Basic Statistics and Probability (2018-19)
Science & Engineering Program Boston University-Faculty of Science UAM

FINAL EXAM. SOLUTIONS

---

1. Let the mean ionized calcium concentration in blood platelets of a normotensive person be μ_1 and the mean ionized calcium concentration in blood platelets of a hypertensive person be μ_2. Then $n_1 = 39$, $n_2 = 44$, $\bar{x} = 108.1$, $\bar{y} = 168.3$, $s_1 = 17.1$ and $s_2 = 25.6$. Since both sample sizes n_1 and n_2 are "large" (> 20), there is no need to assume normality.

a) $\text{CI}_{90\%}(\mu_1) = \left(108.1 \mp 1.645\dfrac{17.1}{\sqrt{39}}\right) = (108.1 \mp 4.5)$

b) $\text{CI}_{95\%}(\mu_1 - \mu_2) = \left(108.1 - 168.3 \mp 1.96\sqrt{\dfrac{17.1^2}{39} + \dfrac{25.6^2}{44}}\right) = (-60.2 \mp 9.3)$

c) At the significance level $\alpha = .05$, we want to test $H_0 : \mu_1 - \mu_2 = 0$ vs $H_1 : \mu_1 - \mu_2 \neq 0$. The rejection region of this test is $R = \{|z| > z_{\alpha/2}\}$, where $z_{\alpha/2} = z_{0.025} = 1.96$ and the test statistic is

$$z = \frac{108.1 - 168.3}{\sqrt{\frac{17.1^2}{39} + \frac{25.6^2}{44}}} = -12.72.$$

Since $|z| = 12.72 > z_{\alpha/2} = 1.96$, there is enough sample evidence to reject H_0 at the 5% significance level. The mean ionized calcium concentration is different in the two groups.

---

2. a) Bernoulli experiment = {
Success: A diseased tree improves its condition after receiving the new treatment.
Failure A diseased tree does not improve its condition after receiving the new treatment.

The parameter of interest is p, the real proportion of diseased trees improving their condition when receiving the new treatment. A point estimate of p is $\hat{p} = 66/84 = 0.7857$.

b) We want to test $H_0 : p \leq 0.7$ vs $H_1 : p > 0.7$ at the level $\alpha = 0.01$. The rejection region for this test is $R = \{z > z_\alpha\}$, where

$$z = \frac{0.7857 - 0.7}{\sqrt{\frac{0.7\,0.3}{84}}} = 1.714 \quad \text{and} \quad z_\alpha = z_{0.01} = 2.331.$$

So there is no sample evidence to reject H_0 at the 1% significance level.

c) The p-value is the lowest significance level α for which we reject the null hypothesis. Since we reject $H_0 : p \leq 0.7$ at $\alpha = 0.01$, we conclude that the p-value is lower than 1%.

d) The true proportion of trees improving their condition after the treatment is $p = 0.6$. We apply the treatment to $n = 5$ diseased trees. The probability distribution of X = "the number of trees of those 5 that improve their condition after being treated" is binomial B($n = 5$,$p = 0.6$). The probability that 3 or less of the 5 trees will improve is

$$P\{X \leq 3\} = 0.663 \quad \text{(we have used the binomial table).}$$

The probability that 3 or more of the 5 trees will improve is

$$P\{X \geq 3\} = 1 - P\{X \leq 2\} = 1 - 0.317 = 0.683 \quad \text{(we have used the binomial table).}$$

---

3. a) The z-score of X is $Z = \dfrac{X-7}{10}$, which follows a $N(0,1)$ distribution.

b)

$$\begin{aligned} P\{X > 11\} &= P\left\{\frac{X-7}{10} > \frac{11-7}{10}\right\} = P\{Z > 0.4\} \\ &= 0.5 - P\{0 < Z < 0.4\} = 0.5 - 0.1554 = 0.3446. \end{aligned}$$

c)

$$\begin{aligned} P\{2 < X < 12\} &= P\left\{\frac{2-7}{10} < \frac{X-7}{10} < \frac{12-7}{10}\right\} = P\{-0.5 < z < 0.5\} \\ &= 2\,P\{0 < Z < 0.5\} = 2 \cdot 0.1915 = 0.383 \end{aligned}$$

d)

$$0.3413 = P\{7 \le X \le c\} = P\left\{\frac{7-7}{10} < \frac{X-7}{10} < \frac{c-7}{10}\right\} = P\left\{0 < Z < \frac{c-7}{10}\right\}$$

Since $P\{0 < Z < 1\} = 0.3413$ (see the normal table), we deduce that $(c-7)/10 = 1$, that is, $c = 17$.

---

4. a) The probability of high inflation is $0.16 + 0.24 = 0.4$. The probability of high unemployment is $0.16 + 0.36 = 0.52$.

b) The probability of an scenario with simultaneously high inflation **and** high unemployment is $P(\text{high inflation} \cap \text{high unemployment}) = 0.16$. High inflation and high unemployment are **dependent** events. They would be independent if and only if $P(\text{high inflation} \cap \text{high unemployment}) = P(\text{high inflation}) \cdot P(\text{high unemployment})$, but this equality does not hold.

c) The probability of high inflation conditional on unemployment being high is

$$P\{\text{high inflation}|\text{high unemployment}\} = \frac{\text{high inflation} \cap \text{high unemployment}}{\text{high unemployment}} = \frac{0.16}{0.52} = 0.3077.$$

d) The probability of an scenario with high inflation **or** high unemployment is

$$\begin{aligned} &P(\text{high inflation} \cup \text{high unemployment}) \\ &\quad = P(\text{high inflation}) + P(\text{high unemployment}) - P(\text{high inflation} \cap \text{high unemployment}) \\ &\quad = 0.4 + 0.52 - 0.16 = 0.76. \end{aligned}$$

e) Events "high inflation" and "high unemployment" would be mutually exclusive if their intersection were empty: $\{\text{high inflation}\} \cap \{\text{high unemployment}\} = \emptyset$. However, in (b) we saw that $P(\text{high inflation} \cap \text{high unemployment}) = 0.16$, so their intersection is not empty.

---

5. a) The variable is the weight (in g) of a box of raisins. It is quantitative.

b) $\bar{x} = 320/10 = 32$ g

c) Median $M = 32$ (the thick middle line in the boxplot).

d) Q_L = Left side of the box = 29, Q_U = Right side of the box = 35

e) The sample is symmetric because the boxplot is symmetric with respect to the median.

f) If I substitute the value 38 by 100, the sample mean is $(320 - 38 + 100)/10 = 38.2$. Since the sample size is even, the sample median is the average of the two most central observations in the arranged sample and it does not change when substituting the largest observation (38) by 100.
