The Count Loves Counting with Pearson's Chi-Square

Ben Babcock
University of Minnesota

Chi-square Vocabulary

Expected value: the number of people (monkeys, cucumbers, ...) that you think will fall in each cell.

Cell: A space in a table. Each cell contains an observed count.

L_1	L_2	L_3	L_4
Cell 1	Cell 2	Cell 3	Cell 4

Table: a grid that contains the number of people that fall into a given category. We describe table size by saying the number of rows with data followed by the number of columns with data.

The table above is a 1 x 4 table.

Counts Instead of Cont's

Suppose that you had a variable or several variables that had discrete categories. You want to know if the number of people you observe falling in each category is different than what you _____. You could put the observed counts into a table.

1 Variable: Which ice cream flavor do you prefer?

Vanilla	Chocolate	Strawberry	Hate Ice Cream
$Count_1$	$Count_2$	$Count_3$	$Count_4$

2 Variables: What is your sex and which snack do you prefer?

	Female	Male
Chocolate	$Count_1$	$Count_2$
Chips and Salsa	$Count_3$	$Count_4$

Counts Instead of Cont's

What we have:

1) a table with observed counts in cells
2) expected values for each cell

What we want to know:

The solution to this is Pearson's chi-square (χ^2) test.

Common types of χ^2 tests:

Goodness-of-Fit Test (Expected from a distribution you specify)

Test for Independence (Expected from marginal counts)

The χ^2 Equation

The equation for Pearson's χ^2 statistic for a table is

$$\chi^2(df) = \sum_{i=1}^{C} \frac{(O_i - E_i)^2}{E_i}$$

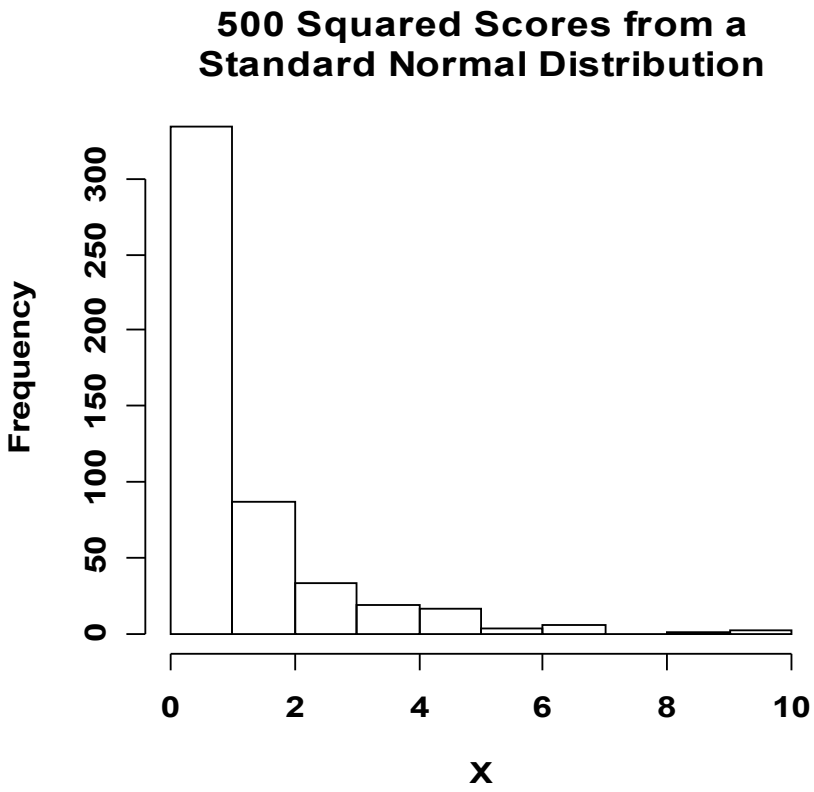
where i is an index for the cell number
df is the appropriate degrees of freedom,
O is the observed count
E is the expected count and
C is the total number of cells.

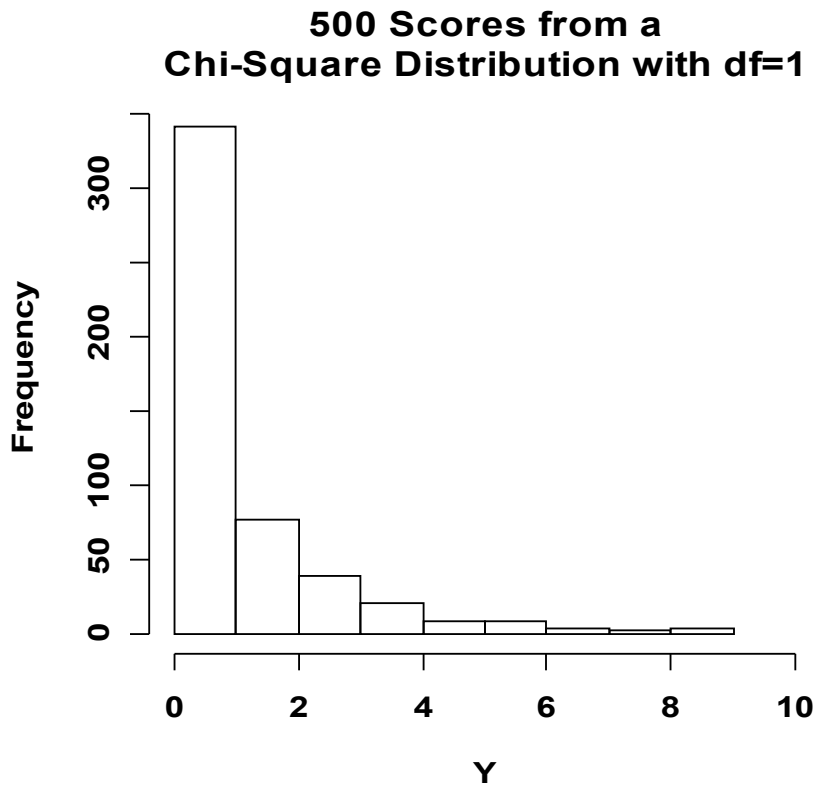
Only 1 row (Goodness of Fit): df = # columns - 1
More than 1 row (Test for Ind): df = (# rows – 1)·(# columns - 1)

The χ^2 Distribution

The χ^2 distributions are the sum of squared *z*-scores!

χ^2 with $df = 1$ is just like squaring scores from the standard normal distribution.





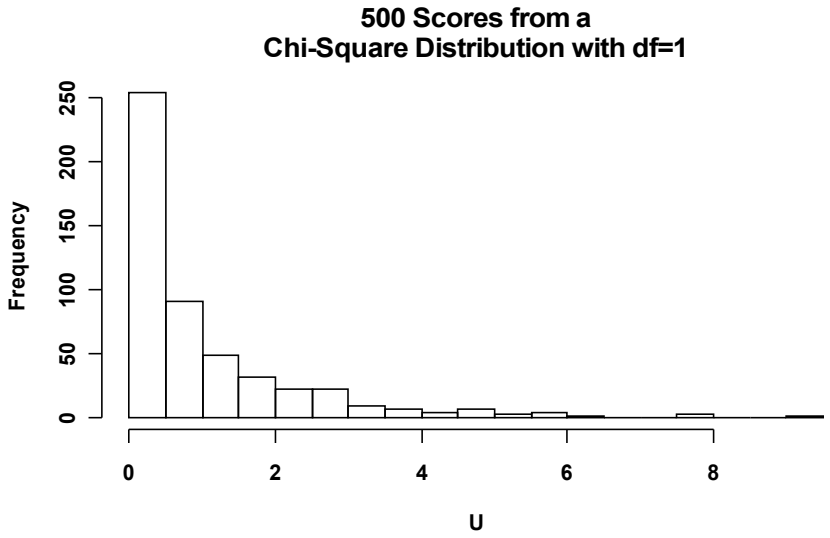
With $df = 1$, guess the χ^2 value where p=.05.

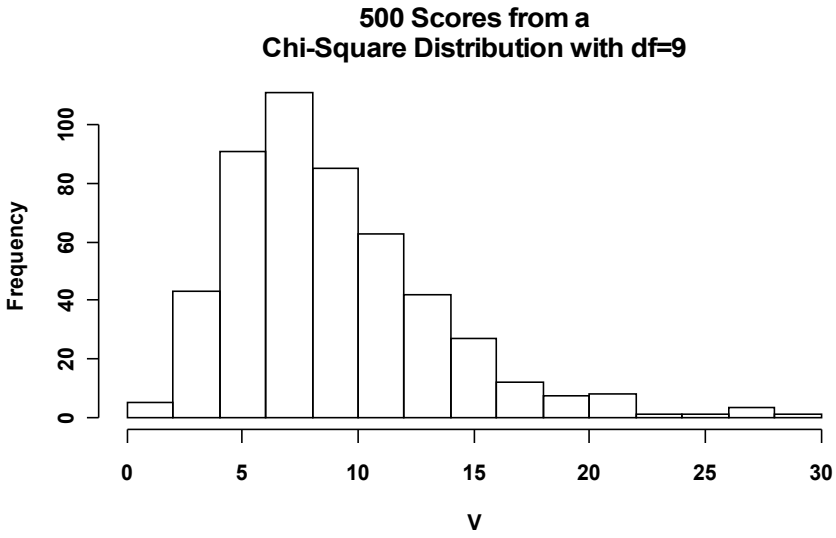
The χ^2 Distribution

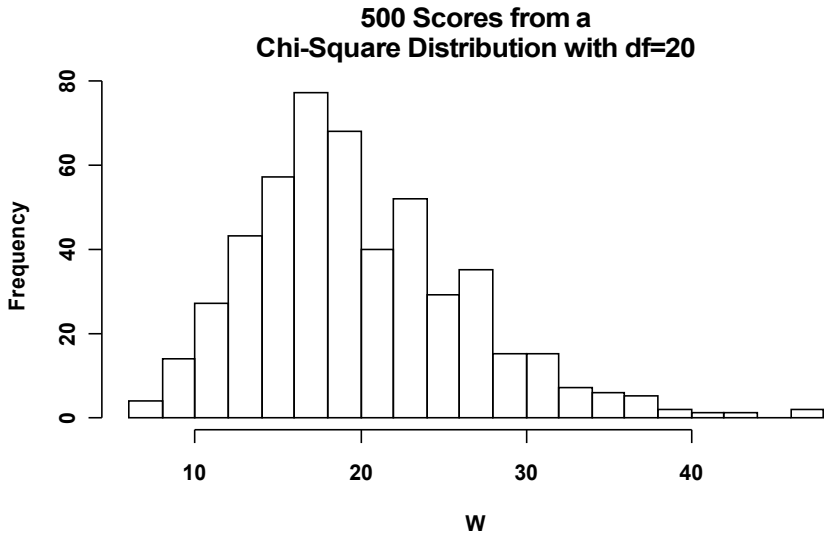
Just like the *t* and *F* distributions, the χ^2 distribution is a family of distributions.

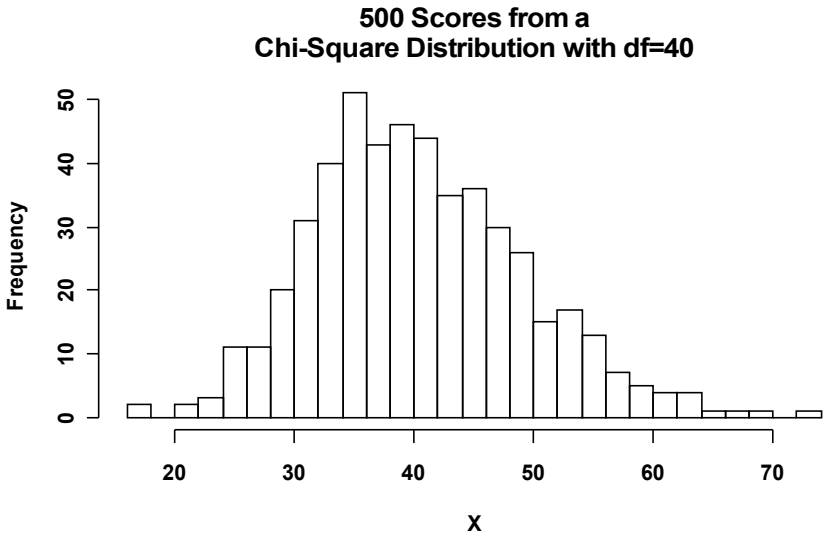
Unlike the *t* distribution, the χ^2 distribution can look wildly different depending on the _____.

The distributions start off with a great deal of _____, which becomes smaller and smaller as the degrees of freedom increases.


500 Scores from a
Chi-Square Distribution with df=1
Frequency
0
50
100
150
200
250
0
2
4
6
8
U


500 Scores from a
Chi-Square Distribution with df=9
Frequency
0
20
40
60
80
100
0
5
10
15
20
25
30
V


500 Scores from a
Chi-Square Distribution with df=20
Frequency
0
20
40
60
80
10
20
30
40
W


500 Scores from a
Chi-Square Distribution with df=40
Frequency
0
10
20
30
40
50
20
30
40
50
60
70
X

Steps to Calculating Chi-square

To the white board!

What does a significant result mean?

What does non-significance mean?

What level (NOIR) are the chi-square count data?