Social Robotics

Albert-Ludwigs-Universität Freiburg

Bernhard Nebel, Felix Lindner, Thorsten Engesser,
Barbara Kuhnert, Laura Wächter

WS 2017/18

χ^2 Tests

Agenda

- Learn about the analysis of categorical data using χ^2 tests:
  - Pearson's Goodness of Fit
  - Pearson's Test of Independence
  - McNemar's Significance of Change

Hypothesis Testing for Categorical Data

- We again consider Robo-One and Robo-Two. Over a fixed time period, we invite people to either interact with Robo-One or with Robo-Two (they can choose which robot they prefer).
- H_1: The robots differ in the number of interactions they have with people.
- H_0: People interact equally likely with Robo-One and Robo-Two.
- This time, we do not consider means but total frequencies of interactions: In total, there were 155 interactions with Robo-One and 195 interactions with Robo-Two. Our raw data may look like this: `RoboOne, RoboOne, RoboTwo, RoboOne, ...`
- Can we reject H_0 in favor of H_1?

Pearson's Goodness of Fit

- Given k observed frequencies $O_1, \ldots, O_k$ of events from levels $1, \ldots, k$ (of one categorical variable).
- Given expectations $E_1, \ldots, E_k$ about these frequencies given by the H_0 hypothesis.
- $\sum_i^k O_i = \sum_i^k E_i = n$

Example

- Level 1: #interactions with Robo-One, Level 2: #interactions with Robo-Two
- $O_1 = 155, O_2 = 195, \sum_i^2 O_i = 350$
- H_0 states that $P(RoboOne) = P(RoboTwo) = 0.5$.
- $E_1 = P(RoboOne) \times 350 = E_2 = P(RoboTwo) \times 350 = 175$

Pearson's χ^2 Statistics

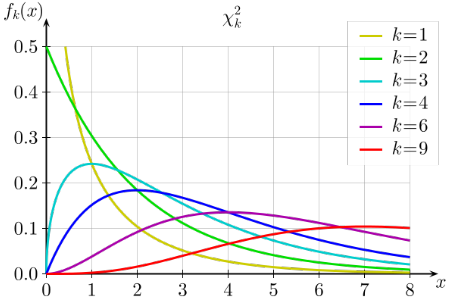
- $\chi^2 = \sum_i^k \frac{(O_i - E_i)^2}{E_i}$
- Intuition:
  - O_i can be thought of as being Poisson distributed variables with mean E_i (per H_0), i.e., $O_i \sim Pois(E_i)$.
  - Thus, $Mean(O_i) = E_i = Var(O_i)$.
  - For $E_i \geq 5$, $Pois(E_i)$ looks roughly normal ($\Rightarrow$Pearson's χ^2 assumes at least 80% of E_is to be 5 or greater).
  - Then, $z = \frac{(O_i - E_i)}{\sqrt{E_i}}$ is just the z-Score for O_i, and $z \sim \mathcal{N}(0, 1)$.
  - Thus, χ^2 is just the sum of squared standardized normally distributed variables.

Example

$$\chi^2 = \frac{(155-175)^2}{175} + \frac{(195-175)^2}{175} = 2.28 + 2.28 = 4.57$$

χ^2 Distribution

- Let $Z_1, \ldots, Z_k$ be standardized normally distributed variables, i.e., $Z_i \sim \mathcal{N}(0, 1)$.
- The sum of the squared variables is chi-squared distributed with degree of freedom k, i.e., $\sum_i^k Z_i^2 \sim \chi_k^2$



Looking up the Critical Value

1. Decide on significance level α.
2. Determine the degree of freedom *df*.
3. Determine whether $P(z \geq \chi^2) \leq \alpha$ (in this case, H_0 can be rejected).

Example

- Let $\alpha = 0.05$ (critical value $\chi^2_{1;95\%} = 3.84$)
- Given n fixed, there is one variable that can vary freely, therefore, $df = 1$.
  - *Note*: If there are n levels, then $df = n - 1$.
- $P(z \geq 4.57) = 1 - P(z < 4.57) = 0.03 \leq 0.05$.

Reporting Result

Report

We have found that Robo-One and Robo-Two significantly differ in numbers of interactions they have with people ($\chi^2(1) = 4.57, p = 0.03$).

Note: Directional Hypotheses

- At first, it seems not quite possible to test directional hypotheses, because by squaring the test statistics, the directionality gets lost. Moreover, the χ^2 distribution only has one tail.
- Trick (only works if *df* = 1), cf. Bortz & Schuster (2010)
  1. If you hypothesize that Level1 is more probable than Level2, first check if $O_1 > E_1$. In case this fails, H_0 may not be rejected. If test succeeds goto (2).
  2. Check whether the computed χ^2 statistics is bigger than the critical value $\chi^2_{1;90\%}$.

Summary

- Observed frequencies of events falling in k different levels.
- Null-Hypothesis H_0 about the expected frequencies, viz., usually that there is no difference in probability.
- Alternative hypothesis H_1 usually is a statement about the probabilities of the k levels to differ.
- χ^2 statistics represents the sum of standardized deviations of the observations from the expectations due to H_0.
- The χ^2-distribution is then used to judge the χ^2-statistics as too big or not. χ^2-distribution actually is a family of distributions, i.e., the p% quartile depends on the degree of freedom.

Test of Independence

- Pearson's statistics can also be used to test whether two variables are independent given some observations. (Here, stochastic independence is meant!)

Example (Fictional!)

Consider the following exmperiment: We are interested in whether the age of a person affects her liking RoboOne or RoboTwo. To test this, for both RoboOne and RoboTwo a picture was uploaded to facebook. People were asked to write in the comments, which of the two robots they like better. In facebook it was easy to also determine the age of each person.

Experiment: Data

	0 – 45	46 – 99	
RoboOne	65	30	**95**
RoboTwo	100	25	**125**
	165	**55**	$n = 220$

Hypotheses

- H_1: There is a relationship between Robot Type and Age.
- H_0: There is no such relationship, i.e., the variables are independent.

Experiment: Calculating χ^2

	0 – 45	46 – 99	
RoboOne	65	30	**95**
RoboTwo	100	25	**125**
	165	**55**	n = 220

- $E_1 = P(RoboOne)P(Age0045) \times n = \frac{95}{220}\frac{165}{220} \times 220 = 71.25$
- $E_2 = P(RoboOne)P(Age4699) \times n = \frac{95}{220}\frac{55}{220} \times 220 = 23.75$
- $E_3 = P(RoboTwo)P(Age0045) \times n = \frac{125}{220}\frac{165}{220} \times 220 = 93.75$
- $E_4 = P(RoboTwo)P(Age4699) \times n = \frac{125}{220}\frac{55}{220} \times 220 = 31.25$

Experiment: Calculating χ^2

	0 – 45	46 – 99	
RoboOne	65 71.25	30 23.75	**95**
RoboTwo	100 93.75	25 31.25	**125**
	165	**55**	$n = 220$

$$\chi^2 = \frac{(65-71.25)^2}{71.25} + \frac{(30-23.75)^2}{23.75} + \frac{(100-93.75)^2}{93.75} + \frac{(25-31.25)^2}{31.25} = 3.859$$

Experiment: Rejecting H_0

- Degree of freedom for 2×2 tables is 1, because, given the totals, knowing one frequency implies all of the other three frequencies.
- Significance level is assumed to be $\alpha = 0.05$.
- To decide whether or not to reject H_0, it remains to check if $\chi^2 = 3.859$ is in the upper 5% interval of χ^2_1: $\chi^2_{1;95\%} = 3.841 < 3.858$ ☺.
- The p-value is $1 - P(z < 3.859) = 0.049 < 0.05$.

Reporting Result

Report

We have found that there is a significant dependency between age and the robot type people like ($\chi^2(1) = 3.859, p = 0.049$).

Generalization to $n \times m$

	0 – 33	34 – 67	68 – 99	
RoboOne	10	20	10	**40**
RoboTwo	20	40	20	**80**
RoboThree	30	20	10	**60**
	60	**80**	**40**	$n = 180$

- The computations of the χ^2 statistics for multiple number of levels are exactly the same as for 2×2 tables.
- The degrees of freedom can be determined by $(r-1)(c-1)$ with r being the number of rows and c being the number of columns.

A Note on R's chisq.test

- No `alternative` parameter: Testing directional hypotheses is not supported.
  - However, it is possible to test directional hypotheses in case of 2×2 tables. E.g., One can test whether people at age 0–45 like RoboTwo better than people at age 46–99: First check $O_3 = 100 > E_3 = 93.75$, then check if $\chi^2_{1;90\%} = 2.71 \leq \chi^2 = 3.859$.
- Yate's correction set TRUE by default: $\frac{(|O_i - E_i| - 0.5)^2}{E_i}$. More conservative. Field, Miles, Field (2012): "it's probably best ignored."

McNemar Test

- Sometimes, we might be interested in whether some manipulation yields some significant change.
- In this case, we test persons twice, i.e., before the manipulation and afterwards.
- To test if the manipulation results in a significant change, the McNemar-χ^2 can be used.

Example

We are interested in whether there is a significant change in the number of people who belief or disbelief in climate change after they have had a discussion with our climate-expert robot.

McNemar Test: Hypothesis & Data

- H_1: The discussion with the robot yields a significant change in people's mind. I.e., the distribution of Belief resp. Disbelief changes, $H_1 : P(c) \neq P(b)$.
- H_0: There is no significant change. I.e., the number of those who switch from Belief to Disbelief should equal the number of those who change from Disbelief to Belief: $H_0 : P(c) = P(b)$,

	Belief	Disbelief
Belief	a = 50	b = 10
Disbelief	c = 30	d = 30

Tabelle: Left variable: Before, Top variable: Afterwards

McNemar Test: Data Analysis

	Belief	Disbelief
Belief	a = 50	b = 10 20
Disbelief	c = 30 20	d = 30

- Assuming H_0, we should expect as many changers from Disbelief to Belief as from Belief to Disbelief. (Such that, all in all, things stay the same.)
- We can now compute the χ^2 statistics, where our observations are the number of changers, and our expectations are just the average of the total number of changers. I.e.:
  - $O_b = 10, O_c = 30, E_b = E_c = \frac{30+10}{2} = 20$
  - $\chi^2 = \frac{(10-20)^2}{20} + \frac{(30-20)^2}{20} = 10.0$

Reporting

Report

Running a McNemar test, we have found a significant change of attitute towards climate change due to discussions with our robotic climate-change expert ($\chi^2(1) = 10.0, p < 0.0016$).

Directional McNemar Test

- The directional test works as follows (cf., Bortz & Schuster 2010): Given that we had the H_1 that the discussion increases the belief in climate change ($H_1 : P(c) > P(b), H_0 : P(c) \leq P(b)$):
  1. Check whether $c > b$. If not, H_0 may not be rejected. If yes, goto (2).
  2. Run McNemar test, confidence level 90% is sufficient ($\chi^2_{1;90\%} = 2.71$).

Summary

- χ^2 tests are the method of choice for the analysis of categorical data.
- Goodness of Fit: Compare observed frequencies to expected frequencies.
- Test of Independence: Check whether or not the observed frequencies for two or more categorical variables are likely if the variables were statistically independent of each other.
- Significance of Change Test: Check whether the distribution of a categorical variable changes due to manipulation.

Sketches

Intentionally left blank :-)