac{a_i(c_i + d_i)}{n_i}}{\sum \frac{c_i(a_i + b_i)}{n_i}} \qquad \hat{OR}_{CMH} = \frac{\sum \frac{a_i d_i}{n_i}}{\sum \frac{b_i c_i}{n_i}}$$

The Mantel – Haenszel Test, CVD Example

	Age ‹ 50					Age ≥ 50		
	CVD	No CVD	Total			CVD	No CVD	Total
Obese	10	90	100		Obese	36	164	200
Not Obese	35	465	500		Not Obese	25	175	200
Total	45	555	600		Total	61	339	400

$$\widehat{OR} = \frac{a/c}{b/d} = \frac{ad}{bc} = \frac{10(465)}{90(35)} = \frac{4650}{3150} = \mathbf{1.48} \qquad OR = \frac{a/c}{b/d} = \frac{ad}{bc} = \frac{36(175)}{164(25)} = \frac{6300}{4100} = \mathbf{1.54}$$

$$\widehat{OR}_{cmh} = \frac{\sum \frac{a_i d_i}{n_i}}{\sum \frac{b_i c_i}{n_i}} = \frac{\frac{10(465)}{600} + \frac{36(175)}{400}}{\frac{90(35)}{600} + \frac{164(25)}{400}} = \frac{7.75 + 15.75}{5.25 + 10.25} = \mathbf{1.52}$$

$$\widehat{RR}_{cmh} = \frac{\sum \frac{a_i(c_i + d_i)}{n_i}}{\sum \frac{c_i(a_i + b_i)}{n_i}} = \frac{\frac{10(35+465)}{600} + \frac{36(25+175)}{400}}{\frac{35(10+90)}{600} + \frac{25(36+164)}{400}} = \frac{8.33 + 18.00}{5.83 + 12.50} = \mathbf{1.44}$$

Summary

❖ The chi-square test is equivalent to the normal Z test for one proportion or two proportions in that they always yield the same p-values. The method used is usually a matter of convenience.

❖ Tests for contingency tables:

Pearson's Chi-square test (Fisher's exact test for small n)
McNemar's test for paired data
Cohan-Mental-Haenszel test for stratified data.

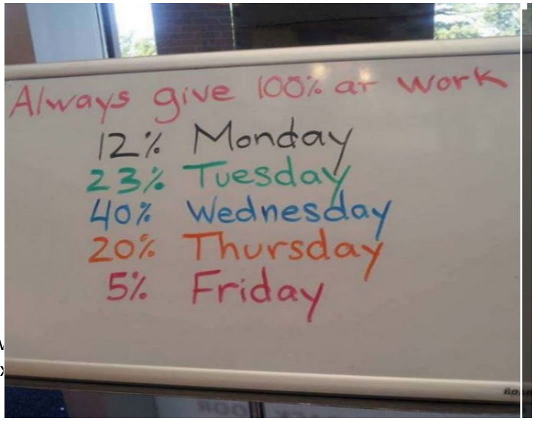
❖ Odds ratio further measures the strength of association between the row and column variables.

❖ The odds ratio is a valid measure of association for cohort as well as matched and unmatched case-control studies.

Yes, these are all for Categorical variables!

Outcome	Explanatory Variable	Nonparametric Method
Dichotomous	Dichotomous	Z test for two proportions 2x2 Chi-square test (Fisher's exact test for small n)
Nominal (R rows)	Nominal (C columns)	RxC Chi-square test (Fisher's exact test for small n)
Dichotomous	Dichotomous Paired /Matched	McNemar Test
Dichotomous	2^{nd} Factor : Dichotomous Stratified by a 3^{rd} Factor: Nominal	C-M-H test

It is all about proportions!



https://w
search/?

References

https://www.jmp.com/en_us/statistics-knowledge-portal/chi-square-test.html

https://www.youtube.com/watch?v=7_cs1YlZoug&ab_channel=CrashCourse

https://www.youtube.com/watch?v=o7DLuXlxGHc&ab_channel=M.MustafaBio-Statistician

https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_multivariable/bs704_multivariable_print.html

https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717_ComparingFrequencies/PH717_ComparingFrequencies8.html

https://www.displayr.com/what-is-the-chi-square-test-of-homogeneity/

https://documentation.sas.com/doc/en/pgmsascdc/v_010/procstat/procstat_freq_examples07.htm