Sepehr Akbari

April 26th, 2024

## Calculating in the x^2-distribution

### Problem 1

Compute $P(x^2 < 7.1)$ in $x^2(8)$.

**Answer:**

```r
pchisq(7.1, 8)
```

```
[1] 0.4741171
```

### Problem 2

Compute $P(x^2 > 12)$ in $x^2(15)$.

**Answer:**

```r
1 - pchisq(12, 15)
```

```
[1] 0.6790291
```

## Goodness of fit testing

For each of the following problems,

(a) Write hypotheses appropriate to the research question.

(b) Compute expected counts for the specified categorical variable.

(c) Compute the $x^2$ test statistic for the given sample.

(d) Identify the number of degrees of freedom of the $x^2$ distribution.

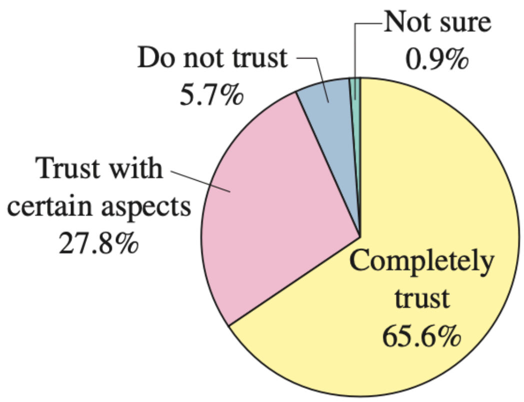
(e) Compute the p-value of the test.

(f) Make a decision at significance level $\alpha = 0.05$. State your conclusion in non-technical language.

(g) Confirm your results with the `chisq.test` function.

## Problem 1

The pie chart shows the distribution of how much married U.S. male adults trust their spouses to manage their finances. A financial services company claims that the distribution of how much married U.S. female adults trust their spouses to manage their finances is the same as the distribution for married U.S. male adults. To test this claim, you randomly select 400 married U.S. female adults and ask each how much she trusts her spouse to manage their finances. The table shows the results. At $a = 0.10$, test the company's claim.




| Survey results | |
|---|---|
| **Response** | **Frequency, *f*** |
| Completely trust | 243 |
| Trust with certain aspects | 108 |
| Do not trust | 36 |
| Not sure | 13 |

**Answer:**

(a) Hypotheses

Null Hypothesis (H0): The distribution of trust in spouses' finances is the same for married U.S. males and females.

Alternative Hypothesis (H1): The distribution of trust in spouses' finances is different for married U.S. males and females.

b) Expected Counts

```
male_data <- c(0.656, 0.278, 0.057, 0.009)

female_expected <- male_data * 400
print(female_expected)
```

```
[1] 262.4 111.2  22.8   3.6
```

c) X-squared statistic

```r
female_data <- c(243, 108, 36, 13)
female_expected <- c(262.4, 111.2, 22.8, 3.6)

x2 <- sum((female_data - female_expected) ^ 2 / female_expected)
print(x2)
```

```
[1] 33.71293
```

d) Degrees of Freedom

```r
df <- length(female_data) - 1
print(df)
```

```
[1] 3
```

e) p-value

```r
p_value <- pchisq(x2, df)
print(p_value)
```

```
[1] 0.9999998
```

f) conclusion

The p-value in this case is very high, and more than 0.1 (alpha), therefore we can say the probability that this will happen is high, and the null hypotheses is not rejected.

g) Validation

```r
chisq.test(female_data)
```

```

	Chi-squared test for given probabilities

data:  female_data
X-squared = 321.78, df = 3, p-value < 2.2e-16
```

## Problem 2

A doctor claims that the number of births by day of the week is uniformly distributed. To test this claim, you randomly select 700 births from a recent year and record the day of the week on which each takes place. The table shows the results. At $a = 0.10$, test the doctor's claim.

| Day | Frequency, $f$ |
|---|---|
| Sunday | 65 |
| Monday | 107 |
| Tuesday | 117 |
| Wednesday | 115 |
| Thursday | 114 |
| Friday | 109 |
| Saturday | 73 |

**Answer:**

a) Hypotheses

Null Hypothesis (H0): The number of births is uniformly distributed across all days of the week.

Alternative Hypothesis (H1): The number of births is not uniformly distributed across all days of the week.

b) Expected Counts

```
exp <- 700 / 7 # 7 is days of the week
print(exp)
```

```
[1] 100
```

c) X-squared statistic

```
expected_value <- c(700/7, 700/7, 700/7,700/7, 700/7, 700/7, 700/7)
observed_value <- c(65, 107, 117, 115, 114, 109, 73)

x2 <- sum((observed_value - expected_value) ^ 2 / expected_value)
print(x2)
```

```
[1] 27.94
```

d) Degrees of Freedom

```r
df <- length(observed_value) - 1
print(df)
```

```
[1] 6
```

e) p-value

```r
p_value <- 1 - pchisq(x2, df)
print(p_value)
```

```
[1] 9.643914e-05
```

f) conclusion

As our p-value is signicantly lower than 0.05, we easily reject the null-hypotheses. Moreover, as p-value is extremely small, we can say number of births in a week being uniform is almost impossible.

g) Validation

```r
chisq.test(observed_value)
```

```

	Chi-squared test for given probabilities

data:  observed_value
X-squared = 27.94, df = 6, p-value = 9.644e-05
```