College Prep Stats
Extra Practice and Review

1. The author drilled a hole in a die and filled it with a lead weight, then proceeded to roll it 200 times. Here are the observed frequencies for the outcomes of 1, 2, 3, 4, 5, and 6, respectively: 27, 31, 42, 40, 28, 32. Use a 0.05 significance level to test the claim that the outcomes are not equally likely.

Outcomes	Observed Frequency	Expected Proportion	Expected Frequency, *E*
1	27	1/6	33.333
2	31	1/6	33.333
3	42	1/6	33.333
4	40	1/6	33.333
5	28	1/6	33.333
6	32	1/6	33.333

Requirements Check: 1) Random Sample 2) Observed Frequencies are all given 3) All Expected Frequencies are > 5

a) State the null hypothesis and the alternative hypothesis.

H_0 : $p_1 = p_2 = p_3 = p_4 = p_5 = p_6$
H_1 : At lease one of the probability is different (original claim)

b) Calculate the test statistic.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 5.86$$

c) Determine the *p*-value.

P-Value = 0.3201 > 0.05, Fail to reject H_0

d) What is the conclusion?

There is not sufficient evidence to support the claim that the outcomes are not equally likely.

2. Among the four northwestern states, Washington has 51% of the total population, Oregon has 30%, Idaho has 11%, and Montana has 8%. A market researcher selects a sample of 1000 subjects, with 450 in Washington, 340 in Oregon, 150 in Idaho, and 60 in Montana. At the 0.05 significance level, test the claim that the sample of 1000 subjects has a distribution that agrees with the distribution of state populations.

Northwestern States' Population	Observed Frequency	Expected Proportion	Expected Frequency, *E*
Washington	450	0.51	510
Oregon	340	0.30	300
Idaho	150	0.11	110
Montana	60	0.08	80

Requirements Check: 1) Random Sample 2) Observed Frequencies are all given 3) All Expected Frequencies are > 5

a) State the null hypothesis and the alternative hypothesis.

H_0 : p_1 = 0.51, p_2 = 0.30, p_3 = 0.11, p_4 = 0.08 (original claim)
H_1 : At lease one of the proportion is different from the given claimed value

b) Calculate the test statistic.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 31.9376$$

c) Determine the *p*-value.

P-Value = 5.3943E-7 < 0.05, Reject H_0

d) What is the conclusion?

There is sufficient evidence to warrant rejection of the claim that the sample of 1000 subjects has a distribution that agrees with the distribution of state populations.

3. Researchers investigated the issue of race and equality of access to clinical trials. The table below shows the population distribution and the numbers of participants in clinical trials involving lung cancer (based on data from "Participation in Cancer Clinical Trials," by Murthy, Krumholz, and Gross, Journal of the American Medical Association, Vol. 291, No. 22). Use a 0.01 significance level to test the claim that the distribution of clinical trial participants fits well with the population distribution.

Race/Ethnicity	White Non-Hispanic	Hispanic	Black	Asian/Pacific Islander	American Indian/Alaskan Native
Distribution of Population	75.6%	9.1%	10.8%	3.8%	0.7%
Number in Lung Cancer Clinical Trials	3855	60	316	54	12

Race/Ethnicity	Observed Frequency	Expected Distribution of Population	Expected Frequency, *E*
White Non-Hispanic	3855	0.756	3248.5
Hispanic	60	0.091	391.03
Black	316	0.108	464.08
Asia/Pacific Islander	54	0.038	163.29
American Indian/Alaskan Native	12	0.007	30.079

Requirements Check: 1) Random Sample   2) Observed Frequencies are all given   3) All Expected Frequencies are > 5

a) State the null hypothesis and the alternative hypothesis.

H_0 : p_1 = 0.756, p_2 = 0.091, p_3 = 0.108, p_4 = 0.038, p_5 = 0.007   (original claim)
H_1 : At lease one of the proportion is different from the given claimed value

b) Calculate the test statistic.

$\chi^2 = \sum \frac{(O-E)^2}{E} = 524.7132$

c) Determine the *p*-value.

P-Value = 0 < 0.01, Reject H_0

d) What is the conclusion?

There is sufficient evidence to warrant rejection of the claim that the distribution of clinical trial participants fits well with the population distribution.

4. Randomly selected nonfatal occupational injuries and illnesses are categorized according to the day of the week that they first occurred, and the results are listed below (based on data from the Bureau of Labor Statistics). Use a 0.05 significance level to test the claim that such injuries and illnesses occur with equal frequency on the different days of the week.

Day	Mon	Tues	Weds	Thurs	Fri
Number	23	23	21	21	19

Day	Observed Frequency *O*	Expected Proportion	Expected Frequency *E*
Mon	23	0.2	21.4
Tues	23	0.2	21.4
Weds	21	0.2	21.4
Thurs	21	0.2	21.4
Fri	19	0.2	21.4

Requirements Check: 1) Random Sample 2) Observed Frequencies are all given 3) All Expected Frequencies are > 5

a) State the null hypothesis and the alternative hypothesis.

H_0 : $p_1 = p_2 = p_3 = p_4 = p_5$ (original claim)
H_1 : At lease one of the proportion is different

b) Calculate the test statistic.

$\chi^2 = \sum \frac{(O-E)^2}{E} = 0.5234$

c) Determine the *p*-value.

P-Value = 0.9712 > 0.05, Fail to reject H_0

d) What is the conclusion?

There is not sufficient evidence to warrant rejection of the claim that such injuries and illnesses occur with equal frequency on the different days of the week.

5. Records of randomly selected births were obtained and categorized according to the day of the week that they occurred (based on data from the National Center for Health Statistics). Because babies are unfamiliar with our schedule of weekdays, a reasonable claim is that births occur on the different days with equal frequency. Use a 0.01 significance level to test that claim.

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

Day	Observed Frequency O	Expected Proportion	Expected Frequency E
Sun	77	1/7	110.43
Mon	110	1/7	110.43
Tues	124	1/7	110.43
Weds	122	1/7	110.43
Thurs	120	1/7	110.43
Fri	123	1/7	110.43
Sat	97	1/7	110.43

Requirements Check: 1) Random Sample 2) Observed Frequencies are all given 3) All Expected Frequencies are > 5

a) State the null hypothesis and the alternative hypothesis.

H_0 : $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7$ (original claim)
H_1 : At lease one of the proportion is different

b) Calculate the test statistic.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 16.8952$$

c) Determine the *p*-value.

P-Value = 0.0097 < 0.01, Reject H_0

d) What is the conclusion?

There is sufficient evidence to warrant rejection of the claim that births occur on the different days with equal frequency.

6. Mars, Inc. 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. Refer to the table below and use the sample data to test the claim that the color distribution is as claimed by Mars, Inc. Use a 0.05 significance level.

Color	Observed Frequency O
Green	19
Orange	25
Yellow	8
Blue	27
Red	13
Brown	8

Color	Observed Frequency O	Expected Proportion	Expected Frequency E
Green	19	0.16	16
Orange	25	0.20	20
Yellow	8	0.14	14
Blue	27	0.24	24
Red	13	0.13	13
Brown	8	0.13	13

Requirements Check: 1) Random Sample 2) Observed Frequencies are all given 3) All Expected Frequencies are > 5

a) State the null hypothesis and the alternative hypothesis.

H_0 : $p_1 = 0.16$, $p_2 = 0.20$, $p_3 = 0.14$, $p_4 = 0.24$, $p_5 = 0.13$, $p_6 = 0.13$ (original claim)
H_1 : At lease one of the proportion is different from the given claimed value

b) Calculate the test statistic.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 6.6820$$

c) Determine the *p*-value.

P-Value = 0.2454 > 0.05, Fail to reject H_0

d) What is the conclusion?

There is not sufficient evidence to warrant rejection of the claim that the color distribution is as claimed by Mars, Inc.

According to Benford's law, a variety of different data sets include numbers with leading (first) digits that follow the distribution shown in the table below. In Numbers 2 & 3, test for goodness-of-fit with Benford's law.

Leading digit	1	2	3	4	5	6	7	8	9
Benford's Law: distribution of leading digits	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%

7. When working for the Brooklyn District Attorney, investigator Robert Burton analyzed the leading digits of the amounts from 784 checks issued by seven suspect companies. The frequencies were found to be 0, 15, 0, 76, 479, 183, 8, 23, and 0, and those digits correspond to the leading digits of 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively. If the observed frequencies are substantially different from the frequencies expected with Benford's law, the check amounts appear to result from fraud. Use a 0.01 significance level to test for goodness-of-fit with Benford's law.

Leading digit	Observed Frequency O	Benford's Law
1	0	0.301
2	15	0.176
3	0	0.125
4	76	0.097
5	479	0.079
6	183	0.067
7	8	0.058
8	23	0.051
9	0	0.046

Leading digit	Observed Frequency O	Expected Proportion	Expected Frequency E
1	0	0.301	235.08
2	15	0.176	137.46
3	0	0.125	97.625
4	76	0.097	75.757
5	479	0.079	61.699
6	183	0.067	52.327
7	8	0.058	45.298
8	23	0.051	39.831
9	0	0.046	35.926

Requirements Check: 1) Random Sample 2) Observed Frequencies are all given 3) All Expected Frequencies are > 5

a) State the null hypothesis and the alternative hypothesis.

H_0 : $p_1 = 0.301$, $p_2 = 0.176$, $p_3 = 0.125$, $p_4 = 0.097$, $p_5 = 0.079$, $p_6 = 0.067$, $p_7 = 0.058$, $p_8 = 0.051$, $p_9 = 0.046$ (original claim)
H_1 : At lease one of the proportion is different from the given claimed value

b) Calculate the test statistic.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 3650.2514$$

c) Determine the *p*-value.

P-Value = $0 < 0.01$, Reject H_0

d) What is the conclusion?

There is sufficient evidence to warrant rejection of claim that the Benford's law is fit. Or, those check amounts appear to result from fraud.

8. Amounts of recent political contributions are randomly selected, and the leading digits are found to have frequencies of 52, 40, 23, 20, 21, 9, 8, 9, and 30. (Those observed frequencies correspond to the leading digits of 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively, and they are based on data from "Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance," by Cho and Gaines, American Statistician, Vol. 61, No. 3.) Using a 0.01 significance level, test the observed frequencies for goodness-of-fit with Benford's law.

Leading digit	Observed Frequency O	Benford's Law
1	52	0.301
2	40	0.176
3	23	0.125
4	20	0.097
5	21	0.079
6	9	0.067
7	8	0.058
8	9	0.051
9	30	0.046

Leading digit	Observed Frequency *O*	Expected Proportion	Expected Frequency *E*
1	52	0.301	63.812
2	40	0.176	37.312
3	23	0.125	26.5
4	20	0.097	20.564
5	21	0.079	16.748
6	9	0.067	14.204
7	8	0.058	12.296
8	9	0.051	10.812
9	30	0.046	9.752

Requirements Check: 1) Random Sample 2) Observed Frequencies are all given 3) All Expected Frequencies are > 5

a) State the null hypothesis and the alternative hypothesis.

H_0 : $p_1 = 0.301$, $p_2 = 0.176$, $p_3 = 0.125$, $p_4 = 0.097$, $p_5 = 0.079$, $p_6 = 0.067$, $p_7 = 0.058$, $p_8 = 0.051$, $p_9 = 0.046$ (original claim)
H_1 : At lease one of the proportion is different from the given claimed value

b) Calculate the test statistic.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 49.6894$$

c) Determine the *p*-value.

P-Value = 4.6887 E-8 < 0.01, Reject H_0

d) What is the conclusion?

There is sufficient evidence to warrant rejection of the claim that the Benford's law is fit. Or, there exists Statistical Fraud Detection in Campaign Finance.