

Firstly, the “main” plagiarizer is Yuqi Mai, who copied the contents and paste them on the project. And he doesn’t tell us.

Secondly, I actually did a lot of things on his part to make his part correct. Things include:

1. In 3.4 step 4, I find that degree of freedom should be 11 instead of 6. So I change the χ_6^2 to χ_{11}^2 .
2. In his origin step, he use both fisher test and NHST, that is to say, he calculated both p-value and reject region for $\alpha = 0.05$. (That’s because his github source did so.) However, I find it meaningless, so I eliminate one step.
3. He used average num in the table, while use total num in histogram. I find it out and ask him to change it back.

Similarly, we also do our test on months.

Through raw data processing, we have the following data table and graph:

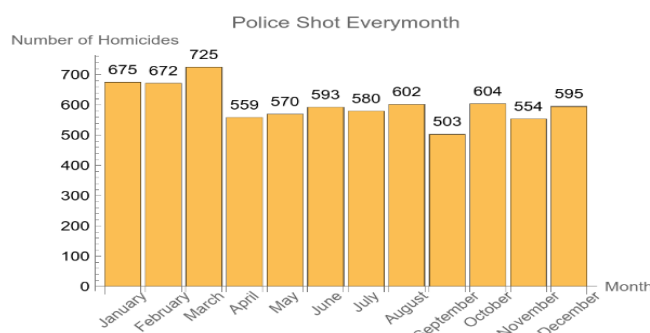


Figure 3: Number of Occurrence for Different Months

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Observe	84.38	84.00	90.63	79.86	81.43	84.71	82.86	86.00	71.88	86.29	79.14	85.00
Expect	83.01	83.01	83.01	83.01	83.01	83.01	83.01	83.01	83.01	83.01	83.01	83.01

Table 4: Average Observed and Expected Occurrence

Figure 1: experiment setup

4. He used average number (per year) instead of total number in the chi-squared goodness-of-fit test, which is wrong because we should use observed data. At first I feel confused, and as a result I asked on piazza. After I understand I change all the calculation process to correct.
5. I also changed the table style to make it more professional.
6. Most importantly, I explain everything to him in the wechat group about my change to him.

I can send you the wechat message mentioned above.

Since we use latex overleaf to cooperate, we have a history record, and I recovered an older version of the project, in which I haven't reviewed his work yet. The phenomena in this project paper include:

1. He use the wrong method and get weird result in chi-squared goodness-of-fit test.

Step 3 We then find the Pearson statistic.
It follows a chi-squared distribution with

$$7 - 1 = 6$$

degrees of freedom.

The observed test statistic is

$$\sum_{i=1}^7 \frac{(O_i - E_i)^2}{E_i} = 0.037 \quad (1)$$

Step 4 Fix $\alpha = 0.05$, we test the null hypothesis with the $\chi_{0.05,6}^2 = 12.592$. We found that

$$\chi_0^2 = 0.037 < 12.592 = \chi_{0.05,6}^2$$

0.069 is much less than 12.592. Hence we are unable to reject H_0 .

Step 5 At last, we calculate the P -value of this test.

$$\begin{aligned} p &= P[\chi_0^2 \mid H_0] \\ &\leq P[\chi^2 \geq 0.037] \\ &= 1 - P[\chi^2 \geq 0.037] \\ &= 0.999999 \\ &\approx 1 \end{aligned}$$

Figure 2: wrong method, duplicated steps and weird result

2. His tables are not professional.
3. The degree of freedom he used is 6 in 3.4 step4.

So, from these records, you can see that **I have already tried my best to keep good communication with the whole group.** I've contacted him with everything that he made it wrong(because he copied github). I don't know he copied github at that time, but still, you see, I'm alert from what he wrote and find out many mistakes.

You can see, before the modification, his part is terrible and it indeed violate HC. However, I do not ignore the mistakes, and I correct them.

In conclusion, I think I've done my best at that time. It's just because I don't know he copied github. So I shouldn't be guilty.