DISPARITY IN THE AGE AT FIRST KILL OF SERIAL KILLERS ACROSS MULTIPLE MOTIVES

INTRODUCTION

In the last few years, the number of active serial killers that currently exist has been waning. However, this does not reduce the fact that these "monsters" are still to be dreaded, ever since their sharp rise in the 1980s (Weiss, 2020). A definition of a "serial killer" will be provided for the sake of clarity. A serial killer, according to Oxford's Learners dictionary, is "someone who murders several people one after the other in a similar way" (Oxford Dictionary, 2021). This paper aims to answer the question of whether the average age at the first kill of serial killers differs from motive to motive.

Data used for this analysis was acquired from the Radford/FGCU Serial Killer Database (https://www.fgcu.edu/skdb/). The data is based on a sample of serial killers who committed murders after 1900. Only a select sample of serial killers will the analysed. The following are the key variables in the data set:

- **Age at First Kill**: The age at which serial killers had their first murder.
- **Motive**: The reason(s) why serial killers committed their murders.
- **Age at Last Kill**: The age at which serial killers committed their final murder.
- **Career Duration**: The duration of a serial killer's campaign.

The variables for the age at first kill and the motive will be the major focus in subsequent sections. The sample being analysed has 3 motives; these are:

Angel of Death: Doctors, nurses, or caregivers that claim a life unjustly. (Crime Library, 2008)

Revenge or vigilante justice: Killers who killed for revenge or justice**.**

Anger (including mission-oriented killers): Killers who murdered due to anger**.**

RESULTS

The analysis of the age at first kill for each motive will be performed in this section. However, first, key variables in the data set will be briefly explored, then the motives, followed by their hypothesis tests, and finally, each motive is compared to one another.

EXPLORATION OF THE DATA

Interesting inferences can be drawn from variables in the sample. **Figure 1** shows that the duration of serial killers' careers is mostly short-lived with a few exceptions. It also tells us that the variable is exponentially distributed.

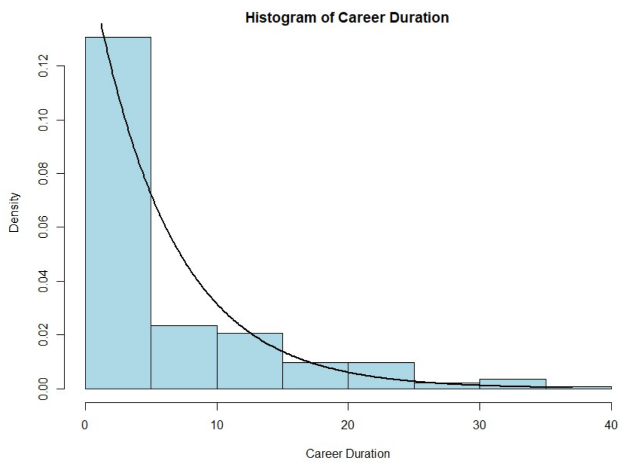

Figure 1: Histogram of Career Duration

Figure 2 shows that most serial killers in this sample commit their final murder between the ages of 20 and 30. The figure also shows the age of last kill of killers is approximately normally distributed.

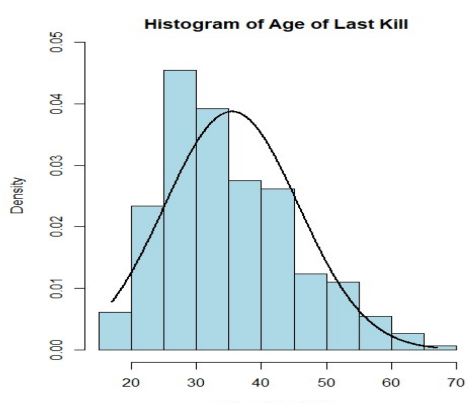

Figure 2: Histogram of serial killers' age of last kill.

EXPLORATION OF MOTIVES

The three motives for this analysis are:

- Anger (including mission-oriented kills)
- Angel of Death
- Revenge or vigilante kills

In **Figure 3**, most of the ages of serial killers at their first murder are between 20 and 30 years with only a few outliers. This is in line with the ages of the most notable serial killers in existence. (Buchanan-Dunne, 2017)

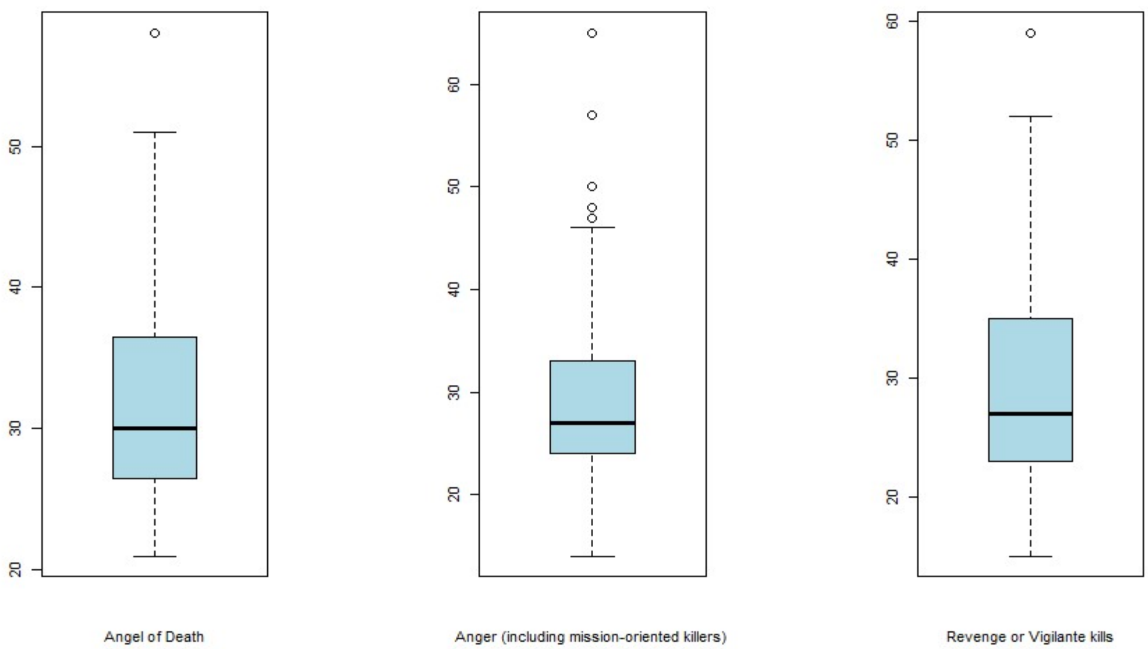

Figure 3: Boxplots of All Motives.

Figure 4, **Figure 5**, and **Figure 6** help us to determine whether the distribution of the motives is normal. From the figures, each motive can be assumed to be approximately normally distributed.

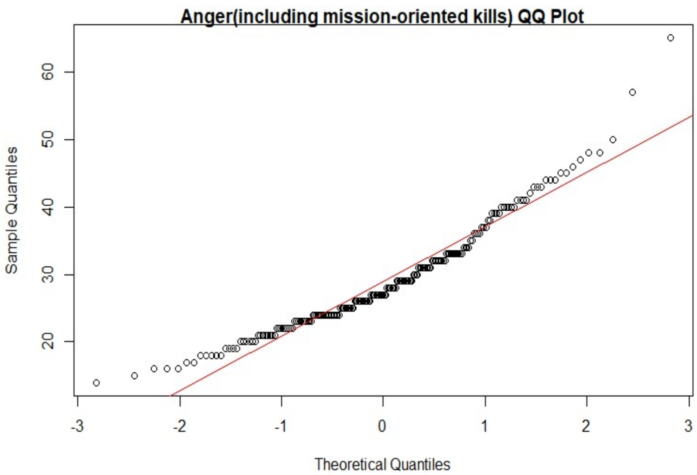

Figure 4: Anger Quantile Plot

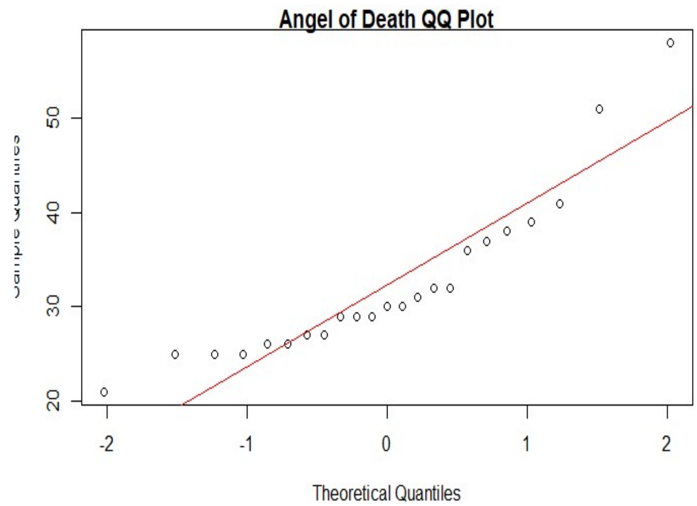

Figure 5: Angel of Death Quantile Plot

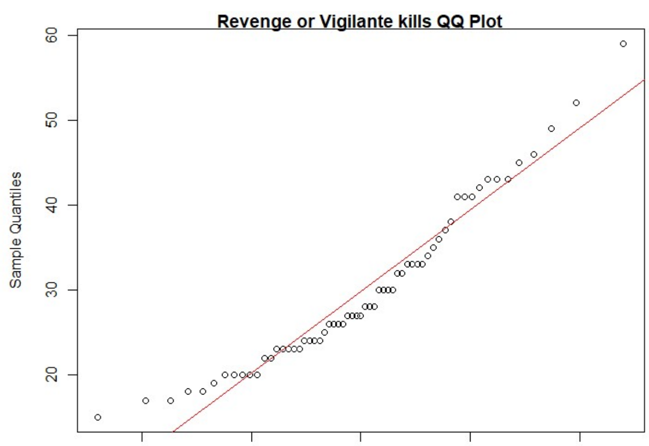

Figure 6: Revenge or Vigilante killings Quantile Plots

Table 1 gives the significant estimates and summaries that will be necessary for performing the hypothesis tests and comparing all motives. For the remainder of the report, motives and their alternative names will be used interchangeably.

Table 1:Summary Statistics of Motives

S/N	Motive	Alternative name (for conciseness)	Sample Size (n)	Sample Mean ($\bar{x}$)	Sample Variance (s^2)
1	Anger (including mission-oriented killers)	Anger	207	29.0	65.1
2	Angel of Death	Angel of Death	23	32.3	92.3
3	Revenge or vigilante justice	Revenge	61	29.9	75.8

HYPOTHESIS TESTING

The goal of this analysis is to determine if the average age at the first kill of serial killers differs from motive to motive. A good way to determine this is to perform hypothesis tests. Previous research has found that the mean age at which serial killers take their first life is 27. Therefore, the hypotheses that the average ages at first kill of serial killers for each motive is equal to the average age at first kill for the whole population of serial killers, along with its alternative, will be tested at the 5% significance level. Before tests for each motive can be determined, assumptions must be made.

Figure 4 shows the **"Anger"** motive is approximately normally distributed, and from **Table 1,** its sample size is large enough. Therefore, a **z-test** will be performed on this motive. Its sample variance will be assumed to be approximately equal to the population variance of the motive due to the sample size being large enough. **Figure 5** and **Figure 6** display the "**Angel of Death**" and "**Revenge**" motives as approximately normally distributed. Also, with their sample sizes being small from **Table 1** and unknown population variances, **t-tests** will be performed on both motives. Table 2 summarises all choices made.

Table 2: Motives' Hypothesis Test

MOTIVE	HYPOTHESIS TEST (5% SIGNIFICANCE LEVEL)
Anger	z-test
Angel of Death	t-test
Revenge	t-test

Figure 7 displays the results of the tests in the form of confidence intervals.

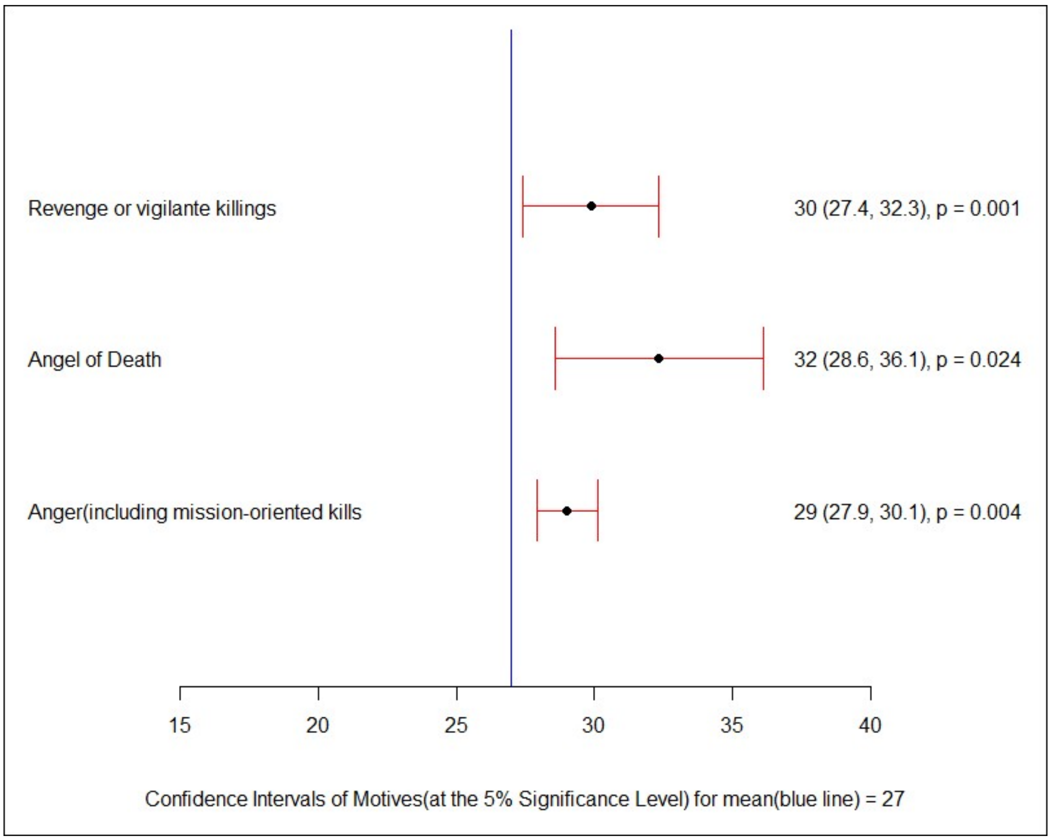

Figure 7: Forest plot of confidence intervals of Motives for the Hypothesis Test: mean = 27

Figure 7 shows that all 3 motives reject the null hypothesis. Consequently, from the confidence intervals, a conclusion can be drawn that the average age at first kill of serial killers for each motive is not equal to the average age at first kill of the whole population of serial killers.

COMPARISON OF MOTIVES

Comparison of the three motives will be done using a two independent sample t-test. Before the test, assumptions on the equality of variances must be made. Each motive will be assumed to have unequal variances. This is because each value for the age at first kill is acquired from different populations. Therefore, the hypotheses that the average ages at first kill of serial killers for each motive are not different from one another, along with its alternative, will be tested. Each test will be performed at the 5% significance level.

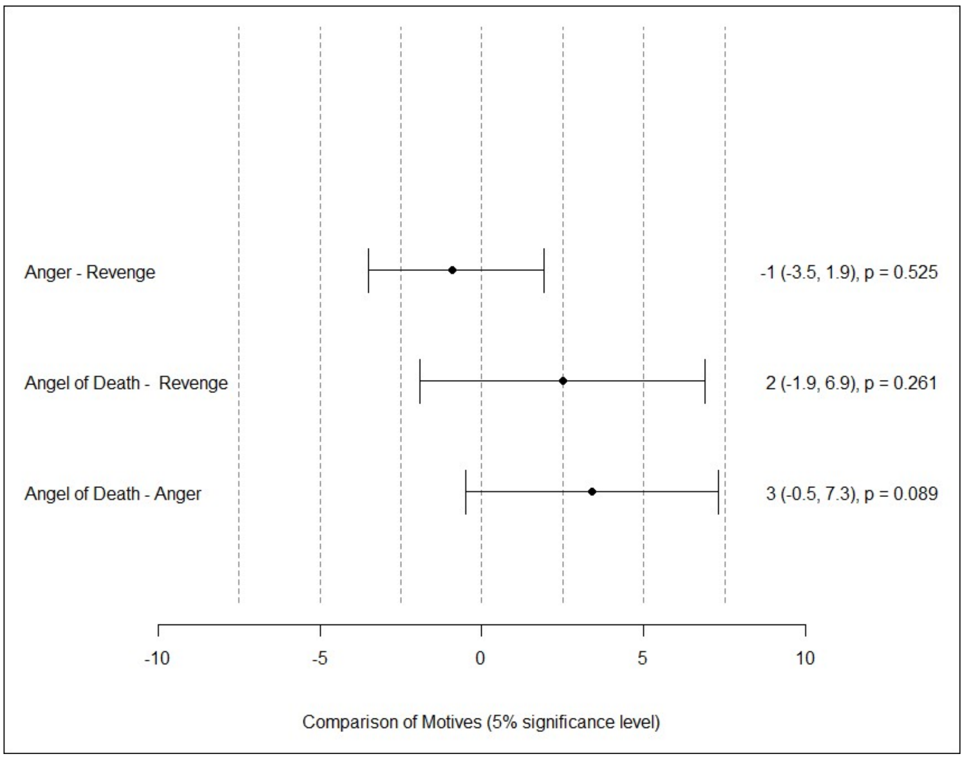

Figure 8: Caterpillar Plot showing the comparisons of motives

From the "**Angel of Death – Anger**" confidence interval in **Figure 8**, the null hypothesis is not rejected. Therefore, the statement that there is a difference in the population mean of motives is also rejected. For the "**Angel of Death – Revenge**" comparison, the confidence interval does not reject the null hypothesis. Hence, it is safe to assume that the population means of both motives are similar.

Finally, the “**Anger – Revenge”** comparison confidence interval does not reject the null hypothesis. So, a conclusion can be drawn that the average age at first kill for both motives is not different.

In summary, based on the confidence intervals, a cumulative assumption can be made that there is no difference among the average age of serial killers at their first kill for the three motives.

DISCUSSIONS AND CONCLUSION

OVERVIEW OF FINDINGS

The conclusion from the hypothesis test shows the average age of serial killers at their first kill for each motive is not equal to the population mean of the ages of the whole population of serial killers at their first kill. Also, after the comparison of motives, a conclusion that the mean age of serial killers at first kill for each motive is not different, was reached. Despite reaching these conclusions, limitations were present that may have affected the report.

LIMITATIONS

Below are the major limitations faced while performing the analysis:

A limited number of samples: With a limited sample size, the analysis carried out may not give a true representation of the serial killer world. It also made determining the distribution of each motive a bit tricky. A data set with more samples, especially for the Angel of Death motive, would have been beneficial to this analysis.

Unknown population variance of the Anger Motive: A z-test was done on the "Anger" motive based on the assumptions made. This led to the use of the sample variance for the z-test. However, there is a chance that the sample variance may not be a true representation of the population of the motive. This is because the samples for the motive may not be large enough to assume normality. Nevertheless, with a limited amount of room to maneuver, the test was still done.

Type I Error Issue: Every hypothesis test was performed at a significance level of 5%. Therefore, there is a chance that our analyses may have produced false positives. This issue could influence the one-sample hypothesis tests of the motives, which all rejected the null hypothesis ($\mu = 27$).

CONCLUSION

This analysis was performed to find if the age serial killers committed their first murder differ between motives. After exploring each motive and performing the hypothesis tests, it was identified that the average age at which serial killers take their first life, for each motive, is not similar to the average age at first kill of the whole population of serial killers. However, after comparing the average ages of first kill of each motive to one another, and observing their confidence intervals, the average ages of serial killers during their first kill for each motive are not disparate from one another.

REFERENCES

Buchanan-Dunne, M. J. (2017, April 4). *Serial Killers - What Age Do They (Usually) Start Killing?* Retrieved from Murder Mile Tours: https://www.murdermiletours.com/blog/serial-killers-what-age-do-they-usually-start-killing

Crime Library. (2008). *Angels of Death*. Retrieved from Tru Crime Library Criminal Minds and Methods: https://web.archive.org/web/20081218133027/http://www.trutv.com/library/crime/notorious_murders/angels/index.html

Oxford Dictionary. (2021). *Oxford Advanced Learner's Dictionary*. Retrieved from Oxford Learner's Dictionary: https://www.oxfordlearnersdictionaries.com/definition/english/serial-killer?q=serial+killer

Weiss, D. C. (2020, December 17). *Serial killings are waning, leading to speculation about the cause*. Retrieved from Aba Journal: https://www.abajournal.com/news/article/serial-killings-are-waning-leading-to-speculation-about-the-cause

APPENDIX

```
setwd("C:/Users/Gbenga Obasa/Desktop/MATH 5741M Coursework")
load(file = "mysample.Rdata")

newtable5 <- mysample
#REMOVING ALL ROWS WITH KILLERS WHO COMMITTED THEIR FIRST MURDER BEFORE 1900
newtable5[newtable5$AgeFirstKill + newtable5$YearBorn < 1900, ]
new_table <- newtable5[!(newtable5$AgeFirstKill + newtable5$YearBorn < 1900), ]
dim(new_table)
#####################################################################################
#REMOVING ALL ROWS WITH NULL/SPECIAL VALUES IN THE SAMPLE
new_table[new_table$AgeFirstKill==99999, ]
new_table2 <- new_table[!(new_table$AgeFirstKill==99999), ]
levels(new_table2$Motive)
new_table2[!complete.cases(new_table2$Motive),]
final_table <- new_table2[!(!complete.cases(new_table2$Motive)),]
#THE TABLE USED FOR THE ANALYSIS
final_table

#CALCULATING CAREER DURATION############################################
y <- final_table$AgeLastKill - final_table$AgeFirstKill
final_table$CareerDuration <- y
final_table

#####HISTOGRAM CAREER DURATION############################################
hist(final_table$CareerDuration, freq = FALSE, main = "Histogram of Career Duration", xlab = "Career Duration" , col = 'light blue')#histogram showing the distribution of the CareerDuration variable
x <- seq(from = min(final_table$CareerDuration), to = max(final_table$CareerDuration), by = 0.1) #x values for normal distribution
```

```
lines(x, dexp(x, rate = 1/mean(final_table$CareerDuration)), lwd = 2, col = "black")#graph showing the distribution of career duration

#####HISTOGRAM AGE LAST KILL####################

hist(final_table$AgeLastKill, freq = FALSE, main = 'Histogram of Age of Last Kill', xlab = "Age of Last Kill", ylim = c(0,0.05), col = "yellow")

w <- seq(from = min(final_table$AgeLastKill), to = max(final_table$AgeLastKill), by = 0.1)

lines(w, dnorm(w, mean = mean(final_table$AgeLastKill), sd = sqrt(var(final_table$AgeLastKill))), lwd = 2, col = "black")


###CREATING INDIVIDUAL TABLES FOR EACH MOTIVE############################################

AOfD <- final_table[final_table$Motive == "Angel of Death", ]

Revenge <- final_table[final_table$Motive == "Revenge or vigilante justice", ]

Anger <- final_table[final_table$Motive == "Anger (including mission-oriented killers)", ]


#BOXPLOTS FOR EACH MOTIVE

par(mfrow = c(1,3))

boxplot(AOfD$AgeFirstKill, xlab = "Angel of Death", cex=1.5, col = “light blue”)

boxplot(Anger$AgeFirstKill, xlab = 'Anger (including mission-oriented killers)', cex=1.5, col = “light blue”)

boxplot(Revenge$AgeFirstKill, xlab ="Revenge or Vigilante kills", cex=1.5, col = “light blue”)


#SAMPLE MEAN OF EACH MOTIVE

mu_aofd <- mean(AOfD$AgeFirstKill)

mu_rev <- mean(Revenge$AgeFirstKill)

mu_ang <- mean(Anger$AgeFirstKill)

#SAMPLE VARIANCE OF EACH MOTIVE

var_aofd <- var(AOfD$AgeFirstKill)

var_rev <- var(Revenge$AgeFirstKill)

var_ang<- var(Anger$AgeFirstKill)

#SD OF EACH MOTIVE

sd_aofd <- sd(AOfD$AgeFirstKill)

sd_rev <- sd(Revenge$AgeFirstKill)

sdr_ang<- sd(Anger$AgeFirstKill)
```

```
#QQ PLOT FOR EACH MOTIVE#############################################
qqnorm(Revenge$AgeFirstKill, main = "Revenge or Vigilante kills QQ Plot")
abline(a = mu_rev,  b = sd_rev, col = "red")
qqnorm(AOfD$AgeFirstKill, main = "Angel of Death QQ Plot")
abline(a = mu_aofd,  b = sd_aofd, col = "red")
qqnorm(Anger$AgeFirstKill, main = "Anger(Including Mission Oriented Kills) QQ Plot")
abline(a = mu_ang,  b = sd_ang, col = "red")

#############HYPOTHESIS TESTS FOR ALL
MOTIVES####################################
t.test(AOfD$AgeFirstKill, mu =27, alternative="two.sided")
t.test(Revenge$AgeFirstKill, mu =27, alternative="two.sided")
z.test(Anger$AgeFirstKill, mu =27, sigma.x = sd_ang, alternative="two.sided")

##################FOREST PLOT OF ONE SAMPLE HYPOTHESES
TESTS########################
plot(x = 0,                              # One point at (0,0).
    xlim = c(15,40), ylim=c(0, 4),        # Axis limits.
    type = "n", xaxt = "n", yaxt="n",       # No points, no axes drawn.
    xlab = NULL, ylab= NULL, ann = FALSE,   # No axis labels or numbers.
    bty="n")

#Plot a line marking the proposed value of the mean
abline(v = 27, col = "blue")
#Plot the N intervals
verticalpos = 1:3
analysis1 = c("Anger(including mission-oriented kills",
           "Angel of Death ",
           "Revenge or vigilante killings")
estimatesw  =  c(29, 32.3, 29.9)
upperw    =  c(30.1, 36.1, 32.3)
lowerw    =  c(27.9, 28.6, 27.4)
pvalw     =  c(0.004,0.024,0.001)
```

```
for(i in 1:3 ){
  lines(c(lowerw[i], upperw[i]), c(verticalpos[i], verticalpos[i]), col = 'red')
  lines(c(lowerw[i], lowerw[i]), c(verticalpos[i] + 0.2, verticalpos[i] - 0.2), col = 'red')
  lines(c(upperw[i], upperw[i]), c(verticalpos[i] + 0.2, verticalpos[i] - 0.2), col ='red')
}
  mtext(text = analysis1,  at = verticalpos,
        side = 2, line = 5, outer = FALSE, las = 1, adj = 0)
  points(estimatesw, verticalpos, pch = 16)
  axis(side = 1, cex.axis = 1)
  mtext("Confidence Intervals of Motives(at the 5% Significance Level) for mean(blue line) = 
27",
        side = 1, line = 4)
  est <- formatC(estimatesw, format='f', digits = 0)
  P <- formatC(pvalw , format = 'f', digits = 3)
  pval <- paste("p =", P)    # Type pval to see what this does.
  L <- formatC(lowerw, format = 'f', digits = 1)
  U <- formatC(upperw, format = 'f', digits = 1)
  interval <- paste("(", L, ", ", U, "),", sep = "")   # Type interval to check.
  # Putting it all together:
  results <- paste(est, interval, pval)
  # Add to the plot:
  mtext(text = results, at = verticalpos,
        side = 4, line = 5, outer = FALSE, las = 1, adj = 1)
  # Like a Christmas present, an R
  # plot belongs in a box:
  box("inner")
#####################TWO SAMPLE T-TESTS OF EACH
MOTIVE#####################################################
t.test(x = AOfD$AgeFirstKill, y = Revenge$AgeFirstKill, mu = 0, paired = FALSE, var.equal = 
FALSE, conf.level = 0.95)
t.test(x = Anger$AgeFirstKill, y = Revenge$AgeFirstKill, mu = 0, paired = FALSE, var.equal = 
FALSE, conf.level = 0.95)
t.test(x = AOfD$AgeFirstKill, y = Anger$AgeFirstKill, mu = 0, paired = FALSE, var.equal = 
FALSE, conf.level = 0.95)
```