Degree of Freedom

What are degrees of freedom in statistics?

Degrees of freedom are the number of independent values that a statistical analysis can estimate. You can also think of it as the number of values that are free to vary as you estimate parameters.

Typically, the degrees of freedom equal your sample size minus the number of parameters you need to calculate during an analysis.

Degrees of freedom is a combination of how much data you have and how many parameters you need to estimate. It indicates how much independent information goes into a parameter estimate.

Example:

Suppose we collect the random sample of observations shown below. Now, imagine we know the mean, but we don't know the value of an observation—the X in the table below.

	Values
	6
	8
	5
	9
	6
	8
	4
	11
	7
	X
Average	6.9
Sum	69

Using simple algebra (64 + X = 69), we know that X must equal 5

As you can see, that last number has no freedom to vary. It is not an independent piece of information because it cannot be any other value. Estimating the

parameter, the mean in this case, imposes a constraint on the freedom to vary. The last value and the mean are entirely dependent on each other. Consequently, after estimating the mean, we have only 9 independent pieces of information, even though our sample size is 10.

DF = N – P

Where:

- N = sample size
- P = the number of parameters or relationships

For example, the degrees of freedom formula for a 1-sample t test equals N – 1 because you're estimating one parameter, the mean. To calculate degrees of freedom for a 2-sample t-test, use N – 2 because there are now two parameters to estimate.

To find the chi-square DF for a table with r rows and c columns, use this formula to calculate degrees of freedom: (r-1) (c-1)

For example, to find the degrees of freedom in a 2 X 2 table, after you enter one value in the table, you can calculate all the remaining cells.

	Category A		**Total**
Category B	**15**	(4)	19
	(10)	(3)	13
Total	25	7	32

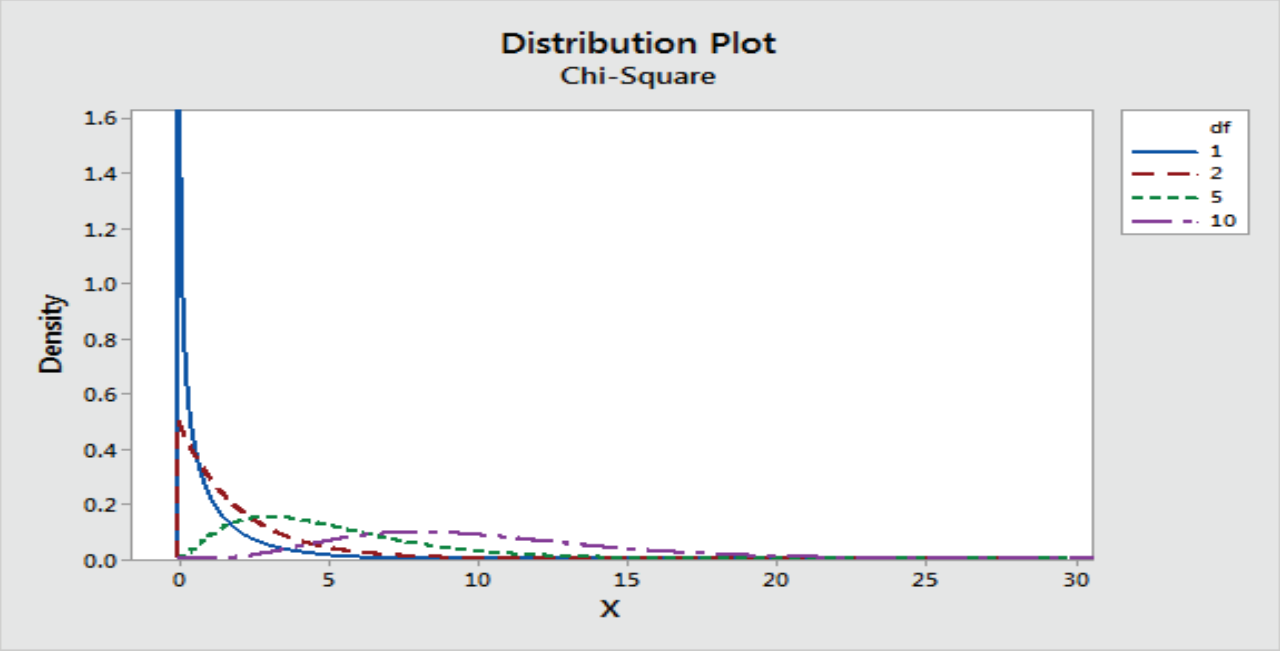
In the table above, I entered the bold 15, and then I can calculate the remaining three values in parentheses. Therefore, this table has 1 DF.

let's try finding degrees of freedom for 3 X 2 table.

	Shirt Color			Total
Status	**129**	**46**	(215)	390
	(7)	(9)	(24)	40
Total	136	55	239	430

In the table, one categorical variable is shirt color, which can be blue, gold, or red. The other categorical variable is status, which can be dead or alive. After I entered the two bolded values, I can calculate all the remaining cells. Consequently, this table has 2 DF.

chi-square distribution is a family of distributions where the DF define the shape.


Distribution Plot
Chi-Square
Density
X
df
1
2
5
10
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
0
5
10
15
20
25
30