EMTH1019 – final statistics assessment Trimester 3 2020

In this assignment I had 10 multiple choice questions where you had to provide the answer only and the then the longer questions. Students had 1 week to complete the assignment.

The design of the final statistics assessment changed in trimester 1 2021. Mainly because I was seeing too many submissions where all the MC were correct and then people not being able to do anything else ☹

Typical feedback is in the last page.

MJ

Multiple Choice questions

Question 1 — Maryjane OCALLAGHAN, 29 November 2020, 12:18 PM — Maryjane OCALLAGHAN, 29 November 2020, 12:19 PM

A random sample of data is taken from a population that is normally distributed but whose population parameters are unknown. The most appropriate distribution to use is:

a. A Normal Distribution.

b. A "t" Distribution.

c. Poisson distribution

Question 2 — Maryjane OCALLAGHAN, 29 November 2020, 12:19 PM — Maryjane OCALLAGH, 29 November 2020, 12:19

For an ordered set of discrete data of size n=24, the position of Q1 is:

a. Located at position 6 as $\frac{1}{4} * 24 = 6$

b. Not possible to determine as the data set is not provided.

c. Located a quarter of the way between position 6 and 7.

Question 3 — Maryjane OCALLAGHAN, 29 November 2020, 12:20 PM — Maryjane OCALLAGH, 29 November 2020, 12:21

The square root of a population variance equals

a. σ

b. s

c. μ

Question 4 — Maryjane OCALLAGHAN, 29 November 2020, 12:21 PM — Maryjane OCALLAGH, 29 November 2020, 12:21

According to a 2019 poll the average individual consumption of avocados is 13 avocados per year. In a statistics question what sort of variable is 13 avocados per year?

a. Poisson.

b. Bernoulli.

c. Normal.

Question 5 — Maryjane OCALLAGHAN, 29 November 2020, 12:22 PM — Maryjane OCALLAGH, 29 November 2020, 12:22

A box and whiskers plot contains the:

a. Minimum, Mode, Median, Mean, Maximum and Outliers.

b. Minimum, Q1, Mean, Q3, Maximum and Outliers.

c. Minimum, Q1, Median, Q3, Maximum and Outliers.

Question 6 — Maryjane OCALLAGHAN, 29 November 2020, 12:23 PM — Maryjane OCALLAGH, 29 November 2020, 2:12 P

A Standard normal distribution has:

a. $\mu = 0$

b. $s = 0$

c. $\sigma = 0$

d. All of the above.

Question 7 Maryjane OCALLAGHAN 29 November 2020, 12:24 PM Maryjane OCALLAGH 29 November 2020, 12:24 P

For a sample of discrete data of n>30, the data distribution is always approximately

a. Symmetrical.

b. Normally distributed.

c. None of the above.

Question 8 Maryjane OCALLAGHAN 29 November 2020, 12:31 PM Maryjane OCALLAGH 29 November 2020, 12:31 P

John has made 10 sheets of square origami paper and wants to determine the population standard deviation. John decides to use the average of the sides of each piece of paper as this will mean he only has to work with 10 bits of data. Taking the average of the sides:

a. Will have no effect on the value of the standard deviation calculated.

b. Will decrease the value of the standard deviation calculated.

c. Will increase the value of the standard deviation calculated.

Question 9 Maryjane OCALLAGHAN 29 November 2020, 12:32 PM Maryjane OCALLAGH 29 November 2020, 12:32 P

Virus protection software is designed to detect malicious programs. If the software correctly identifies a malicious program, this would be an example of:

a. A false negative.

b. A false positive.

c. None of the above.

Question 10 Maryjane OCALLAGHAN 29 November 2020, 12:38 PM Maryjane OCALLAGH 29 November 2020, 12:38

A z-distribution can be used for:

a. A population with unknown μ, known σ and a reasonable belief that the population is normally distributed.

b. A small random sample, unknown σ and a reasonable belief that the population is normally distributed.

c. A population with unknown μ, unknown σ, a simple random sample as long as n > 30.

Long answer questions

Question 11 — Maryjane OCALLAGHAN, 29 November 2020, 12:41 PM — Maryjane OCALLAGHAN, 29 November 2020, 12:41 PM

MJ has a small business selling recycled origami paper. Nina bought some of this paper and is complaining that the paper is not square and wants a refund. Nina claims to have measured all the sides of the first 3 pieces of paper from the pack of paper she bought from MJ. The side length data (in millimetres) that Nina provided in an email are:

{150.3, 152, 155.5, 148.0, 160.65, 156.01, 165.0, 153, 149.2, 152.3, 157.87, 151.99, 185}

a) Determine the mean and standard deviation of the sample data.
b) Determine the 5-number summary.
c) Identify any outliers
d) Construct a box plot
e) Determine the 99% confidence interval for the above data assuming that the population is normally distributed.
f) MJ does NOT want to refund the $2.37 that Nina paid for the recycled origami paper. Use your knowledge of statistics, the information provided and/or calculated to **advise MJ** on what she could say to reject Nina's claim.

Question 12 — Maryjane OCALLAGHAN, 29 November 2020, 12:42 PM — Maryjane OCALLAGHAN, 29 November 2020, 12:42 PM

A thick shake dispenser at a Curtin food van is designed so that the amount dispensed per large thick shake is 650 ml with a standard deviation of 11 ml. Repeated testing has confirmed that the distribution is normal.

a) What is the probability a large thick shake contains:
   i) Less than 650 ml?
   ii) Exactly 670 ml?
   iii) Between 645 ml and 652 ml?
   iv) Above what amount (to the nearest ml) does the top 7% of the distribution lie?
b) A sample of 9 large thick shakes was taken:
   i) Determine the probability that the sample means lies between 645 ml and 652 ml.
   ii) Explain why this probability is different to the answer in a) iii; other than the formula being different.

Question 13

Maryjane OCALLAGHAN
29 November 2020, 12:45 PM

Maryjane OCALLAGHAN
29 November 2020, 12:45 PM

The mean outer diameter of a skateboard wheel bearing is specified to be 22.000 mm. A skateboard manufacturer receives a delivery of the bearings, takes a sample of bearings from the delivery and measures the outer diameter. The skateboard manufacturer will reject the delivery if the sample mean outer diameter is significantly different to the product specification at the 5% significance level.

The sample diameter data in mm is: $\{22.412, 22.508, 21.891, 22.199, 22.189, 22.240\}$

Set up and test an appropriate hypothesis at the specified significance to determine if the sample provides sufficient evidence to support the bearing manufacturer's claim that the outer bearing diameter is 22.000 mm.

Ensure that you state and explain all assumptions, identify the critical region, determine the test statistic and provide a conclusion with reasoning.

Question 14

Maryjane OCALLAGHAN
29 November 2020, 12:47 PM

Maryjane OCALLAGHAN
29 November 2020, 12:47 PM

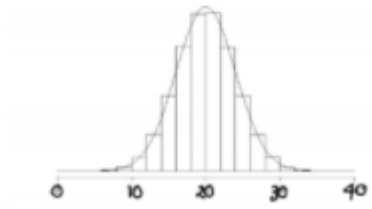
Below are four graphs showing data distributions.

The top graph represents the distribution of a population with $\mu = 20$ and $\sigma = 4$.

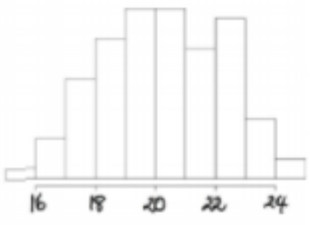
The other 3 graphs labelled A, B and C show distributions of:

1. A single random sample of 100 values from the population shown in the top graph.
2. A distribution of 100 sample means from random samples of size 5 from the population shown in the top graph.
3. A distribution of 100 sample means from random samples of size 25 from the population shown in the top graph.

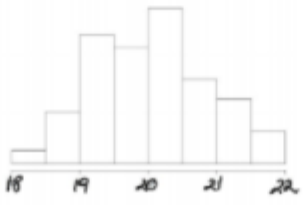
Match each plot **A, B** and **C** with the corresponding description **1, 2** and **3.** Justify your answers.



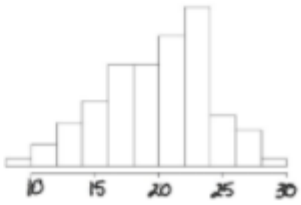
Plot A



Plot B



Plot C



Question 15

Maryjane OCALLAGHAN
29 November 2020, 12:47 PM

Maryjane OCALLAGHAN
29 November 2020, 12:47 PM

According to a website, 34% of adults still sleep with a soft toy. Assuming that this claim is true, what is the probability that at least 1 adult out of a random sample of 20 adults, sleeps with a soft toy?

Typical feedback

MC B C A A C A C B C A
Formula missing or incorrect
Insufficient working
Minus problem
Not attempted.
11a sample mean is x bar and sample sd is s. s= 9.74mm and x bar=156.68mm and n=13. Make sure you declare all key variables and formula.
11b Q1=150.3+0.5(151.99-150.3)= 151.145 mm Q3= 157.87+0.5(160.65-157.87)=159.26 5 number summary {148,151.145, 153,159.26,185}
11c Outlier check Upper limit = Q3+1.5*IQR = 171.43 and Lower Limit Q1-1.5*IQR=138.97. These are not outliers themselves. 185 is an outlier
11d box plots are a graph and a scale needs to be linear and all key quartiles values must be identified. Outliers are represented by a star.
11e the population standard deviation is unknown so you cannot use z but can probably use t. CI=156.678 +/- 8.26mm
11f confidence interval gives you an indication that the population mean lies with a certain range. The truth of a confidence interval depends on the /validity of the assumptions and the quality of the sample data.
11f INina claims 3 papers but there are 14 bits of data, measurements are inconsistent precision, MJ never said it was square paper, sample is not srs, no evidence of accuracy.
12a need to show working and clearly identify the formula you are using. e.g. z=(x-u)/sd
12a work backwards from tables and find that z=+1.48=(x-u)/sd x= 666.28ml
12bneed to write down the formula you are using z=(x-u)/(sd/sqrt(n) Drawing the question often helps me
12b when you take a sample and use x bar in the z formula the mean of the sample eliminates/hides variation in the data e.g. if the side numbers for a square were 630, 620 660 670 then the mean used would be 645mm the mean hides that the square is not square.
13 assumptions need to be clear and justified . It would be reasonable to assume the manufacturing process of bearings is ND. Assuming sample is srs means you are aware of the importance but no guarantee that it is.
13 we have no population data and can used t as long as our assumptions are correct.
13 Null hypothesis = 22.000mm, Alternate hypothesis does not equal 22.000mm. This is definitively a 2 tail distribution. if the diameter is too small or too large the bearings will be wrong.
13 if significance = 5% and 2 tail then alpha=0.025 and dof = 6-1=5 then t = =/- 2.571
13 conclusion needs to state all assumptions, t distribution, sample size, significance, 2 tail, why you are confident in rejecting or accepting null hypothesis as well as referring back to original word question
13 t test = +2.75 is greater than t = 2.571
14 1C, 2A, 3B the most important thing here is the range of the distributions. For the 100 samples of size 25 the mean will have very little variation from the population mean.. The shape of the graph is not th important feature.
15 This is a binomial distribution of p=0.34 P(X=0)=0.0002 P(X>0)=1-0.0002

Marking guide options