14.2 Inference for Two-Way Tables

	Dem	Rep	Other
Male	24	2	1
Female	2	24	1

Gender and party preference are related; not independent

	Dem	Rep	Other
Male	24	2	1
Female	24	2	1

Variables are not related; independent

	Dem	Rep	Other
Male	14	11	2
Female	12	10	5

Unclear; statistical questioning required

14.1 : 14.2 Compares observations to a set of expectations

14.1 - single categorical variable on a single population

14.2 - ⓐ two categorical variables on a single population

ⓑ single categorical variable on two or more populations

Two-way Tables row x column (r x c table)

rows = response variable (Male / Female)

columns = explanatory variable (Dem / Rep)

H_0: The distribution of the response variable (rows) is the same across all populations (columns) (in context)

Expected Counts = $\frac{\text{row total} \times \text{column total}}{n}$ * may not be a whole #

[illegible] How to quit smoking (Informal analysis)

Treatment	Subjects	Successes
Patch	244	40
Drug	244	74
Patch + Drug	245	87
Placebo	160	25

(a)

Treatment	Success	Failures
Patch	40	204
Drug	74	170
Patch + Drug	87	158
Placebo	25	135

(b) $\hat{p}_N \ \frac{40}{244} = .1639$

$\hat{p}_D \ \frac{74}{244} = .3033$

$\hat{p}_{N+D} \ \frac{87}{245} = .3551$

$\hat{p}_P \ \frac{25}{160} = .15625$

(c) (Bar Chart) Best option P+D followed closely by D. Patch is only slightly better than a placebo.

(d) Ho: The success rate is the same for all four treatments

(e) Expected Counts:

Treatment	Successes	Failures
Patch	61.75	182.25
Drug	61.75	182.25
P + D	62	183
Placebo	40.49	119.51

χ^2 Statistic - sum of $\frac{(\text{observed} - \text{expected})^2}{\text{expected}}$ of all cells

Large χ^2 values are evidence against Ho

degrees of freedom $df = (r-1)(c-1)$

14.2 cont'd

Chi-Square Test for Homogeneity - tests whether it is reasonable to believe that, when several different populations are broken down into the same categories, they have the same proportion of members in each category. Such populations are homogeneous

Steps for χ^2 homogeneity test

① Hypotheses

H_0: The distribution of (response variable) is the same in all (population of interest)

H_a: The distributions are not the same

② Procedure:

χ^2 test of homogeneity

③ Conditions:

S - Independent SRS from each population

E - Expected counts are all ≥ 5. or all are ≥ 1 and no more than 20% ≤ 5

Expected counts = $\frac{\text{row total} \times \text{column total}}{n}$

④ Computation:

$\chi^2 = \sum \frac{(O-E)^2}{E}$ $df = (r-1)(c-1)$

p-value = _____

Assume $\alpha = .05$ if not specified

⑤ Conclusion in context

Since the p-value ___ $\leqslant \alpha =$ ___, we reject/fail to reject H_0. conclusion in context

Example: Gallup Organization poll, November 1999.
Results from each country were based on SRS of 1000 adults in each country. "For you personally, do you think it is necessary or not necessary to have a child at some point in life to feel fulfilled?"

Response	US	India	Mexico	Canada	Germany	Total
Yes	460	930	610	590	490	3080
no	510	60	380	370	450	1770
Undecided	30	10	10	40	60	150
Total	1000	1000	1000	1000	1000	5000

① H_0: The distribution of people who need to have children to feel fulfilled is the same for all countries surveyed.
H_a: The distributions are not the same.

② χ^2 test for homogeneity

③ SRS is stated. Expected counts are all ≥ 5

Expected counts:

616	616	616	616	616
354	354	354	354	354
30	30	30	30	30

④ $\frac{(460-616)^2}{616} + \dots \frac{(60-30)^2}{30} = 628.075$ $df = (3-1)(5-1) = 8$ $p\text{-value} \approx 0$ { χ^2CDF Table C

⑤ Since the p-value of $0 < \alpha = .05$, we reject H_0. There is sufficient evidence to conclude different countries have different distributions of people who need to have children to feel fulfilled.

142 cont'd

χ^2 test for homogeneity - Calculator

Add data to Matrix A — 2nd [matrix] x^{-1} Edit A

Calculate χ^2, p-value, df — STAT TESTS χ^2-test
Observed A
Expected values are stored in B ← Expected B
Calculate Draw ↑ curve

$\chi^2 = 628.075$ $p = 0$ $df = 8$

Example 14.9

	Exclusive Territory		
Success	Yes	NO	Total
Yes	108	15	123
NO	34	13	47
Total	142	28	170

H_0: There is no association between exclusive terr. and success

Exclusive territory (column) - explanatory
Success (row) - response

Observations are from a single population in two ways. Use χ^2 test of association/independence

χ^2 test - tests the hypothesis row and column variables are not related to each other

Chi Square test of Association / Independence

Compares observations from a single population classified in two ways.

Steps

① Hypothesis:

H_0: There is no association between _____ & _____

H_a: There is an association between the two variables.

② Procedure:

χ^2 test for association

③ Conditions:

SRS

Expected counts are ≥ 5 Expected = $\frac{(\text{row total})(\text{column total})}{\text{grand total}}$

④ Math

$\chi^2 = \sum \frac{(O-E)^2}{E}$ $df = (r-1)(c-1)$ p-value = _____

⑤ Conclusion

Since the p-value _____ $\leq$ α = _____, we reject / fail to reject H_0. Conclusion in context

Example 14.9

	Exclusive Territory		
Success	Yes	No	Total
Yes	108	15	123
No	34	13	47
Total	142	28	170

14.2 cont'd

① H_0: There is no association between success of the business and the exclusive territory clause.

H_a: There is an association between success and exclusive territory

② χ^2 test for association

③ SRS is assumed

$$E = \begin{bmatrix} 102.741 & 20.259 \\ 39.259 & 7.741 \end{bmatrix}$$ All expected counts are ≥ 5

④ $\chi^2 = \frac{(108-102.741)^2}{102.741} + \dots \frac{(13-7.741)^2}{7.741} = 5.911$

$df = (2-1)(2-1) = 1$ p-value = .015

⑤ Since the p-value $.015 < \alpha = .05$ we reject H_0. There is evidence of an association between the success of a business and exclusive territory clause.