M358K: November 6th, 2023.

Goodness of Fit.

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

The possible outcomes will be categories that 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 all $i = 1..k$

Note: $p_1 + p_2 + \cdots + p_k = 1$

Repeat this multinormal experiment n times independently.

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

Note: $X_1 + X_2 + \cdots + X_k = n$

For our test statistic:

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

Works well for $n \cdot p_i \geq 5$ for all $i = 1, \dots, k$

Test Summary.

H_0: $p_1 = p_1^0, p_2 = p_2^0, \ldots, p_k = p_k^0$

vs.

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

$(\exists i \in \{1, \ldots, k\} \; p_i \neq p_i^0)$

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

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

$$E_i = n \cdot p_i^0$$

Let O_i denote the observed counts.

The observed value of the TS:

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

With R, we simply calculate the p-value: 1-pchisq(q^2, df=k-1)

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

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!

i. H_0: $p_g = \frac{1}{8}$, $p_y = \frac{7}{8}$

vs.

H_a: The color distn is different from the null

($p_g \neq \frac{1}{8}$ or $p_y \neq \frac{7}{8}$)

$k=2$, $n=400$

The expected counts: $E_g = 400 \cdot (\frac{1}{8}) = 50$ and $E_y = 350$

The observed counts: $O_g = 40$ and $O_y = 360$

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

$q^2 \uparrow 9.39$ → Fail to Reject!

ii. $\alpha = 0.05 \longrightarrow \chi^2_{0.05}(df = 2-1=1) = 3.841$

iii. p-value: $1-$pchisq$(q.sq, df=1) = 0.13057 > 0.05 = \alpha$