# Computing Chi-Squared by Hand

First, we selected 3 random indexes to use as our toy-data frame, being:

- 2
- 32
- 67

The data is as follows:

| | Hour Played | Hour Watched | |
|---|---|---|---|
| Random #1 | 20 | 8 | 28 |
| Random #2 | 8 | 4 | 12 |
| Random #3 | 30 | 4 | 34 |
| | 58 | 16 | 74 |

These will be our observed values, now we must calculate our expected values with

$$\text{Exp. Freq} = \frac{\text{total \# row observations} \cdot \text{total \# column observations}}{\text{total \# observations}}$$

| | | |
|---|---|---|
| Random #1 | (28·58) / 74 | (28·16) / 74 |
| Random #2 | (12·58) / 74 | (12·16) / 74 |
| Random #3 | (34·58) / 74 | (34·16) / 74 |

↓

Expected Values

| | Hour Played | Hour Watched |
|---|---|---|
| Random #1 | 21.95 | 6.05 |
| Random #2 | 9.41 | 2.59 |
| Random #3 | 26.65 | 7.35 |

Then to get $X^2$ we do:

$$x^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Now we do our calculations:

Random #1 / Played : $\frac{(20 - 21.95)^2}{21.95} = \frac{-1.95^2}{21.95} = \frac{3.8025}{21.95} = .17$

Random #1 / Watched $\frac{(8 - 6.05)^2}{6.05} = \frac{1.95^2}{6.05} = \frac{3.8025}{6.05} = .63$

Random #2 / Played: $\frac{(8-9.41)^2}{9.41} = \frac{-1.41^2}{9.41} = \frac{1.9881}{9.41} = .21$

Random #2 / Watched: $\frac{(4-2.59)^2}{2.59} = \frac{1.41^2}{2.59} = \frac{1.9881}{2.59} = .77$

Random #3 / Played: $\frac{(30-26.65)^2}{26.65} = \frac{3.35^2}{26.65} = \frac{11.2225}{26.65} = .42$

Random #3 / Watched: $\frac{(4-7.35)^2}{7.35} = \frac{-3.35^2}{7.35} = \frac{11.2225}{7.35} = 1.53$

$$\chi^2 = .17 + .63 + .21 + .77 + .42 + 1.53 = 3.73$$

(this is very close to our python value of 3.7185!)

$$df = (\#\text{ row} - 1) \cdot (\#\text{ column} - 1)$$
$$(3-1) \cdot (2-1)$$
$$2 \cdot 1$$
$$df = 2$$

Using chi-square table $\alpha = 0.05$, $df = 2$ we get a crit value of 5.99

Because our test value is less than our crit value, we can accept our null hypothesis that there is an independence

Hand Calculations for Correlation

| | Hour Played | Hour Watched |
|---|---|---|
| Random #1 | 20 | 8 |
| Random #2 | 8 | 4 |
| Random #3 | 30 | 4 |

$\bar{x} = 20+8+30/3 \rightarrow 58/3 = 19.33$

$\bar{y} = 8+4+4/3 \rightarrow 16/3 = 5.33$

$x - x_{mean}$

$20 - 19.33 = .67$

$8 - 19.33 = -11.33$

$30 - 19.33 = 10.67$

$y - y_{mean}$

$8 - 5.33 = 2.67$

$4 - 5.33 = -1.33$

$4 - 5.33 = -1.33$

$(x - x_{mean}) \cdot (y - y_{mean})$

$.67(2.67) = 1.7889$

$-11.33(-1.33) = 15.0689$

$10.67(-1.33) = -14.1911$

Sum: 2.6667

$(x - x_{mean})^2$

$.67^2 = .4489$ Sum:

$-11.33^2 = 128.3689 = 242.6667$

$10.67^2 = 113.8489$

$(y - y_{mean})^2$

$2.67^2 = 7.1289$ Sum: 10.6667

$-1.33^2 = 1.7689$

$-1.33^2 = 1.7689$

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} = \frac{2.6667}{\sqrt{(242.6667)(10.6667)}}$$

$r = 0.0524$

This does match the correlation coefficient we calculated in python.