Section 4.4 – Goodness-of-Fit

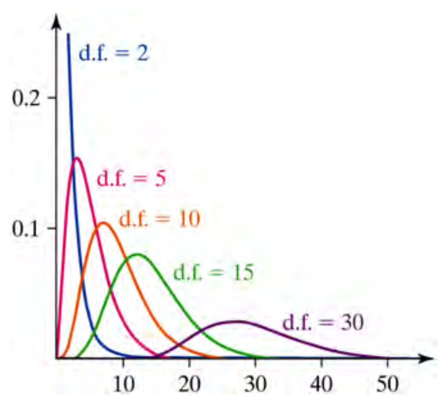
We consider sample data consisting of observed frequency counts arranged in a single row or column (called a one-way frequency table). We will use a hypothesis test for the claim that the observed frequency counts agree with some claimed distribution, so that there is a good fit of the observed data with the claimed distribution.

Definition

A ***goodness-of-fit*** test is used to test the hypothesis that an observed frequency distribution fits (or conforms to) some claimed distribution.

Characteristics of the Chi-Square Distribution

1. It is not symmetric
2. Its shape depends on the degrees of freedom, just like Student's *t*-distribution
3. As the number of degrees of freedom increases, it becomes more nearly symmetric
4. The values of χ^2 are nonnegative. That is, the values of χ^2 are greater than or equal to 0



Test Statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Notation

O represents the ***observed frequency*** of an outcome.
E represents the ***expected frequency*** of an outcome.
k represents the ***number of different categories*** or outcomes.
n represents the total ***number of trials***.

Critical Values

The chi-square distribution with $k-1$ degrees of freedom
Goodness-of-fit hypothesis tests are always *right-tailed*.

Requirements

1. The data have been randomly selected.
2. The sample data consist of frequency counts for each of the different categories.
3. All expected frequencies are greater than or equal to 1 (all $E_i \geq 1$) and
4. No more than 20% of the expected frequencies are less than 5.

CAUTION!

Goodness-of-fit tests are used to test hypotheses regarding the distribution of a variable based on a single population.

Expected Frequencies

If all expected frequencies are <u>equal</u>: $E = \frac{n}{k}$

The sum of all observed frequencies divided by the number of categories

If expected frequencies are <u>not all equal</u>: $E = np$

Each expected frequency is found by multiplying the sum of all observed frequencies by the probability for the category.

Goodness-of-Fit Test

To test the hypotheses regarding a distribution, we use the steps that follow

***Step* 1**: Determine the null and alternative hypotheses.

H_0: The random variable follows a certain distribution

H_1: The random variable does not follow the distribution in the null hypothesis

***Step* 2**: Decide on a level of significance, α, depending on the seriousness of making a Type I error.

***Step* 3:**

a) Calculate the expected counts for each of the k categories. The expected counts are $E_i = np_i$ for $i = 1, 2, \ldots, k$ where n is the number of trials and p_i is the probability of the ith category, assuming that the null hypothesis is true.

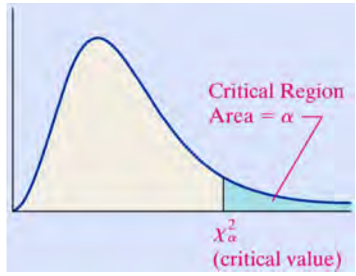
b) Verify that the requirements for the goodness-of-fit test are satisfied.

1. All expected counts are greater than or equal to 1 (all $E_i \geq 1$).
2. No more than 20% of the expected counts are less than 5.

Classical Approach

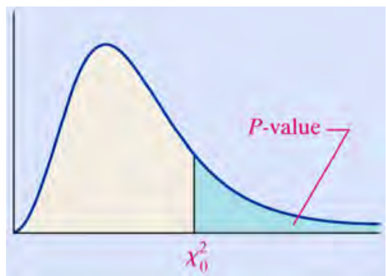
c) Compute the ***test statistic***: $\chi_0^2 = \sum \frac{\left(O_i - E_i\right)^2}{E_i}$

***Step* 4:** Determine the critical value. All goodness-of-fit tests are right-tailed tests, so the critical value is χ^2_α with $k - 1$ degrees of freedom.



Compare the critical value to the test statistic. If $\chi^2_0 > \chi^2_\alpha$ reject the null hypothesis.

d) Use Table to obtain an approximate *P*-value by determining the area under the chi-square distribution with $k - 1$ degrees of freedom to the right of the test statistic.



***Step* 4:** If the *P*-value < α, reject the null hypothesis.

"If the *P* is low, the null must go."
(If the *P*-value is small, reject the null hypothesis that the distribution is as claimed.)

Example

Data Set includes weights from 40 randomly selected adult males and 40 randomly selected adult females. Those weights were obtained as part of the National Health Examination Survey. When obtaining weights of subjects, it is extremely important to actually weigh individuals instead of asking them to report their weights. By analyzing the last digits of weights, researchers can verify that weights were obtained through actual measurements instead of being reported. When people report weights, they typically round to a whole number, so reported weights tend to have many last digits consisting of 0. In contrast, if people are actually weighed with a scale having precision to the nearest 0.1 pound, the weights tend to have last digits that are uniformly distributed, with 0, 1, 2, … , 9 all occurring with roughly the same frequencies. Table shows the frequency distribution of the last digits from 80 weights. (For example, the weight of 201.5 lb has a last digit of 5, and this is one of the data values included in Table)

Last Digit	Frequency
0	7
1	14
2	6
3	10
4	8
5	4
6	5
7	6
8	12
9	8

Test the claim that the sample is from a population of weights in which the last digits do not occur with the same frequency. Based on the results, what can we conclude about the procedure used to obtain the weights?

Solution

Requirements are satisfied: randomly selected subjects, frequency counts, expected frequency is 8 (> 5)

At least one of the probabilities $p_0, p_1, \ldots, p_9$, is different from the others

At least one of the probabilities are the same:

$$p_0 = p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9$$

Null hypothesis contains equality

$$H_0 : p_0 = p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9$$

H_1 : At least one probability is different

No significance specified, use $\alpha = 0.05$

Testing whether a uniform distribution so use goodness-of-fit test: χ^2

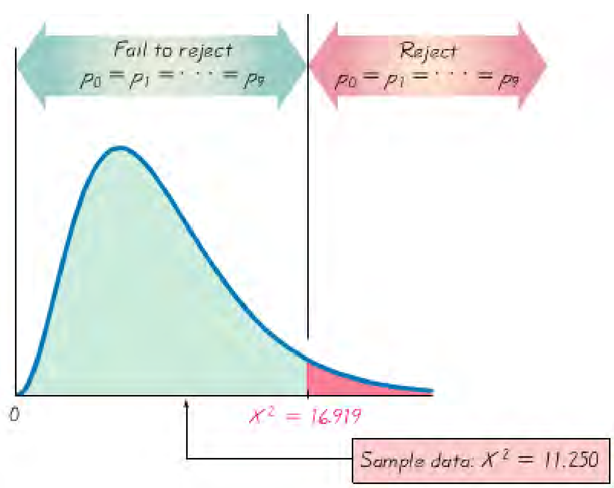
Last Digit	*Observed Frequency O*	*Expected Frequency E*	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
0	7	8	–1	1	0.125
1	14	8	6	36	4.500
2	6	8	–2	4	0.500
3	10	8	2	4	0.500
4	8	8	0	0	0.000
5	4	8	–4	16	2.000
6	5	8	–3	9	1.125
7	6	8	–2	4	0.500
8	12	8	4	16	2.00
9	8	8	0	0	0.000

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 11.250$$

The test statistic $\chi^2 = 11.250$, using $\alpha = 0.05$ and $k - 1 = 9$ degrees of freedom, the critical value is $\chi^2 = 16.919$

Because the test statistic does not fall in the critical region, there is not sufficient evidence to reject the null hypothesis.

There is not sufficient evidence to support the claim that the last digits do not occur with the same relative frequency.



Conclusion

This goodness-of-fit test suggests that the last digits provide a reasonably good fit with the claimed distribution of equally likely frequencies. Instead of asking the subjects how much they weigh, it appears that their weights were actually measured as they should have been.

Example

Table below lists the numbers of games played in the baseball World Series. That table also includes the expected proportions for the numbers of games in a World Series, assuming that in each series, both teams have about the same chance of winning. Use a 0.05 significance level to test the claim that the actual numbers of games fit the distribution indicated by the probabilities.

Games Playe4d	4	5	6	7
Actual World Series Contests	19	21	22	37
Expected Proportion	$\frac{2}{16}$	$\frac{4}{16}$	$\frac{5}{16}$	$\frac{5}{16}$

Solution

Step 1: The original claim: $p_4 = \frac{2}{16},\quad p_5 = \frac{4}{16},\quad p_6 = \frac{5}{16},\quad p_7 = \frac{5}{16}$

Step 2: If the original claim is false, then at least one of the proportions does not have the value as claimed.

Step 3: null hypothesis contains equality

$H_0:\ p_4 = \frac{2}{16},\quad p_5 = \frac{4}{16},\quad p_6 = \frac{5}{16},\quad p_7 = \frac{5}{16}$

$H_1:$ At least one probability is not equal to the given claimed value.

Step 4: The significance level is $\alpha = 0.05$

Step 5: Testing whether the distribution of numbers of games in World Series contests is as claimed, use goodness-of-fit test: χ^2

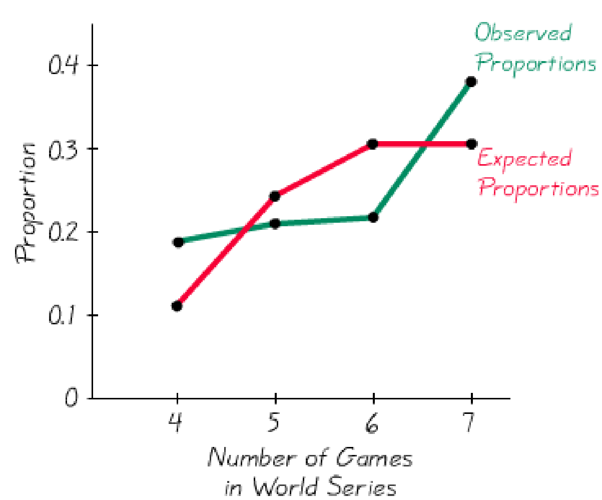
Step 6: Table above shows the computation of the χ^2 test statistic. The test statistic $\chi^2 = 7.885$, using $\alpha = 0.05$ and $k - 1 = 3$ degrees of freedom, the *P*-value is 0.048

Step 7: The *P*-value of 0.048 is less than the significance level of 0.05, so there is sufficient evidence to reject the null hypothesis. Also the test statistic of $\chi^2 = 7.885$ is in critical region bounded by the critical value of 7.185.

Step 8: There is sufficient evidence to warrant rejection of the claim that actual numbers of games in World Series contests fit the distribution indicated by the expected proportions.

Conclusion

This goodness-of-fit test suggests that the numbers of games in World Series contest do not fit the distribution expected from probability calculations



Exercises *Section* 4.4 – Goodness-of-Fit

1. A poll typically involves the selection of random digits to be used for telephone numbers. The New York Times states that "within each (telephone) exchange, random digits were added to form a complete telephone number, thus permitting access to listed and unlisted numbers. "When such digits are randomly generated, what is the distribution of those digits? Given such randomly generated digits, what is a test for "goodness-off-fit"?

2. When generating random digits, we can test the generated digits for goodness-of-fit with the distribution in which all of the digits are equally likely. What does an exceptionally large value of the χ^2 test statistic suggest about the goodness-of-fit? What does an exceptionally small value of the χ^2 test statistic (such as 0.002) suggest about the goodness-of-fit?

3. You purchased a slot machine, and tested it by playing it 1197 times. There are 10 different categories of outcome, including no win, win jackpot, win with three bells, and so on. When testing the claim the observed outcomes agree with the expected frequencies, the author obtained a test statistic of $\chi^2 = 8.185$. Use a 0.05 significance level to test the claim that the actual outcomes agree with the expected frequencies. Does the slot machine appear to be functioning as expected?
Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

4. Do "*A*" students tend to sit in a particular part of the classroom? The author recorded the locations of the students who received grades *A*, with these results: 17 sat in the front, 9 sat in the middle, and 5 sat in the back of the classroom. When testing the assumption that the "*A*" students are distributed evenly throughout the room, the author obtained the test statistic of $\chi^2 = 7.226$. If using a 0.05 significance level, is there sufficient evidence to support the claim that the "*A*" students are not evenly distributed throughout the classroom? If so, does that mean you can increase your likelihood of getting an *A* by sitting in the front of the room? Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

5. Randomly selected nonfat occupational injuries and illnesses are categorized according to the day of the week that they first occurred, and the results are listed below. Use a 0.05 significance level to test the claim that such injuries and illness occur with equal frequency on the different days of the week. Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

Day	*Mon*	*Tues*	*Wed*	*Thurs*	*Fri*
Number	23	23	21	21	19

6. Records of randomly selected births were obtained and categorized according to the day of the week that they occurred. Because babies are unfamiliar with our schedule of weekdays, a reasonable claim is that occur on the different days with equal frequency. Use a 0.01 significance level to test that claim. Can you provide an explanation for the result? Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

Day	*Sun*	*Mon*	*Tues*	*Wed*	*Thurs*	*Fri*	*Sat*
Number of births	77	110	124	122	120	123	97

7. The table below lists the frequency of wins for different post positions in the Kentucky Derby horse race. A post position of 1 is closest to the inside rail, so that horse has the shortest distance to run. (Because the number of horses varies from year to year, only the first ten post positions are included.) Use a 0.05 significance level to test the claim that the likelihood of winning is the same for the different post positions. Based on the result, should bettor consider the post position of a horse racing in the Kentucky Derby? Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

Post Position	1	2	3	4	5	6	7	8	9	10
Wins	19	14	11	14	14	7	8	11	5	11

8. The table below lists the cases of violent crimes are randomly selected and categorized by month. Use a 0.01 significance level to test the claim that the rate of violent crime is the same for each month. Can you explain the result? Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

Month	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Number	786	704	835	826	900	868	920	901	856	862	783	797

9. The table below lists the results of the Advanced Placement Biology class conducted genetics experiments with fruit flies. Use a 0.05 significance level to test the claim that the observed frequencies agree with the proportions that were expected according to principles of genetics Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

Characteristic	***Red eye / normal wing***	***Sepia eye / normal wing***	***Red eye / vestigial wing***	***Sepia eye / vestigial wing***
Frequency	59	15	2	4

10. The table below lists the claims that its M&M plain candies are distributed with the following color percentages: 16% green, 20% orange, 14% yellow, 24% blue, 13% red, and 13% brown. Use a 0.05 significance level to test the claim that the color distribution is as claimed.

Green	Orange	Yellow	Blue	Red	Brown
19	25	8	27	13	8