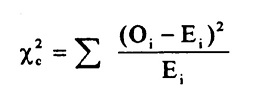
Final Conclusion

From the leaking query, we have the number of leaking places in each class. By going through each revision's log, we check if it mentions "bug(s)", "fix(es)", "error(s)" and "issue(s)". If yes, then we consider this revision is about fixing bugs. We consider the related classes in this revision had bugs once. Everytime we found bug fixing revision. add 1 to bugs number for each class. Right now, we have leaking number and bugs number for each class. we can start to test our assumption: If there is correlation between bugs and leaks? The first test is regression model test. We have 26,018 ckasses and their bugs number and leaks number. We put leaks number as X-axis, bugs number as Y-axis and get figure 1. We fit the data with regression model and get the equation below:

Y= a+b\*X+ c\*lnX + d\*√ X + e\*X^-1 + f\*X^-2

The R-square value of this model is only 0.03. This value explains how well this model fits the data. 0.03 is too small to be a good model. **So we consider that there is no correlation between number of bugs and number of leaks**.

Right now, we want to figure out if leaks have influence on the appearance of bugs. To figure out this question, we us Chi-square test of independence to check if leaks and bugs are two independent variable, or there exist some influence on each other. Our null hypothesis is (H0): There is no relationship between leak and bug. And Ha: There is relationship between leak and bug. We have 31256 classes in total, and each number for leak, non-leak, bug and non-bug, in table 1.



Oi is the number of 3690, 5460, 8292 and 13810 in the table 1. Ei is the estimated number in table 2 correspondingly. By the formula, we have X^2 is 21.365. We choose the 1 degree of freedom and 0.05 level of P-value = 3.841. 21.365 is much bigger than 3.841, which means our null hypothesis is only less 0.05 possibility correct. So we reject null hypothesis. **We conclude that There is relationship between leaks and bugs.**

Combine two conclusions we get above, we know That:

1. No correlation between number of bugs and number of leaks, which means that more leaks does means more bugs in the class.

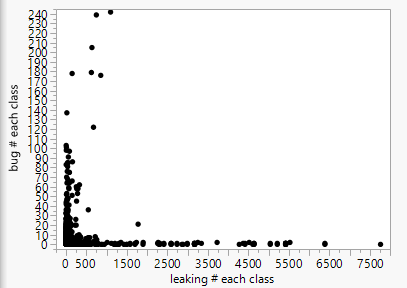
2. There is relationship between leaks and bugs, which means a class with leaks is more likely has bugs compared with the class without leaks.

|  |  |  |  |
| --- | --- | --- | --- |
|  | leak | Non-leak | sum |
| bug | 3690 | 5464 | 9154 |
| Non-bug | 8292 | 13810 | 22102 |
| sum | 11982 | 19274 | 31256 |

Table 1

|  |  |  |
| --- | --- | --- |
|  | estimated leak | estimated nonleak |
| estimated bug | 3509.189532 | 5644.810468 |
| estimated nonbug | 8472.810468 | 13629.18953 |

Table 2

Figure 1

