THE UNIVERSITY OF MELBOURNE

FACULTY OF BUSINESS & ECONOMICS

Helpsheet

Giblin Eunson Library

PROBABILITY VALUE (p-Value)

Use this sheet to help you:

- Interpret p-values
- Calculate p-value using binomial distribution
- Comparison of calculated p-value with defined p-value (significance level - **α)**
- Decision-making –reject or not reject null hypothesis?
- Misunderstandings of p-values.

Author: Carter, D.
Design and layout: Shaw, J.

p-value

In statistical hypothesis testing, the **p-value** is the probability of obtaining a result at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. The fact that p-values are based on this assumption is crucial to their correct interpretation.

Usually p-values of 0.05 or 0.01 are used, corresponding to a 5% chance or 1% of an outcome that extreme, given the null hypothesis. The result of a test of significance is either "statistically significant" or "not statistically significant"; there are no shades of grey.

More technically, a p-value of an experiment is a random variable defined over the sample space of the experiment such that its distribution under the null hypothesis is uniform on the interval [0, 1]. Many p-values can be defined for the same experiment.

Example:
An experiment is performed to determine whether a coin flip is fair (50% chance of landing heads or tails) or unfairly biased, either toward heads (> 50% chance of landing heads) or toward tails (< 50% chance of landing heads). (A bent coin produces biased results).

Since we consider both biased alternatives, a two-tailed test is performed. The null hypothesis is that the coin is fair, and that any deviations from the 50% rate can be ascribed to chance alone.

Suppose that the experimental results show the coin turning up heads 14 times out of 20 total flips. The p-value of this result would be the chance of a fair coin landing on heads at least 14 times out of 20 flips plus the chance of a fair coin landing on tails 14 or more times out of 20 flips. In this case the random variable T has a binomial distribution. The probability that 20 flips of a fair coin would result in 14 or more heads is 0.0577.(See end of sheet for calculation of this result). By symmetry, the probability that 20 flips of the coin would result in 14 or more tails (alternatively, 6 or fewer heads) is the same, 0.0577. Thus, the p-value for the coin turning up the same face 14 times out of 20 total flips is 0.0577 + 0.0577 = 0.1154.

Generally, the null hypothesis is rejected if the p-value is smaller than or equal to the significance level, often represented by the Greek letter α (alpha). If the level is 0.05, then results that are only 5% likely or less are deemed extraordinary, given that the null hypothesis is true.

In the above example we have:

null hypothesis (H_0) — fair coin;

observation (O) — 14 heads out of 20 flips; and

probability (p-value) of observation (O) given H_0:
$p(O \mid H_0) = 0.0577 \times 2$ (two-tailed) = 0.1154 (percentage expressed as 11.54%).

The calculated p-value exceeds 0.05, so the observation is consistent with the null hypothesis — that the observed result of 14 heads out of 20 flips can be ascribed to chance alone — as it falls within the range of what would happen 95% of the time were this in fact the case. In our example, we fail to reject the null hypothesis at the 5% level. Although the coin did not fall evenly, the deviation from expected outcome is just small enough to be reported as being "not statistically significant at the 5% level".

However, had a single extra head been obtained, the resulting p-value (two-tailed) would be 0.0414 (4.14%). This time the null hypothesis - that the observed result of 15 heads out of 20 flips can be ascribed to chance alone - is rejected. Such a finding would be described as being "statistically significant at the 5% level".

Some misunderstandings of p-values:

1. The p-value is not the probability that the null hypothesis is true. (This false conclusion is used to justify the "rule" of considering a result to be significant if its p-value is very small (near zero).

2. The p-value is not the probability of falsely rejecting the null hypothesis.

3. 1 – (p-value) is not the probability of the alternative hypothesis being true.

4. The significance level of the test is not determined by the p-value.

   The significance level of a test is a value that should be decided upon by the researcher interpreting the data before the data are viewed, and is compared against the p-value or any other statistic calculated after the test has been performed.

Calculation of coin experiment:

$$
\begin{aligned}
&{}^{20}C_{14}(0.5)^{14}(0.5)^{6} + {}^{20}C_{15}(0.5)^{15}(0.5)^{5} + {}^{20}C_{16}(0.5)^{16}(0.5)^{4} \\
&+ {}^{20}C_{17}(0.5)^{17}(0.5)^{3} + {}^{20}C_{18}(0.5)^{18}(0.5)^{2} + {}^{20}C_{19}(0.5)^{19}(0.5)^{1} \\
&+ {}^{20}C_{20}(0.5)^{20}(0.5)^{0} \\
&= 0.0577
\end{aligned}
$$