

**UNIVERSITY OF LIFE**

**ECON101**

**DEPARTMENT OF LEARNING STUFF  
LEVEL 4 CLASS TEST**

**Statistics 101**

**7/11/2024**

**DURATION: 2 hours**

**Paper Instructions**

There are **THREE** sections. Candidates should answer **ALL THREE** sections.

Each section is worth the same amount of marks in total, and a mark in square brackets is assigned to each part of a question.

Student is permitted to remove question paper from examination venue.

**Do not turnover until instructed to do so.**

## SECTION 1

1. A random sample of size  $n = 48$  is taken from a large population with known standard deviation  $\sigma = 26$ . If the sample mean is 74, calculate to 2 decimal places the following confidence intervals for the population mean  $\mu$ :

- (a) The 90% confidence interval. [3]  
 (b) The 95% confidence interval. [3]  
 (c) The 99% confidence interval. [3]

1. Solution:

- (a) If the sample of size  $n = 48$  has a mean  $\bar{x} = 74$  and the population has a standard deviation  $\sigma = 26$ , then the sampling distribution of the mean follows a normal distribution:

$$\bar{X} \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right) = N\left(74, \frac{26^2}{48}\right) = N(74, 14.08)$$

Note, we are given the population standard deviation  $\sigma$ , which means that we can use the Normal distribution to calculate the confidence interval regardless of sample size (i.e. we do not need the Central Limit Theorem).

The 90% confidence interval can be constructed using the standard normal distribution, but looking for the value of  $z$  such that:

$$P(-z < Z < z) = 0.9 \iff P(0 \leq Z \leq z) = 0.45$$

Using the table of values, we find  $z = \pm 1.645$ . Using the mean and standard deviation of the distribution of  $\bar{X}$ , we can then recover 90% confidence interval.

$$C.I._{.90} = \bar{x} \pm 1.645 \frac{\sigma}{\sqrt{n}} = 74 \pm 1.645 \cdot 3.75 = (67.83, 80.17)$$

With 90% confidence, the population mean  $\mu$  is between 67.83 and 80.17.

- (b) The 95% confidence interval can be constructed by looking for the value of  $z$  such that:

$$P(-z < Z < z) = 0.95 \iff P(0 \leq Z \leq z) = 0.475$$

Using the table of values, we find  $z = \pm 1.96$ . Using the mean and standard deviation of the distribution of  $\bar{X}$ , we can then recover 95% confidence interval.

$$C.I._{.95} = \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 74 \pm 1.96 \cdot 3.75 = (66.65, 81.35)$$

With 95% confidence, the population mean  $\mu$  is between 66.65 and 81.35.

- (c) The 99% confidence interval can be constructed by looking for the value of  $z$  such that:

$$P(-z < Z < z) = 0.99 \iff P(0 \leq Z \leq z) = 0.495$$

Turn over

Using the table of values, we find  $z = \pm 2.575$ . Using the mean and standard deviation of the distribution of  $\bar{X}$ , we can then recover 99% confidence interval.

$$C.I._{.99} = \bar{x} \pm 2.575 \frac{\sigma}{\sqrt{n}} = 74 \pm 2.575 \cdot 3.75 = (64.34, 83.66)$$

With 99% confidence, the population mean  $\mu$  is between 64.34 and 83.66.

2. A random sample of size  $n = 25$  is taken from a large population. The sample mean is  $\bar{x} = 84$ , and the sample standard deviation is  $s = 32$ .
- (a) Using this information, construct a 95% confidence interval for the population mean  $\mu$ . Enter the interval using square brackets, separating the two values with commas [...]: **[2]**
- (b) Suppose you are told that 32 is also the value of the population standard deviation  $\sigma$ . Construct a new 95% confidence interval and compare this with your previous calculation. **[3]**

2. Solution:

- (a) Given that the sample has a size  $n = 25$ , if we do not know the population standard deviation  $\sigma$ , then we need to (a) replace  $\sigma$  by the sample standard deviation  $s$  and (b) rely on the  $t$ -distribution instead of the standard normal distribution, to allow for additional uncertainty generated by using  $s$ , which is an estimate of  $\sigma$ :

$$C.I._{.95} = \bar{x} \pm t_{0.05}^{n-1} \frac{s}{\sqrt{n}}$$

The critical value  $t_{0.05}^{n-1}$  can be found in the table of values for Student's  $t$ -distribution, for a two-tailed probability of 0.05 (corresponding to 95% confidence) and  $n - 1 = 25 - 1$  degrees of freedom. In this case,  $t_{0.05}^{25-1} = 2.064$ . This gives us the following confidence interval:

$$C.I._{.95} = 84 \pm 2.064 \frac{32}{\sqrt{25}} = [70.790, 97.21]$$

- (b) If the sample of size  $n = 25$  has a mean  $\bar{x} = 84$  and the population has standard deviation  $\sigma = 32$ , then the sampling distribution of the mean follows a normal distribution:

$$\bar{X} \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right) = N\left(84, \frac{32^2}{25}\right) \approx N(84, 6.4^2)$$

Note, because we now know the value of the population standard deviation  $\sigma$ , we are able to calculate confidence intervals using the normal distribution, and we do not need to use Student's  $t$ -distribution. This is only required when (a) we use the sample standard deviation  $s$  as an estimate of  $\sigma$  and (b) the sample size  $n < 30$ , so that there is a lot of uncertainty in the accuracy of  $s$ .

The 95% confidence interval can be constructed using the standard normal distribution, but looking for the value of  $z$  such that:

$$P(-z < Z < z) = 0.95 \iff P(0 \leq Z \leq z) = 0.475$$

Using the table of values, we find  $z = \pm 1.96$ . Using the mean and standard deviation of the distribution of  $\bar{X}$ , we can then calculate 95% confidence interval.

$$C.I._{.95} = \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 84 \pm 1.96 \frac{32}{\sqrt{25}} = [71.46, 96.540]$$

**Turn over**

This is much tighter confidence interval than the previous answer. Even if the actual value of the standard deviation has not changed (still  $p3$ ), the fact that we know for certain the value of the population standard deviation  $\sigma$  rather than have to rely on an estimate  $s$  means we can be much more precise for a given confidence level.

3. A pair of dice is tossed 105 times and the sum is seven 20 times.

- (a) If the dice were fair, what would be the probability of the sum being seven? Round to 3 decimal places if required. **[3]**
- (b) We can reject the hypothesis that the dice are fair using a two tailed test at 5% significance. True or False? **[3]**
- (c) We can reject the hypothesis that the dice are fair using a one tailed test at 5% significance. True or False? Explain which type of test you think is more appropriate in this case. **[3]**

3. Solution:

- (a) With two dice, a throw will result in  $6^2 = 36$  possible pairs of faces. There are 6 possible ways that the sum of the two faces  $X$  can be 7:

$$\begin{array}{ccc} 1 + 6 & 2 + 5 & 3 + 4 \\ 4 + 3 & 5 + 2 & 6 + 1 \end{array}$$

If the dice are fair, all 36 outcomes are equally probable and the probability is therefore:  $P(X = 7) = \frac{6}{36} = \frac{1}{6}$

- (b) The information required to set up the hypothesis test is that the true proportion is  $p = \frac{1}{6}$ , while the sample proportion is  $\hat{p} = 0.190$  with a population standard deviation component  $\sigma = \sqrt{\frac{6}{36} \times \frac{30}{36}} = \sqrt{\frac{5}{36}}$  (remember, we assume that the null is true!). The sample size is  $n = 105$ , therefore we can draw the critical values from the normal distribution.

The hypotheses we want to test are:

$$\begin{cases} H_0 : p = \frac{1}{6} \\ H_1 : p \neq \frac{1}{6} \end{cases}$$

The test statistic is:

$$t = \frac{\hat{p} - p}{\frac{\sigma}{\sqrt{n}}} = \frac{0.190 - \frac{1}{6}}{\sqrt{\frac{5}{36 \times 105}}} = \frac{0.024}{0.036} = 0.655$$

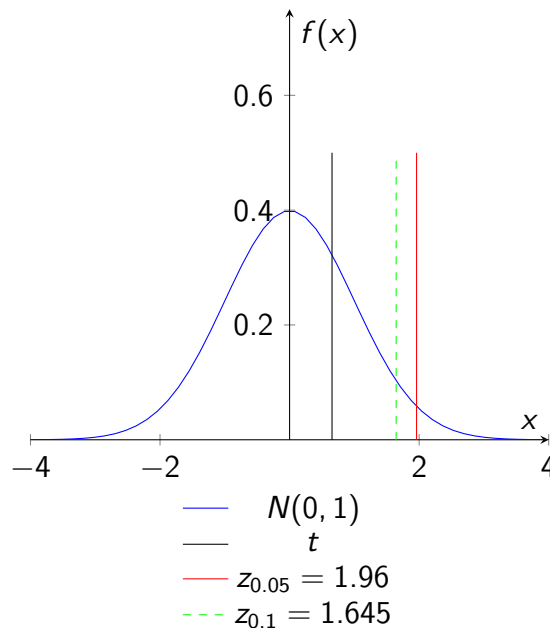
Given that we want 95% confidence on a two-tailed test with a sample size of  $n = 105$ , the critical value of the t-statistic is  $z_{0.05} = 1.96$ . This tells us that  $t < z_{0.05}$ , therefore there is not enough evidence to reject the null hypothesis that the dice are fair. The statement is *false*.

(c) For the one-tailed test, given that  $\hat{p} > p$ , we can set up the following hypotheses:

$$\begin{cases} H_0 : p \leq \frac{1}{6} \\ H_1 : p > \frac{1}{6} \end{cases}$$

The test statistic is the same as before,  $t = 0.655$ . If we want 95% confidence, on a one-tailed test the critical value of the t-statistic becomes  $z_{0.1} = 1.645$ . This tells us that  $t > z_{0.1}$ , therefore we cannot reject the null hypothesis that the dice are fair. The statement is *false*.

The difference here rests in the choice of alternate hypothesis. In a two-tailed test, the alternate hypothesis is  $H_1 : p \neq \frac{1}{6}$ . We are simply stating the sample proportion is not equal to the stated proportion: it could be higher, or it could be lower. In the absence of information about how the dice might be unfair, this is the most conservative approach as it allows for either possibility. In contrast, with the one-tailed test, we have  $H_1 : p > \frac{1}{6}$ : we are explicitly ruling out the possibility that the true proportion is lower than 17.5 out of 105. This assumption allows us to reject the null more easily (i.e. with a smaller deviation from the stated proportion), at the risk of being very wrong if our assumption was incorrect.



4. You are told that a random variable  $X$  has a population standard deviation of 14.
- (a) How large a sample  $n$  would you need in order to be able to estimate the population mean to within  $\pm 3.3$  at the 90% confidence level? Round your answer up to an integer. **[2]**

4. Solution:

- (a) In order to determine how large a sample we need to achieve an error of  $\pm 3.3$  at 90% confidence, we can invert the formula for the confidence interval, where  $z_{90}$  is the critical value of the standard normal distribution. Suppose we pick the upper  $U_{90}$  bound of the confidence interval:

$$\begin{aligned} U_{90} &= \bar{x} + z_{90} \frac{\sigma}{\sqrt{n}} \\ U_{90} - \bar{x} &= z_{90} \frac{\sigma}{\sqrt{n}} \\ \sqrt{n} &= z_{90} \frac{\sigma}{U_{90} - \bar{x}} \\ n &= \left( z_{90} \frac{\sigma}{U_{90} - \bar{x}} \right)^2 \end{aligned}$$

The 90% critical value interval can be constructed using the standard normal distribution, but looking for the value of  $z_{90}$  such that:

$$P(-z_{90} < Z < z_{90}) = 0.9 \quad \Longleftrightarrow \quad P(0 \leq Z \leq z_{90}) = 0.45$$

Using the table of values, we find  $z_{90} = 1.645$ . Replacing this in the equation above, and including the information that the upper bound of the confidence interval should be 3.3 above the mean, we can find the required sample size.

$$n = \left( 1.645 \frac{14}{3.3} \right)^2 \approx (6.979)^2 \approx 48.703$$

We need a sample of at least 49 observations in order to generate a confidence interval of  $\pm 3.3$  around the sample mean.



## SECTION 2

1. A sample of 100 UK households have the following distribution:

Household size	Frequency $f_i$
1	27
2	33
3	18
4	14
5	7
6	1

- (a) Calculate the mean household size  $\bar{x}$ . Round your answer to two decimal places if needed. **[3]**
- (b) Calculate the median household size  $\tilde{x}$ . **[2]**
- (c) Calculate the mode. **[2]**
- (d) Calculate the sample standard deviation  $s$ , rounded to 2 decimal places. **[3]**

1. Solutions:

- (a) The frequency table is given by:

Household size	Frequency $f_i$	$f_i x_i$	$f_i(x_i - \bar{x})^2$
1	27	27	55.99
2	33	66	6.39
3	18	54	5.64
4	14	56	34.07
5	7	35	45.88
6	1	6	12.67
Total	100	244	160.64

$$\bar{x} = \frac{\sum_i f_i x_i}{N} = \frac{244}{100} = 2.44$$

- (b)

$$\tilde{x} = 2$$

- (c)

$$\text{Mode} = 2$$

- (d)

$$s^2 = \frac{\sum_i f_i (x_i - \bar{x})^2}{N - 1} = \frac{160.64}{99} \approx 1.62$$

If the sample variance is  $s^2 \approx 1.62$  then the sample standard deviation is  $s \approx \sqrt{1.62} \approx 1.27$

2. You are told  $P(A) = 0.34$ ,  $P(B) = 0.71$  and  $P(A \text{ or } B) = 0.84$ . Are the following statements true or false? Provide an explanation in each case.

(a)  $A$  and  $B$  are independent. **[3]**

(b)  $A$  and  $B$  are mutually exclusive. **[3]**

(c) If two events that occur with non-zero probabilities are mutually exclusive, then they cannot be independent. **[4]**

2. Solutions:

(a) By the addition rule we know that  $P(A \text{ and } B) = P(A) + P(B) - P(A \text{ or } B)$ . Using the values above, this give us:

$$P(A \text{ and } B) = 0.34 + 0.71 - 0.84 = 0.21$$

Furthermore, we can calculate:

$$P(A) \times P(B) = 0.34 \times 0.71 = 0.24$$

Because  $P(A \text{ and } B) \neq P(A) \times P(B)$ , one can conclude that  $A$  and  $B$  are not independent. *false*

(b)  $A$  and  $B$  cannot be mutually exclusive, as  $P(A \text{ and } B) = 0.21 \neq 0$ . There is a probability that both events will happen together. *False*

(c) If two events  $A$  and  $B$  are mutually exclusive, then  $P(A \text{ and } B) = 0$ . If these two events are also independent, then we also have  $P(A \text{ and } B) = P(A) \times P(B)$ . The combination of the two implies:

$$P(A \text{ and } B) = P(A) \times P(B) = 0$$

This can only be the case if at least one of the events has a zero probability. As we are told that both events have non-zero probability, they cannot be mutually exclusive and independent. *True*.

3. A ball is drawn from a box containing 11 red, 27 white, 21 blue and 17 orange balls. Find the probability (rounded to 2 decimal places if required) that it is:

- (a) Orange or red. **[1]**
- (b) Not red and not blue. **[1]**
- (c) Not blue. **[1]**
- (d) White. **[1]**

3. Solutions:

- (a) There are  $11 + 27 + 21 + 17 = 76$  balls in total in the box. The probabilities can be calculated by counting the number of balls that meet the condition described in the event and dividing by the total number of balls (76). In this case we are using  $P(O \text{ or } R) = P(O) + P(R)$ , as the draw is mutually exclusive: we cannot draw both an orange and a red, so  $P(O \text{ and } R) = 0$ .

$$P(O \text{ or } R) = \frac{17 + 11}{76} \approx 0.37$$

- (b) The most direct route is to realise that the only types of balls that are simultaneously 'not red' and 'not blue' are the orange and white ones, which directly gives:

$$P(\text{not } R \text{ and not } B) = P(O \text{ and } W) = \frac{17 + 27}{76} = \frac{44}{76} \approx 0.58$$

Alternatively, you can also work out the solution using the addition rule. This takes a bit longer and the use of 'not' can complicate thinking about this, but the formula still works:

$$P(\text{not } R \text{ or not } B) = P(\text{not } R) + P(\text{not } B) - P(\text{not } R \text{ and not } B)$$

Rearranging:

$$P(\text{not } R \text{ and not } B) = P(\text{not } R) + P(\text{not } B) - P(\text{not } R \text{ or not } B)$$

A key aspect is to realise that  $P(\text{not } R \text{ or not } B) = 1$ . Any ball you pick is guaranteed to be either 'not red' or 'not blue'. This is obviously true for the orange and white balls, but also the red and blue ones: a red ball is 'not blue' and a blue ball is 'not red'! Using this, and replacing  $P(\text{not } R) = 1 - P(R)$  and  $P(\text{not } B) = 1 - P(B)$  we get:

$$P(\text{not } R \text{ and not } B) = 1 - P(R) + 1 - P(B) - 1 = 1 - P(R) - P(B)$$

$$P(\text{not } R \text{ and not } B) = \frac{76 - 11 - 21}{76} = \frac{44}{76}$$

(c)

$$P(\text{not } B) = 1 - P(B) = \frac{76 - 21}{76} = \frac{55}{76} \approx 0.72$$

(d)

$$P(W) = \frac{27}{76} \approx 0.36$$

**SECTION 3**

1. Suppose 22% of customers at Aldi pay by cash. In a random sample of 13 customers, what is the probability (rounded to 4 decimal places) that:

(a) Two of the customers pay by cash? **[2]**

(b) At least two of the customers pay by cash? **[2]**

1. Solutions:

The probability of a given number of customers  $X$  paying in cash can be modelled using a Binomial distribution, with  $N = 13$  being the number of trials and  $p = 0.22$  the probability of a success (paying in cash):

$$P(X = k \mid n = 13, p = 0.22) = C_k^{13} 0.22^k 0.78^{13-k}$$

(a)

$$P(X = 2) = C_2^{13} 0.22^2 0.78^{13-2} \approx 0.2455$$

(b) To calculate  $P(X \geq 2)$ , one first decomposes the probability as:

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1)$$

Each individual probability can be calculated using the binomial formula:

$$P(X \geq 2) = 1 - C_0^{13} 0.22^0 0.78^{13} - C_1^{13} 0.22^1 0.78^{13-1}$$

$$P(X \geq 2) \approx 1 - 0.0396 - 0.1450$$

$$P(X \geq 2) \approx 0.8154$$

2. What is the probability, rounded to 4 decimal places:

(a) Of getting exactly 7 heads and 1 tails in 8 tosses of a fair coin? **[3]**

(b) Of getting 2 sixes in 8 rolls of a fair dice? **[3]**

2. Solutions:

(a) The probability of getting  $H$  heads can be calculated using a Binomial distribution, with  $N = 8$  being the number of trials and  $p = 1/2$  being the probability of a success (getting a heads):

$$P(H = 7) = C_7^8 \left(\frac{1}{2}\right)^8 \approx 0.0312$$

Note that we define  $H$  as the number of heads, however one could chose instead to count the number of tails. The binomial formula should provide the same answers in both cases, as  $C_7^8 = C_1^8$  and  $p = 1 - p = 1/2$ . Hopefully this is intuitive, the probability of the event should be the same whether you count heads or tails.

(b) The probability of getting  $X$  sixes can be calculated using a Binomial distribution, with  $N = 8$  being the number of trials and  $p = 1/6$  being the probability of a success (getting a six):

$$P(X = 2) = C_2^8 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^6 \approx 0.2605$$

3. If  $Z \sim N(0, 1)$ , find the critical value  $z_0$ , rounded to 2 decimal places, such that:

- (a)  $P(Z \geq z_0) = 0.09$  [2]
- (b)  $P(Z \leq z_0) = 0.25$  [2]
- (c)  $P(-z_0 \leq Z \leq z_0) = 0.53$  [2]

3. Solutions:

Note: It may help you to draw pictures to visualise the area you are being asked to calculate.

- (a) If  $P(Z \geq z_0) = 0.09$ , then  $P(0 \leq Z \leq z_0) = 0.5 - P(Z \geq z_0) = 0.41$ . Looking this value up in the table for the standard normal tells us that  $z_0 = 1.341$ .
- (b) If  $P(Z \leq z_0) = 0.25$ , then  $P(z_0 \leq z \leq 0) = 0.5 - P(Z \leq z_0) = 0.25$ . In the tables of the standard distribution, this corresponds to a value of 0.674. However, we are looking for a value of  $z$  that is smaller than  $z_0$ , therefore to the left of it. Given that the probability is less than 0.5, the critical value  $z_0$  must be negative. We therefore have  $z_0 = -0.674$ . A diagram can be particularly useful for cases like this.
- (c) If  $P(-z_0 \leq Z \leq z_0) = 0.53$  then  $P(0 \leq Z \leq z_0) = \frac{0.53}{2} = 0.265$ . Looking up this value in the table tells us that  $z_0 = 0.722$ .

4. Using the table of areas for the standard Normal distribution, calculate the areas under  $N(0, 1)$ , rounded to 3 decimal places:

(a) Between -1.645 and 1.645 [2]

(b) For values greater than 2.575 [2]

(c) Between -1.25 and 1.25 [2]

(d) Between 0.5 and 1.96. [2]

4. Solutions:

(a)

$$P(-1.645 \leq Z \leq 1.645) = P(-1.645 \leq Z \leq 0) + P(0 \leq Z \leq 1.645)$$

$$P(-1.645 \leq Z \leq 1.645) = 2 \times P(0 \leq Z \leq 1.645)$$

$$P(-1.645 \leq Z \leq 1.645) = 2 \times 0.450 = 0.9$$

(b)

$$P(Z \geq 2.575) = 0.5 - P(0 \leq Z \leq 0.5)$$

$$P(Z \geq 2.575) = 0.5 - 0.495 = 0.005$$

(c)

$$P(-1.25 \leq Z \leq 1.25) = P(-1.25 \leq Z \leq 0) + P(0 \leq Z \leq 1.25)$$

$$P(-1.25 \leq Z \leq 1.25) = 2 \times P(0 \leq Z \leq 1.25)$$

$$P(-1.25 \leq Z \leq 1.25) = 2 \times 0.394 = 0.789$$

(d)

$$P(0.5 \leq Z \leq 1.96) = P(0 \leq Z \leq 1.96) - P(0 \leq Z \leq 0.5)$$

$$P(0.5 \leq Z \leq 1.96) = 0.475 - 0.191 = 0.284$$



**End**