**AGYEMANG ERIC**

**MAT 450 HOMEWORK 5**

**Generated SAS Output**

**QUESTION 6**

The GLM Procedure

|  |  |  |
| --- | --- | --- |
| Class Level Information | | |
| Class | **Levels** | **Values** |
| case | 12 | 1 2 3 4 5 6 7 8 9 10 11 12 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 36 |
| **Number of Observations Used** | 36 |

Dependent Variable: worms

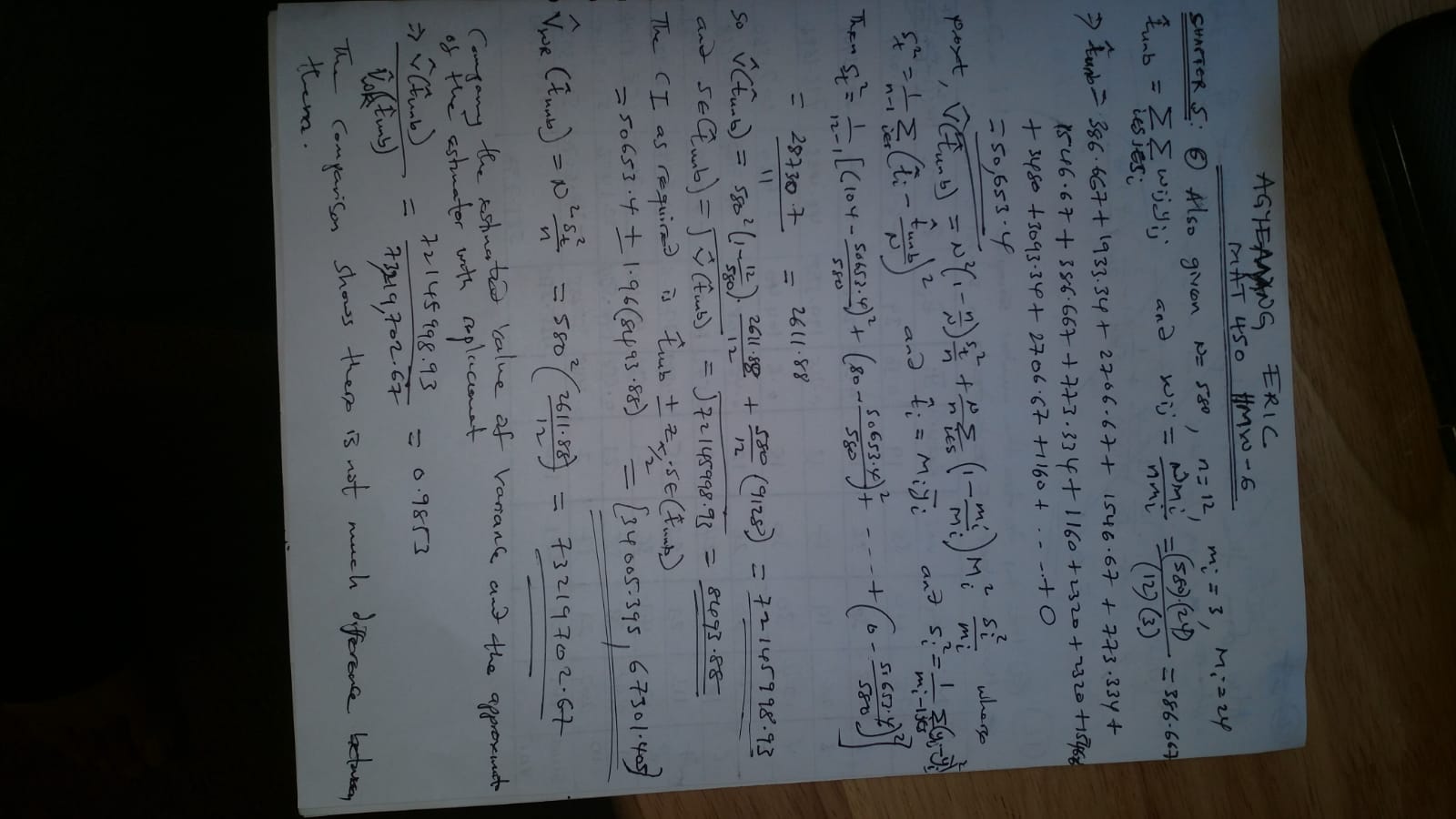
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 11 | 149.6388889 | 13.6035354 | 3.00 | 0.0117 |
| Error | 24 | 108.6666667 | 4.5277778 |  |  |
| Corrected Total | 35 | 258.3055556 |  |  |  |
|  |  |  |  |  |  |
| R-Square | **Coeff Var** | **Root MSE** | **worms Mean** |
| 0.579310 | 58.47547 | 2.127858 | 3.638889 |

|  |  |  |  |
| --- | --- | --- | --- |
| Level of case | N | worms | |
| **Mean** | **Std Dev** |
| 1 | **3** | 4.33333333 | 3.05505046 |
| 2 | **3** | 3.33333333 | 1.15470054 |
| 3 | **3** | 1.00000000 | 1.00000000 |
| 4 | **3** | 5.00000000 | 1.73205081 |
| 5 | **3** | 7.00000000 | 2.64575131 |
| 6 | **3** | 3.33333333 | 3.51188458 |
| 7 | **3** | 3.66666667 | 2.30940108 |
| 8 | **3** | 1.66666667 | 1.52752523 |
| 9 | **3** | 5.00000000 | 2.00000000 |
| 10 | **3** | 2.66666667 | 1.52752523 |
| 11 | **3** | 6.66666667 | 2.51661148 |
| 12 | **3** | 0.00000000 | 0.00000000 |

The SURVEYMEANS Procedure

|  |  |  |  |
| --- | --- | --- | --- |
| Data Summary | | | |
| Number of Clusters | | | 12 |
| Number of Observations | | | 36 |
| Sum of Weights | | | 13920 |
|  | | | |
| Statistics | | | | | |
| Variable | **N** | **Mean** | **Std Error of Mean** | **95% CL for Mean** | |
| worms | 36 | 3.638889 | 0.608324 | 2.29997721 | 4.97780057 |

**We can also calculate the following quantities.**



**QUESTION 16**

**##QUESTION 16A**

> **#The percentage of parents who returned a consent form is given by column ybar\_i in the table meas\_agg below**.

> library(dplyr)

> meas\_agg <-as.data.frame(measles %>%

+ group\_by(`School No`= school, Mi = Mitotal, ki = mi) %>%

+ summarize(Return = sum(returnf==1, na.rm = TRUE),

+ mi = sum(returnf!=9, na.rm = TRUE),

+ ybar\_i = Return/mi))

> meas\_agg

School No Mi ki Return mi ybar\_i

1 1 78 40 19 38 0.5000000

2 2 238 38 19 36 0.5277778

3 3 261 19 13 17 0.7647059

4 4 174 30 18 30 0.6000000

5 5 236 30 12 26 0.4615385

6 6 188 25 13 24 0.5416667

7 7 113 23 15 22 0.6818182

8 8 170 43 21 36 0.5833333

9 9 296 38 23 35 0.6571429

10 10 207 21 7 17 0.4117647

> N<-46

> n<-nrow(meas\_agg)

>

> meas\_agg$si\_sq<-c(.25676,.25635,.19118,.24828,.25846,.25906,.22727,.25,.23193,.25735)

>

> meas\_agg$si\_sq

[1] 0.25676 0.25635 0.19118 0.24828 0.25846 0.25906 0.22727 0.25000 0.23193 0.25735

>

> ###########################################################################

**> #QUESTION 16B**

**> #The sampling weight for each observation is given by the "weight" column in the table**

**meas\_agg1 below.**

> meas\_agg11<-as.data.frame(meas\_agg %>%

+ group\_by(`School No`, Mi, ki, Return, mi, ybar\_i, si\_sq) %>%

+ summarize(est\_ti = Mi\*Return/mi))

>

>

> ybar\_r<-sum(meas\_agg11[,"est\_ti"])/sum(meas\_agg11[, "Mi"])

> ybar\_r

[1] 0.5789482

>

> var\_ybar\_r<-(1/(Mbar^2))\*(((1-n/N)\*(sr\_squared/n))+(sum(final)/(n\*N)))

> var\_ybar\_r

[1] 0.00138099

>

> meas\_agg1 <-as.data.frame(meas\_agg11 %>%

+ group\_by(`School No`, Mi, ki, Return, mi, ybar\_i, est\_ti, si\_sq) %>%

+ summarize(squared\_deviation = (est\_ti-Mi\*ybar\_r)^2,

+ final = (Mi^2)\*(1-mi/Mi)\*(si\_sq/mi)))

> attach(meas\_agg1)

meas\_agg1$weight<-(N/n)\*(Mi/mi)

> meas\_agg1

School No Mi ki Return mi ybar\_i est\_ti si\_sq squared\_deviation final weight

1 1 78 40 19 38 0.5000000 39.00000 0.25676 37.9204848 21.08135 9.442105

2 2 238 38 19 36 0.5277778 125.61111 0.25635 148.3174465 342.34118 30.411111

3 3 261 19 13 17 0.7647059 199.58824 0.19118 2350.5770575 716.18277 70.623529

4 4 174 30 18 30 0.6000000 104.40000 0.24828 13.4176413 207.36346 26.680000

5 5 236 30 12 26 0.4615385 108.92308 0.25846 767.7722103 492.66452 41.753846

6 6 188 25 13 24 0.5416667 101.83333 0.25906 49.1251259 332.80575 36.033333

7 7 113 23 15 22 0.6818182 77.04545 0.22727 135.1244850 106.22806 23.627273

8 8 170 43 21 36 0.5833333 99.16667 0.25000 0.5557246 158.19444 21.722222

9 9 296 38 23 35 0.6571429 194.51429 0.23193 535.7193943 511.94240 38.902857

10 10 207 21 7 17 0.4117647 85.23529 0.25735 1197.6435659 595.38679 56.011765

**#QUESTION 16C**

**> #The overall percentage of parents who received a consent form along with a 95% CI is given**

**below**

> sr\_squared<-sum(meas\_agg1[,"squared\_deviation"])/(nrow(meas\_agg1)-1)

> ybar\_r<-sum(meas\_agg1[,"est\_ti"])/sum(meas\_agg1[, "Mi"])

> attach(meas\_agg1)

N<-46

> Mbar<-sum(Mi)/n

> var\_ybar\_r<-(1/(Mbar^2))\*(((1-n/N)\*(sr\_squared/n))+(sum(final)/(n\*N)))

> var\_ybar\_r

[1] 0.00138099

> ###point estimate of percentage of parents

> ybar\_r

[1] 0.5789482

>

> ##confidence interval for ybar\_r

>

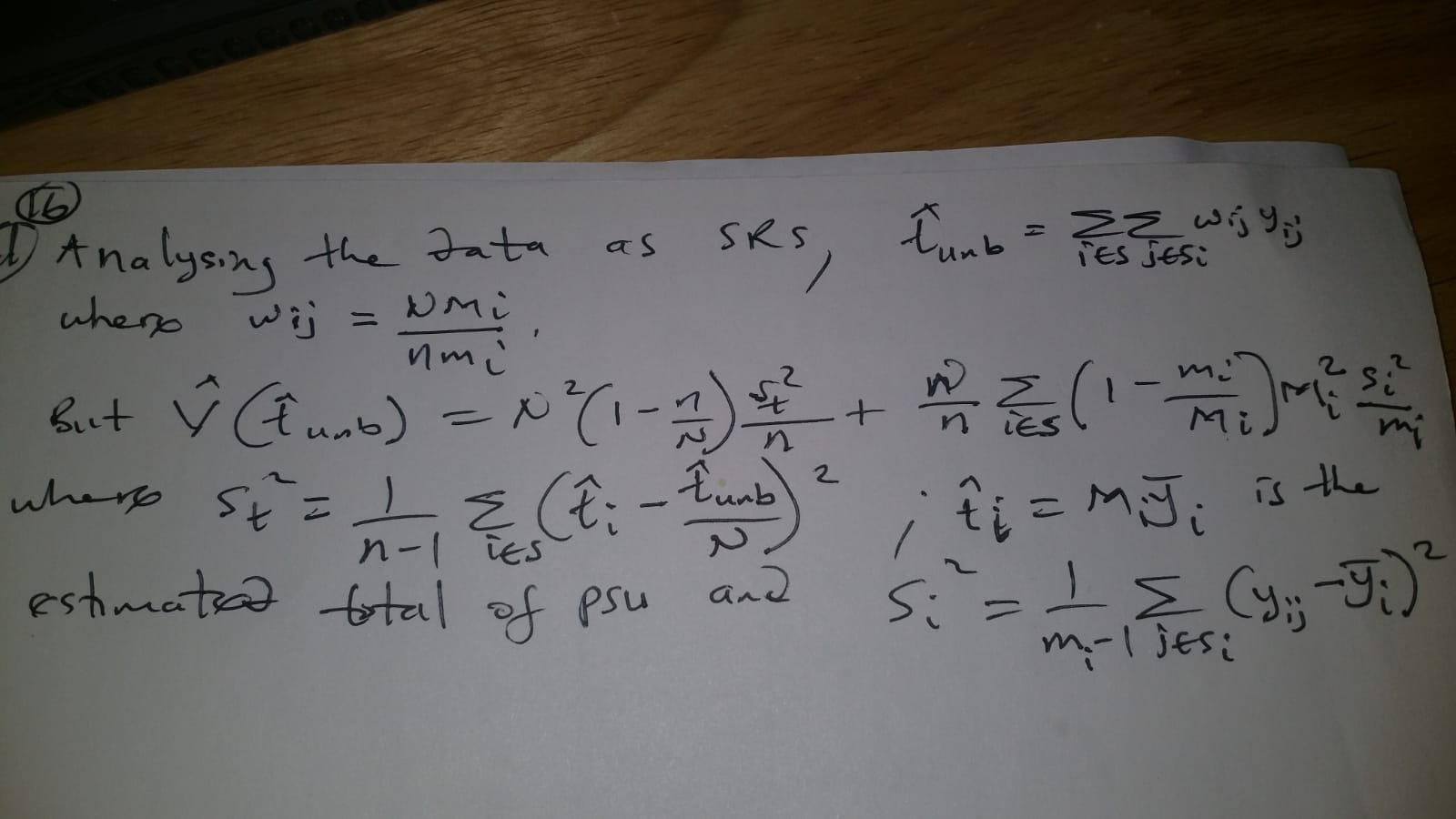
> ybar\_r-1.96\*sqrt(var\_ybar\_r); ybar\_r+1.96\*sqrt(var\_ybar\_r)

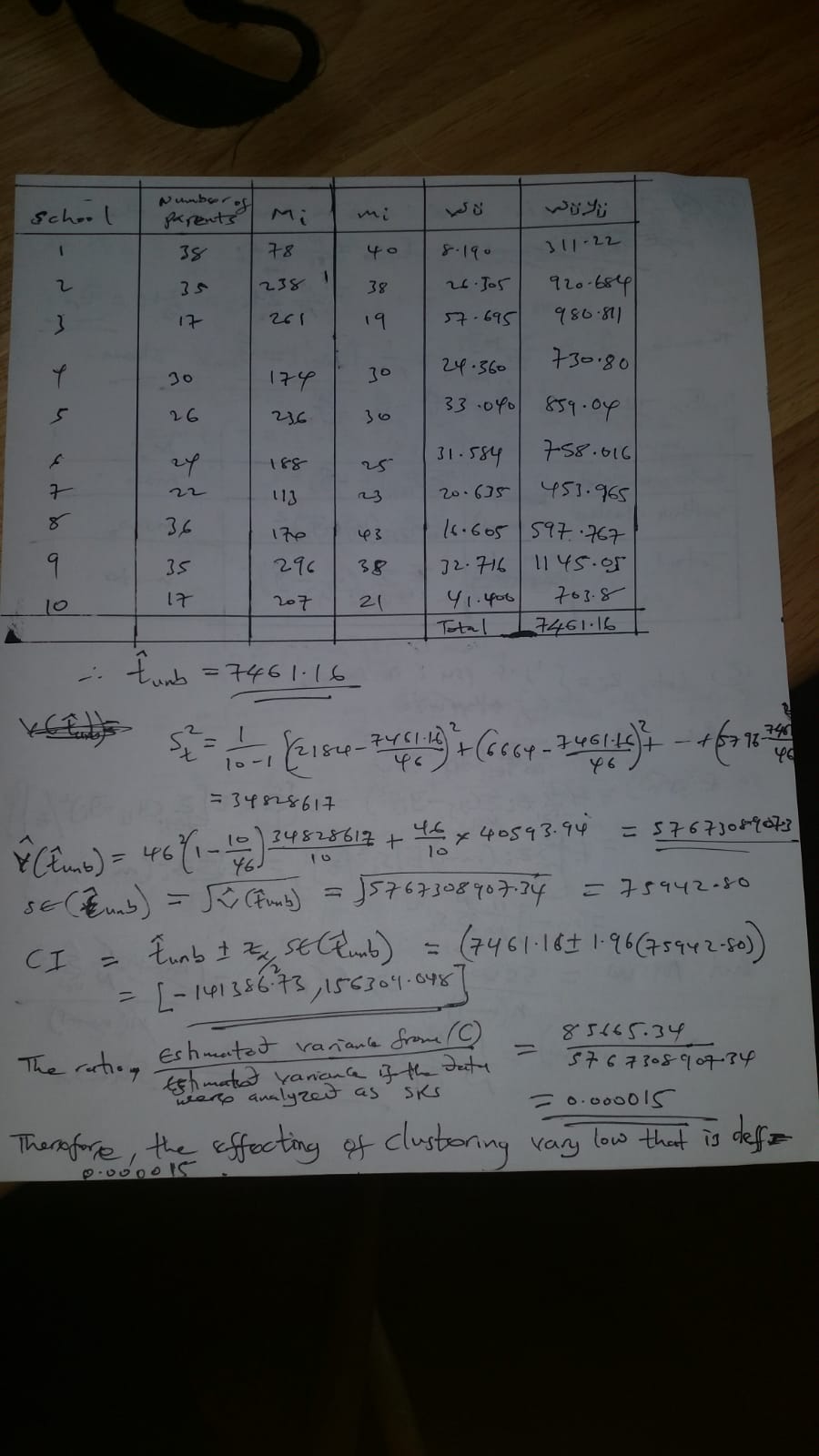
[1] 0.5061113

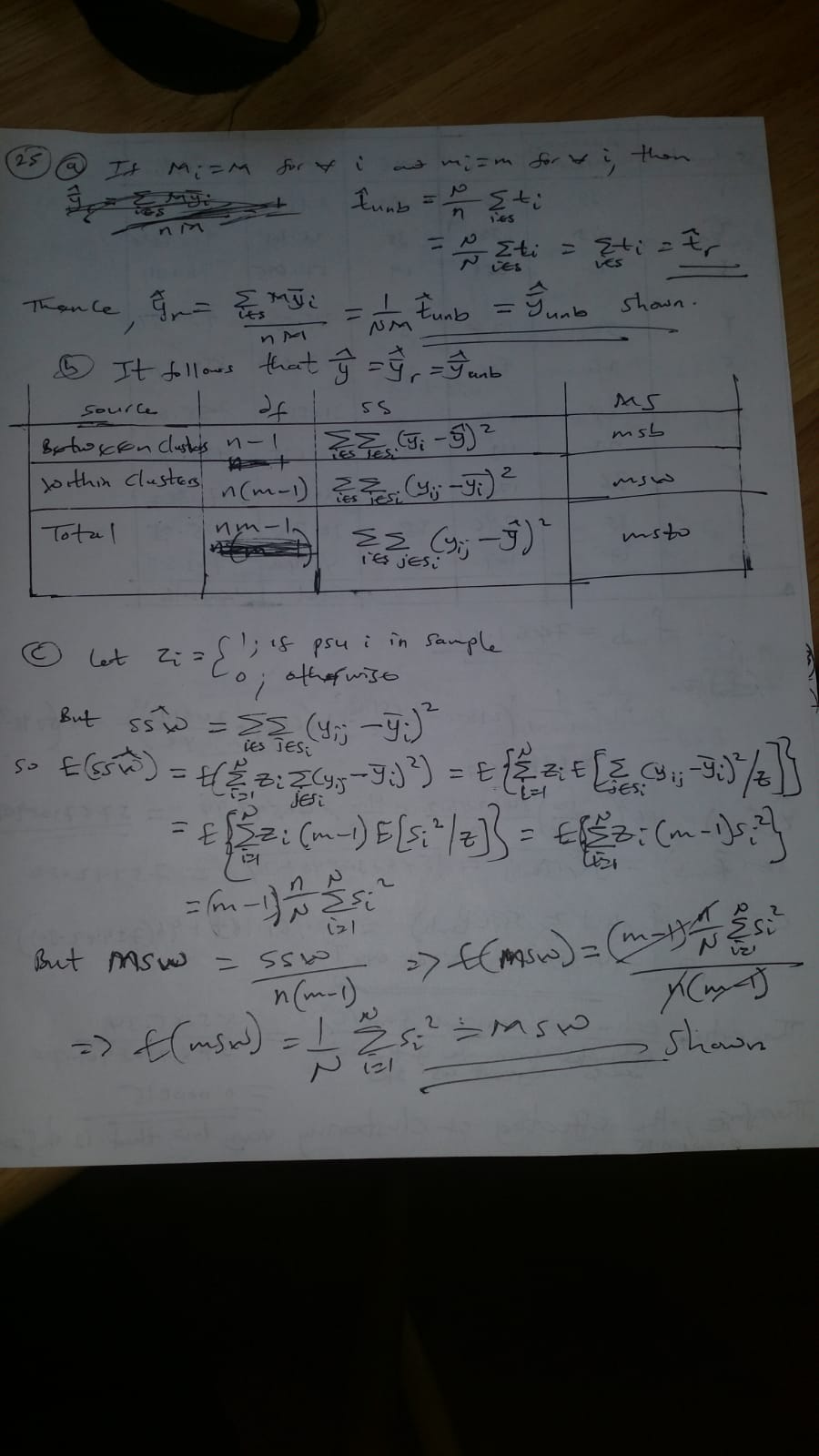
[1] 0.6517851

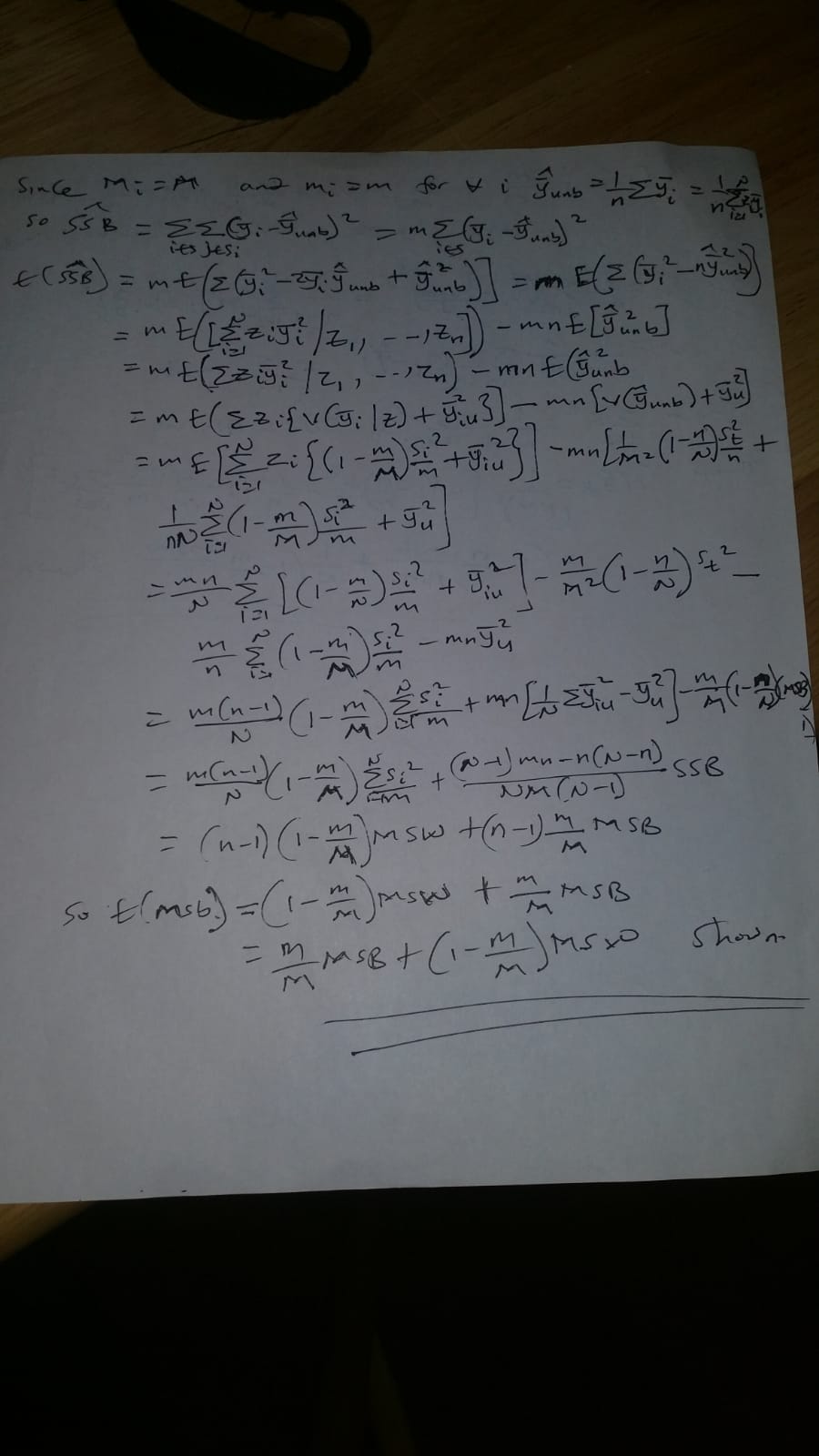
**(0.5061113,0.6517851), As the CI required.**

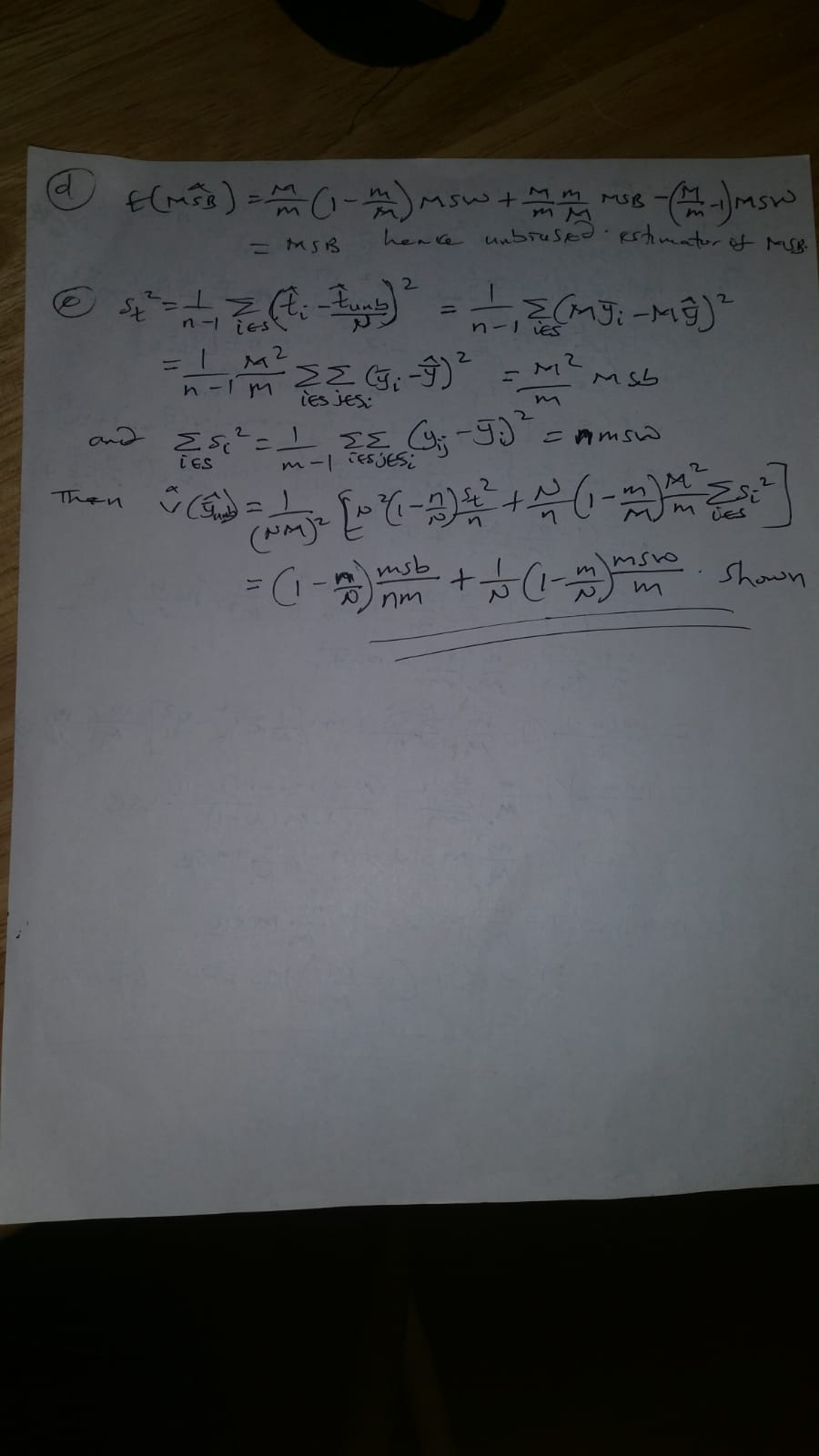
**#QUESTION 16D**

****

****

****





**/\* AGYEMANG ERIC\*/**

**/\*MAT 450 HOMEWORK \*/**

**/\*Generated SAS Code\*/**

**/\*QUESTION 6\*/**

data worms;

do case = 1 to 12;

do can = 1 to 3;

input worms @@;

wt = (580/12)\*(24/3);

output;

end;

end;

cards;

1 5 7

4 2 4

0 1 2

3 6 6

4 9 8

0 7 3

5 5 1

3 0 2

7 3 5

3 1 4

4 7 9

0 0 0

;

proc print data=worms;

run;

proc glm data=worms;

class case;

model worms = case;

mean case;

run;

/\* Due to the 2-stage sampling, SAS do not calculate the extra term for variance \*/

proc surveymeans data=worms total = 580;

weight wt;

cluster case;

var worms;

run;

**#R-CODES**

**##QUESTION 16A**

#The percentage of parents who returned a consent form is given by column ybar\_i in the table meas\_agg below.

library(dplyr)

meas\_agg <-as.data.frame(measles %>%

group\_by(`School No`= school, Mi = Mitotal, ki = mi) %>%

summarize(Return = sum(returnf==1, na.rm = TRUE),

mi = sum(returnf!=9, na.rm = TRUE),

ybar\_i = Return/mi))

meas\_agg

N<-46

n<-nrow(meas\_agg)

meas\_agg$si\_sq<-c(.25676,.25635,.19118,.24828,.25846,.25906,.22727,.25,.23193,.25735)

meas\_agg$si\_sq

####################################################################################################

**#QUESTION 16B**

#The sampling weight for each observation is given by the "weight" column in the table meas\_agg1 below.

meas\_agg11<-as.data.frame(meas\_agg %>%

group\_by(`School No`, Mi, ki, Return, mi, ybar\_i, si\_sq) %>%

summarize(est\_ti = Mi\*Return/mi))

ybar\_r<-sum(meas\_agg11[,"est\_ti"])/sum(meas\_agg11[, "Mi"])

ybar\_r

var\_ybar\_r<-(1/(Mbar^2))\*(((1-n/N)\*(sr\_squared/n))+(sum(final)/(n\*N)))

var\_ybar\_r

meas\_agg1 <-as.data.frame(meas\_agg11 %>%

group\_by(`School No`, Mi, ki, Return, mi, ybar\_i, est\_ti, si\_sq) %>%

summarize(squared\_deviation = (est\_ti-Mi\*ybar\_r)^2,

final = (Mi^2)\*(1-mi/Mi)\*(si\_sq/mi)))

attach(meas\_agg1)

meas\_agg1$weight<-(N/n)\*(Mi/mi)

meas\_agg1

##################################################################################################

**#QUESTION 16C**

#The overall percentage of parents who received a consent form along with a 95% CI is given below

sr\_squared<-sum(meas\_agg1[,"squared\_deviation"])/(nrow(meas\_agg1)-1)

ybar\_r<-sum(meas\_agg1[,"est\_ti"])/sum(meas\_agg1[, "Mi"])

attach(meas\_agg1)

N<-46

Mbar<-sum(Mi)/n

var\_ybar\_r<-(1/(Mbar^2))\*(((1-n/N)\*(sr\_squared/n))+(sum(final)/(n\*N)))

var\_ybar\_r

###point estimate of percentage of parents

ybar\_r

##confidence interval for ybar\_r

ybar\_r-1.96\*sqrt(var\_ybar\_r); ybar\_r+1.96\*sqrt(var\_ybar\_r)

#(0.5061113,0.6517851), As the CI