# MAT1372, Classwork22, Fall2025

## *6.3 Testing for Goodness of Fit Using Chi-square (Conti.)*

6. Conditions for The Chi-square test. Two conditions that must be checked before performing a chi-square test:

*Independence*. Each case that contributes a count to the table must be independent of all the other cases in the table

*Sample size / distribution*. Each particular scenario must have at least 5 expected cases.

(*Warning*: Failing to check conditions may affect the test's error rates.)

7. CHI-SQUARE TEST FOR ONE-WAY TABLE

Suppose we are to evaluate whether there is convincing evidence that a set of observed counts $O_1, O_2, O_3, \ldots, O_k$ are unusually different from what might be expected under a null hypothesis. Call the expected counts $E_1, E_2, \ldots, E_k$ that are based on the null hypothesis.

If each expected count is at least 5 and the null hypothesis ~~is~~ was true, then the test statistic below follows a chi-square distribution with k−1 degree of freedom.

$$X^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} + \cdots + \frac{(O_k-E_k)^2}{E_k}$$

The p-value for this test statistic is found by looking at the upper tail of this chi-square distribution. We consider the upper tail because larger values of $X^2$ would provide greater evidence against the null hypothesis.

## *6.4 Testing for Independence in Two-way Tables*

1. Differences of One-way tables versus Two-way tables.

One-way table: It describes counts for each outcome in a single value

Two-way table: It describes counts for combinations of outcomes for two variables. We offten want to know "are these variable related in any way?"

2. Example of a Two-way table.

In the used iPhone market, what questions should a buyer ask a seller about any underlying problems with what they're selling?

| | General | Positive Assumption | Negative Assumption | Total |
|---|---|---|---|---|
| Disclose Problem | 2 | 23 | 36 | 61 |
| Hide Problem | 71 | 50 | 37 | 158 |
| Total | 73 | 73 | 73 | 219 |

- General: What can you tell me about it?
- Pos. Asm.: It doesn't have any problems, does it?
- Neg. Asm.: What problems does it have?

(a) What is the proportion of all the sellers who disclosed problem of the devices? $61/219 = 27.85\%$

(b) *If there was no difference among the questions* and (a) shows the disclosure proportion of the sellers no matter the questions, what is the expected number of sellers who we would expect to disclose or hide the issue for all groups? (if $E$( ***general*** and ***disclosed***) means expected counts in ***general*** and ***disclosure***)

$E$( general and disclosed) = $(\text{General total}) \times \frac{\text{Disclosure Total}}{\text{total}} = 73 \cdot \frac{61}{219} = 20.33$

$E$( pos. and disclosed) = $(\text{Pos. Total}) \times \frac{\text{Disclosure total}}{\text{total}} = 73 \cdot \frac{61}{219} = 20.33$

$E$(neg. and disclosed) = $(\text{neg Total}) \times \frac{61}{219} = 73 \cdot \frac{61}{219} = 20.33$

$E$( general and hide) = $(\text{General Total}) \times \frac{\text{Hided Total}}{\text{Total}} = 73 \times \frac{158}{219} = 52.67$

$1 - \frac{61}{219} = \frac{158}{219}$

$E$( pos. and hide) = (pos. total) $\times \frac{158}{219} = 52.67$

$E$(neg. and hide) = (neg. total) $\times \frac{158}{219} = 52.67$

(c) The chi-square test statistic for a two-way table is found the same way it is found for a one-way table:

| | General | Positive Assumption | Negative Assumption | Total |
|---|---|---|---|---|
| Disclose Problem | 2 (20.33) | 23 (20.33) | 36 (20.33) | 61 |
| Hide Problem | 71 (52.67) | 50 (52.67) | 37 (52.67) | 158 |
| Total | 73 | 73 | 73 | 219 |

Figure 6.15: The observed counts and the (expected counts).

Let's use some notation to simplify the expression to find the **chi-square test statistic** $X^2$:

2: $O_{g,d}$ is observed counts in general and disclosure and $E_{g,d}$ (20.33) is the expected counts in general and disclosure.

| | | |
|---|---|---|
| $\frac{(O_{g,d}-E_{g,d})^2}{E_{g,d}} = \frac{(2-20.33)^2}{20.33} = 16.53$ | $\frac{(O_{p,d}-E_{p,d})^2}{E_{p,d}} = \frac{(23-20.33)^2}{20.33} = 0.35$ | $\frac{(O_{n,d}-E_{n,d})^2}{E_{n,d}} = \frac{(36-20.33)^2}{20.33} = 12.08$ |
| $\frac{(O_{g,h}-E_{g,h})^2}{E_{g,h}} = \frac{(71-52.67)^2}{52.67} = 6.38$ | $\frac{(O_{p,h}-E_{p,h})^2}{E_{p,h}} = \frac{(50-52.67)^2}{52.67} = 0.14$ | $\frac{(O_{n,h}-E_{n,h})^2}{E_{n,h}} = \frac{(37-52.67)^2}{52.67} = 4.66$ |

$X^2 = 16.53 + 0.35 + 12.08 + 6.38 + 0.14 + 4.66 = 40.14$

(d) What is the degree of freedom (df) for this test statistic $X^2$?

df = (number of rows − 1) × (number of column − 1) = $(2-1)\times(3-1) = 1\times 2 = 2$

(e) What are appropriate hypotheses for this case?

$H_0$: There is no difference in the effectiveness of the 3 questions.

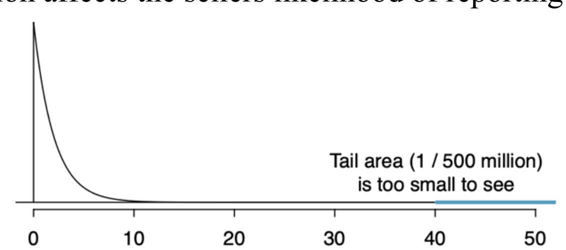
$H_A$: There is some difference of the 3 questions

(f) Find the p-value and draw a conclusion about whether the question affects the sellers likelihood of reporting the problem.

df = 2. $X^2 = 40.14$.

$P(X^2 > 40.14)$ is less than 0.001

(= 0.000000002)




3. The summary of testing in two-way tables.

(1) Build a hypotheses with a significance level $\alpha$.

(2) Compute the expected counts: $\text{Expect Count}_{\text{row } i,\ \text{col } j} = \frac{(\text{row } i \text{ total}) \times (\text{col } j \text{ total})}{\text{table total}}$

(3) Compute the test statistic $X^2$: $X^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$

(4) Compute the degree of freedom (df): df = (# of rows − 1) × (# of columns − 1)

(5) Compute the p-value: Check the table with proper df to find the value or range of p-value

(6) Reject or accept $H_0$: We accept $H_0$ if p-value > $\alpha$