

Worksheet – Sections 10.1-10.2

Hypothesis test of distribution (also called *goodness-of-fit* test): does a random variable have a certain distribution?

- H_0 : the distribution is the same as the proposed distribution
- H_a : the distribution is different

Chi-square statistic measures the discrepancy between the observed frequencies and the expected frequencies:

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

where $O_1, O_2, \dots, O_k$ are the observed frequencies in a sample, and $E_1, E_2, \dots, E_k$ are the expected frequencies of k possible values of the random variable.

X^2 is approximately *chi-square* distributed with $k - 1$ degrees of freedom if the following conditions are satisfied:

- Simple random sampling.
- Large sample: every expected frequency is at least 5.

p-value = $P(X^2 \geq \text{observed value of } X^2)$. Reject H_0 if the p-value is lower than the significance level.

StatCrunch to compute p-value:

- Method 1: Stats $\rightarrow$ Goodness-of-fit $\rightarrow$ Chi-Square Test
- Method 2: Stat $\rightarrow$ Calculators $\rightarrow$ Chi-Square

1) A six-sided dice was rolled 120 times. The outcomes are:

Face	1	2	3	4	5	6
Observed frequency	15	18	25	20	22	20

You wonder if the dice is fair.

a) State the null hypothesis and alternative hypothesis.

b) Find the expected frequencies for the face values.

c) Compute the chi-square statistic X^2 .

d) Check if the conditions guaranteeing that X^2 is chi-square distributed are satisfied.

e) Compute the chi-square statistic X^2 and p-value using Statcrunch.

f) At the 5% significance level, would you reject the assumption that the dice is fair?

2) In a population, blood types are expected in the proportions 40%, 11%, 4%, 45% for A, B, AB, O respectively. A sample of 1000 individuals gives:

Blood type	A	B	AB	O
Observed frequency	300	125	120	455

You began to question whether the blood-type proportions are truly 40%, 11%, 4%, and 45%.

a) State the null hypothesis and alternative hypothesis.

b) Find the expected frequency for each blood type.

c) Check if the conditions guaranteeing that X^2 is chi-square distributed are satisfied.

d) Compute the chi-square statistic X^2 and p-value using Statcrunch.

e) At the 5% significance level, would you reject the null hypothesis?