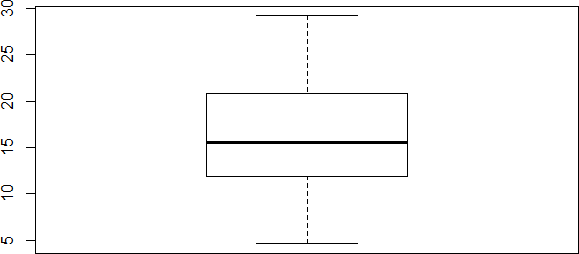


**IMAT5169 Statistics Phase Test 1 2019/20 SOLUTIONS**

**Question 1** [4 marks]

Consider the boxplot shown below.



From this plot, **estimate** the following measures and complete the table below:

|  |  |
| --- | --- |
| Measure | My Answer: |
| Median | Any value around 15 or 16 is fine e.g. 15 [1 mark] |
| Lower Quartile | Any value around 12 [1 mark] |
| Upper Quartile | Any value around 20 or 21 e.g. 21  [1 mark] |
| Inter-Quartile Range | Must equal UQ-LQ e.g. 21-12=9 [1 mark] |

**Question 2** [4 marks]

Consider each of the following two variables:

* Responses to a survey question which falls onto one of five ''Likert'' response categories of

``Strongly Disagree", ``Disagree", ``Neutral", ``Agree" and ``Strongly Agree", which are then coded on the scale of 1, 2, 3, 4 and 5 respectively.

Will the data collected be Nominal, Ordinal or Scale (Numerical)?

My answer:

Ordinal [1 mark]

Indicate whether it would be more sensible to summarise the data using a histogram or a bar chart.

My answer: Bar chart [1 mark]

1. Students' heights (cm) in the Statistics class.

Will the data collected be Nominal, Ordinal or Scale (Numerical)?

My answer:

Scale

[1 mark]

Indicate whether it would be more sensible to summarise the data using a histogram or a bar chart.

My answer: Histogram [1 mark]

**Question 3** [7 marks]

* 1. Using tables of the standard normal distribution (attached), determine the following: i)

P(Z>1.45)

My answer:

You may also wish to write some of your hand worked calculations on the Working out sheet. 0.07353

[1 mark]

ii) P(Z<-0.6)

My answer:

You may also wish to write some of your hand worked calculations on the Working out sheet.

P(Z > -0.6) = 0. 72575 so P(Z < -0.6) =1-0.72575 = 0.27425

[2 marks]

a) A manufacturer of hard wood doors knows that following their manufacture, the width of the doors will vary according to a Normal probability distribution with a mean width of 528mm and a standard deviation of 1.5mm. The width of door frames, that the doors need to fit inside are exactly 530mm wide. Determine **the proportion of doors** that will be greater than 530mm wide and so will need trimming to fit.

My answer:

You may also wish to write some of your hand worked calculations on the Working out sheet. Let X = Door width and assume X ~ N(528, 1.5) so we require P(X > 530).

Standardise: u= (530-528)/1.5= 1.33

Hence P(X > 530) = p(z > 1.33) = 0.09176 (from Standard Normal Tables).

i.e. about 9.2% will need trimming. [4 marks]

**Question 4** [15 marks]

A large online retailing company employs “pickers” who assemble customer orders by locating items in a warehouse and packaging them ready for shipping to customers. The company are undertaking an exercise to assess how its workforce is operating. A random sample of 25 staff was monitored to assess how many orders they assembled in one particular week. Some of the results are shown in the table below (displayed as number of orders x100 per week so that 3.1 = 310 orders etc.). Note that NOT all of the rows of data are shown. This table has the data in the “Orders” column as *x* and then shows the values of *x*2.

|  |  |  |
| --- | --- | --- |
| Staff Member | Orders (x100 per week)  *x* | *x*2 |
| 001 | 3.1 | 9.61 |
| 002 | 4.1 | 16.81 |
| 003 | 3.9 | 15.21 |
| . | . | . |
| . | . | . |
| . | . | . |
| 025 | 3.5 | 12.25 |
| **Totals** | **75.0** | **255.0** |

1. Show, with the help of a calculator and by stating the required calculations in the box below, that the sample mean is 3.0.

My answer:

You may also wish to write some of your hand worked calculations on the Working out sheet.

Sample mean = 75/25=3.0 [1 mark]

1. Show with the help of a calculator and by stating the required calculations in the box below, that the sample standard deviation is 1.118

My answer:

You may also wish to write some of your hand worked calculations on the Working out sheet. Sample variance = 255/24 - 25x3^2/24=10.625-9.375=1.25

Sample s.d.= square root of variance = sq root of 1.25 = 1.118

[4 marks]

1. The company has determined that a fair rate of working is to achieve an average of 250 orders per week per picker (i.e. 2.5 x100 orders per week). Using your results from parts a) to c) or otherwise, undertake a one sample t-test to assess whether there is evidence that the true underlying mean rate is different to 2.5 (x100 orders per week). i.e. test

H0:  =  HA:  ≠ 

Remember to state your conclusions fully.

My answer:

You may also wish to write some of your hand worked calculations on the Working out sheet.

Test statistic is t = (3.0-2.5)/(1.118/5) = 2.236

d.f. = n-1 = 24 so critical t-value is t(24,0.025) =2.064 (for a two-tailed test) Since 2.236 is > 2.064 we know p<0.05

There is evidence the rate that orders are being completed is different to 2.5

The sample mean is in fact 3.0 and so the picking staff are performing over and above the minimum rate.

[5 marks]

1. Calculate a 95% confidence interval for the true population mean and interpret its value.

My answer:

You may also wish to write some of your hand worked calculations on the Working out sheet. 95% CI = 3.0 +/- 2.064\*1.118/5 = 3.0 +/- 0.46 = 2.54 to 3.46

We are 95% confident the true mean rate of picking orders is between 254 and 346 per picker per week.

[3 marks]

1. Explain why the 95% confidence interval from part (d) is consistent with your conclusion from part (c).

My answer:

The 95% CI does NOT include 2.5 and so we are 95% confident that the true rate is not 2.5 which is the conclusion we came to in part e).

[2 marks]

**Question 5** [30 marks]

The management of a national construction company believes that its brick laying staff might work at a different rate in the morning compared to the afternoon. They therefore monitored eight bricklayers to assess how many bricks they lay per hour (person) during one morning and then during one afternoon on site. The results for these eight staff are shown below (displayed as number of bricksx100 per hour so that 1.2 is 120 bricks per hour etc.)

|  |  |  |
| --- | --- | --- |
| Staff | Morning | Afternoon |
| 1 | 1.2 | 0.8 |
| 2 | 1.7 | 1.7 |
| 3 | 2.1 | 1.9 |
| 4 | 0.9 | 1.0 |
| 5 | 1.5 | 0.9 |
| 6 | 1.6 | 1.3 |
| 7 | 1.9 | 1.8 |
| 8 | 1.2 | 1.1 |

1. Using SAS, create a SAS programme to read in the above data using the **input** and **datalines** commands. In the space below copy and paste all your lines of SAS code you have used to complete this.

My SAS Code:

**data** Bricks;

input Staff Morning Afternoon; datalines;

|  |  |  |
| --- | --- | --- |
| 1 | 1.2 | 0.8 |
| 2 | 1.7 | 1.7 |
| 3 | 2.1 | 1.9 |
| 4 | 0.9 | 1.0 |
| 5 | 1.5 | 0.9 |
| 6 | 1.6 | 1.3 |
| 7 | 1.9 | 1.8 |
| 8 | 1.2 | 1.1 |
| ; |  |  |
| **run**; |  |  |

[4 marks]

1. Use **proc means** in SAS to obtain summary statistics for the data for the Morning and the Afternoon. In the space below copy and paste all your lines of SAS code you have used to complete this. There is no need to comment on the results at the moment!

My SAS Code:

**proc means** Data=Bricks; var Morning Afternoon; **run**;

[3 marks]

1. Use SAS to undertake a Paired t-test to assess whether there is any evidence that the mean brick laying rates in the morning and afternoon are different. In the space below copy and paste all your lines of SAS code you have used to complete this. There is no need to comment on the results at the moment!

My SAS Code:

**data** Bricks; set Bricks;

Bricks\_Diff=Morning-Afternoon;

**run**;

**proc ttest** data=Bricks h0=**0**; var Bricks\_Diff;

**run**;

[5 marks]

1. Based on your results obtained in parts (b) and (c), state the values of the relevant t-statistic and p-value provided by SAS, and discuss fully what conclusions you would cometo.

My answer:

t-statistic = 2.49

p-value = 0.0413

Since p<0.05 we reject Ho.

There is some evidence that the true mean difference in the rates at which bricks are being laid is different in the afternoon compared to the morning. Although this is a borderline outcome since p is close to 0.05

The mean rate in the afternoon is lower at 1.31x100 (131) bricks per hour compared to the morning where the mean rate was 1.51x100 (151) bricks per hour.

[10 marks]

1. State the value of the 95% Confidence Interval for the mean difference provided by SAS and interpret this.

My answer:

95% Confidence Interval for the mean difference is 0.0104 to 0.3896

(Feedback be careful you did not choose the other confidence interval for the standard deviation shown by SAS!)

The 95% CI for the mean difference suggests that the true difference could be as low as 0.01x100 (i.e. 1 brick) per hour (hence the reason for the borderline outcome) or as much as 0.39x100 (i.e. 39) bricks per hour.

[4 marks]

1. Comment on any assumptions made in the above analysis.

My answer:

Since the sample size is less than 30 we need to assume that the differences data are normally distributed. Although it is difficult to assess this with such a small sample, the histogram suggests no problems with this and the Q-Q plot shows the data points lie on an approximate straight line which also suggests no problems with this assumption.

[4 marks]