Math 207
Introduction to Statistics

Worksheet: The χ^2 Test

1. Plotting the χ^2 curve.

   a) Use the following to plot a χ^2 curve with 5 degrees of freedom. Make a sketch the result.

   >x <−**seq**(0, 30, **by**=0.2)

   >**plot**(x, **dchisq**(x, **df**=5), type='l', ylim=**c**(0, 0.15))

   b) Repeat part a) for a χ^2 curve with 10 degrees of freedom. Make a sketch of the result using the same axes you used in a).

   c) Repeat part a) for a χ^2 curve with 15 degrees of freedom. Make a sketch of the result using the same axes you used in a).

   d) Describe the effect that increasing the number of degrees of freedom has on the shape of the curve.

   e) (Trick question) Estimate the area under each curve.

2. Find the area under the χ^2-curve with 5 degrees of freedom to the right of the given value. (Hint: 1−**pchisq**(x))

   a) 1.61 b) 9.24 c) 15.09

3. Repeat exercise #1 for a χ^2-curve with 10 degrees of freedom.

4. A die is rolled 60 times. The following frequency table is calculated. Is the die fair?

value	observed frequency
1	5
2	7
3	17
4	16
5	8
6	7

5. A die is rolled 60 times. The following frequency table is calculated. Is the die fair?

value	observed frequency
1	9
2	11
3	10
4	8
5	12
6	10

6. A die is rolled 600 times. The following frequency table is calculated. Is the die fair?

value	observed frequency
1	90
2	110
3	100
4	80
5	120
6	100

7. A die is rolled 60,000 times. The following frequency table is calculated. Is the die fair?

value	observed frequency
1	10,287
2	10,056
3	9,708
4	10,080
5	9,935
6	9,934

8. A four-sided die is rolled several times. The following frequency table is calculated. Is the die fair?

value	observed frequency
1	105
2	90
3	111
4	94

Solutions: 1. d) Increasing the degrees of freedom shifts area to the right and flattens the curve. e) area = 1
2. a) 90% b) 10% c) 1% 3. a) 100% b) 51% c) 13% 4. Probably not: $\chi^2 = 13.2$ and $P = 2.2\%$
5. Yes: $\chi^2 = 1.0$ and $P = 96.3\%$ 6. Maybe: $\chi^2 = 10.0$ and $P = 7.5\%$
7. No: $\chi^2 = 18.6$ and $P = 0.2\%$ 8. Yes: $\chi^2 = 2.82$ and $P = 42.0\%$

Math 207
Introduction to Statistics

Worksheet: The χ^2 Test

1. Plotting the χ^2 curve.

   a) Use the following to plot a χ^2 curve with 5 degrees of freedom. Make a sketch the result.

   ```
   >x <-seq(0, 30, by=0.2)
   >plot(x, dchisq(x, df=5), type='l', ylim=c(0, 0.15))
   ```

There are some nice examples here: https://en.wikipedia.org/wiki/Chi-squared_distribution

   b) Repeat part a) for a χ^2 curve with 10 degrees of freedom. Make a sketch of the result using the same axes you used in a).

   c) Repeat part a) for a χ^2 curve with 15 degrees of freedom. Make a sketch of the result using the same axes you used in a).

   d) Describe the effect that increasing the number of degrees of freedom has on the shape of the curve.

It shifts the peak to the right.

   e) (Trick question) Estimate the area under each curve. The area under any histogram is 1.

2. Find the area under the χ^2-curve with 5 degrees of freedom to the right of the given value. (Hint: 1−**pchisq**(x))

   a) 1.61 b) 9.24 c) 15.09

   a) 90.0% b) 9.99% c) 0.999%

3. Repeat exercise #1 for a χ^2-curve with 10 degrees of freedom.

   a) 99.9% b) 50.9% c) 12.9%

4. A die is rolled 60 times. The following frequency table is calculated. Is the die fair?

For a fair die, the expected frequence is $60 \times \frac{1}{6} = 10$. So,
$\chi^2 = \frac{(5-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(17-10)^2}{10} + \frac{(16-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(7-10)^2}{10}$.
The sum is 13.2. So, the P-value is $1 - \text{pchisq}(13.2, \text{df} = 5)$
which is 2.16%. This supports the alternative hypothesis: the die is not fair.

value	observed frequency
1	5
2	7
3	17
4	16
5	8
6	7

5. A die is rolled 60 times. The following frequency table is calculated. Is the die fair?

For a fair die, the expected frequence is $60 \times \frac{1}{6} = 10$. So,
$\chi^2 = \frac{(9-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(10-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(10-10)^2}{10}$.
The sum is 1.0. So, the P-value is $1 - \text{pchisq}(1.0, \text{df} = 5)$
which is 96.3%. This supports the null hypothesis: the die is fair.

value	observed frequency
1	9
2	11
3	10
4	8
5	12
6	10

6. A die is rolled 600 times. The following frequency table is calculated. Is the die fair?

For a fair die, the expected frequence is $600 \times \frac{1}{6} = 100$. So,
$\chi^2 = \frac{(90-100)^2}{100} + \frac{(110-100)^2}{100} + \frac{(100-100)^2}{100} + \frac{(80-100)^2}{100} + \frac{(120-100)^2}{100} + \frac{(100-10)^2}{100}$.
The sum is 10.0. So, the P-value is $1 - \text{pchisq}(10.0, \text{df} = 5)$
which is 7.5%. This weakly supports the null hypothesis: the die might be fair.

value	observed frequency
1	90
2	110
3	100
4	80
5	120
6	100

7. A die is rolled 60,000 times. The following frequency table is calculated. Is the die fair?

For a fair die, the expected frequence is $60{,}000 \times \frac{1}{6} = 10{,}000$. So,
$\chi^2 = \frac{(10{,}287-10{,}000)^2}{10{,}000} + \frac{(10{,}056-10{,}000)^2}{10{,}000} + \cdots + \frac{(9{,}934-10{,}000)^2}{10{,}000}$.
The sum is 18.6. So, the P-value is $1 - \text{pchisq}(18.6, \text{df} = 5)$
which is 0.2%. This supports the alternative hypothesis: the die is not fair.

value	observed frequency
1	10,287
2	10,056
3	9,708
4	10,080
5	9,935
6	9,934

8. A four-sided die is rolled several times. The following frequency table is calculated. Is the die fair?

The die was rolled a total of 400 times. For a fair die, the expected frequence is $400 \times \frac{1}{4} = 100$. So,
$\chi^2 = \frac{(105-100)^2}{100} + \frac{(90-100)^2}{100} + \frac{(111-100)^2}{100} + + \frac{(94-100)^2}{100}$.
The sum is 2.82. So, the P-value is $1 - \text{pchisq}(2.82, \text{df} = 3)$
which is 0.2%. This supports the null hypothesis: the die is fair.

value	observed frequency
1	105
2	90
3	111
4	94