MODULE 20

GOODNESS-OF-FIT TEST

Contents

Iᴛ ɪs ᴄᴏᴍᴍᴏɴ ᴛᴏ ᴅᴇᴛᴇʀᴍɪɴᴇ ɪғ ᴛʜᴇ ғʀᴇǫᴜᴇɴᴄʏ of individuals in the levels of a categorical response variable follow frequencies suggested by a particular theory or distribution. The simplest of these situations occurs when a researcher is making a hypothesis about the percentage or proportion of individuals in one of two categories. The "distribution" of individuals in two categories comes from the proportion in the hypothesis for one category and one minus the proportion in the hypothesis for the other category. In situations with more than two levels, the "distribution" of individuals into the categories likely comes from the hypothesis that a particular theoretical distribution holds true. For example, a researcher may want to determine if frequencies predicted from a certain genetic theory are upheld by the observed frequencies found in a breeding experiment, if the frequency that a certain animal uses habitats is in proportion to the availability of those habitats, or if the frequency of consumers that show a preference for a certain product (over other comparable products) is non-random.

In each of these cases, the theoretical distribution articulated in the research hypothesis must be converted to statistical hypotheses that will then be used to generate expected frequencies for each level. These expected frequencies will then be statistically compared to the observed frequencies to determine if the theoretical distribution represented in the null hypothesis is supported by the data. The method used for comparing the observed to expected frequencies, where the expected frequencies come from a hypothesized theoretical distribution, is a Goodness-of-Fit Test, the subject of this module.

20.1 Goodness-of-Fit Test Specifics

20.1.1 The Hypotheses

A Goodness-of-Fit Test is used when a single categorical variable has been recorded and the frequency of individuals in the levels of this variable are to be compared to a theoretical distribution. In its most general form the statistical hypotheses for the Goodness-of-Fit Test will be "wordy," relating whether the "distribution" of individuals into the levels of the response variable follows a specific theoretical distribution or not. The null hypothesis will generally be like H_0 : "the distribution of individuals into the levels follows the 'theoretical distribution' ", where 'theoretical distribution' will likely be replaced with more specific language. For example, the research hypothesis that states that "50% of students at Northland are from Wisconsin, 25% are from neighboring states, and 25% are from other states" would be converted to H_0: "the proportion of students from Wisconsin, neighboring states, and other states is 0.50, 0.25, and 0.25, respectively" with an H_A: "the proportion of students from Wisconsin, neighboring states, and other states is NOT 0.50, 0.25, and 0.25, respectively."

The hypotheses are simpler, but you must be more careful, when there are only two levels of the response variable. For example, a research hypothesis of "less than 40% of new-born bear cubs are female" would be converted to H_0: "the proportion of bear cubs that are female and male is 0.40 and 0.60, respectively" with an H_A: "the proportion of bear cubs that are female and male is NOT 0.40 and 0.60, respectively." However, these hypotheses are often simplified to focus on only one level as the other level is implied by subtraction from one. Thus, these hypotheses are more likely to be written as H_0: "the proportion of bear cubs that are female is 0.40" with an H_A: "the proportion of bear cubs that are female is NOT 0.40."

One may also have expected, from the wording of the research hypothesis about the sex of bear cubs, that the alternative hypothesis would have been H_A: "the proportion of bear cubs that are female is LESS THAN 0.40." Recall from Section 19.1, however, that the chi-square test statistic always represents a two-tailed situation. Thus, the H_A here reflects that constraint. The researcher will ultimately be able to determine if the proportion is less than 0.40 if the p-value from the Goodness-of-Fit Test indicates a difference and the observed proportion of female bear cubs is less than 0.40.

20.1.2 The Tables

For a Goodness-of-Fit Test, the data are summarized in an observed frequency table as in Module 5. Additionally, a table of expected frequencies must be constructed from the theoretical distribution in the null hypothesis and the total number of observed individuals (n). Specifically, the expected frequencies are found by multiplying the expected proportions from the theoretical distribution in the null hypothesis by n. For example, consider this situation:

> Bath and Buchanan (1989) surveyed residents of Wyoming by distributing a mailing to random residents and collecting voluntarily returned surveys. One question asked of the respondents was, "Do you strongly agree, agree, neither agree or disagree, disagree, or strongly disagree with this statement? – 'Wolves would have a significant impact on big game hunting opportunities near Yellowstone National Park'." The researchers hypothesized that more than 50% of Wyoming residents would either disagree or strongly disagree with the statement. Of the 371 residents that returned the survey, 153 disagreed and 43 strongly disagreed with the statement.

At first glance it may seem that this variable has five levels – i.e., the levels of agreement offered in the actual survey. However, the researcher's hypothesis collapsed the results of the survey question into two levels: (1) strongly disagree or disagree combined and (2) all other responses. Thus, the statistical hypotheses for this situation are H_0: "the proportion of respondents that disagreed or strongly disagreed is 0.50" and H_A: "the proportion of respondents that disagreed or strongly disagreed is NOT 0.50."

The expected frequencies in each level are derived from the total number of individuals examined and the specific null hypothesis. For example, if the null hypothesis is true, then 50% of the 371 respondents would be expected to disagree or strongly disagree with the statement. In other words, $371*0.50{=}185.5$ individuals would be expected to disagree or strongly disagree. Furthermore, the other 50%, or $371*(1-0.50){=}185.5$ would be expected to "not" disagree or strongly disagree. These expectations are summarized in Table 20.1.

Table 20.1. Expected and observed frequency of respondents that disagreed or strongly disagreed (i.e., labeled as "Disagree") with the given statement in the Wyoming survey example.

Category	Frequency	
	Expected	Observed
"Disagree"	185.5	196
not "Disagree"	185.5	175
Total	371	371

◇ **The expected table should maintain at least one decimal in each cell even though the values represent frequencies.**

Consider the following situation where construction of expected frequencies is bit more complex.

> *Mendel's law of independent assortment predicts that the genotypes (i.e., how they look) of the offspring from mating the offspring of a dihybrid cross of homozygous dominant and homozygous recessive parents should follow a 9:3:3:1 ratio. In an experiment to test this, Mendel crossed a pea plant that produces round, yellow seeds (i.e., all dominant alleles, YYWW) with a pea plant that produces green, wrinkled seeds (i.e., all recessive alleles, yyww) such that only round, yellow heterozygous offspring (i.e., YyWw) were produced. Pairs of these offspring were then bred. Mendel's theory says that $\frac{9}{16}$ of these offspring should be round, yellow; $\frac{3}{16}$ should be round, green; $\frac{3}{16}$ should be wrinkled, yellow; and $\frac{1}{16}$ should be wrinkled, green. Of 566 seeds studied in this experiment, Mendel found that 315 were round, yellow; 108 were round, green; 101 were wrinkled, yellow; and 32 were wrinkled, green. Use these results to determine, at the 5% level, if Mendel's law of independent assortment is supported by these results.*

The statistical hypotheses are as follows,

H_0 : "the proportion of RY, RG, WY, and WG individuals will be $\frac{9}{16}, \frac{3}{16}, \frac{3}{16}$, and $\frac{1}{16}$, respectively"

H_A : "the proportion of RY, RG, WY, and WG individuals will NOT be $\frac{9}{16}, \frac{3}{16}, \frac{3}{16}$, and $\frac{1}{16}$, respectively"

where RY="round, yellow", RG="round, green", WY="wrinkled, yellow", and WG="wrinkled, green." If these proportions are applied to the $n = 566$ observed offspring, then the following frequencies for each genotype would be expected:

- $\frac{9}{16} \cdot 566 = 318.375$ would be expected to be round, yellow.
- $\frac{3}{16} \cdot 566 = 106.125$ would be expected to be round, green.
- $\frac{3}{16} \cdot 566 = 106.125$ would be expected to be wrinkled, yellow.
- $\frac{1}{16} \cdot 566 = 35.375$ would be expected to be wrinkled, green.

These expected frequencies are summarized in Table 20.2.

Table 20.2. Expected and observed frequency of 566 pea seeds in four types.

	Frequency	
Category	Expected	Observed
round, yellow	318.375	314
round, green	106.125	108
wrinkled, yellow	106.125	101
wrinkled, green	35.375	32
Total	566	566

The hypothesis test method developed in the following sections will be used to determine if the differences between the expected and observed frequencies is "large" enough to suggest that the observed frequencies do not support the distribution represented in the null hypothesis.

20.1.3 Specifics

The Goodness-of-Fit Test is characterized by a single categorical response variable. The hypotheses tested usually cannot be converted to mathematical symbols and are thus "wordy." Specifics of the Goodness-of-Fit Test are in Table 20.3.

Table 20.3. Characteristics of a Goodness-of-Fit Test.

- **Hypotheses:** H_0 :"the observed distribution of individuals into the levels follows the 'theoretical distribution' "
  H_A :"the observed distribution of individuals into the levels DOES NOT follow the 'theoretical distribution'."
- **Statistic:** Observed frequency table.
- **Test Statistic:** $\chi^2 = \sum_{cells} \frac{(Observed - Expected)^2}{Expected}$
- **df:** Number of levels minus 1.
- **Assumptions:** Expected value in each level is ≥ 5.
- **Confidence Interval (for one level):** $\hat{p} \pm Z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- **Use with:** Categorical response, one group (or population), comparing to a theoretical distribution.

It is cumbersome to produce a confidence interval in a Goodness-of-Fit Test because there generally is not a single parameter (i.e., there are as many parameters as levels in the response variable). Confidence intervals can be calculated for the proportion in each level as shown below. However, confidence intervals will only be "hand"-calculated when there are two levels. When using R (as discussed in a subsequent section), confidence intervals will be computed for all levels, no matter the number of levels.

Let p be the population proportion in a particular level and $\hat{p}$ be the sample proportion in the same interval. The $\hat{p}$ is computed by dividing the frequency of individuals in this level by the total number of individuals in the sample (i.e., n). The $\hat{p}$ is a statistic that is subject to sampling variable with that sampling variability measured by $SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ for "large" values of n. For "large" values of n the $\hat{p}$ will follow a normal

distribution such that a confidence interval for p is computed using the general confidence interval formula found in Section 15.2 and repeated below:

$$\text{``Statistic''} + \text{``scaling factor''} * SE_{statistic}$$

where the scaling factor is the familiar Z^*. Thus, the confidence interval for p is constructed with

$$\hat{p} \pm Z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Note that one does not need to worry about lower and upper bounds, only confidence intervals will be computed, because of the two-tailed nature of the chi-square test statistic.

In the Wyoming survey example, the proportion of respondents in the sample that either disagreed or strongly disagreed was $\hat{p}=\frac{196}{371}$=0.528. The standard error for this sample proportion is $\sqrt{\frac{0.528(1-0.528)}{371}}$=0.026. For a 95% confidence interval, $Z^*=\pm 1.960$.[1] Thus, the confidence interval is $0.528 \pm 1.960 * 0.026$ or 0.528 ± 0.051 or $(0.477, 0.579)$. Therefore, one is 95% confident that the population proportion that either disagreed or strongly disagreed is between 0.477 and 0.579. Because there are only two levels in this example it can also be said with 95% confidence that the population proportion that did not either disagree or strongly disagree is between 0.421 and 0.523.

20.1.4 Example - $1 Coins

Below are the 11-steps (Section 16.1) for completing a full hypothesis test for the following situation:

> *USA Today (June 14, 1995) reported that 77% of the population opposes replacing $1 bills with $1 coins. To test if this claim holds true for the residents of Ashland a student selected a sample of 80 Ashland residents and found that 54 were opposed to replacing the bills with coins. Develop a hypothesis test (at the 10% level) to determine if the proportion of Ashland residents that are opposed to replacing bills with coins is different from the proportion opposed for the general population.*

1. α=0.10.
2. H_0: "the proportion of Ashland residents that oppose replacing the $1 bill with the $1 coin is 0.77" vs. H_A: "The proportion of Ashland residents that oppose replacing the $1 bill with the $1 coin is NOT 0.77."
3. A Goodness-of-Fit Test is required because (a) a single categorical variable was recorded (opinion about $1 coin), (ii) a single group (or population) was considered (Ashland residents), and (iii) the frequency of responses is being compared to a hypothesized distribution in the null hypothesis.
4. The data appear to be part of an observational study with no clear indication of random selection of individuals.
5. The expected number in the "oppose" level is $80 * 0.77$=61.6. The expected number in the "do not oppose" category is $80 * 0.23$=18.4. These expectations are shown in the table in the next step. The assumption of more than five individual in all cells of the expected table has been met.
6. The observed table is shown below (along with the expected table).

---

[1]This Z^* is computed with `distrib(0.975,type="q")`

Level	Frequency	
	Expected	Observed
"Oppose"	61.6	54
"Do Not Oppose"	18.4	26
Total	80	80

7. $\chi^2 = \frac{(61.6-54)^2}{55} + \frac{(18.4-26)^2}{25} = 0.938 + 3.139 = 4.077$ with $2 - 1 = 1$ df.

8. p-value=0.0435.

9. H_0 is rejected because the p-value $< \alpha = 0.10$).

10. The proportion of Ashland residents that oppose replacing the \$1 bill with the \$1 coin does appear to be different from the proportion (0.77) reported for the general population.

11. I am 90% confident that the proportion of all Ashland residents opposed to the \$1 coin is between 0.596 and 0.767. $[\frac{54}{80} \pm 1.645 * \sqrt{\frac{0.68125*0.31875}{80}} = 0.68125 \pm 1.645 * 0.0521 = 0.68125 \pm 0.0857 = (0.5956, 0.7670).]$

R Appendix:

```
( distrib(4.077,distrib="chisq",df=1,lower.tail=FALSE) )
( distrib(0.95,type="q") )
```

20.2 Goodness-of-Fit Test in R

20.2.1 Data Format

A Goodness-of-Fit Test is conducted in R with `chisq.test()`, which requires an observed table as the first argument. This observed table is entered from summarized data using `c()` or raw data is summarized to a frequency table with `xtabs()` as in Module 5.

For example, suppose that the frequency of shrike observations in the "mid-successional", "open", "scattered trees", "woods", and "wetland" habitats is known to be 43, 1456, 112, 44 and 6, respectively. These summarized values are entered directly into a named vector below.

```
> ( obs <- c(MidSucc=43,Open=1456,ScatTree=112,Woods=6,Wetland=44) )
 MidSucc    Open ScatTree    Woods  Wetland
      43    1456      112        6       44
```

However, instead of having summarized frequencies, suppose that the individual habitat observations were stored in a variable called `hab.use` in the `df` data.frame. These raw data must be summarized into a frequency table.

```
> ( obs <- xtabs(~hab.use,data=shrike.raw) )
hab.use
 MidSucc    Open ScatTree  Wetland    Woods
      43    1456      112        6       44
```

20.2.2 Goodness-of-Fit Test

The Goodness-of-Fit Test is computed with `chisq.test()` with a observed frequencies as the first argument and the following arguments:

- `p=`: a vector of expected proportions for the levels of the theoretical distribution.
- `rescale.p=TRUE`: rescales the values in `p=` to sum to 1. Rescaling is useful if the proportions in `p=` were rounded or are expected frequencies.
- `correct=FALSE`: indicates to not use a "continuity correction."[2]

The results from `chisq.test()` should be assigned to an object so that useful information can be extracted. The chi-square test statistics and p-value are extracted by typing the name of the saved object, the expected values are extracted by appending `$expected` to the object, and a visual of the p-value is obtained by submitting the object to `plot()`. In addition, confidence intervals for the proportions of individuals in each level are constructed by submitting the saved object to `gofCI()`.

20.2.3 Example - Loggerhead Shrikes

Below are the 11-steps (Section 16.1) for completing a full hypothesis test for the following situation:

> Bohall-Wood (1987) *constructed 24 random 16-km transects along roads in counties near Gainesville, FL. Two observers censused each transect once every 2 weeks from 18 October 1981 to 30 October 1982, by driving 32 km/h and scanning both sides of the road for perched and flying shrikes (**Lanius ludovicianus**). The habitat, whether the bird was on the roadside or actually in the habitat, and the perch type were recorded for each shrike observed. Habitats were grouped into five categories. The number of shrikes observed in each habitat was 1456 in open areas, 43 in midsuccessional, 112 in scattered trees, 44 in woods, and 6 in wetlands. Separate analyses were used to construct the proportion of habitat available in each of the five habitat types. These results were as follows: 0.358 open, 0.047 midsuccessional, 0.060 scattered trees, 0.531 woods, and 0.004 wetlands. Use these data to determine, at the 5% level, if shrikes are using the habitat in proportion to its availability.*

1. α=0.05.
2. H_0:"distribution of habitat use by shrikes is the same as the proportions of available habitat" vs. H_A:"distribution of habitat use by shrikes is NOT the same as the proportions of available habitat."
3. A Goodness-of-Fit Test is required because (i) a categorical variable was recorded (habitat use), (ii) a single group (or population) was considered (shrikes in this area), and (iii) the observed distribution is compared to a theoretical distribution.
4. The data appear to be part of an observational study where the individuals were not randomly selected but the transects upon which they were observed were.
5. There are more than five individuals expected in each habitat level (Table 20.4).
6. The statistic is the observed frequency table in Table 20.4.
7. χ^2=2345.1 with 4 df (Table 20.5).
8. p-value< 0.00005 (Table 20.5).
9. H_0 is rejected because the p-value< α.
10. The shrikes do not appear to use habitats in the same proportions as the availability of the habitat.

---

[2]Some statisticians argue that small chi-square tables with small sample sizes should be corrected for the fact that the chi-square distribution is a continuous distribution. This correction is applied by simply subtracting 0.5 from each observed-expected calculation. We will not use the continuity correction in this course so that R calculations will match hand calculations.

11. The 95% confidence intervals for the proportion of use in each habitat level are in Table 20.6. From these results it appears that the shrikes use the "open" habitat much more often and the "woods" habitat mush less often than would be expected if they used all habitats in proportion to their availability.

R Appendix:

```
( obs <- c(Open=1456,MidSucc=43,ScatTree=112,Woods=6,Wetland=44) )
( p.exp <- c(Open=0.358,MidSucc=0.047,ScatTree=0.060,Woods=0.531,Wetland=0.004) )
( shrike.chi <- chisq.test(obs,p=p.exp,rescale.p=TRUE) )
data.frame(obs=shrike.chi$observed,exp=shrike.chi$expected)
gofCI(shrike.chi,digits=3)
```

Table 20.4. Observed and expected frequencies for the Goodness-of-Fit Test for shrike habitat use.

```
          obs      exp
Open     1456 594.638
MidSucc    43  78.067
ScatTree  112  99.660
Woods       6 881.991
Wetland    44   6.644
```

Table 20.5. Results from the Goodness-of-Fit Test for shrike habitat use.

```
X-squared = 2345.071, df = 4, p-value < 2.2e-16
```

Table 20.6. Observed proportions, 95% condidence intervals for the proprtions, and expected proportions for shrike habitat use.

```
         p.obs p.LCI p.UCI p.exp
Open     0.877 0.860 0.892 0.358
MidSucc  0.026 0.019 0.035 0.047
ScatTree 0.067 0.056 0.081 0.060
Woods    0.004 0.002 0.008 0.531
Wetland  0.026 0.020 0.035 0.004
```

20.2.4   Example - Modes of Fishing

The 11-steps (Section 16.1) for a hypothesis test for this example is below:

> Herriges and King (1999) examined modes of fishing for a large number of recreational saltwater users in southern California. One of the questions asked in their Southern California Sportfishing Survey was what "mode" they used for fishing – "from the beach", "from a fishing pier", "on a private boat", or "on a chartered boat." The results to this question, along with other data not used here, are found in FishingModes.csv. One hypothesis of interest states that two-thirds of the users will fish from a boat, split evenly between private and charter boats, while the other one-third will fish from land, also split even between those fishing on the beach and those from a pier. Use the data in the mode variable of the data file to determine if this hypothesis is supported at the 10% level.

1. α=0.10.
2. H_0:"The distribution will follow the proportions of $\frac{1}{3}$, $\frac{1}{3}$, $\frac{1}{6}$, and $\frac{1}{6}$ for private boat, charter boat, beach, and pier modes of fishing, respectively" vs. H_A:"The distribution will NOT follow the proportions of $\frac{1}{3}$, $\frac{1}{3}$, $\frac{1}{6}$, and $\frac{1}{6}$ for private boat, charter boat, beach, and pier modes of fishing, respectively." [*Thought process – the two-thirds for "boat" fishing is split to one-third each for private and charter boats; the one-third, or two-sixths, for "land" fishing is split to one-sixth each for beach and pier fishing.*]
3. A Goodness-of-Fit Test is required because (i) a categorical variable was recorded (mode), (ii) a single group (or population) was considered (Southern California Sportfishers), and (iii) the obseved distribution is compared to a theoretical distribution.
4. The data appear to be part of an observational study where the individuals were not obviously (probably were not) randomly selected.
5. There are more than five individuals expected in each mode (Table 20.7).
6. The statistic is the observed frequency table in Table 20.7.
7. χ^2=31.980 with 3 df (Table 20.8).
8. p-value< 0.00005 (Table 20.8).
9. H_0 is rejected because the p-value< α.
10. The modes of fishing do not appear to match the distribution outlined in the null hypothesis.
11. The 95% confidence intervals for the proportion of use of each mode is in Table 20.9. From these results it is apparent that the users use the beach slightly less than expected and use charter boats slightly more than expected. The use of the pier and private boats are not different from expected.

R Appendix:

```
setwd("c:/data/")
sf <- read.csv("FishingModes.csv")
obs <- xtabs(~mode,data=sf)
p.exp <- c(beach=1/6,boat=1/3,charter=1/3,pier=1/6)
( sf.chi <- chisq.test(obs,p=p.exp,rescale.p=TRUE) )
data.frame(obs=sf.chi$observed,exp=sf.chi$expected)
gofCI(sf.chi,digits=3)
```

Table 20.7. Observed and expected frequencies for the Goodness-of-Fit Test for modes of fishing.

```
        obs.mode obs.Freq exp
beach      beach      134 197
boat        boat      418 394
charter  charter      452 394
pier        pier      178 197
```

Table 20.8. Results from the Goodness-of-Fit Test for modes of fishing.

```
X-squared = 31.9797, df = 3, p-value = 5.285e-07
```

Table 20.9. Observed proportions, 95% condidence intervals for the proprtions, and expected proportions for modes of fishing.

```
        p.obs p.LCI p.UCI p.exp
beach   0.113 0.097 0.133 0.167
boat    0.354 0.327 0.381 0.333
charter 0.382 0.355 0.410 0.333
pier    0.151 0.131 0.172 0.167
```

20.2.5   Example - Mendelian Genetics II

Below are the 11-steps (Section 16.1) for completing a full hypothesis test for the following situation:

> *Geneticists hypothesized that three of every four progeny from a cross between two parent fruit-flies known to possess both a dominant and recessive allele would have red eyes. In a controlled experiment, 82 of 151 randomly selected progeny had red eyes. Test at the 1% level if the percentage of red-eyed progeny in the population of progeny is different than what was hypothesized.*

1. α=0.01.
2. H_0:"The proportion of progeny with red eyes is 0.75" vs. H_A:"The proportion of progeny with red eyes is NOT 0.75."
3. A Goodness-of-Fit Test is required because (i) a categorical variable was recorded (red eye color or not), (ii) a single group (or population) was considered in the experiment, and (iii) the observed distribution is compared to a theoretical distribution.
4. The data appear to be experimental in that a specific cross was made and the environment in which they were raised was controlled. Progeny were randomly selected.
5. There are more than five individuals expected in each eye level (Table 20.10).
6. The appropriate statistic is the observed frequency table in Table 20.10.
7. χ^2=34.49 with 1 df (Table 20.11)
8. p-value< 0.00005 (Table 20.11).
9. H_0 is rejected because the p-value $< \alpha$.
10. The proportion of red-eyed progeny appears to be different than 0.75. Thus, the Mendelian theory is not supported by these results.
11. From the 95% confidence intervals in Table 20.12 it appears that the proportion of progeny with red eyes was between 0.464 and 0.620, which indicates that there were many fewer red-eyed progeny than would be expected from the Mendelian theory.

R Appendix:

```
obs <- c(red=82,nonred=151-82)
p.exp <- c(red=0.75,nonred=0.25)
( m.chi <- chisq.test(obs,p=p.exp,rescale.p=TRUE) )
data.frame(obs=m.chi$observed,exp=m.chi$expected)
gofCI(m.chi,digits=3)
```

Table 20.10. Observed and expected frequencies for the Goodness-of-Fit Test for the genetic cross experiment.

```
        obs     exp
red      82 113.25
nonred   69  37.75
```

Table 20.11. Results from the Goodness-of-Fit Test for the genetic cross experiment.

```
X-squared = 34.4923, df = 1, p-value = 4.279e-09
```

Table 20.12. Observed proportions, 95% condidence intervals for the proprtions, and expected proportions for eye colors in the genetic cross experiment.

```
       p.obs p.LCI p.UCI p.exp
red    0.543 0.464 0.620  0.75
nonred 0.457 0.380 0.536  0.25
```