Math 171 Prelim #2 – Solutions

1. (a)

$$E\left(\overline{X}\right) = E\left(\frac{X_1 + \cdots + X_{16}}{16}\right) = \frac{E(X_1) + \cdots + E(X_{16})}{16} = \frac{16 \cdot 17}{16} = 17$$

and since $X_1, \ldots, X_{16}$ are a SRS

$$\mathrm{Var}\left(\overline{X}\right) = \mathrm{Var}\left(\frac{X_1 + \cdots + X_{16}}{16}\right) = \frac{\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_{16})}{16^2} = \frac{16 \cdot 12^2}{16^2} = \frac{12^2}{16}.$$

Thus,

$$\mathrm{SD}\left(\overline{X}\right) = \frac{12}{4} = 3.$$

(b)

$$E\left(\overline{Y}\right) = E\left(\frac{Y_1 + \cdots + Y_{25}}{25}\right) = \frac{E(Y_1) + \cdots + E(Y_{25})}{25} = \frac{25 \cdot 18}{25} = 18$$

and since $Y_1, \ldots, Y_{25}$ are a SRS

$$\mathrm{Var}\left(\overline{Y}\right) = \mathrm{Var}\left(\frac{Y_1 + \cdots + Y_{25}}{25}\right) = \frac{\mathrm{Var}(Y_1) + \cdots + \mathrm{Var}(Y_{25})}{25^2} = \frac{25 \cdot 20^2}{25^2} = \frac{20^2}{25}.$$

Thus,

$$\mathrm{SD}\left(\overline{X}\right) = \frac{20}{5} = 4.$$

(c)

$$E\left(\overline{X} - \overline{Y}\right) = E\left(\overline{X}\right) - E\left(\overline{Y}\right) = 17 - 18 = -1$$

and since the SRS are done independently

$$\mathrm{SD}\left(\overline{X} - \overline{Y}\right) = \sqrt{\mathrm{Var}\left(\overline{X}\right) + \mathrm{Var}\left(\overline{Y}\right)} = \sqrt{3^2 + 4^2} = 5.$$

2. (a) Let X be the weight of the apple so that X is $\mathcal{N}(5, 0.3)$, and let Y be the weight of the orange so that Y is $\mathcal{N}(6, 0.4)$. Therefore, the probability that the apple is heavier than the orange is $P(X > Y)$. It is reasonable to assume that the apple weight and the orange weight are independent so that

$$\begin{aligned} P(X > Y) = P(X - Y > 0) = P(W > 0) &= P\left(\frac{W - (-1)}{0.5} > \frac{0 - (-1)}{0.5}\right) \\ &= P(Z > 2) = 0.0228 \end{aligned}$$

where $W = X - Y$ has a $\mathcal{N}(5 - 6, \sqrt{0.3^2 + 0.4^2}) = \mathcal{N}(-1, 0.5)$ distribution and Z has a $\mathcal{N}(0, 1)$ distribution.

(b) Let $\overline{X}$ be the average weight of the apples so that $\overline{X}$ is $\mathcal{N}(5, 0.3/\sqrt{3})$, and let $\overline{Y}$ be the average weight of the oranges so that $\overline{Y}$ is $\mathcal{N}(6, 0.4/\sqrt{4})$. (This is analogous to problem 1.) Therefore, the probability that the apple average is heavier than the orange average is $P(\overline{X} > \overline{Y})$. It is reasonable to assume that the apple weights and the orange weights are independent so that

$$\begin{aligned} P(\overline{X} > \overline{Y}) = P(\overline{X} - \overline{Y} > 0) = P(\overline{W} > 0) &= P\left(\frac{\overline{W} - (-1)}{\sqrt{0.07}} > \frac{0 - (-1)}{\sqrt{0.07}}\right) \\ &= P(Z > 3.78) = 0.0001 \end{aligned}$$

where $\overline{W} = \overline{X} - \overline{Y}$ has a $\mathcal{N}\left(5 - 6, \sqrt{(0.3/\sqrt{3})^2 + (0.4/\sqrt{4})^2}\right) = \mathcal{N}(-1, \sqrt{0.07})$ distribution and Z has a $\mathcal{N}(0, 1)$ distribution.

3. (a) Begin by numbering the volunteers from left to right as 0 to 9. Reading the row of random digits from left to right, we assign to the treatment group the first 5 unrepeated numbers that we encounter, namely 9,6,7,3,5, or

treatment group = { Isaiah, Francine, Gene, Charles, Edward }
placebo group = { Abraham, Beth, Danielle, Holly, Jessica }

(b) In order for this experiment to be double-blind it will be necessary for neither the patients, nor the researchers working with them to know who is in the treatment group and who is in the control group. Therefore, suppose that there are two researchers: one named Karl and one named Lana. As in **(a)** Karl will decide who is in the treatment group and who is in the placebo group, and obviously he will not reveal to the volunteers which group each is in. He will prepare the daily doses of calcium and placebo (which should look identical) and label them. He will then give them to Lana. Lana will administer the daily doses and record the results. She will then give this information to Karl for analysis. Since neither Lana nor the volunteers know which patients are in the control group, this experiment is double-blind.

4. (a) The correlation between age and length is

$$\begin{aligned} r &= \frac{1}{n-1} \sum \left(\frac{x_i - \overline{x}}{s_x}\right)\left(\frac{y_i - \overline{y}}{s_y}\right) = \frac{1}{(n-1)\, s_x\, s_y} \sum (x_i - \overline{x})(y_i - \overline{y}) \\ &= \frac{1}{(51-1) \cdot (1.5) \cdot (2.2)} \cdot (121) = 0.733. \end{aligned}$$

(b) The equation of the regression line is

$$\hat{y} = a + bx$$

where

$$b = r\frac{s_y}{s_x} \quad \text{and} \quad a = \overline{y} - b\overline{x}.$$

Thus,

$$b = \frac{(0.733)(2.2)}{1.5} = 1.076 \quad \text{and} \quad a = 23 - (1.076)(24) = -2.813$$

so that

$$\hat{y} = -2.813 + 1.076x.$$

(c) If $x = 28.3$, then $\hat{y} = -2.813 + (1.076)(28.3) = 27.6$, so that the predicted age of this worm is 27.6 months old.

5. (a) It is easily found that $\overline{x} = 37.4$. A 95% confidence interval when σ is known has corresponding critical value $z^* = 1.96$, and is of the form $\overline{x} \pm z^*\sigma/\sqrt{n}$. Thus, the required CI is

$$37.4 \pm 1.96 \cdot \frac{2.6}{\sqrt{10}} = (35.79, 39.01).$$

(b) We take an answer directly from page 333 of the textbook: If we use 95% confidence intervals often, then in the long run, 95% of these intervals will contain the true mean. In the context of this problem, if the experiment of collecting data from 10 owners of the *Elmira* and using that sample mean to produce a confidence interval is repeated many, many times, then we expect that in the long run 95% of these intervals will contain the true, but UNKNOWN, mean.

(c) Let H_0 be the null hypothesis that the new fuel injection of the *Elmira* has no effect on gas mileage; that is, $H_0 : \mu = 33.4$. Let H_a be the alternative hypothesis that the new fuel injection of the *Elmira* has some effect on gas mileage; that is, $H_a : \mu \neq 33.4$.

(d) Since 33.4 does NOT lie in the interval $(35.79, 39.01)$, the "confidence interval and two-sided hypothesis testing duality" (see page 357) tells us that we CAN reject H_0 at the $\alpha = 1 - 95\% = 5\%$ level.

(e) It we want to have 98% confidence and a margin of error of only 0.5, then we need at least

$$\left(z^* \frac{\sigma}{m}\right)^2$$

surveys where z^* is the critical value from Table A corresponding to 98%; that is, $z^* = 2.33$. Hence, we need $n = 13$ surveys. *(Note that n MUST be a whole number, and that we cannot round 12.116 down to 12.)*

6. (a) The probability of a Type I error is

$$P(\text{reject } H_0 \mid \mu = 4) = P(\overline{x} < 3.6 \mid \mu = 4) = P\left(\frac{\overline{x} - 4}{0.5/\sqrt{4}} < \frac{3.6 - 4}{0.5/\sqrt{4}}\right)$$
$$= P(Z < -1.6) = 0.0548.$$

(b) The probability of a Type II error for the alternative $\mu = 3.5$ is

$$P(\text{fail to reject } H_0 \mid \mu = 3.5) = P(\overline{x} \geq 3.6 \mid \mu = 3.5) = P\left(\frac{\overline{x} - 3.5}{0.5/\sqrt{4}} \geq \frac{3.6 - 3.5}{0.5/\sqrt{4}}\right)$$
$$= P(Z \geq 0.4) = 1 - 0.6554 = 0.3446.$$

(c) Since our alternative is one-sided, we find from Table A that if $z^* = -2.33$, then $P(Z < z^*) = 0.01$. Thus, our rejection rule is to reject H_0 if $\frac{\overline{x}-4}{0.5/\sqrt{4}} < -2.33$, or

$$\text{reject } H_0 \text{ if } \overline{x} < 3.4175.$$

(d) With the rejection rule from **(c)**, we have the probability of a Type II error for the alternative $\mu = 3.5$ is

$$P(\text{fail to reject } H_0 \mid \mu = 3.5) = P(\overline{x} \geq 3.4175 \mid \mu = 3.5) = P\left(\frac{\overline{x} - 3.5}{0.5/\sqrt{4}} \geq \frac{3.4175 - 3.5}{0.5/\sqrt{4}}\right)$$
$$= P(Z \geq -0.33) = 0.6293.$$

(e) The solution to this problem depends on your personal opinion to some extent. One acceptable answer is: We should be more concerned about a Type I error becasue if we commit a Type I error, then we would be wrongly accusing the Pizzeria of a crime they did not commit. However, there are other reasonable answers, and any *good, well-thought* explanation is sufficient.