MATH 1342

Section 8.5

Goodness of Fit Test

Suppose we want to make an inference about a group of data (instead of just one or two). Or maybe we want to test counts of categorical data. **Chi-square** (or χ^2) testing allows us to make such inferences.

There are several types of Chi-square tests but in this section we will focus on the goodness-of-fit test. **Goodness-of-fit** test is used to test how well one sample proportions of categories "match-up" with the known population proportions stated in the null hypothesis statement. The Chi-square goodness-of-fit test extends inference on proportions to more than two proportions by enabling us to determine if a particular population distribution has changed from a specified form.

The null and alternative hypotheses do not lend themselves to symbols, so we will define them with words.

H_o: obs is the same as predicted

H_a: obs is different from predicted

For each problem you will make a table with the following headings:

Observed Counts (O)	Expected Counts (E)	$\frac{(O-E)^2}{E}$

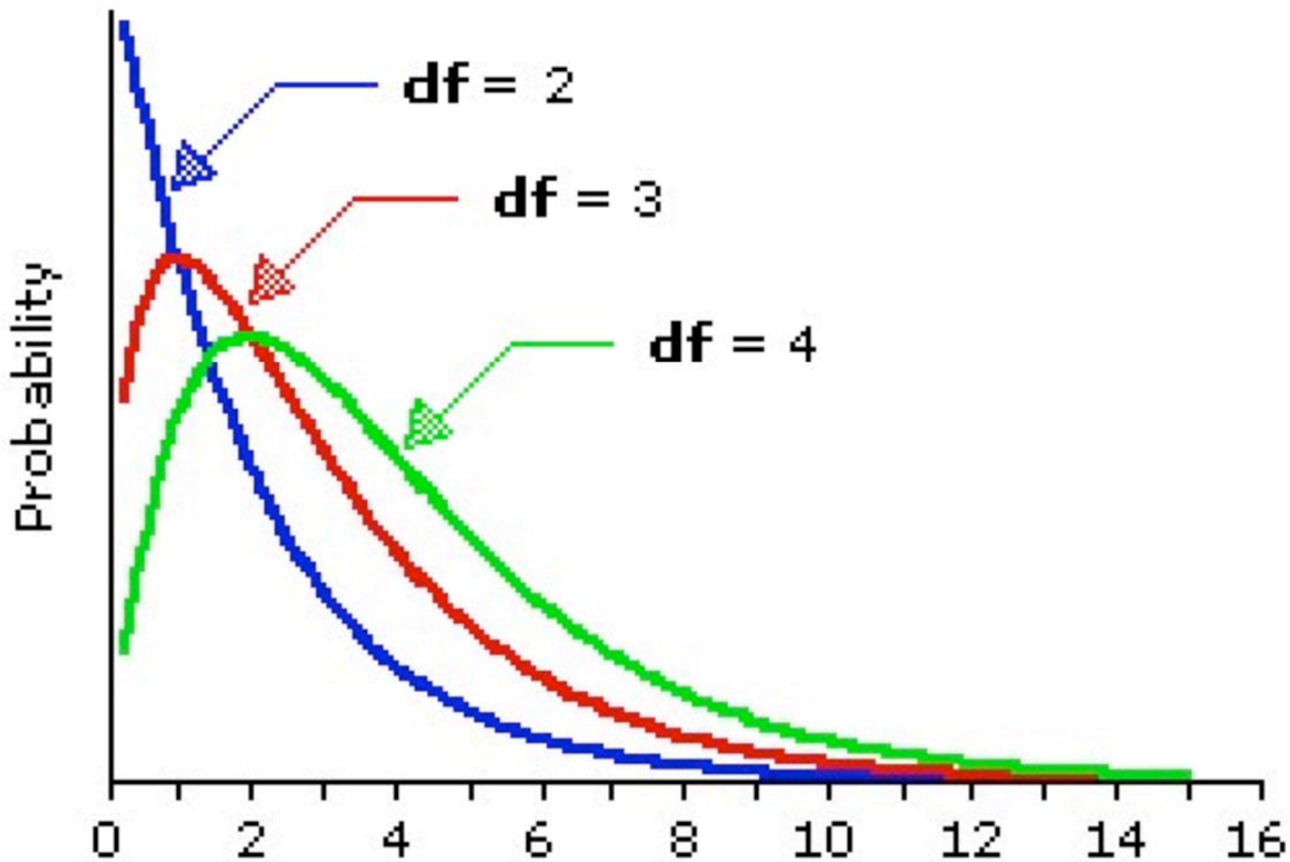
The sum of the third column is called the Chi-square test statistic.

Chi-square

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Table D gives p-values for χ^2 with $n - 1$ degrees of freedom.

Chi-square distributions have only positive values and are skewed right. As the degrees of freedom increase it becomes more normal. The total area under the χ^2 curve is 1.



The assumptions for a Chi-square goodness-of-fit test are:

1. The sample must be an SRS from the populations of interest.
2. The population size is at least ten times the size of the sample.
3. All expected counts must be at least 5.

To find probabilities for χ^2 distributions:

TI-83/84 calculator uses the command χ^2 **cdf** found under the DISTR menu.

R-Studio command is: 1 – pchisq(*test statistic*, *df*)

p-value

Degrees of Freedom (chi-squared test): [number of categories] – 1

Example:

1. The Mixed-Up Nut Company advertises that their nut mix contains (by weight) 40% cashews, 15% Brazil nuts, 20% almonds and only 25% peanuts. The truth-in-advertising investigators took a random sample (of size 50 lbs) of the nut mix and found the distribution to be as follows:

Cashews	*Brazil Nuts*	*Almonds*	*Peanuts*
15 lb	11 lb	13 lb	11 lb

Degrees of Freedom:
4 categories, so we have 3 degrees of freedom

At the 1% level of significance, is the claim made by Mixed-Up Nuts true?

	Exp %	Exp #	Obs	$\frac{(O-E)^2}{E}$
C	40%	20	15	1.25
B	15%	7.5	11	1.633
A	20%	10	13	0.900
P	25%	12.5	11	0.180
Total:			50	3.9633

sum of column

$\chi^2 = 3.9633$

$P = 0.2654 = 26.5\%$

$P > \alpha$ FRH$_0$

R Studio Only (No Table)

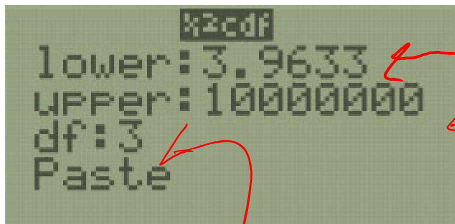
```
> assign("expperc",c(0.40,0.15,0.20,0.25))
> sum(expperc)
[1] 1
> assign("obs",c(15,11,13,11))
> sum(obs)
[1] 50
> expcount=expperc*50
> expcount
[1] 20.0  7.5 10.0 12.5
> (obs-expcount)^2/expcount
[1] 1.250000 1.633333 0.900000 0.180000
> sum((obs-expcount)^2/expcount)
[1] 3.963333
> 1-pchisq(3.963333,3)
[1] 0.2654508
```

Provided in the Question

Highlighted lines are the minimum amount you need to do to get an answer.

Chi-Squared in the Calculator

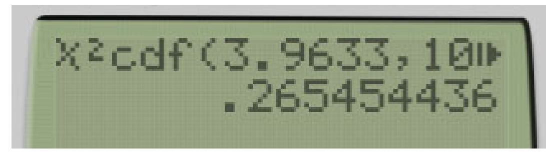
All steps up until the p-value calculation can be done manually or by using the lists in STAT, Edit.



χ^2-value

make up very large number

categories-1



Another Example: Popper 31

If the die were fair, all categories would be equal counts. So: exp = n/[number of categories]

Suppose you rolled a die 60 times and observed 14 ones, 8 twos, 7 threes, 16 fours, 7 fives, and 8 sixes. Is this a fair die?

Test the claim at the 5% significance level.

1. What is the $(O-E)^2/E$ value for rolling a 1?

   a. 14 b. 10 c. 1.6 d. 5

2. What is the $(O-E)^2/E$ value for rolling a 6?

   a. 8 b. 0.4 c. 10 d. 6

3. What is the test statistic?

   a. 7.8 b. 6 c. 5 d. 3.6

4. What is the chi-squared p-value?

   a. 0.5454 b. 0.1676 c. 0.2318 d. 0.9832

5. What is the conclusion?

   a. Reject the Null Hypothesis (Die is not fair)

   b. Do Not Reject the Null Hypothesis (Die is fair)

```
> exp=60/6
> exp
[1] 10
> assign("obs",c(14,8,7,16,7,8))
> (obs-exp)^2/exp
[1] 1.6 0.4 0.9 3.6 0.9 0.4
> sum((obs-exp)^2/exp)
[1] 7.8
> 1-pchisq(7.8,5)
[1] 0.1676079
```

(6 categories) − 1

P = 16.76% α = 5%

P > α FRH.