UNIVERSITY OF LEEDS

School of Mathematics

Declaration of Academic Integrity
for Individual Pieces of Work

I am aware that the University defines plagiarism as presenting someone else's work as your own. Work means any intellectual output, and typically includes text, data, images, sound or performance.

I promise that in the attached submission I have not presented anyone else's work as my own and I have not colluded with others in the preparation of this work. Where I have taken advantage of the work of others, I have given full acknowledgement. I have read and understood the University's published rules on plagiarism and also any more detailed rules specified at School or module level. I know that if I commit plagiarism I can be expelled from the University and that it is my responsibility to be aware of the University's regulations on plagiarism and their importance.

I re-confirm my consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the EU/EEA) to monitor breaches of regulations, to verify whether my work contains plagiarised material, and for quality assurance purposes.

I confirm that I have declared all mitigating circumstances that may be relevant to the assessment of this piece of work and that I wish to have taken into account. I am aware of the School's policy on mitigation and procedures for the submission of statements and evidence of mitigation. I am aware of the penalties imposed for the late submission of coursework.

Student Signature ______________________ Date 06-05-2021

Student Name Eshana Shrivastava Student Number 201484781

Please note.
When you become a registered student of the University at first and any subsequent registration you sign the following authorisation and declaration:
"I confirm that the information I have given on this form is correct. I agree to observe the provisions of the University's Charter, Statutes, Ordinances, Regulations and Codes of Practice for the time being in force. I know that it is my responsibility to be aware of their contents and that I can read them on the University web site. I acknowledge my obligation under the Payment of Fees Section in the Handbook to pay all charges to the University on demand.
I agree to the University processing my personal data (including sensitive data) in accordance with its Code of Practice on Data Protection http://www.leeds.ac.uk/dpa . I consent to the University making available to third parties (who may be based outside the European Economic Area) any of my work in any form for standards and monitoring purposes including verifying the absence of plagiarised material. I agree that third parties may retain copies of my work for these purposes on the understanding that the third party will not disclose my identity."

Motive comparability with Age of first kill

Eshana Shrivastava - 201484781

Introduction

Serial killing is the act of un-lawful murder of at least two people carried out by the same person(s) in separate events occurring at different times and sometimes place. In this coursework we will be researching and analyzing on background and motive behind killings. We are going to analyze the dataset based on a sample of serial killer from the Radford/Florida Gulf Coast University Serial Killer Database. It consists of 14,773 victim profiles and over 500 documents. We will be considering age at which serial killers commit their first murder. By previous research we know population mean of μ=27 years, with a variance of $\sigma2$=74, for serial killers who were active during the 1900s.

Our main research Question here is - Does the average age at first murder differ between killers with different motives? Motives discussed here- 'Enjoyment or power', 'Escape or avoid arrest' & 'Gang, cult or organized crime'

Results

1. **Data Cleaning:**

We will start by cleaning our dataset. First, we will be removing NA rows for motive column, Age of first kill with no values and also removed rows which has data for killers who killed before 1900 for that we calculated the age of each killer by adding year born and age of first kill and then removed the resultant.

For X=86, total dataset sample is 840.For rows which has no data for first age kill we removed 9 rows (831). 0.010%. For rows which has no data for motive we again removed 6 rows (825). 0.007%. Dataset with x=86 had 7 such rows. 0.008% (818) We also removed a row which had age of last kill < age of first kill making career duration as -1, which is an anomaly and not possible. After cleaning, total rows removed – 23, 0.027%. A new column value was introduced to calculate the career duration of each killer. This was done by calculating the number of years between age of first kill and last kill.

2. **Data Exploration:**

Numerically and graphically summarise the distribution of three variables: age at first kill, age at last kill, and career duration.

Summary	Age of first kill	Age of Last kill	Career duration
Q1	24	27	0
Median	29	33	1
Mean	29.871	34.03	4.15
Q3	34.75	40	6
IQR	10.75	13	6
SD	8.371	9.59	5.83
Variance	70.195	92.007	34.066

Table 1: Numerically summarizing the values of the three variables

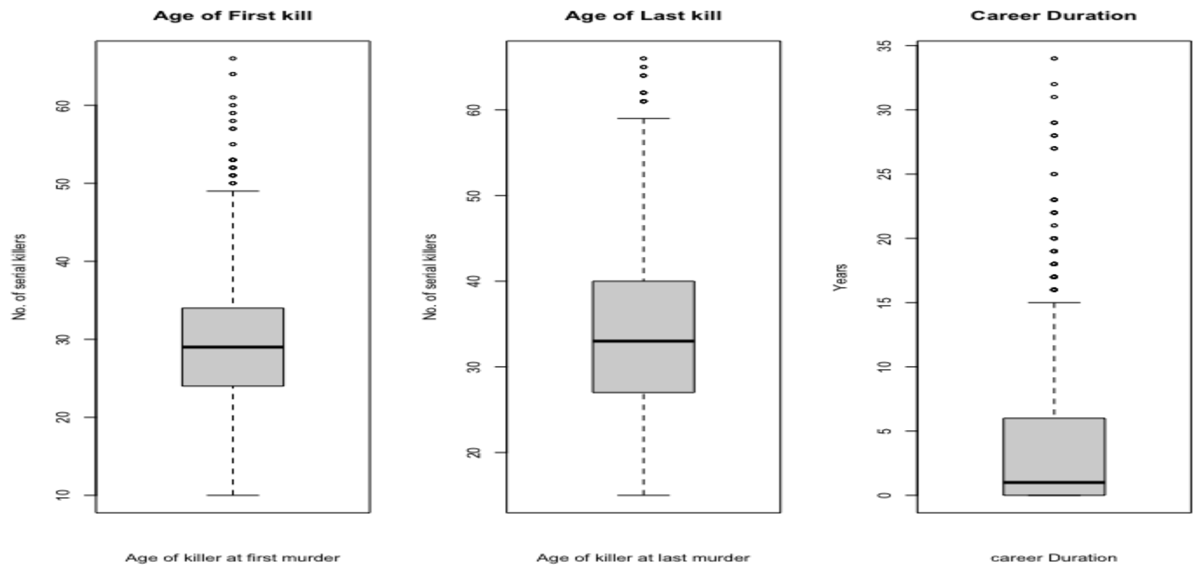

Figure 1: Boxplot for all the three variables

In Table 1 we can see that mean is greater for all values than median, which suggests positive skewness of the data which is verified by figure 1. The difference between Career duration means and median is the highest, the mean is almost 4 times the median. Which means data for career duration is more skewed than others. We do have outliers in figure 1, which means even if the average stats say for example, age of first kill for most killers was 25-30 there were many who started later as well. IQR of first age kill and last kill is almost similar showing that the spread is almost similar. The correlation between age of first kill and last is 0.73(strong corelation) showing strong normal distribution, correlation between age of first kill and career duration -0.12 and between age of last kill and career duration it is 0.50 (moderate correlation)

Figure 2 shows the motives of the murders graphically, clearly 'Enjoyment or power' is the motive for most of the killers from the dataset (X-86).

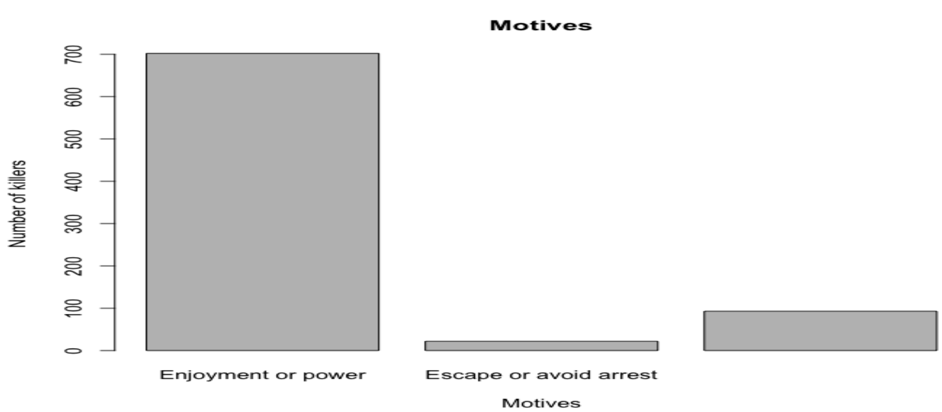

Figure 2: Bar plot representation of all the three motives we will analyze

By exploring the data further, we can see that killers with motive ‘Enjoyment or power’ started killing at a younger age than other two motives. Youngest killer being 10 years old. ‘Escape of avoid arrest’ has the youngest killer at 23 age. Age of last kill is also highest for ‘Enjoyment and power’ In figure 3 we can see the histogram spread of the three columns we are analyzing. We can see that average age of first kill lies between 25-35 range and last kill between 35-45 showing that average career duration of the dataset is around 5-10 years. Which is again, verified by table 1. Career duration mostly is in span of 5-10 years but also 0 years, as in killers killed for less than a year.

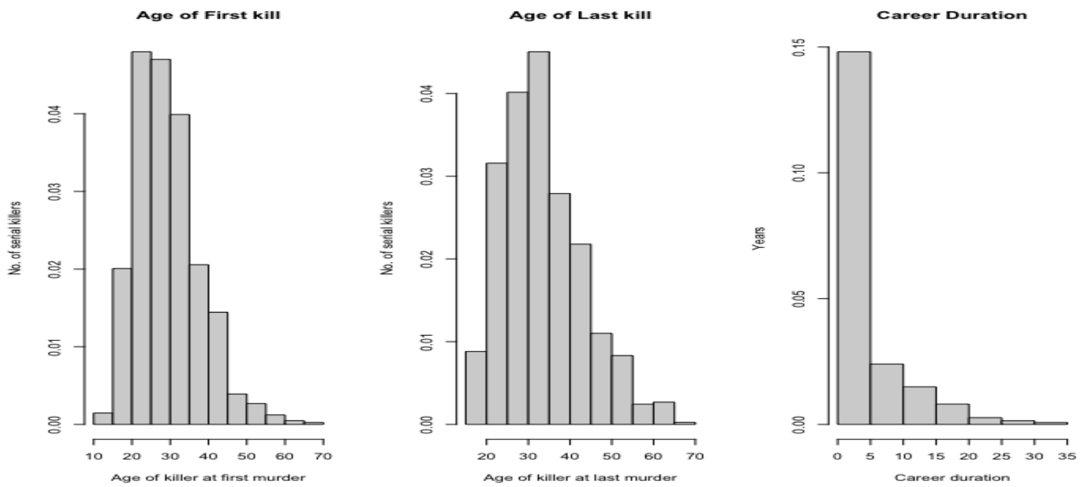

Figure 3: Histogram representation of all the three variables

3. Modelling:

Age of first kill Random variables that can take any value in an interval of ℝ are called “continuous”. Age at first kill is continuous because, in theory, it can take any value in (0,∞).

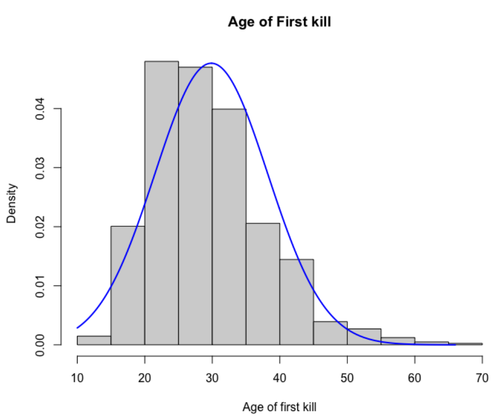

Figure 4: Showing the curve of the graph for age first kill

In figure 4 we can see that there is a positive skewness to data with few outliers. Since age is a continuous variable, it cannot be distinct at a given point in space and figure verifies the shape of curve, bell shaped. Hence, age of first kill is continuous variable with normal distribution. With mean = 29.87164, sd = 8.378261

Age of last kill Here again in figure 5 we can see that data is positively skewed with shape as bell curve and since age again has to be continuous, we would model this for normal distribution as well. mean = 34.02567 & sd = 9.592075

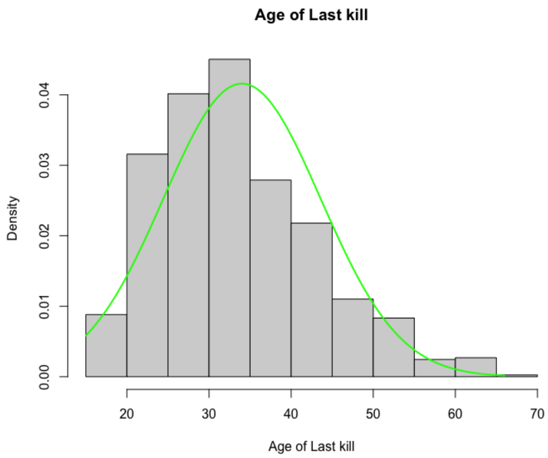

Figure 5: Showing the curve of the graph for age last kill

Career duration

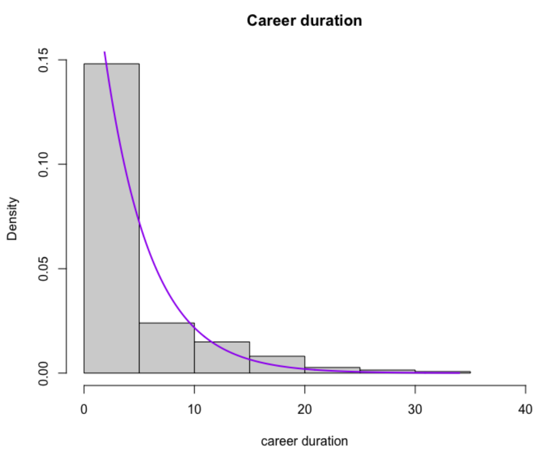

Figure 6: Showing the curve of the graph for career duration

In figure 6 we can see that career duration first of all doesn't have bell curve so it can't be normal distribution, the data is very skewed, with a long tail to the right. Since career span is in terms of years which can only be continuous random variable. We can also see that the data is positively skewed so we will proceed with exponential distribution for this.

4. Estimation:

Since now we have modeled our data for analyzing, we need to verify the parameters we have taken. We ideally need a way to sensibly estimate the parameter from a sample of observations from our distribution. We are going to apply MOM function to do the same. The estimated values calculated by MOM are almost similar to true parameters. We are passing the calculated mean and variance and lambda and estimating the values using a simulation in R.

Age of First Kill

In Figure 7, For age of first kill, we will be estimating the two unknown parameters μ and σ2 using R simulation. **True Parameters- μ- 29.85435 and σ2 -70.03636**. The **mean value- 29.86414** and estimated variance- 69.99056. Sample is a little negative biased, negligible. So, we are assuming estimator shows no bias. Since our n is large, therefore the difference in spread is negligible, making this unbiased estimator the perfect choice**, Variance- 70.126**. Red line and blue line fall on top of each other as n is large and the difference in spread of estimator and true parameter is negligible.

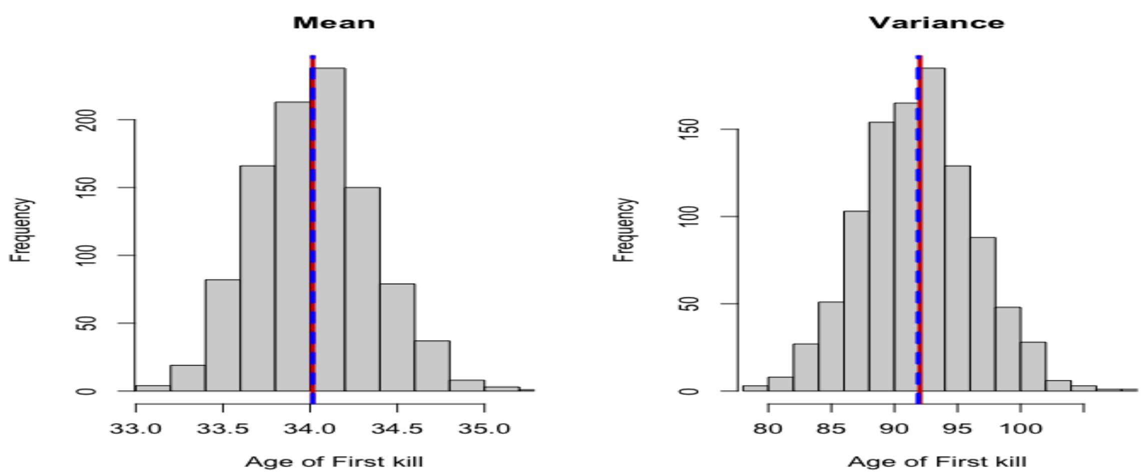

Figure 7: Graphical representation of mean and variance after getting estimated value through MOM for age first kill

Age of Last Kill

Similarly for Age of last kill, **True Parameters- μ- 34.014 and σ2 - 92.021 and Mean value - 34.013 and variance calculated - 91.860** Variance has little bias with estimated value being little less than true parameter given but it's still very close to true parameter so we will consider it. Same is shows graphically in figure 8, both mean and variance estimated are pretty close to the true parameter values.

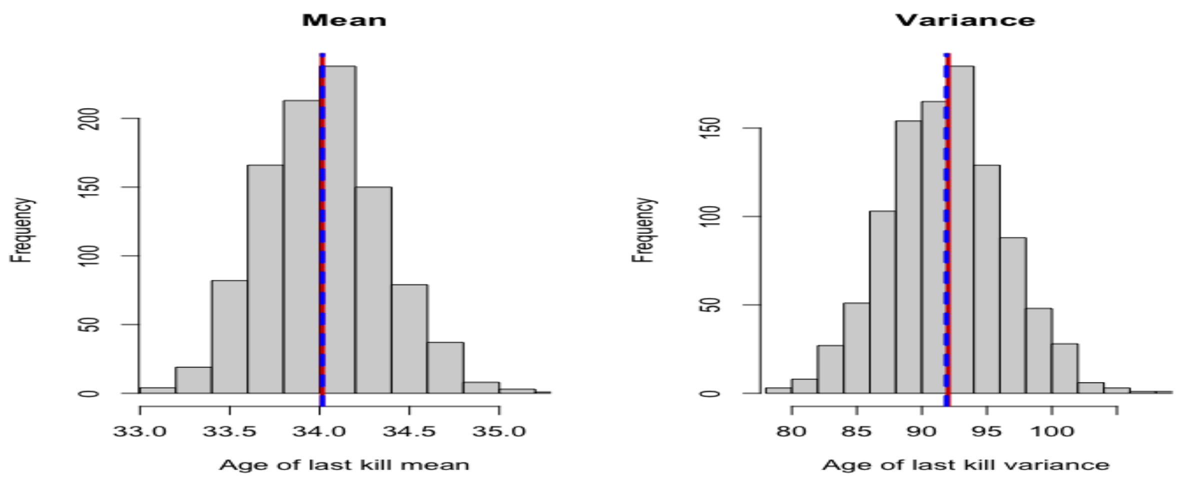

Figure 8: Figure 7 Graphical representation of mean and variance after getting estimated value through MOM for age last kill

Career Duration

Here while applying R function to simulate estimation for this exponential data, we see that we got –**lambda is 0.2403** (1/mean(dataexploration$careerduration)) and **Estimated lamda-0.2429.** If you compare this with the graph figure 9, we can see that blue line (true parameter) is to the left of red line (estimate) which suggests estimates here tends to be the overestimate of the true parameter value.

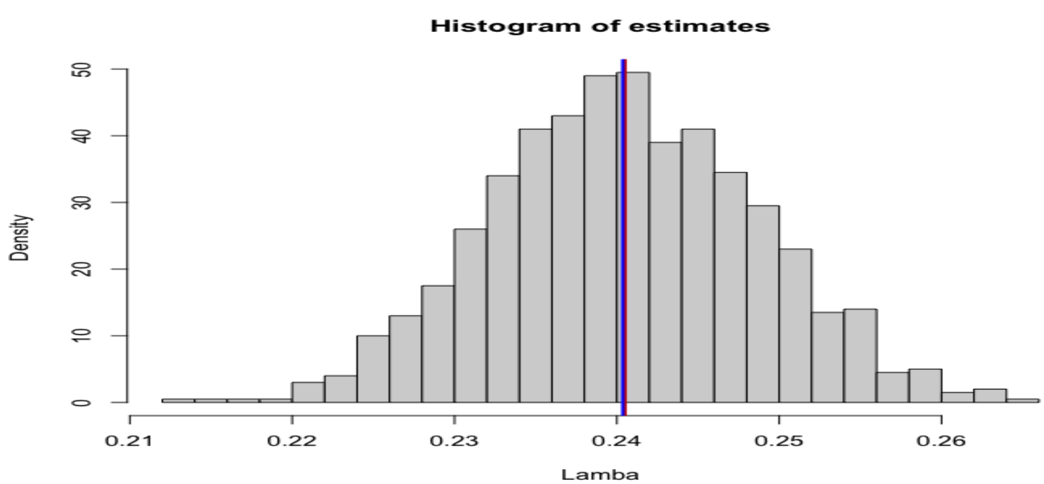

Figure 9: Figure 7 Graphical representation of mean and variance after getting estimated value through MOM for career duration

5. Testing Hypothesis

Before we test the hypothesis, we will first numerically summaries each of the three motives as shown in table2.

Summary	Enjoyment or power	Escape or avoid arrest	Gang, cult or organised crime
Q1	24	28.25	21
Median	30	31	23
Mean	30.5	33.55	24.08
Q3	35	39.5	26
IQR	11	11.25	5
SD	8.43	7.46	5.05
Variance	71.22	55.78	25.5

Table 2: numerical summarization of three motives

We used in built R function QQ-Plot, to plot the three motives and check for normal distribution, as we can see in figure 10, they are all three almost on the straight line.

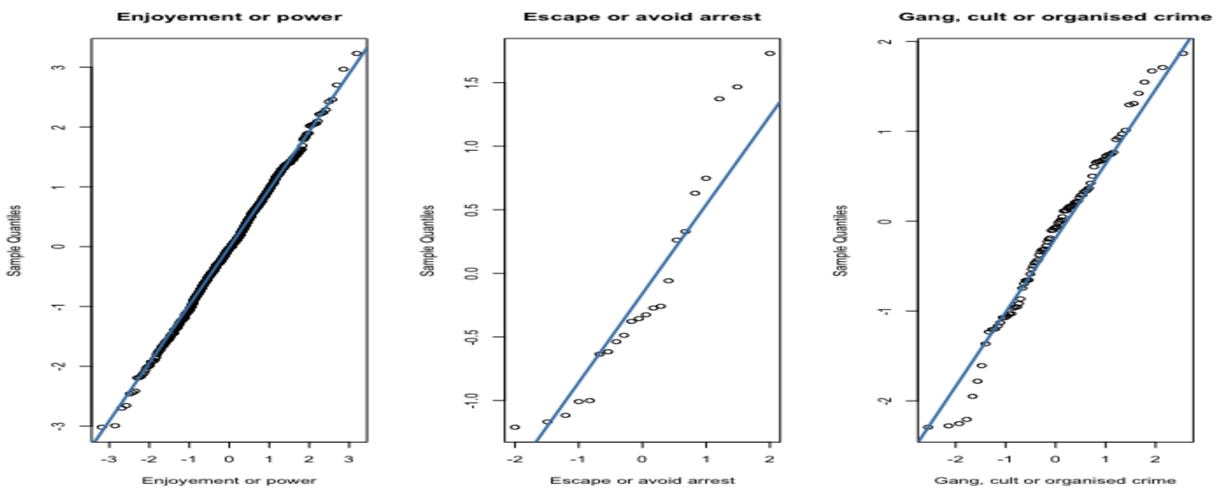

Figure 10: QQ-plot for all three motive, graphical.

Null hypothesis –H_0 Mean age for first kill for all motives is 27. Alternative hypothesis- H_1 Mean age is not 27 across motives. For hypothesis, of 'Enjoyment or power' and 'Gang, cult or organized crime' we are going to assume that the population variance and mean given at the start of document stands correct, i.e., population mean as 27 and variance as 74. In Table 2, we can see that median and mean of all three motives are comparable. IQR for Enjoyment or power and Escape or avoid arrest are similar showing similar spread.

Z-test for - 'Enjoyment or power' and 'Gang, cult or organized crime' the sample size is n=702 and n= 93 respectively similarly, since for our motive 'Escape or avoid arrest' has a sample size of n=22 (which is <30) we will go ahead with t-test for this, herein, we are assuming population variance is not known. For all the three motives are null hypothesis is – Average age of first kill for each motive is 27 and for z tests variance is 74.

Motive	Sample mean	Confidence Interval	p-value 5% (0.05)	Test	Hypothesis Result
Enjoyment or power (n 702)	30.50427	(29.86792 ,31.14062)	< 2.2e-16	Z test	Rejecting the null hypothesis
Escape or avoid arrest (n 22)	33.54545	(30.23396 ,36.85695)	0.0004989	T test	Rejecting the null hypothesis
Gang, cult or organised crime (n 93)	24.07527	(22.32694 25.82360)	0.001043	Z Test	Rejecting the null hypothesis

Table 3: Tabular representation of hypothesis test result

We performed z-test and t-test for our motives using in-build R function in the package – ‘BSDA’. We assumed a significance level of 5% for our hypothesis testing. All the three motives’ confidence interval doesn’t include 27 and the p values are less than 5% significance level for each three. We are rejecting all three hypotheses. P% for ‘Enjoyment or power’ is very small, below 5% for motive’ ‘Escape or avoid arrest’ it is 0.4% and same for ‘Gang, cult or organized crime ‘it is 0.1% so all three are below significance level of 5% also the confidence interval which is based on significance level doesn’t include population mean value, hence forth rejecting the null hypothesis for the three motives. Not only are they not 27, but they also vary among each other as well. The average age of first kill for Gang, cult or organized crime is 24 which is very young compared to the average age of first kill for motives- ‘Enjoyment or power’ and ‘Escape or avoid interest’ as 30 and 34. As collaborated in table 3 as well, all the three-confidence interval does not contain mean 27. So, we reject the hypothesis – that population mean is 27

Pair-wise comparison as seen in figure 10 & table 2 all the three motives follow normal distribution, so we are going ahead and applying – t-test two sample testing with not known variance. We chose t-test comparison testing over z-test because firstly the three motives are normally distributed and secondly, it’s an ideal method to find mean difference between two data sets. We are performing two-sided t- test because our interest doesn’t lie in the finding the greater among the two but to have mean value closest to each other at significance level 5%. Also, we have kept paired as FALSE because for that both the samples should be of same size, which isn’t the case in our dataset. We are treating each sample as independent. We are assuming variance to be not equal for all three comparison.

Null hypothesis – H_0 Average age doesn’t is same for first kill across motives. Alternative hypothesis- H_1Average age is not same across motives. Mane age for first kill for all motives is 27. Alternative hypothesis- Mean age is not 27 across motives We do so by using inbuilt R function to perform pair-wise t test. Table 4 shows the result for our analysis. For all the three comparisons we did, our null hypothesis here is that to check if mean age at first kill differs between motives is 0. (i.e., mean age doesn’t differ)

For comparing the motive- ‘Enjoyment or power’ and ‘Escape or avoid interest’ the p value is 7% which is in the confidence interval which is set at significance level of 5% so we fail to reject the null hypothesis here, i.e., the average age of first kill for motives - ‘Enjoyment or power’ and ‘Escape or avoid interest ‘are similar.

For other two comparisons, both the p values both our p values are extremely small and doesn’t cross our significance level. So, we reject our null hypothesis here, that is to say for motives- Enjoyment or power vs Gang, cult or organised crime and Enjoyment or power vs Gang, cult or organised crime the mean age differs. (mean difference is not 0)

Motive - Comparisions	timated mean differenc	Confidence Interval	p-value 5%(0.05)	Test	Hypothesis Result
Enjoyment or power vs Escape or avoid arrest	3.04118	(-0.320478 , 6.402840)	0.07403	T- test	Fail to reject the null hypothesis
Enjoyment or power vs Gang, cult or organised crime	9.47018	(6.022745, 12.917627)	6.35E-06	T- test	Rejecting the null hypothesis
Enjoyment or power vs Gang, cult or organised crime	6.429	(5.219007, 7.639002)	< 2.2e-16	T- test	Rejecting the null hypothesis

Table 4: Numerical summary of pair wise testing

Discussion

Interpretation

From the above analysis, of modelling, estimating our model and then testing our individual motives for age of first kill, we are concluding that average age each motive is not 27, same as population mean we assumed. The assumption that averages first kill age for each motive is 27 is not appropriate. For each motive our age of first kill is either below or above this. We can see for motives like 'Enjoyment and power' and 'Gang or cult' it's in early to mid 30s where to escape or avoid arrest it's mid 20s.

The majority killers have the motive of 'Enjoyment or power', about 700 or so and the average age is 30. For motive 'Escape or avoid arrest', number of killers in dataset are 22 and average first age kill is 33. And for motive- 'Gang, cult or organized crime', average age is 24. When we did comparison between each motive we saw for 'Enjoyment or power' and 'Escape or avoid arrest' our average mean age of first kill was almost with no difference, similar to each other. Given we did so by keeping significance level as 5%.

So, our real research question- **Does the average age at first murder differ between killers with different motives?** For the two motives- 'Enjoyment or power' and 'Escape or avoid arrest' our average mean is almost same to each other. But for our motive - Gang, cult or organized crime the average age is younger than rest 22. So, for two motives it's same and 1 it's different.

Real life implications of our analysis of the dataset. Average career duration for almost 29% of killers is 0, out of which maximum are with motive 'Enjoyment or power' (201/818) 24% killer with this motive killed for less than a year. Out of 22 killers with motive and 'Escape or avoid arrest' 15 have killed for less than a year (68%). This shows either the killers don't kill after a year mostly or are caught before.

Female killers make up to mere 1.71% of total population. With majority as motive- Enjoyment or power and all belonging to race as 'white' with average first age kills as 27. So, either there are less female killer populations, or they are good at not being caught. In total only three killers who pleased Insanity plea. All having motive as Enjoyment or power and average age of first kill as 42. Enjoyment and power seem to be the most prevalent reason why killer kill irrespective of race gender and age.

In general, we see average age first kill across the database differs based on different criteria.

Drawbacks

The variance assumption we made while performing hypothesis testing might be questionable as we are doing so based on our understanding and that could or could not be correct. We also didn't clean data properly there is a null value in sentence and insanity plea, though one could argue we are not analyzing our dataset based on that column, but it could skew the overall dataset's truthfulness. We had a data where age of last kill is less than age of first kill. Us assuming that variance is not as population variance would also hamper our final analysis because the confidence interval changes and so does the value. We could have analyzed more accurately had we not made so many assumptions regarding variance and mean value for population and while comparing the hypothesis.

References

-Benjamin Thorpe (b.thorpe@leeds.ac.uk) Monita Baruah(m.baruah@leeds.ac.uk)[PDF document accessed through Minerva]. MATH5741M: Statistical Theory and Methods University of Leeds.

-https://www.britannica.com/topic/serial-murder

Appendix

```
##Data Loading

load("/Users/eshanashrivastava/Desktop/Semester1/Stats/killersandmotives.rdata")

createsample(86)

View(mysample)

dataexploration <-mysample


## Data Cleaning

dataexploration=dataexploration[dataexploration$AgeFirstKill!="99999",]

View(dataexploration)

dataexploration = dataexploration[!is.na(dataexploration$Motive), ]

View(dataexploration)

dataexploration=dataexploration[dataexploration$YearBorn+dataexploration$AgeFirstKill
>=1900, ]

dataexploration$careerduration= dataexploration$AgeLastKill -
dataexploration$AgeFirstKill

dataexploration=dataexploration[dataexploration$AgeLastKill -
dataexploration$AgeFirstKill>=0, ]

View(dataexploration)


## Data Exploration

#Numerically

summary(dataexploration$AgeFirstKill)

summary(dataexploration$AgeLastKill)

summary(dataexploration$careerduration)

IQR(dataexploration$AgeFirstKill)

IQR(dataexploration$AgeLastKill)
```

```
IQR(dataexploration$careerduration)

var(dataexploration$AgeFirstKill)

var(dataexploration$AgeLastKill)

var(dataexploration$careerduration)


# Graphical

par(mfrow = c(1,3))

boxplot(dataexploration$AgeFirstKill, main= "Age of First kill", xlab = "Age of killer at first murder", ylab = "No. of serial killers")

boxplot(dataexploration$AgeLastKill, main="Age of Last kill", xlab = "Age of killer at last murder", ylab = "No. of serial killers")

boxplot(dataexploration$careerduration, main="Career Duration", xlab = "career Duration", ylab = "Years")

par(mfrow = c(1,3))

hist(dataexploration$AgeFirstKill, main= "Age of First kill", freq = FALSE, xlab = "Age of killer at first murder", ylab = "No. of serial killers")

hist(dataexploration$AgeLastKill, main="Age of Last kill", freq = FALSE, xlab = "Age of killer at last murder", ylab = "No. of serial killers")

hist(dataexploration$careerduration, main="Career Duration", freq = FALSE, xlab = "Career duration", ylab = "Years")

par(mfrow = c(1,1))

barplot(table(dataexploration$Motive), main="Motives",xlab = "Motives", ylab = "Number of killers")


#modelling

#age of first kill - SND

par(mfrow=c(1,2))

hist(dataexploration$AgeFirstKill, freq = FALSE, main='Age of First Kill',xlab = "Age of First kill")

hist(dataexploration$AgeFirstKill, freq = FALSE, main='Curve',xlab = "Age of first kill")
```

```
x <- seq(from = min(dataexploration$AgeFirstKill), to = max(dataexploration$AgeFirstKill), by = 0.1)

lines(x, dnorm(x, mean = mean(dataexploration$AgeFirstKill), sd = sd(dataexploration$AgeFirstKill)), lwd = 2, col = "blue")



#age of last kill

par(mfrow=c(1,2))

hist(dataexploration$AgeLastKill, freq = FALSE, main='Age of Last kill',xlab = "Age of Last kill")

hist(dataexploration$AgeLastKill, freq = FALSE, main='Curve',xlab = "Age of Last kill")

x <- seq(from = min(dataexploration$AgeLastKill), to = max(dataexploration$AgeLastKill), by = 0.1)

lines(x, dnorm(x, mean = mean(dataexploration$AgeLastKill), sd = sd(dataexploration$AgeLastKill)), lwd = 2, col = "green")


#career duration

mean(dataexploration$careerduration)

rate = mean(dataexploration$careerduration)^-1

par(mfrow=c(1,2))

hist(dataexploration$careerduration, freq = FALSE, xlim =c(0, 40), breaks = 10, main='Career Duration',xlab = "Career Duration")

hist(dataexploration$careerduration, freq = FALSE, xlim =c(0, 40), breaks = 10, main='Curve',xlab = "career duration")

x <- seq(from = min(dataexploration$careerduration), to = max(dataexploration$careerduration), by = 0.1)

lines( x, dexp(x, rate), lwd = 2,col = "purple")


#estimation

#age of first kill

#estimating mean
```

```
mu     <- mean(dataexploration$AgeFirstKill)
sigma   <- sd(dataexploration$AgeFirstKill)
mu
sigma
muhat1  <- rep(NA, 1000)

for(i in 1:1000){

  x <- rnorm(n = 817, mean = mu, sd = sigma)
  muhat1[i] <- mean(x)

}
mean(muhat1)
#variance
mu     <- mean(dataexploration$AgeFirstKill)
sigma   <- sd(dataexploration$AgeFirstKill)

sigma2hat1 <- rep(NA, 1000)

for(i in 1:1000){

  x <- rnorm(n = 817, mean = mu, sd = sigma)

  sigma2hat1[i] <- sd(x)^2

}

mean(sigma2hat1)
```

```
sigma^2
mu


par(mfrow = c(1, 2))
hist(muhat1, xlim = range(c(muhat1)),main='Mean',xlab = "Age of First kill")
abline(v = mu, col = "red3", lwd = 3)
abline(v = mean(muhat1), col = "blue", lty = 2, lwd = 3)
hist(sigma2hat1, xlim = range(c(sigma2hat1)), main='Variance',,xlab = "Age of First kill")
abline(v = sigma^2, col = "red3", lwd = 3)
abline(v = mean(sigma2hat1), col = "blue", lty = 2, lwd = 3)


#age of last kill
#estimating mean
mu      <- mean(dataexploration$AgeLastKill)
sigma   <- sd(dataexploration$AgeLastKill)
mu
sigma^2
muhat1  <- rep(NA, 1000)

for(i in 1:1000){


  x <- rnorm(n = 817, mean = mu, sd = sigma)
  muhat1[i] <- mean(x)


}
mean(muhat1)
#variance
mu      <- mean(dataexploration$AgeLastKill)
```

```
sigma   <- sd(dataexploration$AgeLastKill)


sigma2hat1 <- rep(NA, 1000)


for(i in 1:1000){


  x <- rnorm(n = 817, mean = mu, sd = sigma)


  sigma2hat1[i] <- sd(x)^2
}


mean(sigma2hat1)
sigma^2


par(mfrow = c(1, 2))
hist(muhat1, xlim = range(c(muhat1)),main='Mean',xlab="Age of last kill mean")
abline(v = mu, col = "red3", lwd = 3)
abline(v = mean(muhat1), col = "blue", lty = 2, lwd = 3)
hist(sigma2hat1, xlim = range(c(sigma2hat1)), main='Variance', xlab="Age of last kill
variance")
abline(v = sigma^2, col = "red3", lwd = 3)
abline(v = mean(sigma2hat1), col = "blue", lty = 2, lwd = 3)



#career duration
n <- 817
mean(dataexploration$careerduration)
lambda <- 1/mean(dataexploration$careerduration)
```

```
x <- rexp(n, lambda)      # Simulate a sample.

lambdahat <- 1/mean(x)    # Calculate the MLE.

lambdahat                 # View the MLE.

lambda


estimates <- rep(NA, 1000)   # Empty vector to store estimates.


for(i in 1:1000){


  x <- rexp(n, lambda)       # Sample in the i-th experiment.


  lambdahat <- 1/mean(x)     # Estimate in the i-th experiment.


  estimates[i] <- lambdahat  # Store the lambdahat in our vector.

}

par(mfrow = c(1, 1))

hist(estimates, breaks = "FD", freq = F, xlab="Lamba")

mean(estimates)

sd(estimates)

min(estimates)

max(estimates)

abline(v = mean(estimates), col = "red3", lw=2)

abline(v = lambda, col = "blue", lw=2)


#hypothesis

motive1 = dataexploration[dataexploration$Motive=="Enjoyment or
power","AgeFirstKill"]

motive2= dataexploration[dataexploration$Motive=="Escape or avoid
arrest","AgeFirstKill"]
```

```
motive3= dataexploration[dataexploration$Motive=="Gang, cult or organised
crime","AgeFirstKill"]

par(mfrow = c(1, 3))

set.seed(42)

x <- rnorm(motive1)

qqnorm(x, main="Enjoyement or power", xlab="Enjoyement or power"); qqline(x,
col="steelblue", lw=2)

y<- rnorm(motive2)

qqnorm(y, main="Escape or avoid arrest", xlab="Escape or avoid arrest"); qqline(y,
col="steelblue", lw=2)

z <- rnorm(motive3)

qqnorm(z, main="Gang, cult or organised crime", xlab="Gang, cult or organised crime");
qqline(z, col="steelblue", lw=2)

summary(motive1)

summary(motive2)

summary((motive3))

IQR(motive1)

IQR(motive2)

IQR(motive3)

sd(motive1)

sd(motive2)

sd(motive3)

var(motive1)

var(motive2)

var(motive3)

install.packages("BSDA")

library(BSDA)

z1 <- z.test(x = motive1, mu = 27, sigma.x = 74^(1/2), conf.level = 0.95)
```

```
z1

t1 <- t.test(x = motive2, mu = 27, sigma.x = 74^(1/2), conf.level = 0.95)

t1

z2 <- z.test(x = motive3, mu = 27, sigma.x = 74^(1/2), conf.level = 0.95)

z2

#Comparisons

EnjoymentorEscape <- t.test(x = motive2, y = motive1,  conf.level = 0.95, paired=FALSE
,alternative = "two.sided",var.equal = FALSE)

EnjoymentorEscape

EscapeorGang <- t.test(x = motive2, y = motive3,  conf.level = 0.95, paired=FALSE
,alternative = "two.sided",var.equal = FALSE)

EscapeorGang

GangorEnjoyment <- t.test(x = motive1, y = motive3,  conf.level = 0.95, paired=FALSE
,alternative = "two.sided",var.equal = FALSE)

GangorEnjoyment
```