Chi-Square

Goodness of Fit Test

Course: Statistics 1

Lecturer: Dr. Courtney Pindling

Parametric and Nonparametric Tests

Parametric:

- Normal and homogeneous distributions
- Assumptions about parameters
- Interval or ratio scales

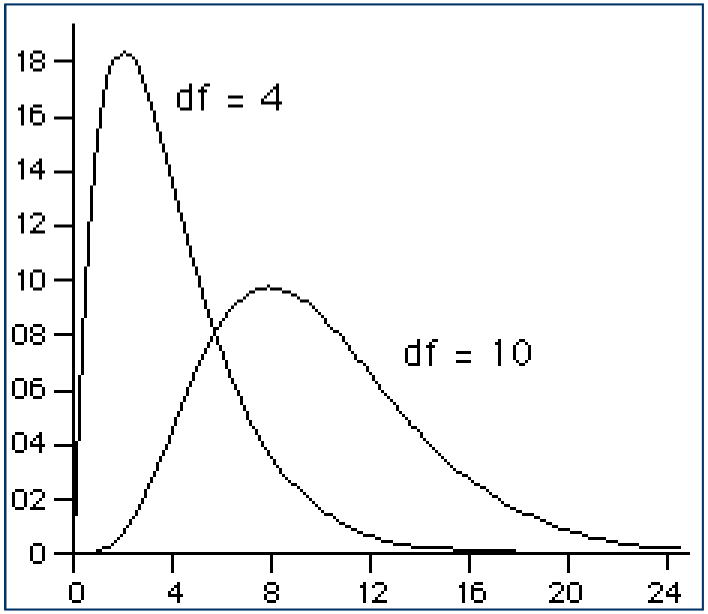
Nonparametric:

- Distribution independent
- Frequencies, ranks, or nonparametric measures
- Nominal or ordinal scales

Chi-Square Distribution, χ^2

Ch-Square Distribution:
Degree of freedom,
df = ***C - 1***, C is number of categories

1. Shape of Chi-Square depends on ***df***
2. As number of categories increase, the mode of the distribution has a larger chi-square value
3. Family of chi-square distributions (*df*)



Goodness of Fit

- Uses sample data to test hypothesis about the shape or proportion of a population distribution
- Test how well the sample distribution fits the population distribution specified by H_0
- Null Hypothesis, H_0:
  - ***No Preference****: The proportion is equally divided among the categories* ***or***
  - ***No Difference from Know Population****: The proportion of one population is no different from the proportion of another*

Frequencies

- **Observed Frequency**, f_o:

  The number of individuals from the sample who are classified in a particular category. Each individual is counted as one-and-only one category

- **Expected Frequency**, f_e:

  For each category, is the frequency value that is predicted from the H_0 and the sample size (n).

  $f_e = pn$, where p is the proportion stated by H_0

Example of f_e

n = 60	Category A	Category B	Category C	Category D
H_0:	25%	20%	30%	25%
p	0.25	0.20	0.30	0.25
$f_e = pn$	*15*	*12*	*18*	*15*

Chi-Square Statistics

Steps to calculate χ^2

1. Find $f_o - f_e$ for each category
2. Square the difference
3. Divide Step 1 by f_e
4. Add values from all categories, this is the χ^2 *statistics*

$$\text{chi-square} = \chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

Sample of Goodness of Fit

- A researcher uses a chi-square test for goodness of fit with 389 people to determine if there are any preferences among four different fruit juices.

n = 389 ***df*** = 4 -1 = **3**	Drink A	Drink B	Drink C	Drink D
H_0:	25%	25%	25%	25%
p	0.25	0.25	0.25	0.25
Responses	*89*	*107*	*104*	*89*

H_0: No Preference or equally likely proportion, p = ¼ = **0.25**

Goodness of Fit Table

Criterion	f_o	p	$f_e = pn$	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
Drink A	89	0.25	97.25	-8.25	68.0625	0.70
Drink B	107	0.25	97.25	9.75	95.0625	0.98
Drink C	104	0.25	97.25	6.75	45.5625	0.47
Drink D	89	0.25	97.25	-8.25	68.0625	0.70
Totals	**389**	**1**	**389**	**0**	X^2 =	**2.85**

Chi-square statistics = 2.85; restriction: no more than 20% of cells should have $f_e < 5$

Chi-Square: Critical Value

df	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.10}$	$\chi^2_{0.90}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$	$\chi^2_{0.995}$
1	0.000039	0.00016	0.00098	0.0039	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.01	0.0201	0.0506	0.1026	0.2107	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.8

- The critical region of the chi-square test is the region above *1- a*; so for a = 0.05 and *df* = 4 – 1 = 3, $\chi^2 = 7.81$ ($\chi^2_{0.95}$)

Decision and Conclusion

- Chi-Square statistics of 2.85 < Chi-Square Critical or *2.85* < *7.81*
- Do <u>not reject</u> H_0 and so
- Conclude that the four *drinks are equally likely to be preferred*.