ECON10072 Advanced Statistics

Exercise Sheet 8: Solutions

1. Each of the sample random variables Y_1, Y_2, Y_3 has expected value μ and variance σ^2.

   (a) Expected values are:

$$\begin{aligned} E[A_1] &= \frac{1}{3}(E[Y_1] + E[Y_2] + E[Y_3]) = \frac{3\mu}{3} = \mu; \\ E[A_2] &= \frac{1}{2}(E[Y_1] + E[Y_2]) = \frac{2\mu}{2} = \mu; \\ E[A_3] &= \frac{1}{2}(E[Y_1] + E[Y_2] + E[Y_3]) = \frac{3\mu}{2} = 1.5\mu; \\ E[A_4] &= 0.75E[Y_1] + 0.75E[Y_2] - 0.5E[Y_3] = \mu. \end{aligned}$$

   (b) A_1, A_2 and A_4 are unbiased estimators of μ; A_3 is a biased estimator of μ.

   (c) The variances of the unbiased estimators are, recalling independence of Y_1, Y_2, Y_3 from random sampling,

$$\begin{aligned} \text{var}[A_1] &= \frac{1}{9}(\text{var}[Y_1] + \text{var}[Y_2] + \text{var}[Y_3]) = \frac{3\sigma^2}{9} = 0.333\sigma^2 \\ \text{var}[A_2] &= \frac{1}{4}(\text{var}[Y_1] + \text{var}[Y_2]) = \frac{2\sigma^2}{4} = 0.5\sigma^2 \\ \text{var}[A_4] &= (0.75)^2\,\text{var}[Y_1] + (0.75)^2\,\text{var}[Y_2] + (0.5)^2\,\text{var}[Y_3] = 1.375\sigma^2. \end{aligned}$$

   Hence A_1 is the most efficient.

   (d) Efficiency relative to A_1 is defined by

$$\text{eff}(A_i, A_1) = \frac{\text{var}[A_1]}{\text{var}[A_i]}, \qquad i = 2, 4.$$

   Then,

$$\begin{aligned} \text{eff}(A_2, A_1) &= \frac{\text{var}[A_1]}{\text{var}[A_2]} = \frac{(0.333\sigma^2)}{(0.5\sigma^2)} = 0.667 \\ \text{eff}(A_4, A_1) &= \frac{\text{var}[A_1]}{\text{var}[A_4]} = \frac{(0.333\sigma^2)}{(1.375\sigma^2)} = 0.242. \end{aligned}$$

   A_2 and A_4 are inefficient relative to A_1; and A_4 is the least efficient relative to A_1.

2. Denote unemployment duration by X, $X \sim N(\mu, 129.6)$. For a random sample of size 20, $\bar{X} \sim N\left(\mu, \frac{129.6}{20}\right)$. A confidence level of 98% implies $\alpha = 0.02$, where

$$98\% = 100(1-\alpha)\%.$$

We require the percentage point $z_{\alpha/2} = z_{0.01}$ from the $N(0,1)$ distribution such that

$$\Pr(Z > z_{0.01}) = 0.01$$

and hence, by symmetry,

$$\Pr(Z < -z_{0.01}) = 0.01.$$

From the standard normal table, $z_{0.01} = 2.326$. The 98% confidence interval $[c_L, c_U]$ for μ (the mean unemployment duration of women) is given by

$$\begin{aligned} \bar{x} \pm z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} &= 14.7 \pm (2.326)\sqrt{\frac{129.6}{20}} \\ &= 14.7 \pm (2.326)(2.4558) \\ &= [8.778,\ 20.622]. \end{aligned}$$

3. The question does not state that sampling takes place from a normal distribution. However, the sample size is small and progress cannot be made without this assumption. So, assume $X \sim N(\mu, \sigma^2)$ with both μ and σ^2 unknown. An estimate of σ^2 is $s^2 = 32.5$.

   (a) Construct the confidence interval using

$$T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t_{n-1}.$$

   Degrees of freedom here are $n - 1 = 14$. For a 95% confidence interval we need the percentage point $t_{n-1,\alpha/2} = t_{14,0.025}$ such that

$$\Pr(T \leqslant t_{14,0.025}) = 0.975$$

   and, from the t-distribution table,

$$t_{14,\,0.025} = 2.145.$$

   The 95% confidence interval $[c_L, c_U]$ for μ is

$$\begin{aligned} \bar{x} \pm t_{n-1,\alpha/2}\sqrt{\frac{s^2}{n}} &= 107.3 \pm (2.145)\sqrt{\frac{32.5}{15}} \\ &= 107.3 \pm (2.145)(2.1667) \\ &= [104.143,\ 110.457]. \end{aligned}$$

   (b) Since 113 is outside the confidence interval, this value is not plausible or "likely" for the true mean IQ. Therefore, we would not be happy with the parent's claim.
   Formally: set up the hypotheses

$$\begin{aligned} H_0 &: \mu = 113 \\ H_A &: \mu \neq 113 \end{aligned}$$

Since $\mu = 113$ is outside the 95% confidence interval in (a), H_0 is rejected at the significance level of $100 - 95 = 5\%$.

4. Denote the true proportion not satisfied with Council services as π.

   (a) A 99% confidence interval for π is based on the approximate sampling distribution for the sample proportion P:

   $$\frac{P - \pi}{\sqrt{\frac{P(1-P)}{n}}} \sim N(0,1), \qquad \text{approximately.}$$

   For a 99% confidence interval we need the percentage point $z_{\alpha/2} = z_{0.005}$ such that

   $$\Pr(Z > z_{0.005}) = 0.005.$$

   From tables,

   $$z_{0.005} = 2.576.$$

   The 99% confidence interval $[c_L, c_U]$ for π is then

   $$\begin{aligned} p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} &= 0.34 \pm (2.576)\sqrt{\frac{0.34(1-0.34)}{100}} \\ &= 0.34 \pm (2.576)(0.04737) \\ &= [0.218,\ 0.462]. \end{aligned}$$

   (b) Council claims all but 20% are satisfied with services & hence that true dissatisfaction is $\pi = 0.2$. Test this claim through hypotheses

   $$\begin{aligned} H_0 &: \pi = 0.2 \\ H_A &: \pi \neq 0.2 \end{aligned}$$

   Since is $\pi = 0.2$ is outside the 99% confidence interval in (a), the Council's claim is rejected at the 1% level of signfiicance.

5. 72.40 is the claim about the population mean μ, so the null hypothesis is

   $$H_0 : \mu = 72.40.$$

   Since $X \sim N(\mu, 2.1^2)$, then if H_0 is true,

   $$\bar{X} \sim N\left(72.40, \frac{4.41}{35}\right)$$

   and the sample mean $\overline{x} = 73.2$ yields

   $$z = \frac{73.2 - 72.4}{\sqrt{4.41/35}} = 2.25.$$

(a) Consider first the two sided alternative hypothesis

$$H_A : \mu \neq 72.40.$$

For this two-sided alternative, the p-value is probability of obtaining a sample statistic *at least as extreme* as $\pm z = \pm 2.25$. From the standard normal distribution

$$\begin{aligned} p &= \Pr[|Z| \geq 2.25] \\ &= 2 \times 0.012 = 0.024. \end{aligned}$$

For the upper one-sided alternative

$$H_A : \mu > 72.40$$

the p-value is probability of obtaining a sample statistic *at least as extreme* as +2.25. [Only the upper direction is relevant under H_A.] Hence for this alternative

$$\begin{aligned} p &= \Pr[Z \geq 2.25] \\ &= 0.012. \end{aligned}$$

(b) Both p-values are relatively small, indicating that the sample mean $\overline{x} = 73.2$ provides little support for $H_0 : \mu = 72.40$ against either alternative hypothesis.

(c) For performing a classical test at the 5% level of significance, H_0 will be rejected against the relevant H_A when $p < 0.05$. For the two-sided H_A, we have $p = 0.024 < 0.05$ while for the one-sided H_A, $p = 0.012 < 0.05$. Therefore, the null hypothesis is rejected at the 5% level of significance against either alternative hypthesis.
Question 2, Exercise Sheet 7 comes to the same conclusion for the two-sided alternative hypothesis.