Statistical Inference - Confidence Intervals and Hypothesis Tests

ANSWERS

Exercise 1 - Confidence interval

In order to define the confidence interval, we need to find the values at the lower and upper boundaries of a range, symmetrical with respect to the measured mean, in the *sampling distribution*. The range must include a percentage of values such that the percentage is equal to the confidence level, which in our case is 95%. The upper and lower boundaries are, hence, values on the x-axis of the distribution that 'fence out' the remaining 5% of possible measured means (2.5% on each tail of the distribution). The boundary values can be read out of the standardised normal distribution percentage point table. This table is given in the appendix in the lab sheet.

Upper-tail percentage points of the standard normal distribution

The table gives the values of z for which $P(Z > z) = p$, where the distribution of Z is $N(0, 1)$.

p	z	p	z	p	z	p	z	p	z
.50	0.000	.15	1.036	.025	1.960	.010	2.326	$.0^3 4$	3.353
.45	0.126	.14	1.080	.024	1.977	.009	2.366	$.0^3 3$	3.432
.40	0.253	.13	1.126	.023	1.995	.008	2.409	$.0^3 2$	3.540
.35	0.385	.12	1.175	.022	2.014	.007	2.457	$.0^3 1$	3.719
.30	0.524	.11	1.227	.021	2.034	.006	2.512	$.0^4 5$	3.891

The z-value (an x-axis value in the standardised normal distribution) that is lower than 2.5% of all possible values is 1.96 (see the picture). As the standardised normal distribution is symmetrical around the mean of 0, the value that is higher than 2.5% of all possible values is -1.96.

The other calculation that we must do is to find the *standard error*, which is the standard deviation of the sampling distribution. It is equal to $\dfrac{s}{\sqrt{n}}$, where s is the standard deviation.

Finally, we can say, with a confidence of 95%, that the mean amount of shampoo in the bottles is in the range:

$$200ml \pm \frac{2ml}{10} \times 1.96$$

or

$$200ml \pm 0.39ml$$

Exercise 2 - Hypothesis test

In this hypothesis test the null hypothesis is that the mean value of the amount of shampoo in the bottles is 200ml:

$$H_0 : \mu = 200ml$$

$$H_a : \mu \neq 200ml$$

H_a is the alternative hypothesis, that the mean is different from 200ml.

For the hypothesis test we need to identify the lower and upper boundaries of the *sampling distribution* range symmetrical around the measured mean that 'fences out' the percentage of values equal to the significance level, which in our case is 1%. Similarly to the look-up in the previous problem, for the upper boundary we need to find the z-value that corresponds to a p-value of 0.005 (0.5%), since the lower boundary will 'fence out' the other 0.5% of the 1% in the negative (left) end tail of the distribution.

Upper-tail percentage points of the standard normal distribution

The table gives the values of z for which $P(Z > z) = p$, where the distribution of Z is $N(0, 1)$.

p	z	p	z	p	z	p	z	p	z
.50	0.000	.15	1.036	.025	1.960	.010	2.326	$.0^3 4$	3.353
.45	0.126	.14	1.080	.024	1.977	.009	2.366	$.0^3 3$	3.432
.40	0.253	.13	1.126	.023	1.995	.008	2.409	$.0^3 2$	3.540
.35	0.385	.12	1.175	.022	2.014	.007	2.457	$.0^3 1$	3.719
.30	0.524	.11	1.227	.021	2.034	.006	2.512	$.0^4 5$	3.891
.25	0.674	.10	1.282	.020	2.054	.005	2.576	$.0^4 1$	4.265
.24	0.706	.09	1.341	.019	2.075	.004	2.652	$.0^5 5$	4.417
.23	0.739	.08	1.405	.018	2.097	.003	2.748	$.0^5 1$	4.753
.22	0.772	.07	1.476	.017	2.120	.002	2.878	$.0^6 5$	4.892
.21	0.806	.06	1.555	.016	2.144	.001	3.090	$.0^6 1$	5.199
.20	0.842	.050	1.645	.015	2.170	$.0^3 9$	3.121	$.0^7 5$	5.327
.19	0.878	.045	1.695	.014	2.197	$.0^3 8$	3.156	$.0^7 1$	5.612
.18	0.915	.040	1.751	.013	2.226	$.0^3 7$	3.195	$.0^8 5$	5.731
.17	0.954	.035	1.812	.012	2.257	$.0^3 6$	3.239	$.0^8 1$	5.998
.16	0.994	.030	1.881	.011	2.290	$.0^3 5$	3.291	$.0^9 5$	6.109

The z-value we are looking for is 2.576. This is the upper boundary of the 'acceptable' range.

Now we calculate the z-value corresponding to the measured mean, in order to see where it falls with respect to the acceptable range. This we call the test statistic and calculate as:

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

where $\bar{x}$ is the mean calculated from the sample (197ml), μ_0 is the hypothesized population mean (200ml) and the fraction acting as nominator is the *standard error* i.e. the standard deviation of the sampling distribution.

We calculate the statistic:
$$T = \frac{197ml - 200ml}{\frac{2ml}{\sqrt{100}}} = 15$$

Note that the units (ml) have been cancelled out and that we have a dimensionless number, i.e. a z-value, which has a standardised normal distribution.

The last thing we need to do is to see whether our calculated test statistic falls within the 'acceptable' range:

$-2.576 \leq T \leq 2.576$ **?**

$-2.576 \leq 15 \leq 2.576$ **FALSE!**

It does not fall in the range, so we *reject* the null hypothesis.