Click on 'Bookmarks' in the Left-Hand menu and then click on the required section - answers to Q questions at end of section.

14.4 Test for Proportion – using Normal Distribution

14.4.1 Introduction

A simple proportion, where the outcome is either *Y* or *N* (=$\overline{Y}$), is a direct binary choice, and the associated probabilities are determined by the statistics of the binomial distribution (Unit 8.4).

14.4.2 True Proportion

The unknown *true* proportion of outcomes, *Y*, for a system is a *parameter* of the system and we will describe it by the Greek symbol, Π, (capital pi). The true value, Π, is the proportion of *Y* outcomes that would be achieved if the whole *population* of the system were measured.

The experimentally measured proportion, *P*, from a *sample* is the *best estimate* for the unknown true proportion, Π.
If *r* outcomes, giving *Y*, are returned from a sample size, *n*:

$$P = \frac{r}{n} \qquad [14.6]$$

The unknown *true* mean, μ, of the number, *r*, in a binomial distribution (where each trial has a probability *p* of giving the outcome *Y*) is given by [8.33]:

$$\mu = n \times p \qquad [14.7]$$

The unknown *true* proportion, Π, for the number of '*Y* outcomes will be given by:

$$\Pi = \frac{\mu}{n} = p \qquad [14.8]$$

The *true* standard deviation for the number of *Y* outcomes, *r*, is given by [8.20]:

$$\sigma = \sqrt{n \times p \times (1 - p)} \qquad [14.9]$$

Using the fact that the experimental proportion, *P*, is the best estimate for Π, together with [14.8] gives:

$$p \approx P \qquad [14.10]$$

We can then write, combining [14.9] and [14.10]:

$$\sigma \approx \sqrt{n \times P \times (1 - P)} \qquad [14.11]$$

The best estimate of the true standard deviation of the *proportion* is then given by:

$$\sigma_P = \frac{\sigma}{n} \approx \frac{\sqrt{n \times P \times (1-P)}}{n} = \sqrt{\frac{P \times (1-P)}{n}} \qquad [14.12]$$

The 95% Confidence Interval for the value of Π is given by (see [8.22]):

$$CI(\Pi, 95\%) = P \pm 1.96 \times \sqrt{\frac{P \times (1-P)}{n}} \qquad [14.13]$$

14.4.2 One Sample Test

In a one sample test, the measured proportion, *P*, is compared with a particular value for the proportion, Π_O.

A test statistic, *z*, can be calculated:

$$z = \frac{P - \Pi_O}{\sigma_P} = \frac{P - \Pi_O}{\sqrt{\frac{P(1-P)}{n}}} \qquad [14.14]$$

This is equivalent to the *t*-statistic in [10.1]

The critical value $z_{T,\alpha}$ is equivalent to $t_{T,\alpha,\infty}$, which is the equivalent *t*-value evaluated for *df* = ∞ (infinity) - see 8.1a (under Additional Materials).

E14.11
A country is due to have a referendum with only two choices - Yes or No.
In anticipation of the referendum, 1000 people are selected at random in an 'opinion poll' and it is found that 520 will vote Yes and 480 will vote No.
Using this information, calculate whether it is possible to be confident at a 95% level that the result of the actual referendum will be 'Yes'.

The hypotheses for the test become:

Null Hypothesis, H_O	True proportion of Yes votes ≤ 50%
Proposed Hypothesis, H_1	True proportion of Yes votes > 50%

This is a one sample, 1-tailed, test to compare the proportion, *P* = 0.52 returned in one sample with a specific target proportion, Π = 0.5
Using [14.24]

$$z = \frac{0.520 - 0.500}{\sqrt{\frac{0.52(1-0.52)}{1000}}} = \frac{0.020}{0.0158} = 1.27$$

The critical value, $z_{T,\alpha}$, can be calculated from the 1-tailed *t*-value *df* = ∞:

$$z_{1,0.05} = t_{1,0.05,\infty} = 1.65$$

(see Appendix III)

Since

$$z < z_{T,\alpha}$$

we cannot reject the Null Hypothesis and it is not possible to be 95% confident that the vote will return a 'Yes'.

Q14.12 A random sample of 50 frogs is taken from a lake, and it is found that 37 are female and 13 are male
Is this experimental result significantly (at 0.05) different from the expected proportion of 60% females to 40% males?

Answer H_O, the proportion of females is 0.6
H_1, the proportion of females is not 0.6
Expected proportion of females $\Pi_0 = 0.6$
Actual proportion $P = 37/50 = 0.74$
Number in sample, $n = 50$
Using [10.25]

$$z = \frac{P - \Pi_0}{\sqrt{\frac{P(1-P)}{n}}} = \frac{0.74 - 0.60}{\sqrt{\frac{0.74(1-0.74)}{50}}} = \frac{0.14}{0.0620} = 2.26$$

for 2-tails
The critical value of z, $z_{2,0.05} = 1.96$
As $z > z_{2,0.05}$ Accept the Proposed Hypothesis H_1
The proportion of females in the sample is significantly greater than 0.6

14.4.2 Two Sample Test

A two sample test compares the measured proportions, P_A and P_B. The relevant statistic is:
(this is very similar to the *t*-statistic [10.3])

$$z = \frac{P_A - P_B}{\sqrt{P'(1-P') \times (1/n_A + 1/n_B)}} \qquad [14.15]$$

where P' is a 'pooled' value for the proportions of the two samples.
(this is very similar to pooled standard deviation [10.2])

$$P' = \frac{n_A P_A + n_B P_B}{n_A + n_B} \qquad [14.16]$$

Q14.13 A student compares the germination of similar seeds under two different conditions. In condition A, 85 out of 100 seeds germinate, and, in condition B, 60 out of 80 seeds germinate.

Test, at 0.05 significance, whether the proportion of seeds germinating in condition A is significantly different to the proportion in condition B.

Answer

	A	**B**
***r* =**	85	60
***n* =**	100	80
***P* =**	0.85	0.75

Using [14.16], P' = (100 × 0.85 + 80 × 0.75)/(100 + 80) = 0.80556

Using [14.15], z = 1.68447

2-tailed z_{CRIT} from tables = 1.96 ($t_{2,0.05,\infty}$)

Since $z < z_{CRIT}$, do not reject H_0