

# STAT230 Assignment 3

Rin Meng

Student ID: 51940633

March 14, 2024

1. (a) *Proof.* Given:  $X$  is a continuous variable that follows a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ . Which also means

$$X \sim N(\mu, \sigma^2)$$

We want to show 95% of the area under the normal density curve lies between 2 standard deviation of the mean. By definition of the given statement, we can safely assume that

$$Z = \frac{X - \mu}{\sigma}$$

and standardized normal distribution, we will have

$$Z \sim N(0, 1)$$

and it is also true that to compute  $P(X \leq x)$ , we can use the fact that

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right)$$

.

$$= P\left(z \leq \frac{x - \mu}{\sigma}\right)$$

$$\Rightarrow P(X \leq \pm 2\sigma) = P\left(z \leq \pm \frac{2\sigma - \mu}{\sigma}\right)$$

the  $\pm$  is because we want to show "between"  $2\sigma$ . Now, if we substitute  $\mu = 0$  and  $\sigma = 1$  we will have the following:

$$P(X \leq \pm 2(1)) = P\left(z \leq \pm \frac{2(1) - 0}{1}\right) = P(z \leq \pm 2)$$

Because we want a between value, we would need to subtract  $P(z \leq 2)$  from  $P(z \leq -2)$  as per  $z$  table. So now we have,

$$\begin{aligned} P(X \leq 2\sigma) &= P(z \leq 2) - P(z \leq -2) \\ &= 0.9772 - 0.0228 = 0.9544 \simeq 0.95 \end{aligned}$$

With this value, 0.95, it implies that the 95% of the area lies between  $P(z \leq 2)$  and  $P(z \leq -2)$ .

$\therefore$  It is true that, 95% of the area under the normal density curve lies between  $2\sigma$  of the mean.  $\square$

- (b) *Proof.* Given:  $Y_1$  and  $Y_2$  are independent random variables,  $U_1 = Y_1 + Y_2$  and  $U_2 = Y_1 - Y_2$ .

We want to show  $Cov(Y_1, Y_2)$  in terms of  $Y_1$  and  $Y_2$ . The covariance formula is given by:

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E[XY] - E[X]E[Y]$$

Since  $Y_1$  and  $Y_2$  are independent random variables, it is true that,

$$Cov(Y_1, Y_2) = 0 = E[Y_1 Y_2] - E[Y_1]E[Y_2]$$

Rearranging our  $U_1$ , and  $U_2$ , we find out that

$$\begin{aligned} Y_1 &= U_1 - Y_2 \\ Y_1 &= U_2 + Y_2 \\ U_1 - Y_2 &= U_2 + Y_2 \\ U_1 - U_2 &= 2Y_2 \\ Y_2 &= \frac{U_1 - U_2}{2} \\ \Rightarrow Y_1 &= U_2 + \frac{U_1 - U_2}{2} = \frac{2U_2}{2} + \frac{U_1 - U_2}{2} \\ Y_1 &= \frac{U_1 + U_2}{2} \end{aligned}$$

Following our formula for covariance above, we have to find  $E[Y_1]$  and  $E[Y_2]$  to solve it. Let us remind ourselves that  $E[aX] = aE[X]$

and that  $E[aX + bY + c] = aE[X] + bE[Y] + c$  where  $a$ ,  $b$ , and  $c$  are some arbitrary constant. Then we will have:

$$\begin{aligned} E[Y_1] &= E\left[\frac{U_1 + U_2}{2}\right] = E\left[\frac{1}{2}(U_1 + U_2)\right] \\ &= \frac{1}{2}E[(U_1 + U_2)] \end{aligned}$$

and  $E[Y_2]$  will follow the same concept as follows:

$$\begin{aligned} E[Y_2] &= E\left[\frac{U_1 - U_2}{2}\right] = E\left[\frac{1}{2}(U_1 - U_2)\right] \\ &= \frac{1}{2}E[(U_1 - U_2)] \end{aligned}$$

Now we substitute it into the formula

$$\begin{aligned} Cov(Y_1, Y_2) &= E[Y_1 Y_2] - E[Y_1]E[Y_2] \\ &= E\left[\frac{U_1 + U_2}{2} \times \frac{U_1 - U_2}{2}\right] - \frac{1}{2}E[(U_1 + U_2)]\frac{1}{2}E[(U_1 - U_2)] \\ &= E\left[\frac{U_1^2 - U_2^2}{4}\right] - \frac{1}{4}E[(U_1 + U_2)]E[(U_1 - U_2)] \\ &= \frac{1}{4}E[U_1^2 - U_2^2] - \frac{1}{4}((E[U_1] + E[U_2]) \times (E[U_1] - E[U_2])) \\ &= \frac{1}{4}E[U_1^2 - U_2^2] - \frac{1}{4}(E[U_1]^2 - E[U_2]^2) \\ &= \frac{1}{4}(E[U_1^2 - U_2^2] - (E[U_1]^2 - E[U_2]^2)) \\ &= \frac{1}{4}(E[U_1^2] - E[U_2^2] - E[U_1]^2 + E[U_2]^2) \\ &= \frac{1}{4}((E[U_1^2] - E[U_1]^2) - (E[U_2^2] - E[U_2]^2)) \end{aligned}$$

We recall that  $Var[X] = E[X^2] - E[X]^2$  so then we have,

$$Cov(Y_1, Y_2) = \frac{1}{4}(Var[U_1] - Var[U_2])$$

$\therefore$  We have expressed  $Cov(Y_1, Y_2)$  in terms of the variance of  $Y_1$  and  $Y_2$ .  $\square$

2. The  $Cov(X, Y) = E[XY] - E[X]E[Y]$  and  $Cov(X, Y) = 0$  when  $X$  and  $Y$  are independent but it is also given that the pdf is,

$$f(x, y) = \begin{cases} 2x & \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

Which ultimately implies that the variable  $x$  and  $y$  is independent because  $f(x, y)$  is not affected for any value of  $y$  in  $0 \leq y \leq 1$ . Which leads to our conclusion that  $Cov(X, Y) = 0$ .

3. *Proof.* Let us denote the given joint probability distribution table with  $\phi$ . We want to show that  $X$  and  $Y$  are dependent, then we need to proof the following

$$\exists(X, Y) \in \phi : P(X, Y) \neq P(X)P(Y)$$

so then we can take the pair  $(0, 0) \in \phi$  where

$$P(0, 0) = 0$$

Now, we find the marginal probability of the following,

$$P(X = 0) = \frac{3}{16} + 0 + \frac{3}{16} = \frac{6}{16}$$

$$P(Y = 0) = \frac{3}{16} + 0 + \frac{3}{16} = \frac{6}{16}$$

Then,

$$P(X = 0)P(Y = 0) = \frac{6}{16} \times \frac{6}{16} = \frac{36}{256}$$

So,

$$P(0, 0) \neq P(X = 0)P(Y = 0) \Rightarrow 0 \neq \frac{36}{256}$$

Hence, we have proven that  $X$  and  $Y$  are not independent.

Now we want to show that  $X$  and  $Y$  have zero covariance, then we need to proof the following

$$\forall(X, Y) \in \phi, C(X, Y) = E[XY] - E[X]E[Y] = 0$$

First we find the marginal sum of

$$E[X] = \sum xP(X = x)$$

$$\begin{aligned}
&= -1\left(\frac{1}{16} + \frac{3}{16} + \frac{1}{16}\right) + 0\left(\frac{3}{16} + 0 + \frac{3}{16}\right) + 1\left(\frac{1}{16} + \frac{3}{16} + \frac{1}{16}\right) \\
&= -\frac{5}{16} + \frac{5}{16} = 0 \\
E[Y] &= \sum yP(Y = y) \\
&= -1\left(\frac{1}{16} + \frac{3}{16} + \frac{1}{16}\right) + 0\left(\frac{3}{16} + 0 + \frac{3}{16}\right) + 1\left(\frac{1}{16} + \frac{3}{16} + \frac{1}{16}\right) \\
&= -\frac{5}{16} + \frac{5}{16} = 0
\end{aligned}$$

and now we find,

$$\begin{aligned}
E[XY] &= \sum xyP(X = x, Y = y) \\
&= (-1)(-1)\left(\frac{1}{16}\right) + (0)(-1)\left(\frac{3}{16}\right) + (1)(-1)\left(\frac{1}{16}\right) \\
&\quad + (-1)(0)\left(\frac{3}{16}\right) + (0)(0)(0) + (1)(0)\left(\frac{3}{16}\right) \\
&\quad + (-1)(1)\left(\frac{1}{16}\right) + (0)(1)\left(\frac{3}{16}\right) + (1)(1)\left(\frac{1}{16}\right) \\
&= \frac{1}{16} + 0 - \frac{1}{16} + 0 + 0 + 0 - \frac{1}{16} + 0 + \frac{1}{16} = 0
\end{aligned}$$

Now, we can finally substitute it in the equation,

$$C(X, Y) = E[XY] - E[X]E[Y] = 0 - 0 * 0 = 0$$

Hence, we have proven that the  $Cov(X, Y) = 0$ .

$\therefore$  We have shown that  $X$  and  $Y$  are dependent, but they have zero covariance.  $\square$

4. Let us denote the data, sorted by ascending order, for the past weeks by  $x$ . So we have,

$$x = \{585, 590, 593, 593, 598, 602, 604\}$$

So then,  $\bar{x} = \frac{4165}{7} = 595$ ,  $\mu_0 = 600$ ,  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ .

$$\begin{aligned}
\sigma &= \sqrt{\frac{\sum (x - \bar{x})^2}{n}} = 6.23 \\
\sigma_{\bar{x}} &= \frac{6.23}{\sqrt{7}} = 0.89
\end{aligned}$$

- (a)  $H_0 : \mu = \mu_0 = 600$ , because  $\mu < \mu_0$ , we will use one-sided (lower-tail) test where  $H_1 : \mu < \mu_0 \Rightarrow \text{p-value} = P(Z \leq z_{obs})$ , since it is 5% level of significant, then,  $CI = 95\% \Rightarrow \alpha = 0.05$ .

$$z_{obs} = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{595 - 600}{0.89} = -5.62$$

$$P(Z \leq -5.62) \simeq 0$$

and it follows that the p-value  $< \alpha$  because p-value = 0  $< 0.05 = \alpha$ . Therefore, we can conclude that we will reject the null hypothesis and yes, the average yield is less than 600 tons.

- (b) Some assumptions that I made throughout this observations are:
- i. The data  $x$  follows a normal distribution.
  - ii. The data  $x$  are independent of each other.
  - iii. The data  $x$  are continuous.
- (c) If  $\alpha = 0.01$ , it is making it harder to reject the null hypothesis but we would still have, p-value = 0  $< 0.01 = \alpha$ . Which still implies we need to reject the null hypothesis.
- (d) Then, our p-value comparison turns into,

$$2P(Z \geq |z_{obs}|) = 2(0) = 0$$

since the z-distribution is symmetric and still, p-value = 0  $< 0.05 = \alpha$ . Which still implies we need to reject the null hypothesis.

- (e) The confidence interval will be calculated as follows:

$$CI_{95\%} = \bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 595 \pm (0.025) \frac{6.23}{\sqrt{7}} \simeq 595 \pm 0.0589$$

$$CI_{95\%} = (595.0589, 594.9411)$$

This means that since our interval is (595.0589, 594.9411), it is impossible for the company to produce, on average, 600 tons of chemical per day, as 600 does not lie anywhere in our calculated interval (595.0589, 594.9411).

- (f) Our result is perfectly consistent with part (a) because we already knew that we are going to reject the null hypothesis where it is believed that the company is able to make 600 tons of chemical per day.