Solution Sheet — Estimation Exercises

Solution Exercise 1: Cholesterol Levels
We are given grouped data from a random sample of $n = 100$ employees.

Cholesterol (class center)	120	160	200	240	280	320
Frequency n_i	9	22	25	21	16	7

1. **Sample Mean and Sample Standard Deviation**

   The sample mean is:

$$\overline{x} = \frac{1}{n} \sum_i n_i x_i$$

$$\sum n_i x_i = 9(120) + 22(160) + 25(200) + 21(240) + 16(280) + 7(320) = 21360$$

$$\boxed{\overline{x} = \frac{21360}{100} = 213.6}$$

   The sample variance is:

$$S^2 = \frac{1}{n} \sum_i n_i x_i^2 - \overline{x}^2$$

$$\sum n_i x_i^2 = 9(120^2) + \cdots + 7(320^2) = 4\,855\,200$$

$$S^2 = \frac{4\,855\,200}{100} - (213.6)^2 = 2916.64$$

$$\boxed{S = \sqrt{2916.64} \approx 54.0}$$

2. **Unbiased Point Estimators**

   We assume that the random variable X has expectation $\mathbb{E}(X) = \mu$ and variance $\mathrm{Var}(X) = \sigma^2$, where both parameters are unknown.

   - Unbiased estimator of population mean μ is approximated by Sample Mean $\overline{x}$ (unbiased and consistent):

$$\mathbb{E}(\overline{X}) = \mathbb{E}(X) = \mu \text{ by the law of large numbers}$$

     and

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} \to 0 \,\text{as}\, n \to \infty\,.$$

     Then, $\mu \approx 213.6$

- Unbiased and consistent estimator of variance:

$$S^{*2} = \frac{n}{n-1} S^2 = \frac{100}{99} \times 2916.64 \approx 2946.3$$

$$\hat{\sigma} = \sqrt{2946.3} \approx 54.3$$

3. **Confidence Interval for the Mean**

   Since $n = 100 > 30$, we use the normal approximation.

$$IC_{0.95}(m) = \overline{x} \pm 1.96 \frac{S^*}{\sqrt{n}}$$

$$IC_{0.95}(m) = 213.6 \pm 1.96 \frac{54.3}{10}$$

$$\boxed{IC_{0.95}(m) = [202.96\,;224.24]}$$

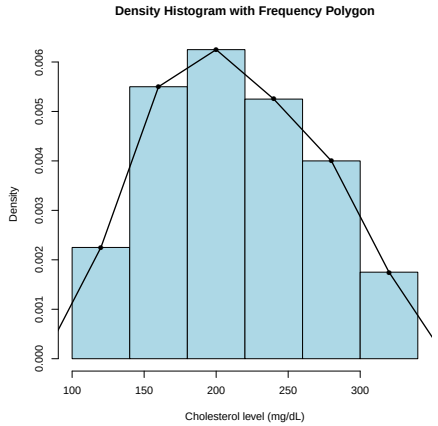
4. **Minimum Sample Size**

   We want the width of the confidence interval to be less than 10:

$$2 \times 1.96 \frac{S^*}{\sqrt{n}} < 10$$

$$n > \left(\frac{2 \times 1.96 \times 54.3}{10} \right)^2 \approx 452.1.$$

$$\boxed{n_{\min} > 452}$$



—

Exercise 2: Coronary Velocity

Observed data ($n = 18$):

1. **Sample Mean and Variance**

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1266}{18} \approx 70.33$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2 = \frac{89508}{18} - 70.33^2 \approx 26.37$$

2. **Unbiased Estimators**

   Unbiased mean:

$$\hat{m} \approx \overline{x} = 70.33$$

   Unbiased variance:

$$\hat{\sigma}^2 \approx S^{*2} = \frac{n}{n-1}S^2 = \frac{18}{17} \times 26.37 \approx 27.92$$

$$\hat{\sigma} \approx S^* = \sqrt{S^{*2}} \approx 5.28$$

3. **Confidence Interval for m (Known Variance)**

   Given $\sigma^2 = 26$, confidence level 98%:

$$IC_{0.98}(m) = \overline{x} \pm z_{0.99}\frac{\sigma}{\sqrt{n}}$$

   where $z_{0.99}$ is the 0.99 quantile of the standard normal distribution (corresponding to the upper tail of 1% for a two-tailed 98% confidence level):

$$z_{0.99} \approx 2.326$$

   The population standard deviation is:

$$\sigma = \sqrt{26} \approx 5.099$$

   The standard error of the mean is:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{5.099}{\sqrt{18}} \approx 1.202$$

   The margin of error is:

$$ME = z_{0.99} \times SE \approx 2.326 \times 1.202 \approx 2.80$$

   Thus, the 98% confidence interval for the mean is:

$$IC_{0.98}(m) \approx 70.33 \pm 2.80 = [67.53, 73.13]$$

4. **Confidence Intervals (Unknown Parameters)**

- IC(Mean) (99%):

$$IC_{0.99}(m) = \overline{x} \pm t_{0.995,17} \frac{S^*}{\sqrt{n}}$$

$$t_{0.995,17} \approx 2.898$$

$$\frac{S^*}{\sqrt{n}} = \frac{5.28}{\sqrt{18}} \approx 1.245$$

Margin of error: $ME = 2.898 \times 1.245 \approx 3.61$

$$\boxed{IC_{0.99}(m) \approx [70.33 - 3.61,\ 70.33 + 3.61] = [66.72, 73.94]}$$

- **IC(Variance) (99%):**

$$IC_{0.99}(\sigma^2) = \left[\frac{(n-1)S^{*2}}{\chi^2_{0.995,\,17}},\ \ \frac{(n-1)S^{*2}}{\chi^2_{0.005,\,17}} \right] = \left[\frac{(n)S^2}{\chi^2_{0.995,\,17}},\ \ \frac{(n)S^2}{\chi^2_{0.005,\,17}} \right]$$

where S^2 is the sample variance, and $\chi^2_{\alpha,\nu}$ denotes the α quantile of the chi-square distribution with ν degrees of freedom.

$$(n-1)S^2 = 17 \times 26.37 \approx 448.29$$

Using chi-square critical values for 17 d.f.:

$$\chi^2_{0.995,17} \approx 6.54, \quad \chi^2_{0.005,17} \approx 32.85$$

Thus, the 99% confidence interval is:

$$IC_{0.99}(\sigma^2) \approx \left[\frac{448.29}{32.85},\ \frac{448.29}{6.54} \right] \approx [13.65, 68.54]$$

Exercise 3: Uterine cancer prevalence

1. **Point estimate**

   Let p denote the true prevalence of uterine cancer in the population of suspected women. The point estimator of p is the sample proportion

$$\hat{p} \approx f = \frac{n}{N} = \frac{25}{100} = 0.25.$$

2. **Confidence interval at the 5% risk level**

   Since

$$n\hat{p} = 25 > 5 \quad \text{and} \quad n(1-\hat{p}) = 75 > 5,$$

   the normal approximation can be used.

   The standard error of $\hat{p}$ is

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.25 \times 0.75}{100}} = 0.0433.$$

For a 95% confidence level, $z_{0.975} = 1.96$. Thus,

$$IC_{95\%}(p) = \hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.25 \pm 0.0849.$$

$$\boxed{IC_{95\%}(p) = [0.165\,;\,0.335]}$$

Exercise 4: Number of individuals to be treated for high cholesterol

A sample of $n = 10\,000$ individuals is considered. The estimated proportion of individuals requiring treatment is

$$\hat{p} \approx f = 7.5\% = 0.075.$$

Let X denote the number of individuals requiring treatment among the 10,000.

$$X \sim \mathcal{B}(10\,000,\ 0.075).$$

Since n is large, a normal approximation is appropriate:

$$X \approx \mathcal{N}(np,\ np(1-p)).$$

Mean and variance

$$\mathbb{E}(X) = np = 10\,000 \times 0.075 = 750,$$
$$Var(X) = np(1-p) = 10\,000 \times 0.075 \times 0.925 = 693.75,$$
$$\sigma_X = \sqrt{693.75} \approx 26.34.$$

95% **confidence interval**

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

With a 95% confidence level, $z_{0.975} = 1.96$, and an observed proportion $\hat{p} = 0.075$, the confidence interval is

$$IC_{95\%}(p) = 0.075 \pm 1.96\sqrt{\frac{0.075(1-0.075)}{10000}} = 0.075 \pm 1.96\sqrt{\frac{0.069375}{10000}}$$

$$IC_{95\%}(p) = 0.075 \pm 1.96\sqrt{\frac{0.069375}{10000}}$$

$$IC_{95\%}(p) = 0.075 \pm 0.00516$$

$$\boxed{IC_{95\%}(p) \approx [0.06984\,;\,0.08016]}$$

Interpretation With a confidence level of 95%, the true number of individuals requiring treatment for high cholesterol among the 10,000 people lies between 698 and 802.