Test answers - Statistical inference

Year: University Year 1 · Duration: 60 minutes · /20

Answer to Question 1 (4 marks)

a. The sample is large ($n = 50$), so the normal approximation can be used with s in place of σ. The standard error is $\frac{4.5}{\sqrt{50}} = 0.636$ (1). The interval is $31.2 \pm 1.96 \times 0.636 = 31.2 \pm 1.25$ (1), that is $(29.95,\ 32.45)$ minutes (1).

b. If the sampling procedure were repeated many times, about 95% of the intervals constructed in this way would contain the true mean commuting time (1). It does not mean that μ has a 95% probability of lying in this particular interval.

Answer to Question 2 (4 marks)

Since σ is known, use a z-test. The standard error is $\frac{10}{\sqrt{40}} = 1.58$, so $z = \frac{78.5 - 75}{1.58} = \frac{3.5}{10/\sqrt{40}} = 2.21$ (2). The alternative is two-sided, so $p = 2(1 - \Phi(2.21)) \approx 0.027$, which is less than 0.05 (1). Equivalently, $2.21 > 1.96$. We reject H_0: at the 5% level there is evidence that the school's mean mark differs from 75, and the sample suggests it is higher (1).

A 95% confidence interval, $78.5 \pm 1.96 \times 1.58 = (75.4,\ 81.6)$, excludes 75, which agrees with rejecting H_0.

Answer to Question 3 (4 marks)

σ is unknown and the sample is small, so use a t-test. The estimated standard error is $\frac{1.2}{\sqrt{9}} = 0.4$, so $t = \frac{10.8 - 10}{0.4} = \frac{0.8}{1.2/3} = 2.0$ (2). There are $9 - 1 = 8$ degrees of freedom and the test is one-sided, so the critical value is $t_{8,\,0.95} = 1.860$ (1). Since $2.0 > 1.860$, the statistic is in the rejection region: we reject H_0 and conclude that there is evidence at the 5% level that $\mu > 10$ (1).

Answer to Question 4 (4 marks)

a. Each expected frequency is (row total $\times$ column total) / grand total. The row totals are 60 and 60 and the column totals 70 and 50, so the expected frequencies are $\frac{60 \times 70}{120} = 35$ and $\frac{60 \times 50}{120} = 25$ in each row: $35,\ 25,\ 35,\ 25$ (1).

b. Every $|O - E| = 5$, so $\chi^2 = \frac{25}{35} + \frac{25}{25} + \frac{25}{35} + \frac{25}{25} \approx 3.43$ (2). A 2×2 table has

$(2-1)(2-1)=1$ degree of freedom, with critical value 3.841. Since $3.43 < 3.841$, we do not reject independence: the data do not show a significant link between attending and passing (1).

Answer to Question 5 (4 marks)

a. The gradient is $b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{63}{42} = 1.5$ (1). The means are $\bar{x} = \dfrac{40}{8} = 5$ and $\bar{y} = \dfrac{96}{8} = 12$, and the line passes through $(\bar{x}, \bar{y})$, so $a = 12 - 1.5 \times 5 = 4.5$. The regression line is $y = 4.5 + 1.5x$ (1).

b. At $x = 6$: $y = 4.5 + 1.5 \times 6 = 13.5$ (1). This is interpolation, so it is reliable. At $x = 20$, far outside the observed range 1 to 9, the prediction would be extrapolation: the linear pattern may not continue, so it is unreliable (1).