MATH10282 Introduction to Statistics
Semester 2, 2019/2020
Example Sheet 9 - Solutions

1. A sample of $n = 81$ students took a test. In the sample of scores, $\bar{x} = 74.6$ and $s = 11.3$. Note that here σ is unknown. We are not told that the data are normally distributed. Hence we use the approximate $100(1-\alpha)\%$ confidence interval based on large n asymptotic results, whose end points are

$$\bar{x} \pm \frac{z_{1-\alpha/2}s}{\sqrt{n}},$$

Here $\alpha = 0.1$, and $z_{1-\alpha/2} = z_{0.95} = \Phi^{-1}(0.95)$ is the 0.95 quantile of a $N(0,1)$ distribution, i.e. $z_{0.95} = 1.6449$. Thus the end points of the approximate 90% CI are

$$74.6 \pm 1.6449 \times 11.3/\sqrt{81}.$$

Hence an approximate 90% CI for the mean score of all students is (72.53, 76.67).

2. Note that $n = 30$, $\bar{x} = \sum_{i=1}^{n} x_i/n = 1568.45/30 = 52.2817$, and also

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \\ &= \frac{1}{29}(83006.73 - 30 \times 52.2817^2) \\ &= 34.67\,. \end{aligned}$$

Moreover we are told that the data are normally distributed.

(i) If it is known that $\sigma^2 = 30$, then the end points of a 95% CI for μ are given by

$$\bar{x} \pm 1.96\sigma/\sqrt{n} = 52.5817 \pm 1.96\sqrt{30/30}\,.$$

Hence a 95% CI for μ is (50.32, 54.24).

(ii) If σ^2 is unknown, we estimate it via $s^2 = 34.67$. Hence a 95% CI for μ has end points

$$\bar{x} \pm t_{1-\alpha/2}s/\sqrt{n} = 52.2817 \pm 2.0452 \times \sqrt{34.67/30}\,,$$

since $t_{1-\alpha/2} = t_{0.025} = 2.0452$ is the 0.975 quantile of a t distribution on $n-1 = 29$ degrees of freedom. Thus the 95% CI for μ is (50.09, 54.48).

The 95% CI for σ^2 is

$$\left(\frac{(n-1)s^2}{\chi^2_{0.975}}, \frac{(n-1)s^2}{\chi^2_{0.025}}\right),$$

where $\chi^2_{0.975} = 45.7223$ is the 0.975 point of a $\chi^2(29)$ distribution, i.e. the 0.975 quantile. Moreover $\chi^2_{0.025} = 16.0471$ is the 0.025 point of a $\chi^2(29)$ distribution, i.e. the 0.025 quantile. These values have been looked up in tables. Hence the 95% CI for σ^2 is

$$\left(\frac{29 \times 34.67}{45.722}, \frac{29 \times 34.67}{16.047} \right) ,$$

i.e. (21.99, 62.66).

3. The number of households in the sample having three or more TV sets is $X \sim \text{Bi}(n, p)$, where $n = 500$ and p is the proportion of all households having three or more TV sets. An approximate $100(1-\alpha)\%$ CI for p has end points

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} ,$$

where $\hat{p} = X/n = 80/500 = 0.16$. Here $\alpha = 0.1$, and so $z_{1-\alpha/2} = z_{0.95} = 1.6449$. Thus the end points are

$$0.16 \pm 1.6449 \sqrt{\frac{0.16 \times 0.84}{500}} .$$

Hence the 90% CI for p is $(0.133, 0.187)$.

4. $X_1, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ distribution where μ is unknown but σ^2 is known. The $100(1-\alpha)\%$ confidence interval for μ is

$$\left[\bar{x} - \frac{z_{1-\alpha/2}\, \sigma}{\sqrt{n}} , \bar{x} + \frac{z_{1-\alpha/2}\, \sigma}{\sqrt{n}} \right] ,$$

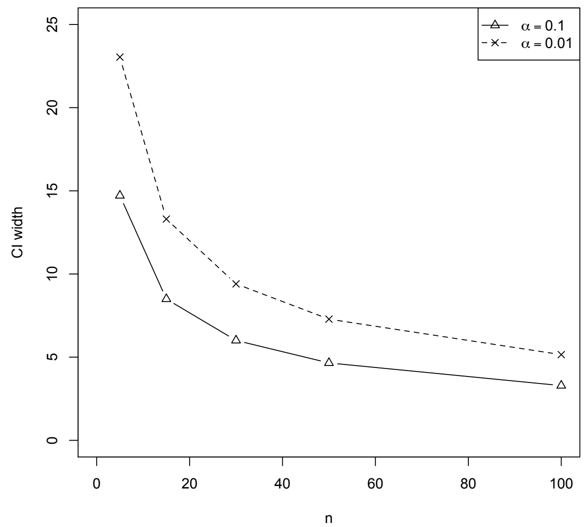
where $\Phi(z_{1-\alpha/2}) = 1-\alpha/2$. This interval has width $2z_{1-\alpha/2}\sigma/\sqrt{n}$, which depends on the values of:

- α, or equivalently the confidence level $1-\alpha$: the higher the confidence level, the wider the confidence interval.
- the standard deviation, σ, of the underlying distribution: for larger σ, the confidence interval is wider
- the sample size, n: the larger the sample size, the narrower the confidence interval

The width of this confidence interval is tabulated below, assuming $\sigma = 10$:

n	5	15	30	50	100	Confidence level
$\alpha = 0.1$	14.71	8.49	6.01	4.65	3.29	90%
$\alpha = 0.01$	23.04	13.30	9.41	7.29	5.15	99%

Note that the relevant z-values are $z_{0.95} = 1.645$ and $z_{0.995} = 2.576$. The results are plotted below:



5. Since $\sigma = 0.05$, the width of the $100(1-\alpha)\%$ confidence interval is $0.1 z_{1-\alpha/2}/\sqrt{n}$.

(i) Here $\alpha = 0.05$ and so the width is less than 0.02 when

$$0.1 \times z_{0.975}/\sqrt{n} \le 0.02 \iff 0.196/\sqrt{n} \le 0.02 \iff n \ge (0.196/0.02)^2 = 96.04\,.$$

Hence at least 97 observations are required.

(ii) Here $\alpha = 0.01$ and so the width is less than 0.02 when

$$0.1 \times z_{0.995}/\sqrt{n} \le 0.02 \iff n \ge (0.258/0.02)^2 = 166.41.$$

Hence at least 167 observations are required.

6. We have a random sample of size $n = 30$ from $N(\mu, \sigma^2)$, with μ and σ^2 unknown, and $\bar{x} = 127.442$, $s^2 = 228.661$.

The 98% CI for μ has end points

$$127.442 \pm t_{0.99}\sqrt{\frac{228.661}{30}} = 127.442 \pm 2.462\sqrt{\frac{228.661}{30}}\,,$$

where $t_{0.99} = 2.462$ is the 0.99 point of a t distribution on $n - 1 = 29$ degrees of freedom. Thus the 98% CI for μ is (120.64, 134.24).

The 98% CI for σ^2 is

$$\left(\frac{(n-1)s^2}{\chi^2_{0.99}}, \frac{(n-1)s^2}{\chi^2_{0.01}}\right) = \left(\frac{29 \times 228.661}{49.588}, \frac{29 \times 228.661}{14.256}\right)$$
$$= (133.73, 465.15)\,,$$

where $\chi^2_{0.99} = 49.588$ and $\chi^2_{0.01} = 14.256$ are respectively the 0.99 point and 0.01 point of a χ^2 distribution on 29 degrees of freedom.