Chi-Square (χ^2)

- Commonly used in a test called "goodness of fit"
- GOF for uniform, Poisson, binomial, normal, a-priori
- GOF for too much variation, too little variation

Example GOF Problem

- Taken from Barron's *EZ 101 Study Keys for Statistics, by Martin Sternstein, 2nd Edition, (Amazon $8.99)*
- *Page 189*

- *A pet food manufacturer runs an experiment to determine whether three brands of dog food are equally preferred. In the experiment, 150 dogs are individually set looses in front of three dishes of food and their choices are noted. Tabulations show that 62 dogs went to brand A, 43 to brand B, and 45 to brand C. Is there sufficient evidence to say that dogs have a preferences among brand? Test at the 2.5% significance level.*

Classical Test of Hypothesis

- State the “null Hypothesis” in terms of a population parameter and an equal (=) sign.
- State the Alternative Hypothesis in terms of the same population parameter and one of three inequality signs.
- State the level of significance to which you wish to test the hypothesis.
- Identify the test statistic as either Z calculated, T calculated, Chi-squared calculated or F-calculated.
- Define the rejection criteria in terms of the calculated value and the criteria value.
- Evaluate the Test statistic and compare it to the criteria.
- Write out the conclusion.

Classical Test of Hypothesis

- State the "null Hypothesis" in terms of a population parameter and an equal (=) sign. $P_A = P_B = P_c$
- State the Alternative Hypothesis in terms of the same population parameter and one of three inequality signs. $P_A \neq P_B \neq P_c$
- State the level of significance to which you wish to test the hypothesis. $\alpha = 0.025$
- Identify the test statistic as either Z calculated, T calculated, Chi-squared calculated or F-calculated.
- Define the rejection criteria in terms of the calculated **value** and the criteria value. Reject null hypothesis when :
- Chi-squared calculated > Chi-square criteria (7.378)
- Evaluate the Test statistic and compare it to the criteria.
- Write out the conclusion.

Chi-squared calculated

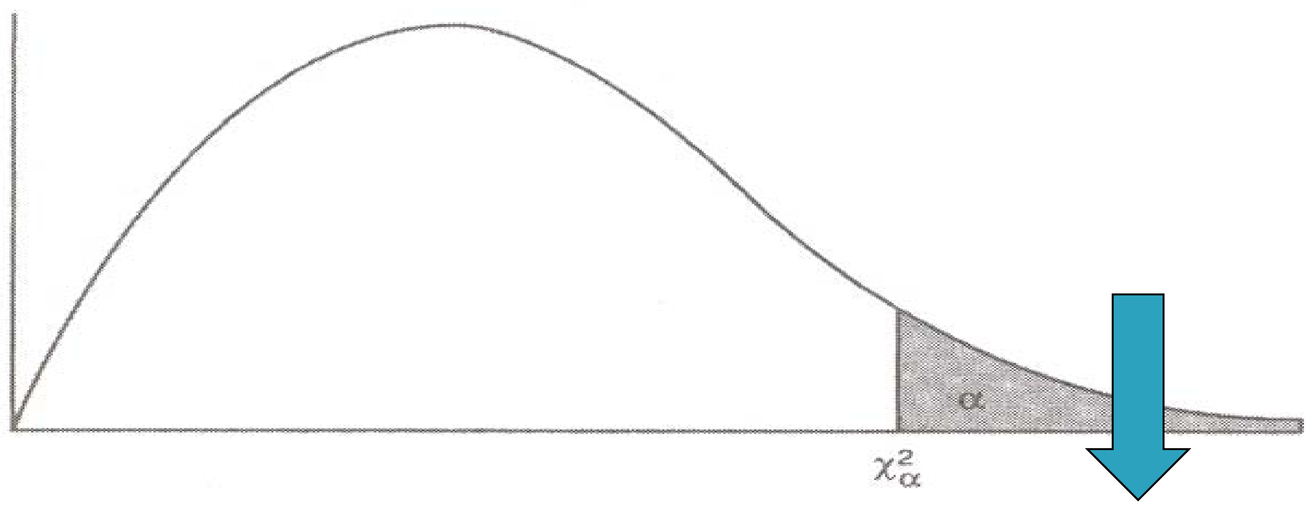
$$\chi^2 = \sum \frac{(actual.value - predicted.value)^2}{predicted.value}$$

$$\chi^2 = \frac{(62-50)^2}{50} + \frac{(43-50)^2}{50} + \frac{(45-50)^2}{50} = 4.36$$

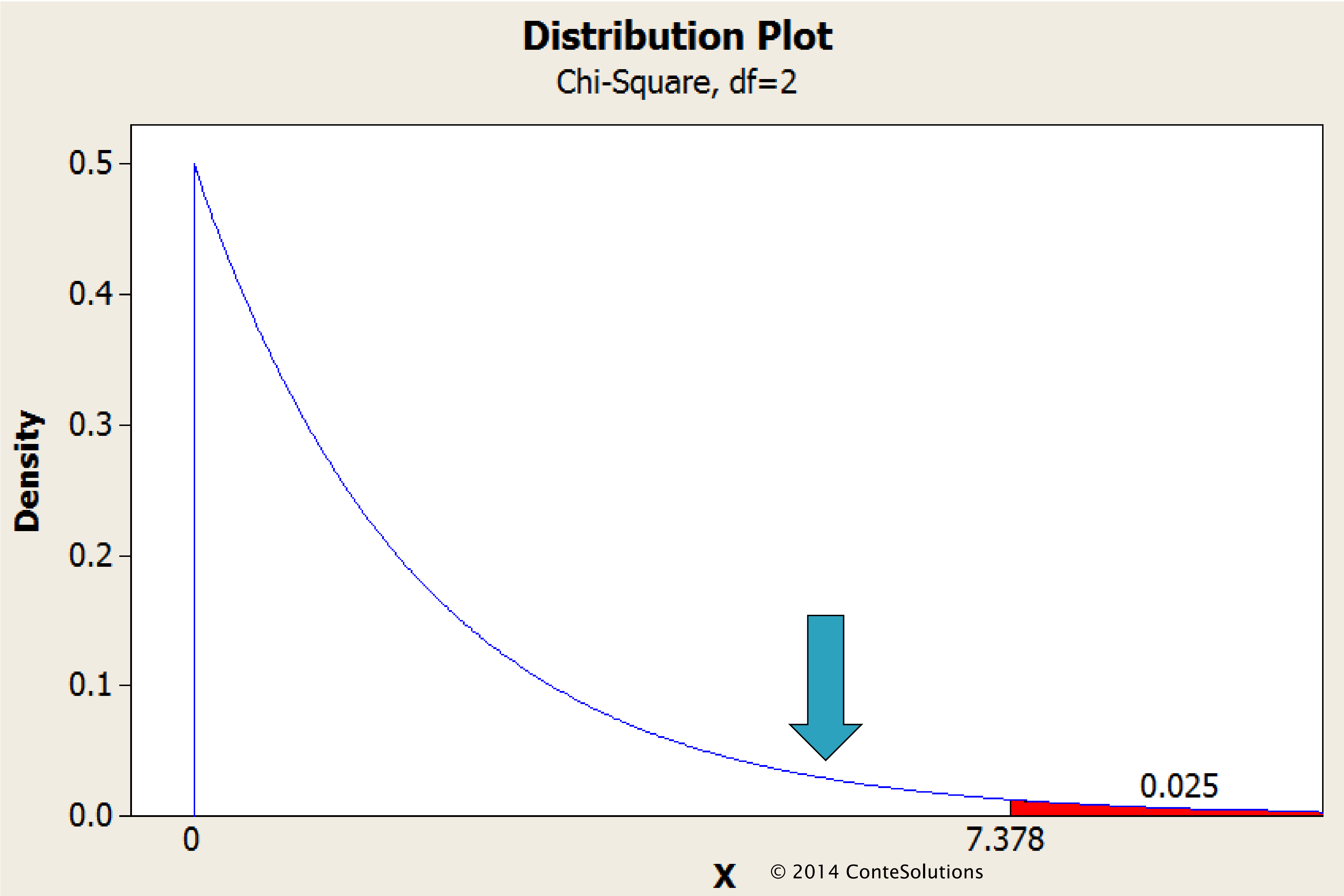
Chi-squared Criteria value= 7.378, α =0.025, d.f. = 2,

Table C

The χ^2-distribution



df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	.0000	.0002	.0010	.0039	.0158	2.706	3.841	5.024	6.635	7.879
2	.0100	.0201	.0506	.1026	.2107	4.605	5.991	7.378	9.210	10.60
3	.0717	.1148	.2158	.3518	.5844	6.251	7.815	9.348	11.34	12.84
4	.2070	.2971	.4844	.7107	1.064	7.779	9.448	11.14	13.28	14.86
5	.4117	.5543	.8312	1.145	1.610	9.236	11.07	12.83	15.09	16.75
6	.6757	.8721	1.237	1.635	2.204	10.64	12.59	14.45	16.81	18.55
7	.9893	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48	20.28
8	1.344	1.647	2.180	2.732	3.490	13.36	15.51	17.53	20.09	21.95
9	1.735	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67	23.59
10	2.156	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21	25.19



Distribution Plot
Chi-Square, df=2
Density
0.5
0.4
0.3
0.2
0.1
0.0
0
7.378
0.025
X
© 2014 ConteSolutions

Conclusion

- Calculated value = 4.36
- Criteria value = 7.378
- Therefore we fail to reject the null hypothesis
- "There is not sufficient evidence to say that the dogs have a preference among the three given brands."
- How do we get sufficient evidence?

GOF on too little variation

- We have been using only small, upper-tail probabilities of the chi-square distribution. There are cases, however, where we are interested in the probability that a chi-square value is less than or equal to a critical value.

To good to be true?

- Suppose your mathematics instructor gives, as a homework assignment, the problem of testing the fairness of a certain die. You are asked to roll the die 6000 times and note how often it comes up 1,2,3,4,5, and 6. The assignment tires you quickly, so you decide to simply invent some "reasonable" data. Being careful to make the total 6000, you write:

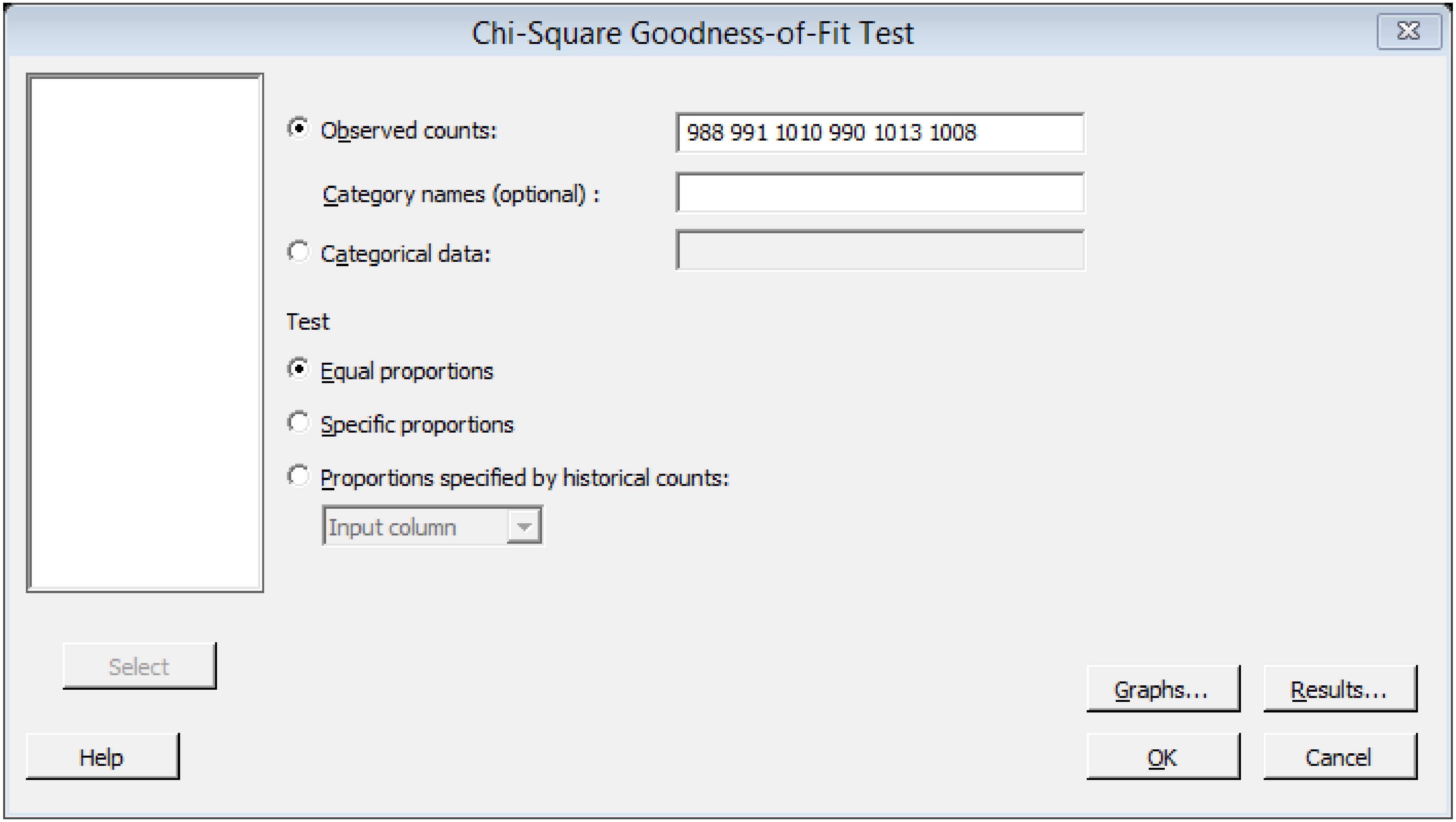
1	2	3	4		
988	991	1010	990	1013	1008

- Null hypothesis?

H_0: good fit with uniform distribution, that is, the die if fair

The "expected" values are all 1000, and so

$$\chi^2 = \frac{12^2}{1000} + \frac{9^2}{1000} + \frac{10^2}{1000} + \frac{10^2}{1000} + \frac{13^2}{1000} + \frac{8^2}{1000} = 0.658$$


Chi-Square Goodness-of-Fit Test
Observed counts:
988 991 1010 990 1013 1008
Category names (optional) :
Categorical data:
Test
Equal proportions
Specific proportions
Proportions specified by historical counts:
Input column
Select
Help
Graphs...
Results...
OK
Cancel

10/20/2014 7:54:39 AM

Welcome to Minitab, press F1 for help.

Chi-Square Goodness-of-Fit Test for Observed Counts

Category	Observed	Test Proportion	Expected	Contribution to Chi-Sq
1	988	0.166667	1000	0.144
2	991	0.166667	1000	0.081
3	1010	0.166667	1000	0.100
4	990	0.166667	1000	0.100
5	1013	0.166667	1000	0.169
6	1008	0.166667	1000	0.064

N	DF	Chi-Sq	P-Value
6000	5	0.658	0.985

98.52% of probability space to right of 0.658