Practice questions

1. The measurements of ten random athlete heights in centimeters are

$$152, 163, 188, 201, 192, 176, 194, 166, 215, 184.$$

   (a) Assuming the heights are independent normal random variables with known standard deviation $\sigma = 20$, give a 95% confidence interval for the mean height.

   Solution: The (symmetric) 95% confidence interval for the mean height of normal samples with known standard deviation σ is $[\overline{X} - z\frac{\sigma}{\sqrt{n}}, \overline{X} + z\frac{\sigma}{\sqrt{n}}]$, where z is chosen so that $\mathrm{P}(-z \leq \mathrm{Normal}(0,1) \leq z) = 95\%$. The latter condition holds for $z = 1.96$. We calculate

$$\overline{X} = \frac{152 + 163 + 188 + 201 + 192 + 176 + 194 + 166 + 215 + 184}{10} = 183.1$$

   to obtain the confidence interval $[170.7, 195.5]$.

   (b) How many samples do you need for a 95% confidence interval of width 5cm?

   Solution: The width of the confidence interval is $\frac{2\sigma}{\sqrt{n}} \cdot z$, so to ensure width at most w, n should be $(2\sigma z/w)^2$. When $w = 5$ and $z = 1.96$, $(2\sigma z/w)^2$, $(2\sigma z/w)^2 = 245.8624$, so $n = 246$ samples are sufficient.

2. A large company conducts a job satisfaction survey among its 6250 employees. Out of 250 employees that are sampled (with repetition), 142 are satisfied with their jobs.

   (a) Calculate a 99% confidence interval for the number of employees that are satisfied with their job.

   Solution: We first come up with a confidence interval for the fraction of satistfied employees p given $n = 250$ Indicator(p) (also known as Bernoulli(p)) independent samples. The sample mean is

$$\overline{X} = \frac{142}{250} = 0.568.$$

   The "traditional" formula based on the normal approximation of a sum of indicator samples gives the 99% confidence interval $[A - zB, A + zB]$ for p, where $z = 2.576$ is the 99 percentile two-sided threshold for the Normal$(0, 1)$ random variable, and

$$A = \frac{\overline{X} + z^2/2n}{1 + z^2/n} \approx 0.567, \qquad B = \frac{\sqrt{\overline{X}(1 - \overline{X})/n + z^2/4n^2}}{1 + z^2/n} \approx 0.031$$

   The 99%-confidence interval for p is $[A - zB, A + zB] \approx [0.487, 0.646]$. The 99%-confidence interval for the number $6250p$ of satisfied employees is $[3041, 4036]$.
   The "simplified" formula gives the 99%-confidence interval

$$\left[\overline{X} - z\sqrt{\overline{X}(1 - \overline{X})/n}, \overline{X} + z\sqrt{\overline{X}(1 - \overline{X})/n}\right] \approx [0.487, 0.648]$$

   for p and the corresponding interval $[3045, 4054]$ for $6250p$.

(b) Find a confidence interval of width 100 for the number of satisfied employees and estimate the confidence level for it.

Solution: If we use the "simplified" formula, the width is $w = N \cdot 2z\sqrt{\overline{X}(1-\overline{X})/n}$, where $N = 6250$ is the total number of employees. If we set $w = 100$, for $\overline{X} = 0.568$ and $n = 250$ we get

$$z = \frac{w\sqrt{n}}{2N\sqrt{\overline{X}(1-\overline{X})}} \approx 0.255$$

which gives a confidence level of $\mathrm{P}(-z \leq \mathrm{Normal}(0,1) \leq z) \approx 0.201$, or only about 20% for the interval $[N\overline{X} - 50, N\overline{X} + 50] = [3500, 3600]$.

3. The midterm test scores of six random students are 81, 84 ,83, 73, 76, 83.

(a) What is the sample variance?

Solution: The sample mean is $\overline{X} = \frac{81+84+83+73+76+83}{6} = 80$, so the sample variance is

$$V = \frac{(81-80)^2 + (84-80)^2 + (83-80)^2 + (73-80)^2 + (76-80)^2 + (83-80)^2}{6} = \frac{50}{3}.$$

(b) Assuming their scores are independent Normal(μ, σ) random variables. Give as large a value for $\hat{\Theta}_-$ as you can so that $(\hat{\Theta}_-, 10)$ is a 95% confidence interval for σ.

Solution: The adjusted sample variance is $S^2 = 50 \cdot 6/(6-1) = 20$. The confidence interval formula for the standard deviation of a normal random variable is $(\sqrt{(n-1)S^2/z_+}, \sqrt{(n-1)S^2/z_-})$, where z_- and z_+ should be chosen so that

$$\mathrm{P}(z_- \leq \chi^2(n-1) \leq z_+) = 0.95.$$

For the right bound to equal 10 we should choose $z_- = 1$. As $\mathrm{P}(\chi^2(n-1) < z_-) \approx 0.03743$ we are looking for z_+ with $\mathrm{P}(\chi^2(n-1) \leq z_+) \approx 0.95 + 0.03743 \approx 0.98743$, which gives $z_+ \approx 14.53$ and $\hat{\Theta}_- = \sqrt{(n-1)S^2/z_+} \approx 2.6234$.

4. A food processing company packages honey in glass jars. The volume of honey in millilitre in a random jar is a Normal(μ, σ) random variable. 5 random jars are picked and the volume of honey inside them in millilitre are 108, 101, 103, 109 and 104.

(a) Suppose μ is unknown and σ is known to be 5. Give a 95% confidence interval for μ.

Solution: The sample mean is $\overline{X} = \frac{108+101+103+109+104}{5} = 105$, so the confidence interval is $(\overline{X} - z\sigma/\sqrt{n}, \overline{X} + z\sigma/\sqrt{n}) \approx (100.62, 109.38)$, where $z \approx 1.96$ is chosen so that $\mathrm{P}(-z \leq \mathrm{Normal}(0,1) \leq z) = 95\%$.

(b) Suppose μ and σ are both unknown. Give a 95% confidence interval for μ.

Solution: The adjusted sample variance is:

$$S^2 = \frac{(108-105)^2 + (101-105)^2 + (103-105)^2 + (109-105)^2 + (104-105)^2}{5-1} = \frac{23}{2}$$

The confidence interval is now of the form $(\overline{X} - zS/\sqrt{n}, \overline{X} + zS/\sqrt{n})$, where z is chosen so that $\mathrm{P}(-z \leq t(4) \leq z) = 95\%$, which gives $z \approx 2.78$ and the confidence interval $(100.78, 109.22)$.

(c) Suppose μ and σ are both unknown. Give a 95% prediction interval for the next sample.

Solution: The prediction interval is of the form $(\overline{X} - zS\sqrt{1+1/n}, \overline{X} + zS\sqrt{1+1/n})$ with the same z as in part (b), giving the answer $(94.67, 115.33)$.