Chi square test

step by step

	pineapple		
OS_new	Idon'tmind	No	Yes
Apple	12	20	9
Microsoft	27	26	27

Contingency table

table(OS_new,pineapple)

pineapple/ new OS	Idon'tmind	No	Yes	Row totals
Apple	12	20	9	
Microsoft	27	26	27	
Column totals				

Column and row sums

please calculate and tell me what to fill in!

pineapple/ new OS	Idon'tmind	No	Yes	Row totals
Apple	12	20	9	41
Microsoft	27	26	27	80
Sum totals	39	46	36	121

Column and row sums

pineapple/ new OS	Idon'tmind	No	Yes	Row totals
Apple	12	20	9	R
Microsoft	27	26	27	R
Column totals	C	C	C	G

Expected values for each cell

E = R *C / G

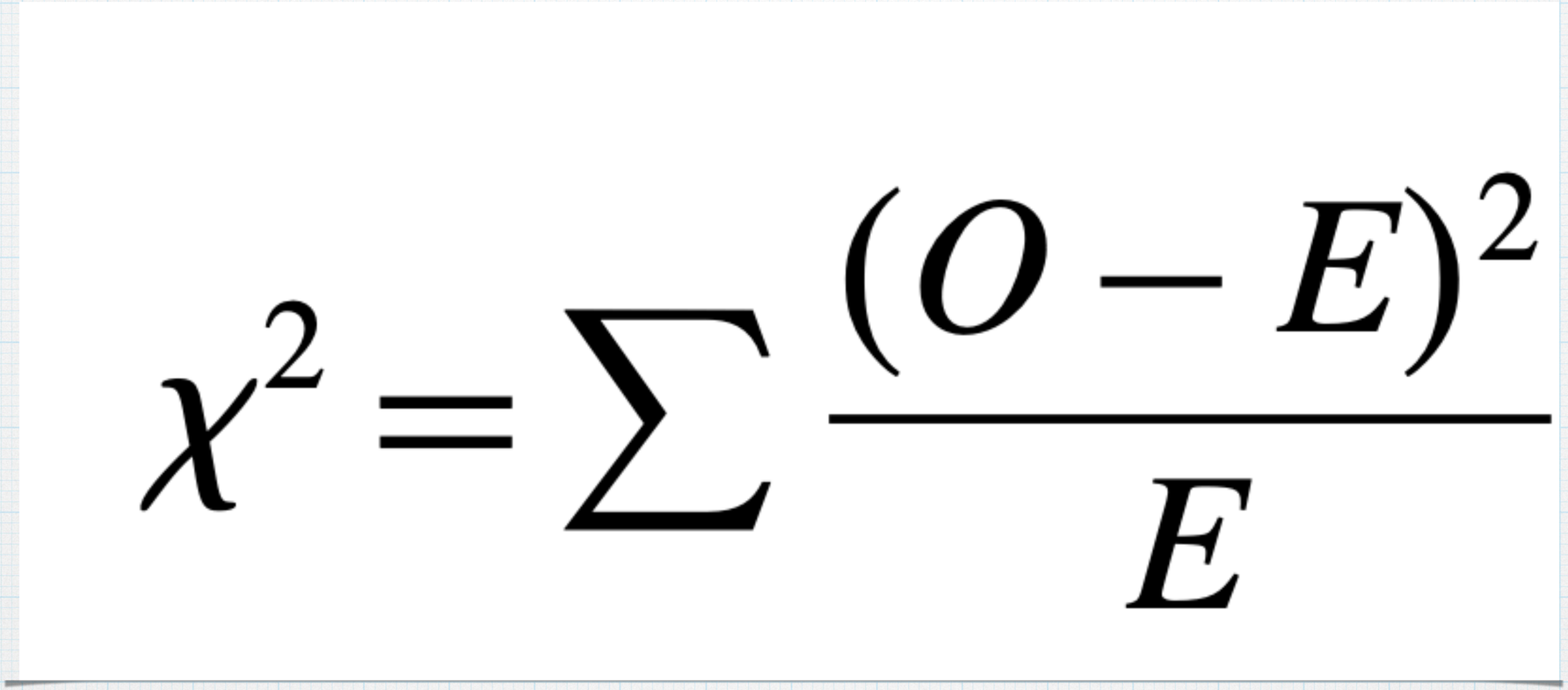
pineapple/ new OS	Idon'tmind	No	Yes	Row totals
Apple	39*41/121	46*41/121	41*36/121	41
Microsoft	80*39/121	80*46/121	80*36/121	80
Column totals	39	46	36	121

Expected values for each cell

E = R *C / G

```
> chisq.test(table(OS_new,pineapple))$expected
           pineapple
OS_new      Idon'tmind       No      Yes
  Apple       13.21488 15.58678 12.19835
  Microsoft   25.78512 30.41322 23.80165
```

Expected values



Chi square value

Observed minus expected frequencies divided by expected frequency

	O	E	(O-E)^2	$\frac{(O-E)^2}{E}$
Apple I don't mind	12	13.21		
Apple No	20	15.58		
Apple Yes	9	12.19		
Microsoft I don't mind	27	25.78		
Microsoft No	26	30.41		
Microsoft Yes	27	23.80		

Observed and Expected values

taking the sum of the last column = Chi square value

	O	E	(O-E)^2	$\frac{(O-E)^2}{E}$
Apple I don't mind	12	13.21	1.47	0.11
Apple No	20	15.58		
Apple Yes	9	12.19		
Microsoft I don't mind	27	25.78		
Microsoft No	26	30.41		
Microsoft Yes	27	23.80		

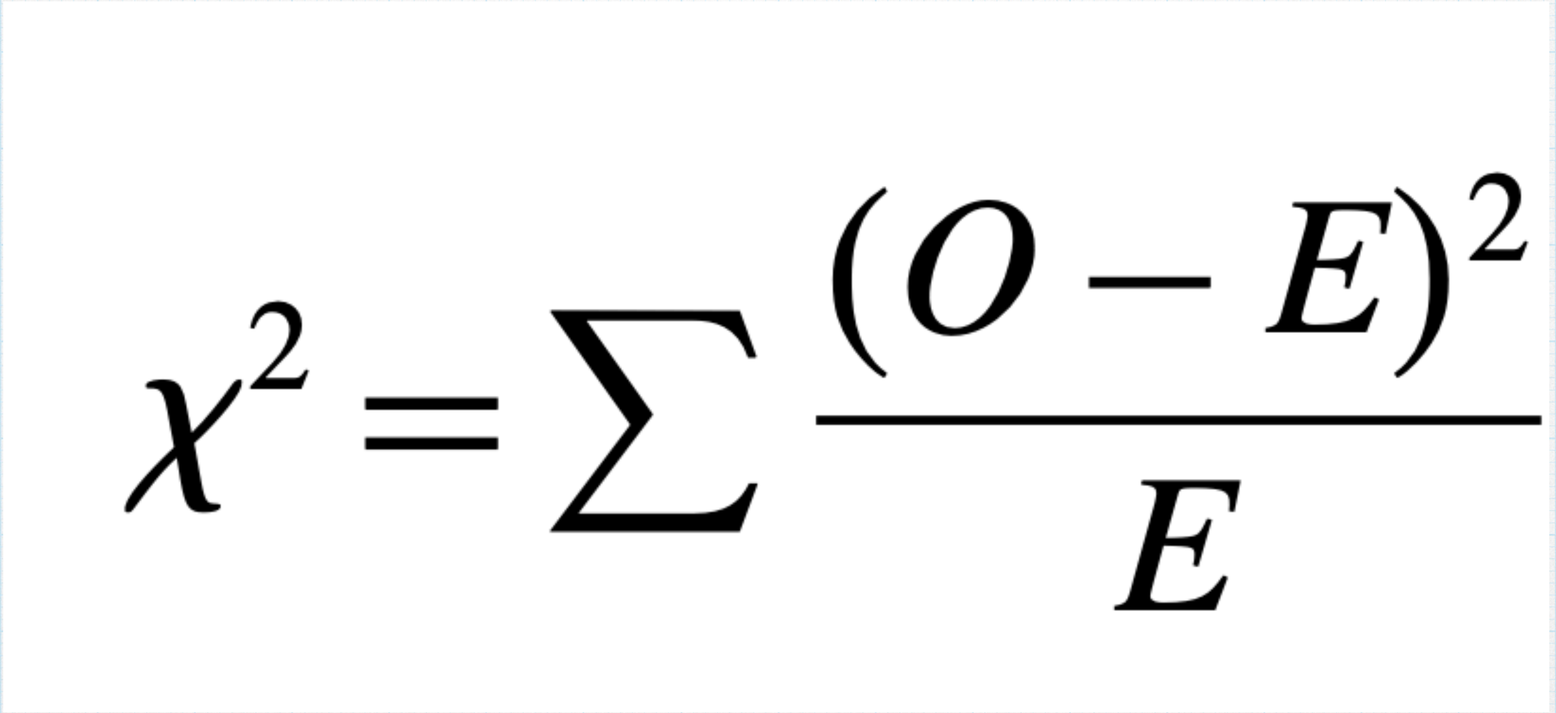
Observed and Expected values

taking the sum of the last column = Chi square value

	O	E	(O-E)^2	$\frac{(O-E)^2}{E}$
Apple I don't mind	12	13.21	1.47	0.11
Apple No	20	15.58	19.48	1.25
Apple Yes	9	12.19	10.23	0.84
Microsoft I don't mind	27	25.78	1.47	0.06
Microsoft No	26	30.41	19.48	0.64
Microsoft Yes	27	23.80	10.23	0.43

Observed and Expected values

taking the sum of the last column = Chi square value



Chi square value

Observed minus expected frequencies divided by expected frequency

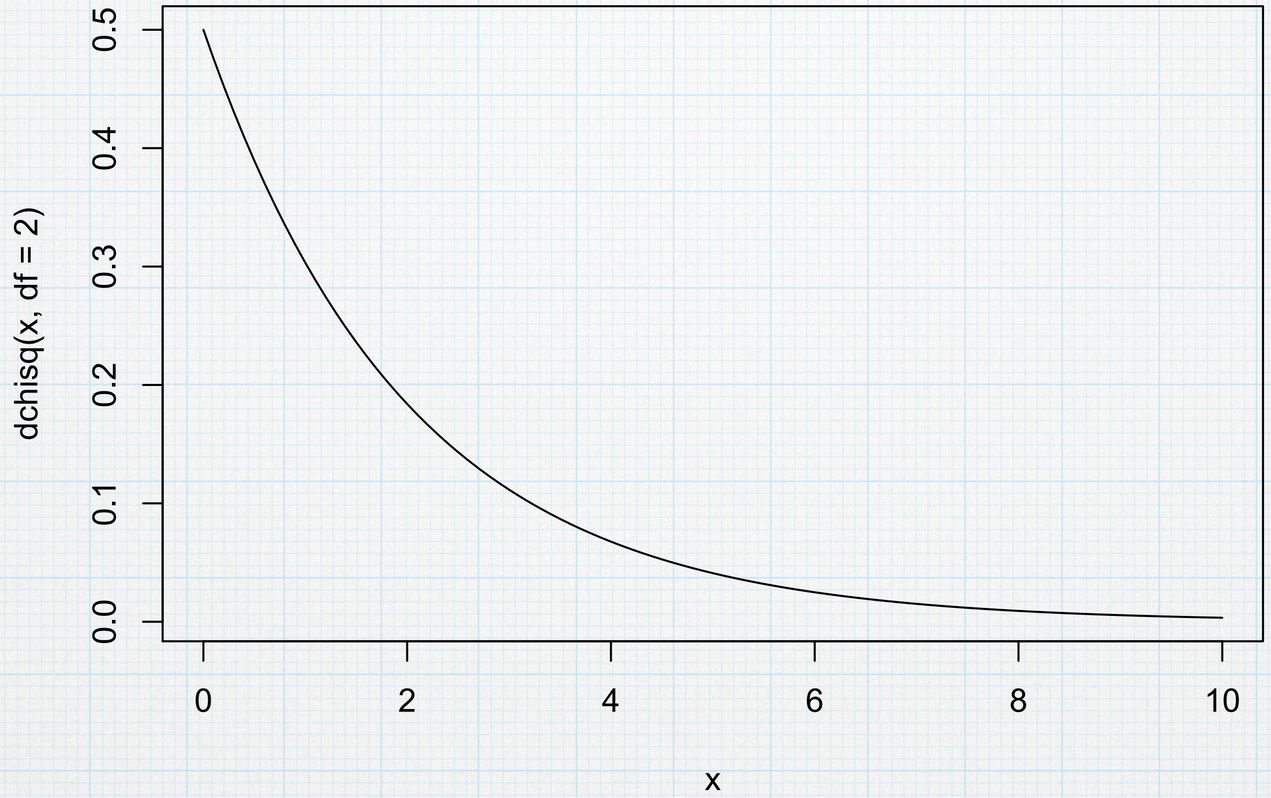
3.327247

```
> chisq.test(table(OS_new,pineapple))

   Pearson's Chi-squared test

data:  table(OS_new, pineapple)
X-squared = 3.3272, df = 2, p-value = 0.1895
```

how did we get the p-value?

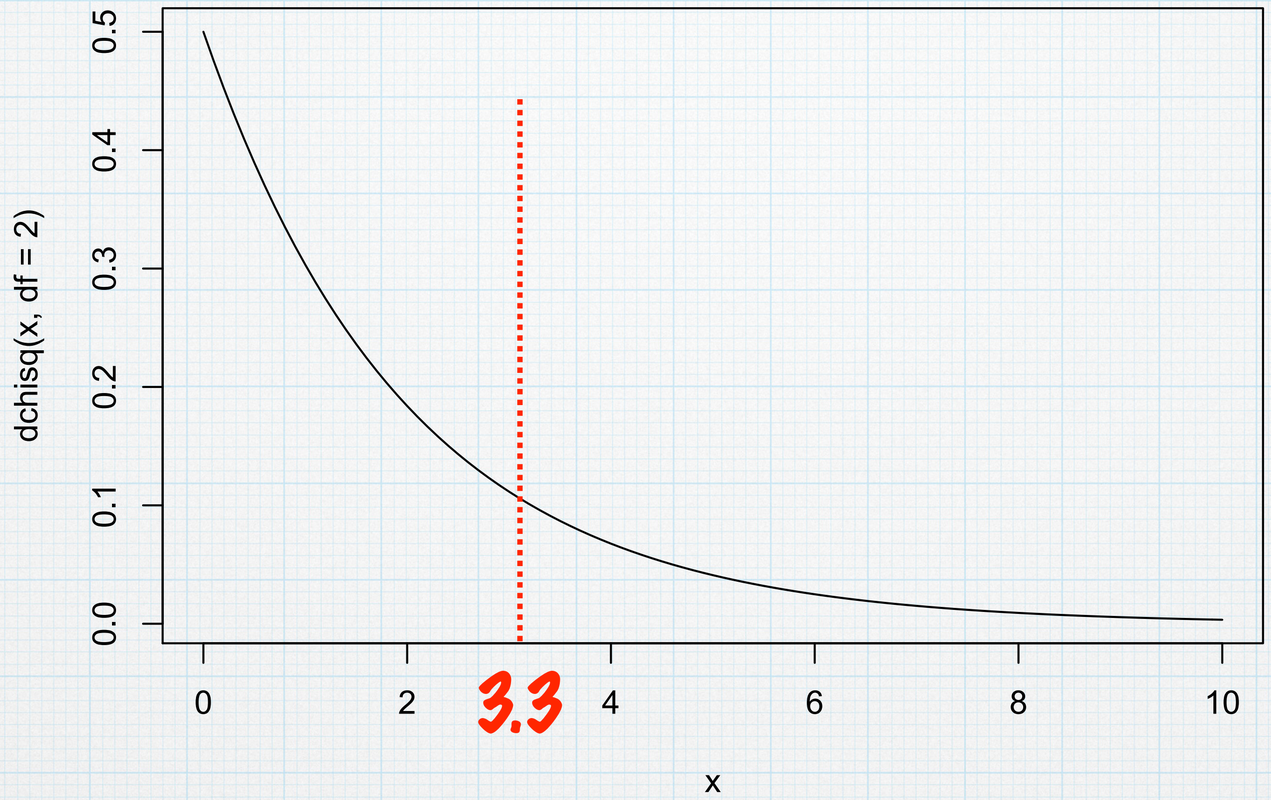
degrees of freedom = (r-1) * (c-1)

(3-1) * (2-1) = 2



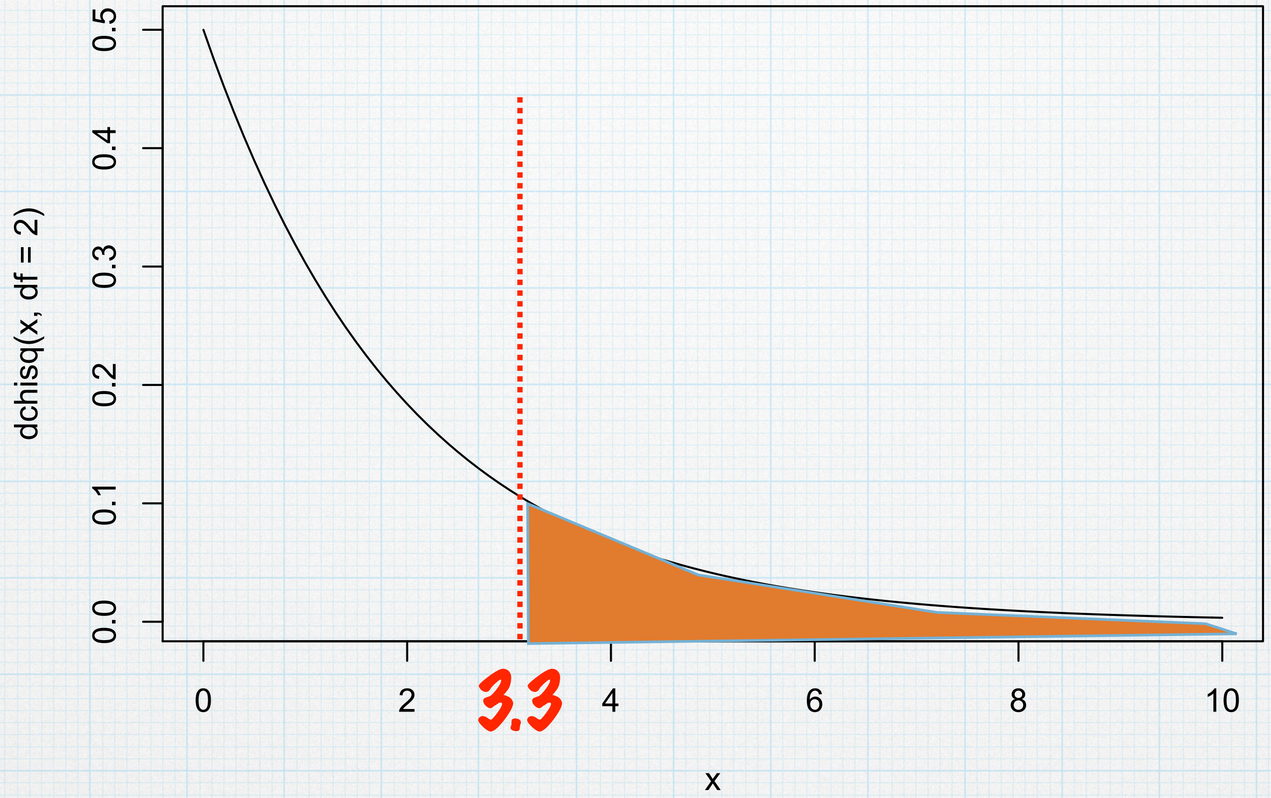
Probability density function

now we have to calculate the area under the curve of our test statistic...



the p-value is the area under the curve to the right of your value!

our Chi Square value = 3.3



the p-value is the area under the curve to the right of your value!

p=0.1895

Upper-tail critical values of chi-square distribution with ν degrees of freedom

	Probability less than the critical value				
ν	0.90	0.95	0.975	0.99	0.999
1	2.706	3.841	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588
11	17.275	19.675	21.920	24.725	31.264

Table with critical values

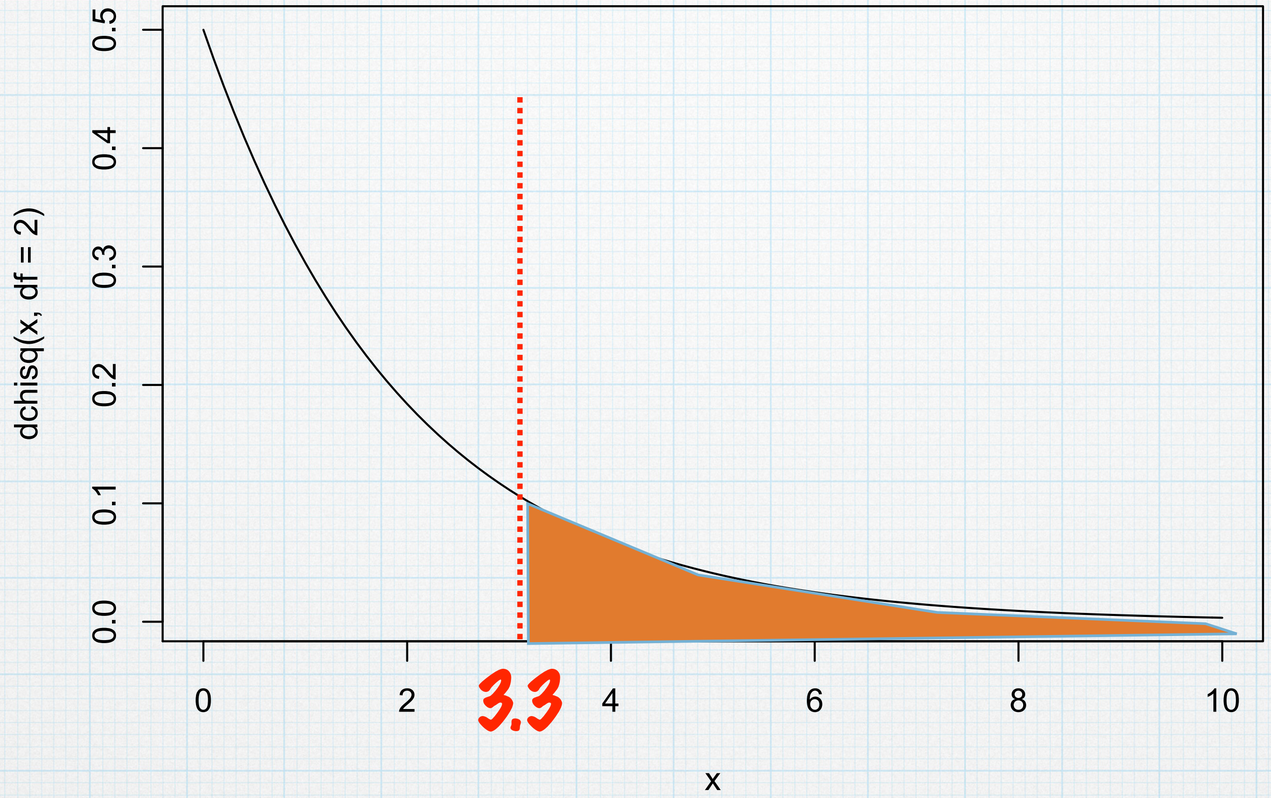
Luckily we do not have to calculate the p-value.
Now we have software... and even before that, we had tables of critical values.

qchisq(0.95,2)

how to calculate the critical value in R
0.95 = certainty level and 2 = degrees of freedom

=5.991465

this is our critical value and since our value is smaller than the critical value we have to except the Null-hypothesis—> no relationship between pineapple on pizza and OS.



the p-value is the area under the curve to the right of your value!

p=0.1895

Take home message

* data format and distribution determine your test statistic
* p-value always has the same meaning across tests
* we are looking for p-values smaller than 0.05
* the way a p-value is calculated is based on the probability density function