"I have pledged my honor that I have abided by the Stevens Honor System."
Cindy Zhang

HOMEWORK 8

MA 222 Probability and Statistics

Problem One. Suppose the sediment density of randomly selected specimen from a certain region is normally distributed with the given:

- Mean = 2.65
- Standard Deviation = 0.85

Suggested in "Modeling Sediment and Water Column Interactions for Hydrophobic Pollutants," Water Research 1984: 1169-1174)

a. If a random sample of 25 specimen in selected, what is the probability that the sample average sediment density is at most 3.00? Between 2.65 and 3.00?

Mean (μ) $= 2.65$

Standard Deviation sub x (σ_x) $= \frac{1}{\sqrt{n}}(\sigma) = \frac{1}{\sqrt{25}}(0.85) = 0.17$

$P(X \leq 3) = P(\frac{X - \mu_x}{\sigma_x} \leq \frac{3-2.65}{0.17}) = P(Z \leq 2.06) = 0.9803$

$P(2.65 \leq X \leq 3) = P(X \leq 3) - P(X \leq 2.65)$

$= P(Z \leq 2.06) - P(Z \leq 0)$

$= 0.4803$

b. How large a sample size would be required to ensure that the first probability in part(a) is at least 0.99?

$P(X \leq 3) = 0.99 = P(\frac{X - \mu_x}{\sigma_x} \leq \frac{3-2.65}{\frac{0.85}{\sqrt{n}}}) = P(Z \leq \frac{0.35}{\frac{0.85}{\sqrt{n}}}) = 0.99$

$$\frac{0.35}{\frac{0.85}{\sqrt{n}}} = 2.33$$

$n = 32.02 \rightarrow 33$

Problem Two. A binary communication channel transmits a sequence of "bits" (1s and 0s). Suppose that for any particular bit transmitted, there is a 10% change of a transmission error (a 0 becoming 1 or a 1 becoming a 0). Assume that bit error occurs independently of one another.

a. Consider transmitting 1000 bits. What is the approximate probability that at most 125 transmission errors occur?

   Given:

   - $p = 0.1$
   - $n = 1000$

   Normal Approximation of binomial distribution needs:

   - $np \geq 5 \rightarrow np = (1000)(0.1) = 100 \geq 5$, TRUE
   - $nq \geq 5$, where q is $(1 - p) \rightarrow nq = (1000)(1 - 0.1) = 900 \geq 5$, TRUE

   Since this is true, we can approximate by normal distribution:

$$z = \frac{x - np}{\sqrt{np(1-p)}} = \frac{125.5 - 1000(0.1)}{\sqrt{1000(0.1)(1 - 0.1)}} = 2.69$$

$$P(X \leq 125) = P(X < 125.5) = P(Z < 2.69) = 0.9964$$

b. Suppose the same 1000-bit message is sent two different times independently of one another. What is the approximate probability that the number of errors in the first transmission is within 50 of number of errors in the second?

$$\mu = np = (1000)(0.1) = 100$$

$$\sigma = \sqrt{npq} = \sqrt{np(1-p)} = \sqrt{(1000)(0.1)(1 - 0.1)} = \sqrt{90} = 9.4868$$

$$\sigma_{X-Y} = \sqrt{\sigma_X^2 + \sigma_Y^2} = \sqrt{(\sqrt{90})^2 + (\sqrt{90})^2} = \sqrt{90+90} = \sqrt{180} = 13.142$$

$$z = \frac{x - \mu_{X-Y}}{\sigma_{X-Y}} = \frac{\pm 50.5 - 0}{13.142} = \pm 3.76$$

$$P(|X-Y| \leq 50) = P(|X-Y| < 50.5) = P(-50.5 < X-Y < 50.5)$$
$$= P(-3.76 < Z < 3.76) = P(Z < 3.76) - P(Z < -3.76) = 1 - 0 = 1$$

   Or $P(|X - Y| \leq 50) =$ 0.9998

Problem Three. Garbage trucks entering a particular waste-management facility are weighed prior to offloading their contents. Let X = the total processing time for a randomly selected truck at this facility (waiting, weighing, and offloading). The article "Estimating Waste Transfer Station Delays Using GPS" (Waste Mgmt., 2008: 1742-1750) suggests the plausibility of a normal distribution with mean 13 min and standard deviation 4 min for X. Assume that this is in fact the correct distribution.

a. What is the probability that a single truck's processing time is between 12 and 15 minutes?

Mean = 13
Std = 4

$$z = \frac{\text{x - mean}}{\text{std}} = \frac{12 - 13}{4} = -0.25$$

$$z = \frac{\text{x - mean}}{\text{std}} = \frac{15 - 13}{4} = 0.50$$

$$P(12 < X < 15) = P(-0.25 < Z < 0.50) = P(Z < 0.50) - P(Z < -0.25)$$
$$= 0.6915 - 0.4013 = 0.2902$$

b. Consider a random sample of 16 trucks. What is the probability that the sample mean processing time is between 12 and 15 minutes?

Mean = 13
Std = 4
Number = 16

New Std = std / sqrt (number) = 4 / sqrt (16) = 4 / 4 = 1

$$z = \frac{\text{x - mean}}{\text{new std}} = \frac{12 - 13}{1} = -1$$

$$z = \frac{\text{x - mean}}{\text{new std}} = \frac{15 - 13}{1} = 2$$

$$P(12 < X < 15) = P(-1 < Z < 2) = P(Z < 2) - P(Z < -1) = 0.9772 - 0.1587$$
$$= 0.8185$$

c. Why is the probability in (b) much larger than the probability in (a)?

The sample mean of 16 trucks is less than a single truck meaning it is more concentrated.

d. What is the probability that the sample mean processing time for a random sample of 16 trucks will be at least 20 minutes?

$z = \frac{\text{x - mean}}{\text{std}} = \frac{20 - 13}{1} = 7$ → P(X > 20) = P(Z > 7) = 1 – P(Z < 7) = 1 – 1 = 0

"I have pledged my honor that I have abided by the Stevens Honor System."
Cindy Zhang

Problem Four. Suppose the expected tensile strength of type-A steel is 105 ksi and the standard deviation of tensile strength is 8 ksi. For type-B steel, suppose the expected tensile streagth and standard deviation of average tensile strength are 100 ksi, respectively. Let X be the sample average tensile strength of a random sample of 40 type-A specimens and let Y be the sample average tensile strength of a random sample 35 type-B specimens.

a. What is the approximate distribution of X? Of Y?

$\mu_X = 105, \quad \mu_Y = 100, \quad \sigma_X = 8, \quad \sigma_Y = 6, \quad n_X = 40, \quad n_Y = 35$

$$\sigma_x = \frac{\sigma_X}{\sqrt{n}} = \frac{8}{\sqrt{40}} = \frac{2\sqrt{10}}{5} = 1.265$$

$$\sigma_y = \frac{\sigma_Y}{\sqrt{n}} = \frac{6}{\sqrt{35}} = \frac{6\sqrt{35}}{35} = 1.014$$

b. What is the approximate distribution of X – Y? Justify your answer.

Linear Combination: $W = aX_1 + bX_2$

$$\mu w = a\mu 1 + b\mu 2$$

$\sigma_W^2 = a^2\sigma_1^2 + b^2\sigma_2^2$

$\sigma_W = \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2}$(If X_1 and X_2 are independent.)

Let's assume they are independent:

$mu_x - mu_y = 105\text{-}100 = 5$

$\sigma_{x\text{-}y} = \sqrt{1.265^2 + 1.014^2} = 1.621$

c. Calculate approximately $P(-1 \le X - Y \le 1)$.

$$z = \frac{x \text{ - mean}}{\text{std}} = \frac{1 - 5}{1.621} = -2.47$$

$$z = \frac{x \text{ - mean}}{\text{std}} = \frac{-1\text{-}5}{1.621} = -3.70$$

$$P(-1 \le X\text{-}Y \le 1) = P(-3.70 < Z < -2.47) = P(Z < -2.47) - P(Z < -3.70)$$
$$= 0.0068 - 0 = 0.0068$$

d. Calculate $P(X - Y \ge 10)$. If you actually observed $X - Y \ge 10$. would you doubt that $\mu_1 - \mu_2 = 5$? → Yes, the probability is very small given below:

$$z = \frac{x \text{ - mean}}{\text{std}} = \frac{10 - 5}{1.621} = 3.08, P(X\text{-}Y \ge 10) = P(Z > 3.08) = 1\text{-}P(Z < 3.08)$$
$$= 1\text{-}0.9990 = 0.0010$$

Problem Five. Two cars with six-cylinder engines and three with four-cylinder engines are to e driven over a 300-mile course. Let X_1,.X_5 denote the resulting fuel efficiencies (mpg). Consider the linear combination.

$$Y = \frac{1}{2}X_1 + \frac{1}{2}X_2 - \frac{1}{3}X_3 - \frac{1}{3}X_4 - \frac{1}{3}X_5$$

$$E(Y) = E(\frac{1}{2}X_1 + \frac{1}{2}X_2 - \frac{1}{3}X_3 - \frac{1}{3}X_4 - \frac{1}{3}X_5)$$
$$= \frac{1}{2}E(X_1) + \frac{1}{2}E(X_2) - \frac{1}{3}E(X_3) - \frac{1}{3}E(X_4) - \frac{1}{3}E(X_5) = -1$$

$$V(Y) = V(\frac{1}{2}X_1 + \frac{1}{2}X_2 - \frac{1}{3}X_3 - \frac{1}{3}X_4 - \frac{1}{3}X_5)$$
$$= \frac{1}{4}V(X_1) + \frac{1}{4}V(X_2 + \frac{1}{9}V(X_3) + \frac{1}{9}V(X_4) + \frac{1}{9}V(X_5) = 3.167$$

$$\sigma_Y = \sqrt{V(X)} = \sqrt{3.167} = 1.7795$$

$$P(Y \geq 0) = P(\frac{Y - E(Y)}{\sigma_Y} \geq \frac{0 - (-1)}{1.7795}) = P(Z \geq 0.56) = 1 - P(Z < 0.56) = 0.2877$$

$$P(-1 \leq Y \leq 1) = P\left(\frac{-1 - (-1)}{1.7795} \leq Z \leq \frac{1 - (-1)}{1.7795}\right) = P(0 \leq Z \leq 1.12) == 0.3686$$