Editing file

Feedback

Degrees of freedom

Color Index:

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

Revised by

MED439
KING SAUD UNIVERSITY

Research Team

RESEARCH TEAM
438
Keep it simple, Keep it focused

Degrees of freedom

Estimates of parameters can be based upon different amounts of information. The number of independent pieces of information that go into the estimate of a parameter is called the degrees of freedom (df). In general, the degrees of freedom of an estimate is equal to the number of independent scores that go into the estimate minus the number of parameters estimated as intermediate steps in the estimation of the parameter itself. For example, if the variance, σ2, is to be estimated from a random sample of N independent scores, then the degrees of freedom is equal to the number of independent scores (N) minus the number of parameters estimated as intermediate steps and is therefore equal to N-1.

Simplified definition of degrees of freedom: How many observations are independent and how many are dependent.

Quantifying uncertainty

Standard deviation: measures the variation of a variable in the sample.

Technically,

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Degrees of freedom for one sample of data

Example

Data: X = {6, 10, 5, 4, 9, 8}; N = 6

X	$X-\bar{X}$ Deviation	$(X-\bar{X})^2$
6	6-7(mean)= -1	1
10	10-7= 3	9
5	5-7= -2	4
4	4-7= -3	9
9	9-7= 2	4
8	8-7= 1	1
Total: 42	Total: 0	Total : 28

5 observations are **independent** (because we need to calculate each one separately and they are independent of the other values)

1 observation is **dependent** (because we can calculate it by subtracting the total of the previous values from the whole total (0 - - 1=1) which makes it a dependent value and that's why the degrees of freedom is calculated as n-1 (1 which is a dependent value on the total) df=6(n)-1(dependent values)= 5

Mathematical principle: always the total of the sum of the deviations from the mean (total "data-mean") is zero. This is one mathematical restriction that we are imposing in this values and it's one meaning of degrees of freedom.

Mean:

$$\bar{x} = \frac{\sum X}{N} = \frac{42}{6} = 7$$

Variance:

$$S^2 = \frac{\sum(X-\bar{x})^2}{N} = \frac{28}{6} = 4.67$$

Standard Deviation:

$$S = \sqrt{S^2} = \sqrt{4.67} = 2.16$$

Chi-Square Test

	MI	Non-MI	
Smoker	29 O / E 22.5	21 O / E 27.5	50
Non smoker	16 O / E 22.5	34 O / E 27.5	50
	45	55	100

O → Observed frequency
E → Expected frequency

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

If you know the value of 3 cells you can calculate the 4th cell by subtracting the sum of the 3 cell from the whole total (100). So we have 3 dependent cells and 1 independent. That's why the degrees of freedom = (r-1)(c-1)=(2-1)(2-1)=1. The degrees of freedom for all 2x2 tables equals 1.

The difference between this exercise (chi-square test) and the previous exercise is that here we considered the dimensions of the table the **rows and columns** and the previous exercise we considered the sample size.

A mistake that the students commonly do in chi square test and mcnemar's test is using the sample size in degrees of freedom calculation.

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 Observed frequency (82) Expected frequency E= (96x940)/1100 I need to do the multiplication making this value an independent value	20 (14) We can calculate the expected frequency by subtracting 82 from 96 (total) making it a dependent value	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

Only these 4 values are **independent** of the marginal total (because we need to do the multiplication of each one separately) can't be calculated by subtraction unlike the rest of the cells which are dependent on the total

A mistake that the students commonly do here is considering the marginal total as a row and the age subdivisions as a column. There is 5 rows in this table **not** 6 and 2 columns.
df= ((r-1)(c-1)=(5-1)(2-1)=4 (the degrees of freedom calculated from the equation shows us how many independent v values we have)

Degree of freedom:

- **Qualitative (categorical)**
  (look at the dimensions of the tables)
  - **Chi-square test:** (R-1)x(C-1)
  - **Fisher's Exact test:** (R-1)x(C-1)
  - **Mcnemar's test:** (R-1)x(C-1)
- **Qualitative:**
  - **Student's t test (one sample):** n – 1
  - **Student's t test (two samples):** n1 + n2 – 2

Research Team

RESEARCH TEAM
438
Keep it simple, Keep it focused

Thank you for checking our work!

Leaders:

Shuaa Khdary Sarah AlQuwayz

Abdulrhman Alsuhaibany

Contact us:

Research4390@gmail.com