

Problem Set 3 Solutions

ECON 340: Economic Research Methods

Instructor: Div Bhagia

1. (3 pts) Suppose you are given that X follows a normal distribution with mean 20 and variance 49 i.e. $X \sim N(20, 49)$.

- (a) What is the probability that X is equal to 15?

$P(X = 15) = 0$ because X follows a continuous distribution and hence can take infinitely many values. It is unlikely that out of all these infinitely many values we will see a realization of exactly 15.

- (b) What is the probability that X is less than 15?

To find $P(X < 15)$, we will first convert X to a standard normal random variable and then find the probability from the standard normal table.

$$\begin{aligned} P(X < 15) &= P\left(\frac{X - 20}{7} < \frac{15 - 20}{7}\right) \\ &= P(Z < -0.71) \\ &= 1 - 0.7611 = 0.2389 \end{aligned}$$

- (c) What is the probability that X is between 15 and 25?

We will proceed in the same way as the last part.

$$\begin{aligned} P(15 < X < 25) &= P\left(\frac{15 - 20}{7} < \frac{X - 20}{7} < \frac{25 - 20}{7}\right) \\ &= P(-0.71 < Z < 0.71) = 1 - 2P(Z < -0.71) \\ &= 1 - 2 \times 0.2389 = 1 - 0.4778 = 0.5222 \end{aligned}$$

- (d) The probability that X is more than x_0 is 0.8. What is x_0 ?

We are given $P(X > x_0) = 0.8$, we can find x_0 as follows

$$P(X > x_0) = P\left(\frac{X - 20}{7} > \frac{x_0 - 20}{7}\right) = 0.8$$

Define $z_0 = \frac{x_0 - 20}{7}$, then we have $P(Z > z_0) = 0.8$. From the standard normal table $z_0 = -0.85$. Using how we defined z_0 , we have

$$x_0 = 20 + 7 \times z_0 = 20 - 7 \times 0.84 = 14.12$$

- (e) Your friend is surprised that you were quickly able to solve complex integrals to answer these questions. Can you explain to your friend what you did instead?

Instead of dealing with complicated integrals, we simply utilized the standard normal table. This is feasible because any normal distribution's area under the curve can be converted to a standard normal distribution.

For instance, suppose we have $X_1 \sim N(10, 9)$ and $X_2 \sim N(20, 49)$. Then, the area between one standard deviation from the mean for X_1 and X_2 is identical. The only distinction is that one standard deviation from the mean for X_1 is 3, whereas one standard deviation from the mean for X_2 is 7. For the standard normal distribution, one standard deviation from the mean is, of course, 1. Thus, the probability that X_1 deviates from its mean of 10 by 3 is the same as the probability that X_2 deviates from its mean of 20 by 7 and Z deviates from its mean of 0 by 1.

2. (3 pts) The table below gives us the joint probability distribution of time taken to read five pages of a book (Y) and caffeine consumption (X).

	Caffeine ($X = 1$)	No Caffeine ($X = 0$)	Total
10-mins ($Y = 10$)	0.2	0.1	0.3
20-mins ($Y = 20$)	0.3	0.4	0.7
Total	0.5	0.5	1

- (a) Find $E(Y)$, which is the (unconditional) expected value of time taken to read

five pages.

$$\begin{aligned} E(Y) &= \sum_y y \Pr(Y = y) = 10 \cdot \Pr(Y = 10) + 20 \cdot \Pr(Y = 20) \\ &= 10(0.3) + 20(0.7) = 17 \end{aligned}$$

(b) Now calculate $E(Y|X = 1)$ and $E(Y|X = 0)$. (Show your work.)

Before we calculate the conditional expectations, we need the following conditional probabilities:

$$\begin{aligned} \Pr(Y = 10|X = 1) &= \frac{\Pr(Y = 10, X = 1)}{\Pr(X = 1)} = \frac{2}{5} = 0.4 \\ \Pr(Y = 20|X = 1) &= \frac{\Pr(Y = 20, X = 1)}{\Pr(X = 1)} = \frac{3}{5} = 0.6 \\ \Pr(Y = 10|X = 0) &= \frac{\Pr(Y = 10, X = 0)}{\Pr(X = 0)} = \frac{1}{5} = 0.2 \\ \Pr(Y = 20|X = 0) &= \frac{\Pr(Y = 20, X = 0)}{\Pr(X = 0)} = \frac{4}{5} = 0.8 \end{aligned}$$

Now we can find $E(Y|X = 1)$ and $E(Y|X = 0)$ as follows:

$$\begin{aligned} E(Y|X = 1) &= \sum_y y \Pr(Y = y|X = 1) = 10(0.4) + 20(0.6) = 16 \\ E(Y|X = 0) &= \sum_y y \Pr(Y = y|X = 0) = 10(0.2) + 20(0.8) = 18 \end{aligned}$$

(c) Given your answer in (b), are X and Y independent variables? Explain.

X and Y are not independent because $E(Y|X = 1) \neq E(Y|X = 0)$.

(d) What can you say about the impact of caffeine on reading speed from your answer in (b)?

Note that,

- $E[Y|X = 1] = 16$ minutes: This is the expected time taken to read five pages when caffeine is consumed.
- $E[Y|X = 0] = 18$ minutes: This is the expected time taken to read five pages when caffeine is not consumed.

From these conditional expectations, we can infer that, on average, it takes two minutes less to read five pages when caffeine is consumed as compared

to when it is not consumed. Therefore, caffeine appears to have a positive impact on reading speed.

3. (2 pts) We have reasons to believe that the distribution of household income for the US population is right-skewed. I am thinking about taking a random sample of 10,000 individuals from the US population and calculating the mean.

- (a) Is it possible to ascertain whether the distribution of the sample mean in this case will be normal? If yes, how did you come to this conclusion?

Yes, we can invoke the Central limit theorem to conclude that the distribution of the sample mean will be normal as we have a large enough sample ($n \geq 100$).

- (b) Say the true population mean of income is \$90,000 and the true standard deviation is \$68,000. What is the expectation and the variance of the sample mean in this case?

$$E(\bar{X}) = \mu = \$90,000$$

$$Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{68,000^2}{10000} = 462400$$

4. (2 pts) We took a random sample of 25 CSUF students and asked them how many hours did they spend in front of a screen in the past one week. Say we know that hours in front of the screen for CSUF students is normally distributed with unknown mean and variance $\sigma^2 = 36$. We found that average hours in front of a screen in our sample was 40. Construct a 95 percent confidence interval for the population mean.

Since the underlying population is normal, we can conclude that $\bar{X} \sim N(\mu, \sigma^2/n)$. Here $n = 25$ and $\sigma^2 = 36$, so the variance of the sample mean:

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = \frac{36}{25} = 1.44$$

In which case, $\bar{X} \sim N(\mu, 1.44)$ and $\sigma_{\bar{X}} = 1.2$.

Given our realization of \bar{X} , $\bar{x} = 40$, the 95% interval is given by:

$$\bar{x} \pm z_{\alpha/2} \sigma_{\bar{X}} = 40 \pm 1.96 \times 1.2 = 40 \pm 2.35$$

This implies that we are 95% confident that the true average hours in front of a screen for CSUF students last week lie somewhere between 37.65 and 42.35 hours.