భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Biostatistics BT2023

Lecture 17
Chi-square test and distribution

Himanshu Joshi
6 October 2023

Gamma Function

$$\Gamma n = \int_0^\infty x^{n-1} e^{-x} dx$$

$$\Gamma n = (n-1)\Gamma(n-1)$$

$$\Gamma \frac{1}{2} = \sqrt{\pi}$$

Gaussian function

$$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2}$$

Most accurate value of pi

The most accurate value of pi is 62,831,853,071,796 digits, and was achieved by University of Applied Sciences of the Grisons (Switzerland) in Chur, Switzerland, on 19 August 2021.

Team DAViS of UAS Grisons, headed by Heiko Rölke and Thomas Keller, computed an additional 12.8 trillion digits of pi. They employed a computer with one terabyte of RAM and 510 terabytes of disk space. The system completed the calculation within 108 days. The record run was part of a tuning and testing phase of a server which will be used in areas such as RNA analysis and machine translation

https://www.guinnessworldrecords.com/world-records/66179-most-accurate-value-of-pi

Chi-square χ^2 test

It is used when the distribution of a categorical variable in a sample often needs to be compared with the distribution of a categorical variable in another sample

Table 8.1 Distribution by socioeconomic class of patients admitted to self poisoning (sample A) and gastroenterological (sample B) units

Socioeconomic class	Samples		Total	Proportion in group A
	A	B		
	a	b	n = a + b	p = a/n
I	17	5	22	0.77
II	25	21	46	0.54
III	39	34	73	0.53
IV	42	49	91	0.46
V	32	25	57	0.56
Total	**155**	**134**	**289**	

The psychiatrist wants to investigate whether the distribution of the patients by social class differed in these two units. She therefore erects the null hypothesis that there is no difference between the two distributions. This is what is tested by the chi squared (χ^2) test (pronounced with a hard ch as in "sky").

By default, all χ^2 tests are two sided.

https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/8-chi-squared-tests

Compute the chi^2 tables

Table 8.2 Calculation of the χ^2 test on figures in table 8.1

Class (I)	Expected numbers		0 - E		$(0-E)^2/E$	
	A (2)	B (3)	A (4)	B (5)	A (6)	B (7)
I	11.80	10.20	5.20	-5.20	2.292	2.651
II	24.67	21.33	0.33	-0.33	0.004	0.005
III	39.15	33.85	-0.15	0.15	0.001	0.001
IV	48.81	42.19	-6.81	6.81	0.950	1.009
V	30.57	26.43	1.43	-1.43	0.067	0.077
Total	**30.57**	**134.00**	**0**	**0**	**3314**	**3.833**

χ^2 = 3.314 + 3.833 = 7.147. d.f. = 4. 0.10<P<0.50.

- The sum of these differences always equals zero in each column.
- Each difference for sample A is matched by the same figure, but with opposite sign, for sample B.

Hypothesis testing and confidence interval

If your chi-square calculated value is greater than the chi-square critical value, then you reject your null hypothesis.

If your chi-square calculated value is less than the chi-square critical value, then you "fail to reject" your null hypothesis.

Chi-square Table

Distribution of μ						
	Probability					
d.f.	0.5	0.10	0.05	0.02	0.01	0.001
1	0.455	2.706	3.841	5.412	6.635	10.827
2	1.386	4.605	5.991	7.824	9.210	13.815
3	2.366	6.251	7.815	9.837	11.345	16.268
4	3.357	7.779	9.488	11.668	13.277	18.465
5	4.351	9.236	11.070	13.388	15.086	20.517
6	5.348	10.645	12.592	15.033	16.812	22.457
7	6.346	12.017	14.067	16.622	18.475	24.322
8	7.344	13.362	15.507	18.168	20.090	26.125
9	8.343	14.684	16.919	19.679	21.666	27.877
10	9.342	15.987	18.307	21.161	23.209	29.588
11	10.341	17.275	19.675	22.618	24.725	31.264
12	11.340	18.549	21.026	24.054	26.217	32.909
13	12.340	19.812	22.362	25.472	27.688	34.528
14	13.339	21.064	23.685	26.873	29.141	36.123
15	14. 339	22.307	24.996	28.259	30.578	37.697

Degrees of freedom df = (number of rows -1) (number of columns- 1)

Normal distribution table

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Chi-square χ^2 test

When we have two ‘categorical variables’, we use chi-square test to Test the null hypothesis

$$\chi^2 = \sum_{k=1}^{N} \frac{(O_k - E_k)^2}{E_k}$$

Example Gender vs Education

Observed

	Female	Male
Without gradualtion	6	7
College	13	16
Bachelor	16	15
Master	8	11
Total	**43**	**49**

Expected

	Female	Male
Without gradualtion	6.08	6.92
College	13.55	15.45
Bachelor	14.49	16.51
Master	8.88	11.12
Total	**43**	**49**

$$\chi^2 = 0.504$$

Chi-square χ^2 test

	Observed	Expected
One Car	73	60
Two car	38	28
Three car	18	12
Total	**129**	

भारतीय सांकेतिक विज्ञान संस्थ हैदराबाद
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Chi-square distribution

Null Hypothesis H_0 ==> *Gender* and *Education* does not have any correlation

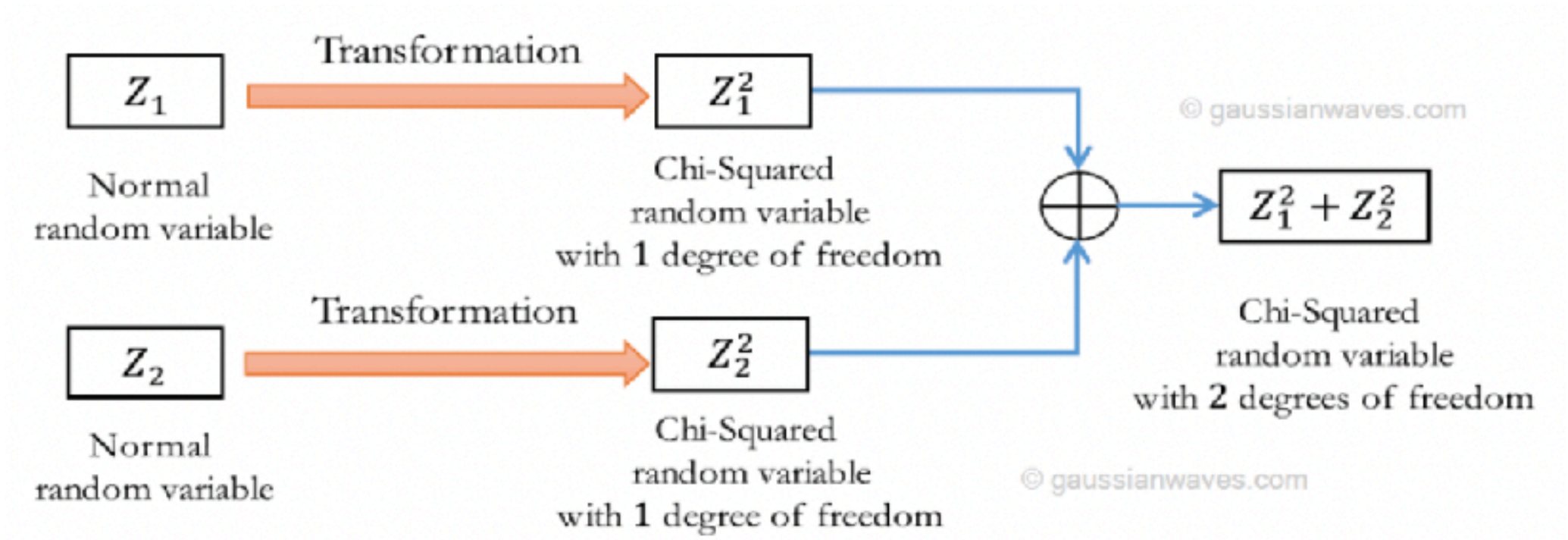
Alternative Hypothesis H_1 ==> *Gender* and *Education* do have any correlation

Degrees of freedom df = (number of rows -1) (number of columns- 1)

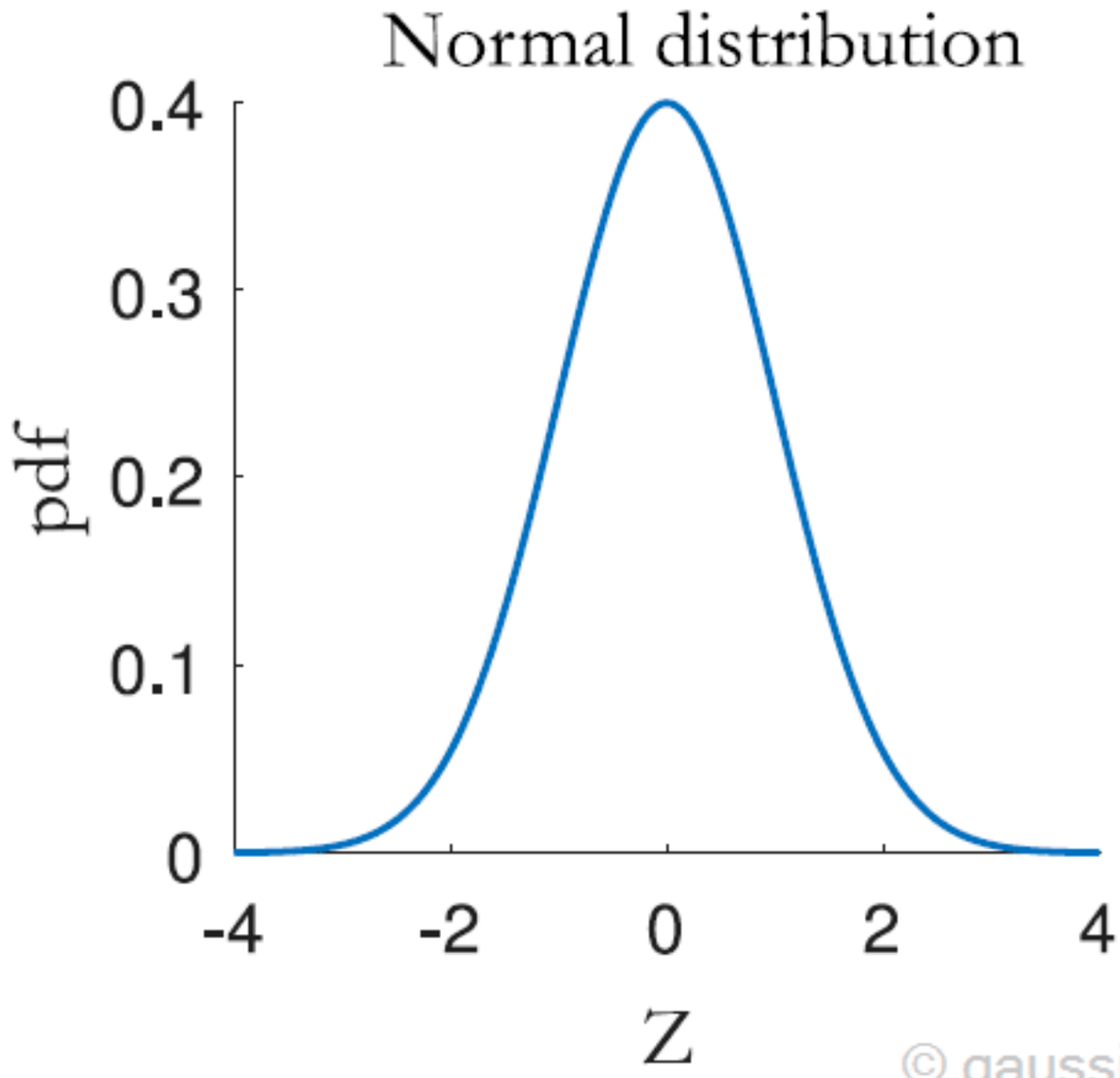
$$f(x,k) = \begin{cases} \frac{x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})}; & x > 0 \\ 0, & otherwise \end{cases}$$

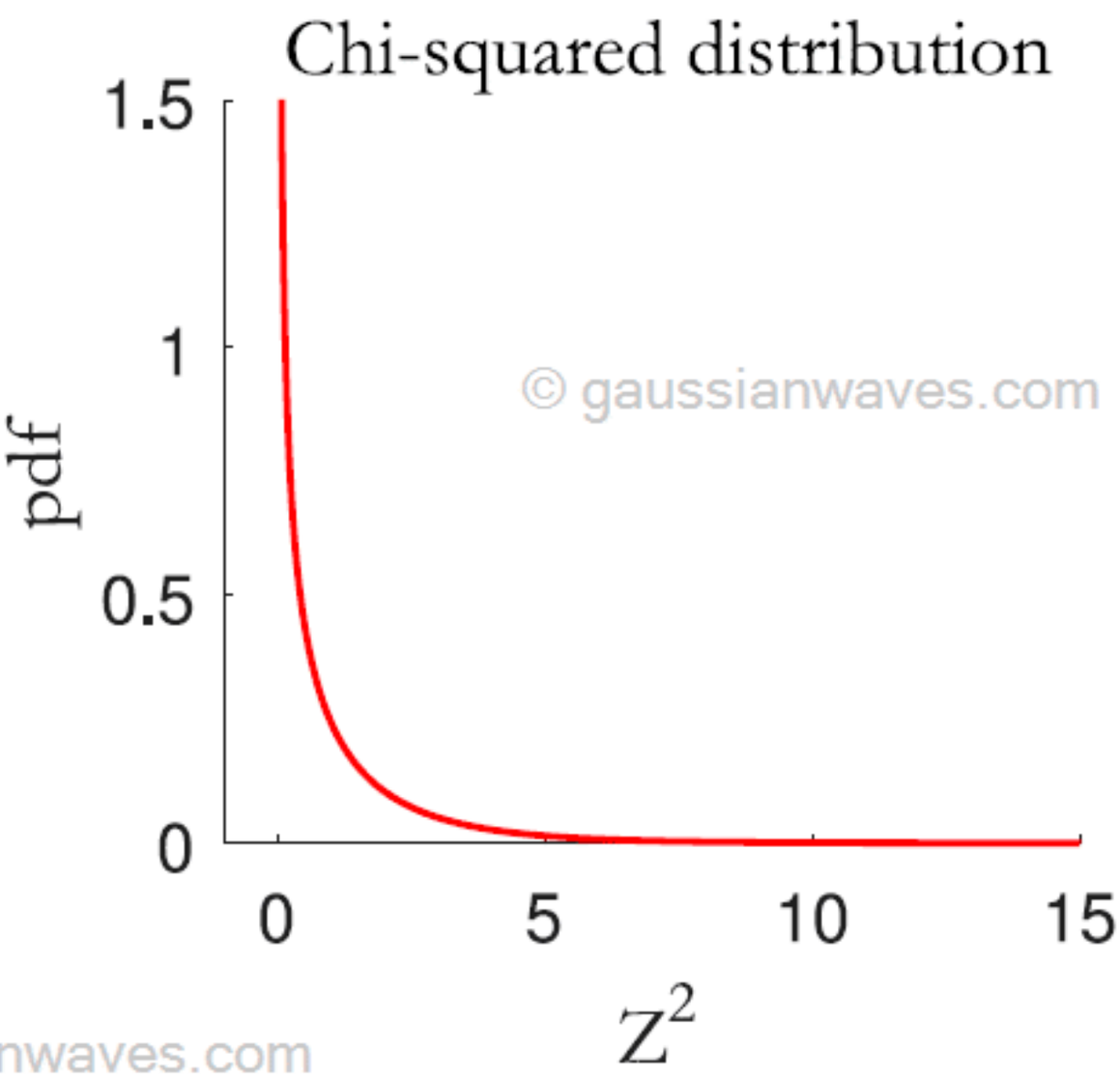
All kinds of variables in natural and social sciences are normally or approximately normally distributed.

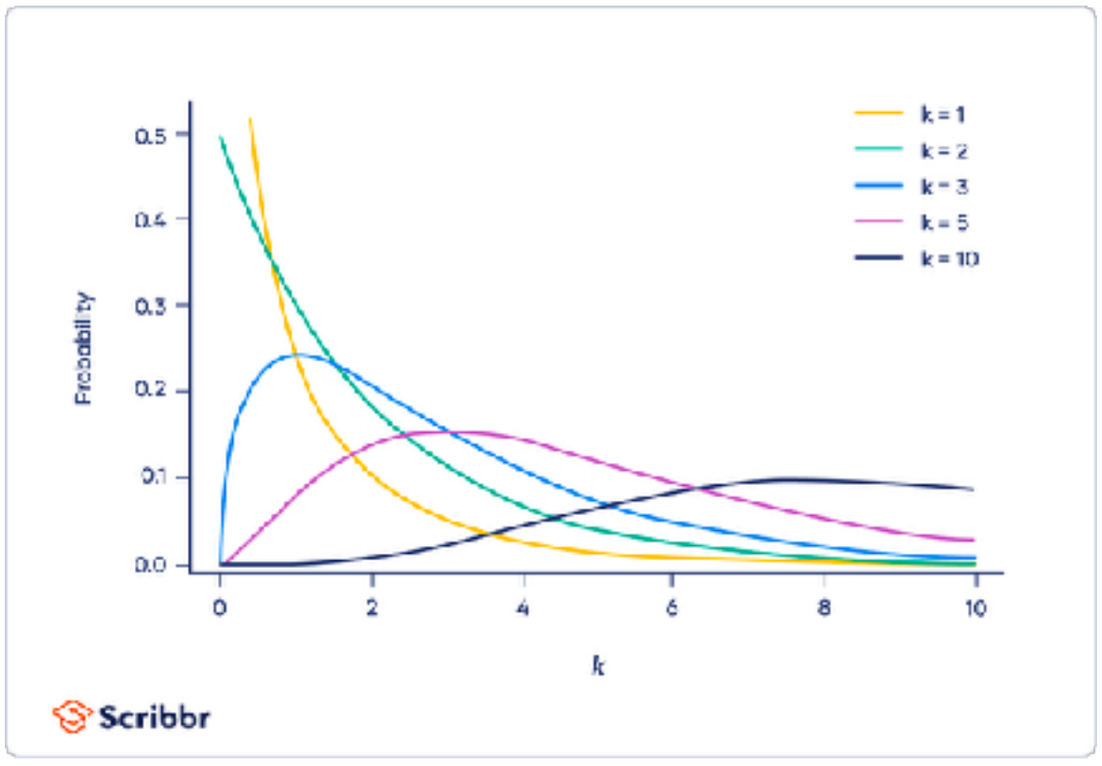
Chi-square distribution

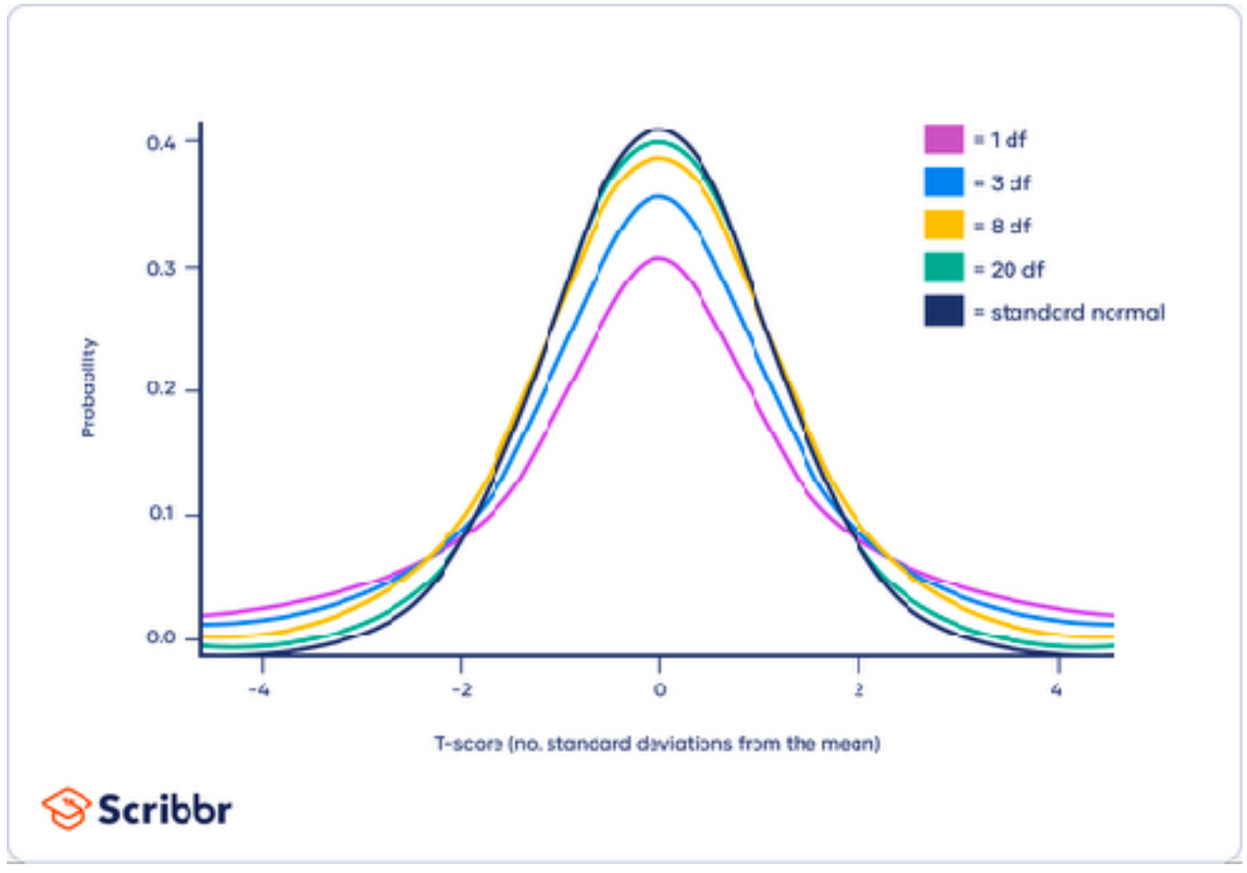


Chi-square distribution






k = 1
k = 2
k = 3
k = 5
k = 10
0.5
0.4
0.3
0.2
0.1
0.0
0
2
4
6
8
10
k
Scribbr


= 1 df
= 3 df
= 8 df
= 20 df
= standard normal
0.4
0.3
0.2
0.1
0.0
-4
-2
0
2
4
T-score (no. standard deviations from the mean)
Scribbr

R programming

Installing in ubuntu using *deb*
Interactive mode
Programming with script
RStudio

Frequently used R objects

- Vectors
- List
- Matrices
- Arrays
- Factors
- Data Frames