M358K: November 10th, 2021.

Goodness · of · Fit.

We are studying a multinomial experiment, possibly w/ categorical descriptions of possible outcomes.

The possible outcomes of the experiment will be categories which are mutually exclusive and exhaustive.

Represent the categories as events: $A_1, A_2, \dots, A_k$.

In our probabilistic model, the parameters are

$$p_1, p_2, \dots, p_k$$

w/ $p_i = \mathbb{P}[A_i]$ for $i = 1, 2, \dots, k$.

Note: $\boxed{p_1 + p_2 + \dots + p_k = 1}$ ☆☆

Repeat the same multinomial experiment n times independently.

Let X_i denote the number of times the outcome i occurred, for $i = 1, 2, \dots, k$.

Note: $\boxed{X_1 + X_2 + \dots + X_k = n}$ ☆

For our test statistic:

$$\boxed{Q^2 := \sum_{i=1}^{k} \frac{(X_i - n \cdot p_i)^2}{n p_i} \approx \chi^2(df = k-1)}$$

Works for $np_i \geq 5$ for all $i = 1, 2, \dots, k$

Test Summary.

$H_0: p_1 = p_1^\circ, p_2 = p_2^\circ, \dots, p_k = p_k^\circ$

vs.

H_a: At least one of the population probabilities is different from its null value

$(\exists i \in \{1, .., k\} \; p_i \neq p_i^\circ)$

The χ^2-test is always an upper-tailed one!

Let E_i denote the expected counts under the null, i.e.,

$$\boxed{E_i = n \cdot p_i^\circ}$$

Let O_i denote the observed counts.

The observed value of the test statistic:

$$q^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

With R, we can calculate p-value: $\boxed{1 - \text{pchisq}(q^2, df = k-1)}$.

With a significance level α, find $\boxed{\chi^2_\alpha(df = k-1)}$.

If $q^2 \geq \chi^2_\alpha(df = k-1)$, then reject the null hypothesis.

If not, then fail to reject.

UNIVERSITY OF TEXAS AT AUSTIN

Problem Set # 18

Goodness of fit.

Problem 18.1. Gregor Almond, the local horticulturalist, grows 400 progeny from a cross of peas. The cross is hypothesised to have a ratio of 1 green to 7 yellow seeds. Suppose that the cross actually produces 360 yellow and 40 green seeded plants.

(i) Calculate the observed value of the test statistic.
(ii) Using the χ^2−tables, what would your decision be at the significance level $\alpha = 0.05$.
(iii) Using **R** with the observed value of the test statistic, find the p−value.
(iv) Using the command
```
chisq.test()
```
perform the χ^2−test and provide the summary.
(v) In this case, you can test the same hypotheses using the z−test. Do this for practice!

$n = 400$, $k = 2$

H_0: $p_g = 1/8$, $p_y = 7/8$

vs.

H_a: The color dist'n is different from the null.
($p_g \neq 1/8$ or $p_y \neq 7/8$)

The expected counts : $E_g = 400(\frac{1}{8}) = 50$ and $E_y = 350$
The observed counts : $O_g = 40$ and $O_y = 360$

$$\Rightarrow q^2 = \frac{(50-40)^2}{50} + \frac{(350-360)^2}{350} = 2.285714$$

$\alpha = 0.05 \longrightarrow \chi^2_{0.05}(df = 2-1=1) = 3.841$ Fail to Reject!

p-value: 1-pchisq(q^2, df=1) = 0.13057 > 0.05 = α