Statistics II

Problem sheet 7 - Solutions

1. i) If T is the lifetime of one light-bulb, with $T \sim Exp(2)$ then $E(T) = \frac{1}{2}$, $var(T) = \frac{1}{4}$.

For a random sample of 36 light-bulbs:

$$E\left(\sum_{i=1}^{36} T_i\right) = \sum_{i=1}^{36} E(T_i) = 36 \times \frac{1}{2} = \underline{18}$$

$$var\left(\sum_{i=1}^{36} T_i\right) = \sum_{i=1}^{36} var(T_i) \quad \text{(because of independence)}$$

$$= 36 \times \frac{1}{4} = \underline{\underline{9}}$$

ii) Similarly

$$E(\overline{T}) = \frac{1}{36} E\left(\sum_{i=1}^{36} T_i\right) = \underline{\frac{1}{2}}$$

$$var(\overline{T}) = \frac{1}{36^2} var\left(\sum_{i=1}^{36} T_i\right) = \frac{1}{36^2} \frac{36}{4} = \underline{\frac{1}{144}}$$

iii) From CLT we have that (approximately)

$$\overline{T} \sim N\left(\frac{1}{2}, \frac{1}{144}\right)$$

Then:

$$P(\overline{T} > 0.7) = P\left(\frac{\overline{T} - 0.5}{1/12} > \frac{0.7 - 0.5}{1/12}\right)$$

$$= P(Z > 2.4), \quad \text{where } Z \sim N(0,1)$$

$$= 1 - \Phi(2.4) = \underline{\underline{0.0082}}.$$

2. Let X be weight of a bag. Then $X \sim N(1, 0.01^2)$

Now weight of 100 bags is $\sum_{i=1}^{100} X_i \sim N(100, 100 \times 0.01^2)$

i.e. $N(100, 0.01)$.

i)

$\Rightarrow E(\sum X_i) = 100$, $\quad S.D.(\sum X_i) = \sqrt{0.01} = 0.1$

ii)

$$P\left(\sum_{i=1}^{100} X_i > 100.1\right) = P\left(\frac{\sum X_i - 100}{0.1} > \frac{100.1 - 100}{0.1}\right)$$

$$= P(Z > 1) = 1 - \Phi(1)$$

$$= 0.1587$$

3. i) $X \sim Bin(100, 0.3)$

ii) We can approximate the distribution of X with a normal distribution using the CLT.

$$X \overset{\cdot}{\sim} N(100 \times 0.3, 100 \times 0.3 \times 0.7) = N(30, 21)$$

Now $P(X > 36) = P(X \geq 37) = P(X > 36.5)$

using continuity correction.

$$P(X > 36.5) = P\left(\frac{X - 30}{\sqrt{21}} > \frac{36.5 - 30}{\sqrt{21}}\right)$$

$$= P(Z > 1.42) \quad \text{where } Z \sim N(0, 1)$$

$$= 1 - \Phi(1.42) = 1 - 0.9222 = 0.0888$$

4\. i) Sample mean $\bar{x} = \frac{1}{10}\sum x_i = 1.61$ kg.

Variance is known. Population s.d. $= \sqrt{\sigma^2} = 0.5$

$\Rightarrow$ 90% CI is $\left(1.61 - z.\frac{0.5}{\sqrt{10}},\ 1.61 + z.\frac{0.5}{\sqrt{10}}\right)$

where z is the 5% pt of the $N(0,1)$, i.e. $z = 1.64$.

This gives $(1.35, 1.87)$ as the 90% CI for μ.

ii) Now assume σ is not known.

Sample variance, $s^2 = \frac{1}{9}\left(\sum x^2 - \frac{(\sum x)^2}{10}\right)$

$= \frac{1}{9}\left(28.11 - \frac{(16.1)^2}{10}\right) = 0.243$

$\Rightarrow s = \sqrt{0.243} = 0.49.$

For this case, the 90% CI is

$$\left(\bar{x} - t.\frac{0.49}{\sqrt{10}},\ \bar{x} + t.\frac{0.49}{\sqrt{10}}\right)$$

where t is the 5% point of the t-distribution on 9 degrees of freedom. i.e. $t = 1.833$

This gives $(1.33, 1.89)$ as the 90% CI for μ.

iii) If we calculate a large number of such CIs (each time using a different sample), we expect 90% of them to contain the true value of the unknown parameter.

5. From Tut. 6 Qu. 6. we have

$$\bar{x} = 36.33 \quad , \quad s = 13.44 \, ,$$

giving $\left(\bar{x} - t\frac{s}{\sqrt{54}}, \ \bar{x} + t\frac{s}{\sqrt{54}}\right)$ as the 95% CI

where $t = t_{53}(2.5) \approx 2$

$\Rightarrow$ 95% CI is $(32.67, 39.99)$

This is approximate because the stem-and-leaf diagram showed that the data were not normal.

6. We have $\bar{x}_A = 418, \quad \bar{x}_B = 402$

$\sigma_A = 26 \quad , \quad \sigma_B = 22$ (variances known)

Sample sizes $n_A = 40, \quad n_B = 50$

(i)
Therefore an 95% CI for $\mu_A - \mu_B$

$$\left(\bar{x}_A - \bar{x}_B - 1.96\sqrt{\frac{26^2}{40} + \frac{22^2}{50}}, \ \bar{x}_A - \bar{x}_B + 1.96\sqrt{\frac{26^2}{40} + \frac{22^2}{50}}\right)$$

$$= (5.9, 26.1)$$

(ii)
0 does not lie in the confidence interval. Both limits are > 0 $\Rightarrow$ it appears that $\mu_A > \mu_B$. On average bulbs of type A have the longer lifetime.

7. In this question the population variances are unknown. Assuming equal variance for the two populations we calculate the pooled estimate of variance as

(i) $$s_p^2 = \frac{1}{(n_1+n_2-2)}\left((n_1-1)s_1^2 + (n_2-1)s_2^2\right)$$

where $n_1 = 12$, $n_2 = 15$, $s_1^2 = 1.2^2$, $s_2^2 = 1.5^2$.

$\Rightarrow \quad s_p^2 = 1.8936 \quad \Rightarrow \quad s_p = 1.38$

(ii)

We then obtain a 95% CI for $\mu_1 - \mu_2$ as

$$\left(\bar{x}_1 - \bar{x}_2 - t_{25}(2.5)\, s_p\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)},\ \bar{x}_1 - \bar{x}_2 + t_{25}(2.5)\, s_p\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right)$$

$$= \quad (-0.2,\ 2.0)$$

The value 0 lies in the CI so that the evidence of a difference in population means is <u>not</u> <u>strong</u>.

8. For Mine A we have $\sum x_i = 41300$

$$\sum x_i^2 = 341,391,800$$

For Mine B:

$$\sum x_i = 39650$$
$$\sum x_i^2 = 314,595,100$$

Therefore can calculate sample means and variances as.

Mine A: $\bar{x}_A = 8260$, $s_A^2 = \frac{1}{4}\left(\sum x_i^2 - \frac{(\sum x_i)^2}{5}\right)$

$$= 63450$$

Mine B: $\bar{x}_B = 7930$, $s_B^2 = 42650$

This gives a pooled sample variance.

$$s_p^2 = \frac{1}{8}\left(4s_A^2 + 4s_B^2\right) = 53050$$

$$\Rightarrow s_p = \sqrt{53050} = 230.3$$

To get 99% CI for $\mu_A - \mu_B$ we take the interval

$$\left(\bar{x}_A - \bar{x}_B - t \cdot s_p\sqrt{\frac{1}{5}+\frac{1}{5}},\ \bar{x}_A - \bar{x}_B + t \cdot s_p\sqrt{\frac{1}{5}+\frac{1}{5}}\right)$$

where t is the 0.5% pt. of the t_8 distribution i.e. $t = 3.355$

$\Rightarrow$ CI is $(-158.67, 818.67)$.

This contains the value 0 so that there is not strong evidence of a difference in the coal from the 2 mines.