Commentary on candidate evidence

Introduction

Candidate 1

Introduction

British farming produces some of the world's most sustainable meat, with our grass-based grazing systems. In Britain there are currently around 30 million sheep and lambs. In 2020 the sheep sector was worth approximately £290 million and employs around 150,000 people on farms and in related industries. With the daily consumption of meat decreasing, the meat production is on the decline.

This report will look at the home fed production of Sheep in the 35 year period from 1985 to 2020. The data used is numerical data. I am using the data that I have gathered to determine if there is a linear relationship between the year and the Production of Sheep. The raw data was obtained from the gov.co.uk website, which we can assume is reliable and correct as it was from a government website. It can also be assumed that the data is unbiased and influenced as it was sourced from the government however it was produced and written and gathered in the United Kingdom and could be favourable towards the United Kingdom.

Candidate 2

In alabama the first death penalty exucution had been carried out in 1812 and back then the method of kiling the murdeur was hanging. Throughout the years more methods goyt introduced such as the electric chair and by 2002 you could choose whether you would rather get executed, electrocuted or the lethal injection. The death penalty was introcued because they thought that was the only way for people to pay for the crime/crimes they have committed. How ever the alabama muder states with it being are still high as you would think the the death penalty would scare people more to not commit crimes. Assault through Ithe years has also been high in numbers.

This project will look at the numbers of murder in alabama compared to asslut in albama from 1990-2015. Murder has dropped throughout the years but not by much however assult rates are much higher compare to murder rates but assult rate are also dropping according to disatercentre.com. We used this data to determined if there was a correlation between the assault and murder

Candidate 3

Introduction

I am researching the correlation between Crime Rate vs Education and the type of data is numerical the data I am getting this from is the data statics exel doc
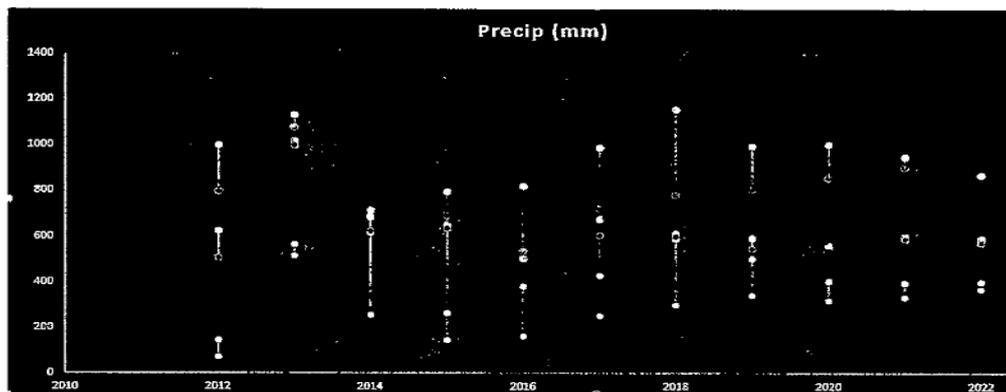
I would say that this data shows that there is a real correlation between crime and education I also made sure to change the graphic to style 1 rather than style 4.

Subjective Impression

Candidate 4

Projects are scanned and will always appear in black and white. Candidates should check that their graphs and charts are readable when printed in black and white.
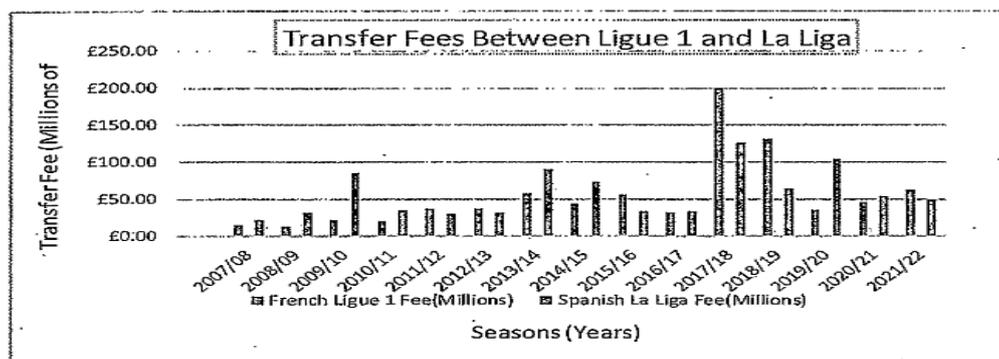
Candidate 5

<u>Subjective Impression</u>

Figure 1 is a comparative bar chart between the Ligue 1 and La Liga for the last 15 season showing their transfer fees each season.
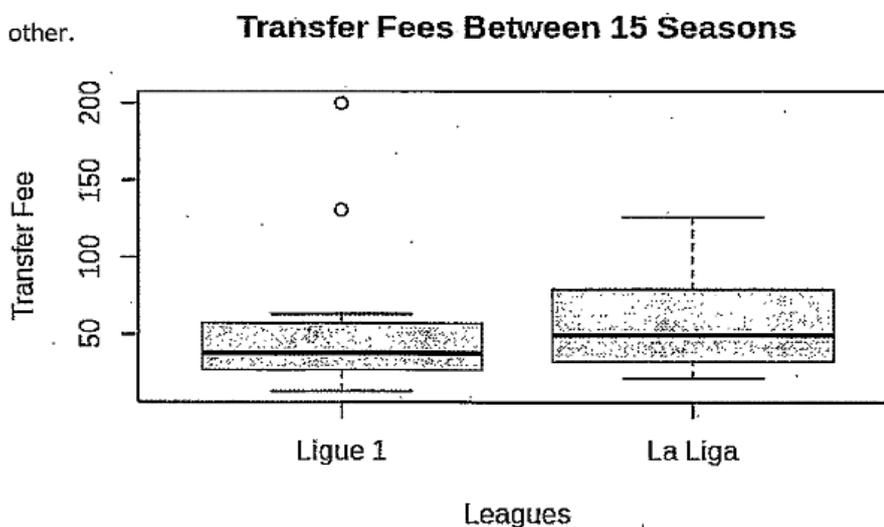
<u>Figure 1:</u>

The bar chart seems to show that the La Liga spent more on average. The Ligue 1s spending was most of the time much lower but on two occasions they spent 'mega bucks', their spending was quite similar through the years but had the two massive transfers.
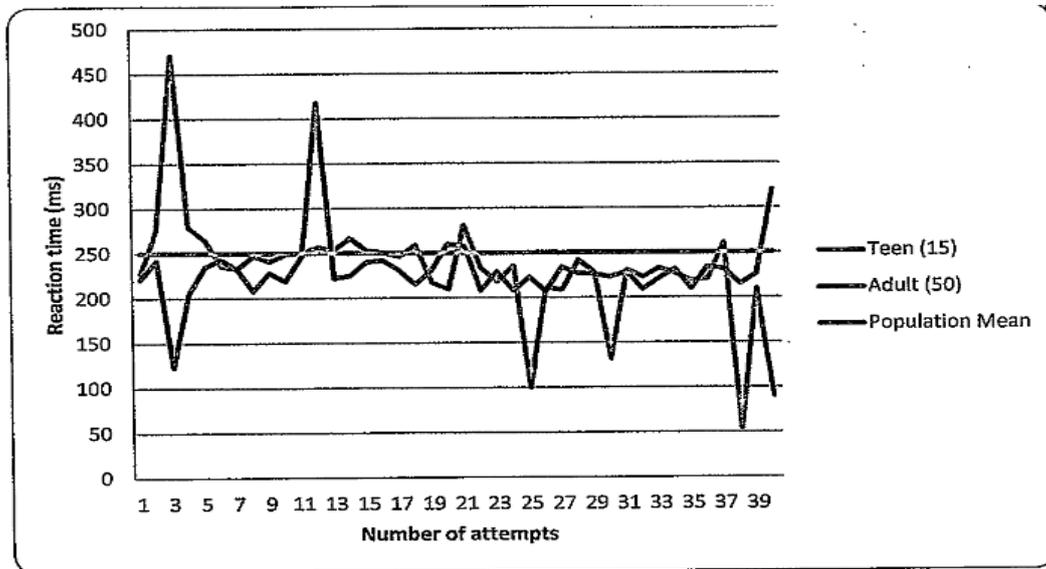
<u>Figure 2:</u>

This comparative box plot shows how the how the two leagues directly compare to each other.

The box plot shows that on average the La Liga spend more money on transfers than the Ligue 1, with exception to the two outliers. In the La Liga the transfers were more varied, which goes back to my earlier assertion that the Ligue 1s transfers were less varied and much closer together without including the two outliers.

Candidate 6

Figure 1: Line graph comparing the reaction times of the two test subjects

The graph shows that as more attempts were made, the adult became more consistent. Interestingly, the adult's reaction times were mostly below the population average, but I will perform a hypothesis test to see if they are significantly outside the average. The teen had less consistency as more attempts were made, but did react faster as the test progressed, suggesting that practise made an impact.

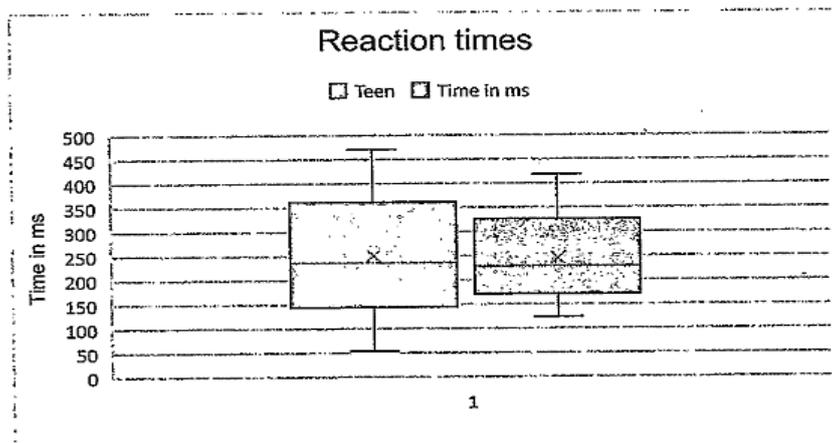## Figure 2: Box plots comparing the spread of results

Figure 2 shows that both the overall range and the interquartile range of the teen's reaction times were much higher than that of the adult. The median for the adult was also slightly faster, which I had predicted due to experience.

Figure 3: Summary data

Teen		Adult
54	Min	123
234.5	Q1	218
236	Median	227.5
253.25	Q3	234.25
471	High	418
231.25	Mean	231.45
62.61	Standard deviation	40.29

The minimum for the teen was outside the normal range of reaction times and could be discounted as an outlier. This is compensated for by using the interquartile range, which does not include the extremes. The mean for both was very similar, with the teen being on average 0.2ms faster.

Candidate 7

Sample

The data from nrscotland.gov.uk has many rows on data from deaths of every drug but I have only taken the statistics from Heroin and Methadone. After I collected all of the relevant data I constructed a Scatter graph on excel showing the difference in deaths from each drug.


```
   X              HEROIN           METHADONE .
Min.  :2000   Min.  :122.0   Min.   : 55.0
1st Qu.:2004   1st Qu.:145.0   1st Qu.: 91.5
Median :2009   Median :193.0   Median :173.0
Mean  :2009   Mean  :208.3   Mean   :196.8
3rd Qu.:2014   3rd Qu.:249.5   3rd Qu.:244.0
Max.  :2018   Max.  :375.0   Max.   :560.0
```

Analysis and Interpretation

Candidate 8

Candidate is working with categorical data.
Evidence for performing and presenting an appropriate test (marks 15-18 only)

t-Test: Paired Two Sample for Means		
	Average of Standard_Rating	Sum of Standard_Rating
Mean	2734.551282	6700.95
Variance	1739.0445	60316616.26
Observations	20	20
Pearson Correlation	0.017821778	
Hypothesized Mean Difference	0	
df	19	
t Stat	-2.28416874	
P(T<=t) one-tail	0.017018907	
t Critical one-tail	1.729132812	
P(T<=t) two-tail	0.034037815	
t Critical two-tail	2.093024054	

Evidence for interpretation in context

After creating a bar chart, I performed a t-test on my data based on the average standard rating and sum of standard rating for each country. As we can see, the p-value for both one-tail and two-tail tests are both less than the standard significance level of 0.05. This means that I can reject the null hypothesis.

Candidate 9

<u>Linear regression model and associated statistics</u>
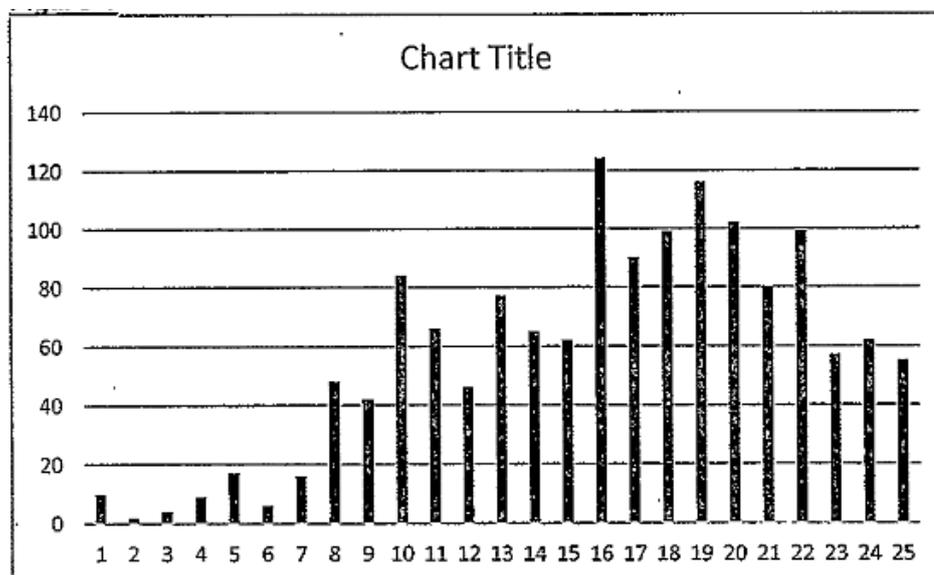The equation of the linear regression model can be calculated as:

Deaths= -7505.592 + 3.765 x year

The R squared Value is calculated at 0.5563, which gives an R value of 0.537. This figure is not very close to one and therefore this suggests that a linear model is not a very good fit for this set of data.

I have calculated the mean for the deaths of journalists over the 25 year period from 1997 to 2021.
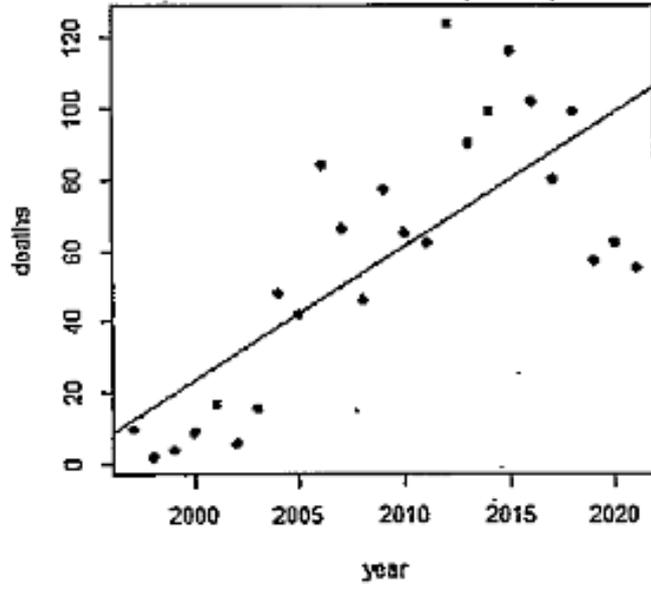
Mean: 57.52

The numbers from the mean tell us that from the 25 year period 58 journalists per year were killed on average.

Chart Title

The bar chart above shows that a linear relationship does not exist for the data gathered.

<u>Figure 2</u>
Below is a scattergraph showing trend over the 25 year period of the data recorded. This figure also shows that there is no linear relationship once again although we can see that death rates are at their highest during the 2012 to 2018 period.

Candidate 10

Analysis and Interpretation

A t-test will be used to determine if there is a difference between the mean male and female deaths caused by dementia. A t-test is used as the sample size is fairly small, with samples over just 20 years. Our null hypothesis would state that there is no difference in means between these two variables – deaths in males and females due to dementia.

Figure 3.1

```
data:  male and female
t = -5.6665, df = 32.049, p-value = 2.842e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2051.7271  -966.7491
sample estimates:
mean of x mean of y
 1245.619  2754.857
```

As our p-value is less than 0.05 we can reject the null hypothesis and state that there is a difference between mean male and female deaths caused by dementia over this time period of 20 years.
As 0 does not fall between our 95% confidence interval of -2051.7271 and -966.7491 we can say with 95% confidence that there is a difference in mean deaths of males and females as there must be a difference in mean deaths between the two groups.

The boxplot (figure 2.1) highlights a great difference in median of deaths in males and females, and using figures 2.2 and 2.3 we can calculate that on average there are 1,336 more female deaths per year during the time period. There is also a higher variability in female deaths, as there is an interquartile range of 2,027 for deaths in females compared to 1,173 in males during this time period.

Conclusion

Candidate 11

Conclusion

In conclusion this report was conducted to find out if the neighbouring countries of Singapore and Malaysia had any correlation to each other and based on the evidence seen, we can say that they have no relationship to each other and don't seem to have any relation on how fast they grow in terms of population. The bar chart shows that Malaysia has had a more steady growth compared to Singapore with Singapore having outliers at the start of when the data was collected. Along with this, the boxplots do not share a similar spread and the descriptive statistics show no similarities except median with significantly different IQRs. When looking at the scatter plot, we can also see that there is no linear relationship between the two countries indicating strongly that there is no correlation.

When looking at the analysis and interpretation, we can see that the r-value is small and that shows that there is no correlation and that the use of a linear model is not a good fit for the data set. The p-value is also higher than 0.05 at 0.675 meaning we fail to reject the null hypothesis and can conclude that there is no linear relationship between Singapore's population growth rate and Malaysia's population growth rate. This confirms that they don't increase together but instead grow at different rates.

Candidate 12

Conclusion

In conclusion the average temperature (in degrees Celsius) of 20 randomly selected cities in the USA were in fact higher than the average temperature (in degrees Celsius) of 20 randomly selected cities in the UK. This was shown in the "histogram of USA temp" as it shows the lowest temperature of one of the randomly selected cities in the USA was above 25 degrees Celsius and the highest temperature of one of the randomly selected cities in the USA was above 40 degrees Celsius, this was also shown in the "histogram of UK temp" as it shows there was 3 cities with an average temperature of 16 degrees in the same month as the USA cities and the highest temperature shown for random UK cities was 21 degrees, this shows how much of a gap there was from the lowest USA temperature and the highest UK temperature- USA's lowest temperature was still

4 degrees Celsius higher than the UK's highest temperature. In the boxplot shown earlier we can see the clear difference between the temperatures in randomly selected cities from both UK and USA, we can see the difference between both of the highest temperatures recorded from the randomly selected cities in both UK and USA is outstandingly huge as the UK's highest temperature shown in the boxplot is just above 22 degrees on average and the USA's highest temperature is just above 42 degrees, this shows the difference between both of the highest temperatures in 20 degrees, USA's highest temperature is 20 degrees higher than the UK's highest temperature. We can also see the clear difference of temperatures from the mean temperature on average randomly selected UK cities (18.68421 degrees Celsius) and of mean temperatures from the randomly selected USA cities (31.47368). We can also reject the null hypothesis as the Pvalue is below 0.05 which means it is not down to chance that USA cities even though they're randomly selected have a higher average temperature than randomly selected UK cities in the same month and year. After looking at all of the data that has been thoroughly researched through a very valid website we can accept that the claim I stated at the start of the project which was that on average and overall USA cities have a higher temperature than UK cities.

Candidate 13

Conclusion

In conclusion, there is a clear difference in in average show attendance between the Broadway shows Chicago and The Lion King. The two-sample t-test proves that there is evidence of a significant difference as the p-value is <0.05, it is 2.2e-16. The confidence interval doesn't span zero, as it is -5668.627 and -5407.097. That supports the fact that there is a significant difference. Using the results of the boxplot as well as mean and standard deviation, we're able to answer our question and say that on average, The Lion King is a more popular show than Chicago.