University of Toronto

Summer 2014

STA304 / 1003H1F:
Surveys, Sampling, and Observational Data

COURSE PROJECT - Cover Page
Written Report Due: June 19th, 6:10pm

Project Title:
A Study On Cash Carrying

All undersigned students have made a significant contribution to this project*:

First and Last Name (Print)	Student Number	Signature
Antong WU	998581798	Antong Wu
Yik Ting Wong	999017889	[illegible]
Ying Liu	998715993	Ying Liu
Yu ZHOU (Ethan)	999937537	[illegible]
Mengna Wang (Mona)	998818490	Mona
Rui Qiu	999292509	Rui Qiu

***Group Members who do not sign this sheet receive a mark of zero.**

<u>PURPOSE OF THE STUDY</u>

1) To estimate the mean amount of cash students on UTSG campus are carrying at night.

2) To discover the possible relationship that exists between amount of cash carrying and gender.

3) To discover the possible relationship that exists between amount of cash and number of credit card carrying.

<u>Target Population:</u>

All students on UTSG campus at night during weekdays.

The result of this study can be useful to organizations such as the campus police, banks, and business on or near campus. For example:

- Campus police would want to have an idea of how much cash students are carrying on campus especially at night.
- If banks think that the amount of cash students are carrying are low, then they may allocate more ATMs on campus that students can access at night.
- If local businesses think that the amounts of cash students are carrying are low, they could rent an ATM machine or/and install credit card payment systems if they have not already done so.

<u>SAMPLING METHODOLOGY</u>

<u>TWO-STAGE CLUSTER SAMPLING:</u>

Two-stage cluster sampling is suitable for our study for the following reasons:

1) We don't have the list of all students on UTSG campus at night. What we can easily obtain is a list of all night classes (including their sizes) and all major library cafeterias on campus that operates at night.

2) If we were going to randomly select some night classes and library cafeterias to sample, then it is not practical to sample everyone in them because of their large sizes. A simple random sample will be applied to each class and library we select.

3) Cluster sampling is cost-effective because students in the classrooms are close to one another and therefore the cost of sampling each student is low and constant.

<u>DATA COLLECTION:</u>

For simplicity, from now on we will call each class or library we sampled as a 'cluster'.

We assumed that how much cash students are carrying are independent of their locations on campus or what classes they are taking. Instead of randomly select clusters; we selected the following clusters that were easier for our team to sample in the 2nd stage of the sampling. More than half of our sampled students are statistics students. We believe that statistics students are more “survey-friendly” in general.

Each cluster above was selected around 6-7pm either on Tuesday or Thursday. Within limited time frame of 15 minutes, we sampled as many students as we could in each cluster. Sampling the entire cluster at the same time is not just cost-efficient but response rate could also be increased significantly. Because of the Herd mentality, students are most likely to respond to the survey if they see other students respond. To further increase response rate, we gave out free candies as a reward, and wrote a privacy disclaimer on the questionnaire as well.

For the students that did not bring their wallet or purse, we treated them as missing data.

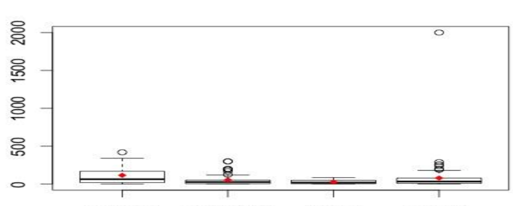
SUMMARY STATISTICS AND ANALYSIS

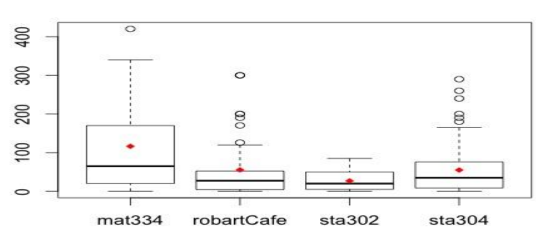
Total number of clusters available on campus on any weekdays (night classes = 70, library cafeterias = 5) = 75

Total number of clusters sampled = 4

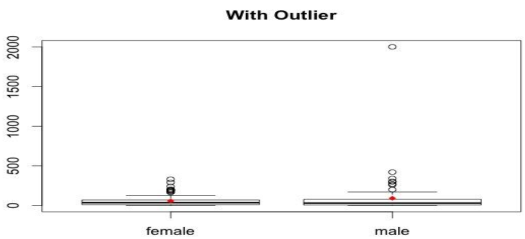
Average cluster size = 101 (it was difficult to know how many students were in the clusters at the time of sampling. However, we did know that the sizes are bounded above by their class sizes. Therefore, we assumed 100% attendance)

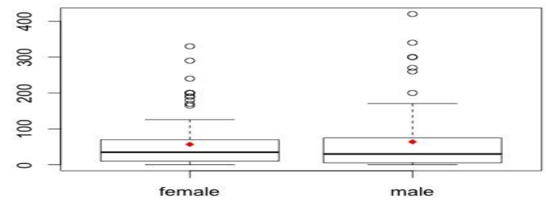
	MAT334	Robarts cafeteria	STA302	STA304
Sample size/total size	21/93	44/60	19/180	77 w outlier, 76 w/o outlier/150
Sample mean	116.44286	55.47727	26.96842	80.27338 /55.01382
Sample variance	15907.4686	6006.6817	669.2412	53102.8865 /4026.3682





Adjusted R-squared: 0.001891 (With Outlier), 0.07472 (Without Outlier)





The low Adjusted R-squared value implies that the amounts of cash students are carrying vary within the clusters but one cluster as whole is just like another cluster. This is important and good news because it suggests that our cluster sampling method and assumption about the population are appropriate.

Ratio estimate w/ outlier	Mean	Variance	95% confidence interval
Male	77.19128	665.2511	26.63806,127.74451
Female	52.28142	362.1823	14.98049, 89.58235
All	64.29222	582.0669	17.00514, 111.57931
Male -Female	24.90986	1027.433	-37.91519, 87.73492
Ratio estimate w/o outlier	**Mean**	**Variance**	**95% confidence interval**
Male	**62.37986**	**546.7696**	**16.54897,108.21074**
Female	**52.28142**	**362.1823**	**14.98049, 89.58235**
All	**56.44764**	**461.1091**	**14.35967, 98.53561**
Male -Female	**10.09844**	**908.9519**	**-48.99327, 69.19014**
Unbiased estimate w/ outlier	Mean	Variance	95% confidence interval
Male	83.95508	689.4655	32.49003, 135.42012
Female	78.49978	252.6164	47.34772, 109.65184
All	76.86422	436.2354	35.92717, 117.80127
Male -Female	5.455297	942.0819	-54.70367, 65.61427
Unbiased estimate w/o outlier	**Mean**	**Variance**	**95% confidence interval**
Male	**67.84582**	**697.2034**	**16.09278, 119.59885**
Female	**78.49978**	**252.6164**	**47.34772, 109.65184**
All	**67.48567**	**266.5387**	**35.48668, 99.48466**
Male -Female	**-10.65396**	**1841.206**	**-71.05949, 49.75156**

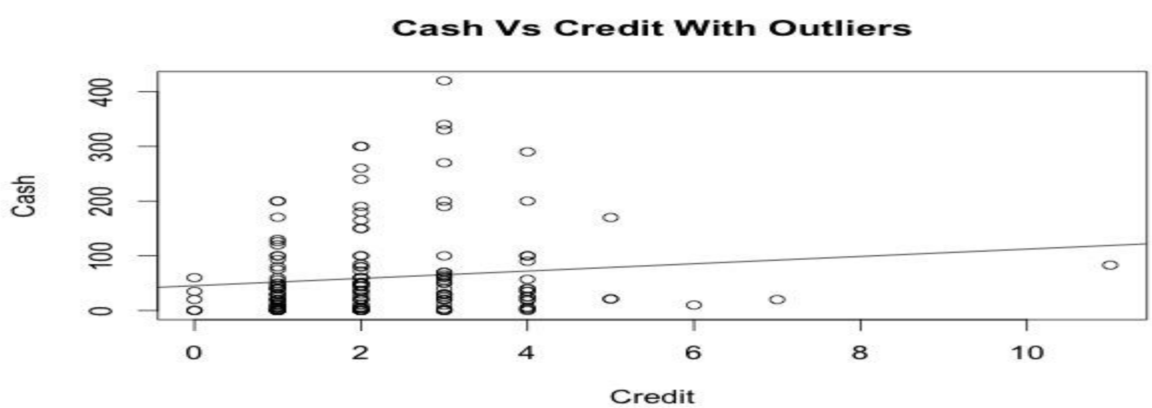
If cluster sizes vary (like in our case), and estimated cluster totals are highly correlated to their sizes, then the unbiased estimator often performs poorly. It is because the variance of the unbiased estimator is largely depended on the variance of the estimated totals; larger cluster often leads to larger estimated total (Lohr, 2010).

However, in our case, the ratio estimation of population mean happens to be less precise compare to the unbiased estimator. This is due to the fact that the cluster sizes and their estimated totals are not correlated at all (r = 0.0478).

By the unbiased estimator, we can say with 95% confidence that the average amount of cash students are carrying on UTSG campus at night is 67.49 ± 32 dollars. In other words, the interval 67.49 ± 32 has a 95% chance of capturing the true population mean. Moreover, we've also discovered that male and female students carry roughly equal amount of cash.

Null hypothesis: there exists a linear relationship between the amount of cash and the amount of credit card carrying.

Alternative hypothesis: the null hypothesis is not true.



- Estimate Std. Error t value Pr(>|t|)

Intercept 45.344 11.304 4.011 9.3e-05

credit 6.680 4.351 1.535 **0.127**

- Multiple R-squared: 0.009194, Adjusted R-squared: 0.002962

The high p-value of 0.127 suggests that assume there exists no relationship between cash and credit card, then there exists an 12.7% probability that we can observe a linear relationship as strong as our sample has obtained, just by random chance. In other words, the linear relationship between cash and credit card is not statistically significant at all.

BIASES

Measurement bias:

Most people “telescoped” – didn’t really check their wallet and only reported how much they think they carried out of convenience. This is a serious bias because we believe that young adults have a tendency to over report their wealth.

Most people rounded their amount to numbers like \$5, \$10, \$20, and \$50. We are not quite sure whether rounding give an overestimated amount or under.

In our questionnaire, we asked about how much cash one carries in CAD, hence considering the existence of many international students, we believe that some individuals are also carrying foreign currency but failed to report them.

Questionnaire also caused some confusion, and as a result, students reported how much credit cards they carry (More detail of this problem will be explained in the ‘Improvement’ section).

Under-coverage:

Our sampling frame only covered all night classes and some major library cafeterias. It did not cover elsewhere on campus, for instance, students in residence not taking night classes, Hart House, Campus Street, and etc. We believe this is not a serious under-coverage because the assumption we made about the population.

Multiple listing:

The schedule of STA304 and STA302 did not overlap, so this means that some students who were taking sta304 could also be taking sta302, therefore, the same individual could be sampled twice. (But this is highly unlikely)

Selection bias:

Did not sample every night of the week, just Tuesday and Thursday. Students may carry more cash on Friday (night life) and not so much on Monday.

Students were sampled around 6-7pm but the average amount of cash carrying would likely to go down as time goes by because some students would by dinner and snacks etc.

Non-response:

Non-response rate is extremely low. Those who didn’t respond were not because of the survey, but the limitation of time. We believe this is not related to how much cash they carry.

IMPROVEMENT

1. We could not convince most of the students to really look into their wallet and count how many cash they really carry. If we had the time we could try to verbally persuade them to do so.
2. The following question of the survey confused a lot of people:

   "How many credit cards (this includes the debit/credit cards) do you have on you right now? (Please check your wallet, purse, or bag) "

   A great proportion of students thought that "debit/credit cards" meant debit OR credit card. What we were really trying to say is that visa-debit cards should also be counted as credit cards. Design the survey questions more carefully and doing a trial survey could avoid this problem.

3. We should have consulted previous study to calculate appropriate sample size to achieve a desired level of precision within certain budget.
4. If we had the time to sample within clusters, we could have used systematic sampling; since we didn't have a list of all students in any cluster, what we did was sampling as many students as could - not a true simple random sample.
5. To avoid multiple listing, a question in the beginning of the survey can weed out those students who have already taken the survey.

CONCLUSION

Our study found that:

1. The average amount of cash students carry on UTSG at night is 67.49 ± 32 dollars.
2. Male and female students carry roughly equal amount of cash.
3. There exists no relationship between amount of cash and number of credit cards carrying.

APPENDIX

Questionnaire

What is your gender?

o Male

o Female

How much cash (in Canadian dollars) do you have **on you right now? (Please check your wallet, purse, or bag)**

How many credit cards (this includes the debit/credit cards) do you have **on you right now? (Please check your wallet, purse, or bag)**

Disclaimer: your information will only be used for our STA304 class project.

R-code and output

```
#GLOBAL VARIABLES(CONSTANT)
n=4 #total number of clusters sampled
N=75 #total numbers of clusters (average 70 night class on a given day plus some library cafeteria)
Mbar = 101 #estimated mean cluster size
Mi = c(93,60,180,150) #sampled cluster size(mat334,robartCafe,sta302,sta304)
#Mi = c(100,200,300,400)
#mi = tapply(dataNoOut$cash, dataNoOut$cluster, length)
#----------WITH ALL outliers
data = subset(clusterdata.complete, cash != "NA")
dataNoOut = subset(clusterdata.complete, cash != "NA"& cash < 2000)
boxplot(data$cash~data$cluster, main = "With Outlier")
means <- tapply(data$cash,data$cluster,mean)
points(means,col="red",pch=18)
dataNoOut1 = subset(clusterdata.complete, cash != "NA" & cash < 2000&credit >=4)
#----------WITHOUT THE 2000 outlier
# dataNoOut = subset(clusterdata.complete, cash != "NA" & cash < 2000)
# boxplot(dataNoOut$cash~dataNoOut$cluster,main = "Without Outlier")
# means <- tapply(dataNoOut$cash,dataNoOut$cluster,mean)
# points(means,col="red",pch=18)
# #----------WITHOUT THE 2000 and 400 outlier
# dataNoOut = subset(clusterdata.complete, cash != "NA" & cash < 400)
# boxplot(dataNoOut$cash~dataNoOut$cluster)
#
# means <- tapply(dataNoOut$cash,dataNoOut$cluster,mean)
# points(means,col="red",pch=18)
#
# #----------WITH ALL outliers
#
#data = subset(clusterdata.complete, cash != "NA")
boxplot(data$cash~data$gender,main = "With Outlier")
means <- tapply(data$cash,data$gender,mean)
points(means,col="red",pch=18)
#
# #----------WITHOUT THE 2000 outlier
```

```
#dataNoOut = subset(clusterdata.complete, cash != "NA" & cash < 2000)
boxplot(dataNoOut$cash~dataNoOut$gender,main = "Without Outlier")
means <- tapply(dataNoOut$cash,dataNoOut$gender,mean)
points(means,col="red",pch=18)
ybari = tapply(dataNoOut$cash,dataNoOut$cluster,mean)
ybarhatr = sum(Mi*ybari)/sum(Mi)
sqi = tapply(dataNoOut$cash,dataNoOut$cluster,var)
bigsum = sum(Mi^2 * (ybari - ybarhatr)^2)
samllsum = sum( Mi^2 *(1-mi/Mi)*(sqi/mi) )
return ((1-(n/N))*(bigsum/(n-1))*(1/(n*Mbar^2)) + (1/(n*N*Mbar^2))*samllsum)
#estimated variance for the ratio estimator
varEstimatedYBarRatio <- function(dataset, n,N,Mi,mi,Mbar) {
        ybari = tapply(dataset$cash,dataset$cluster,mean)
        ybarhatr = sum(Mi*ybari)/sum(Mi)
        sqi = tapply(dataset$cash,dataset$cluster,var)
        bigsum = sum(Mi^2 * (ybari - ybarhatr)^2)
        samllsum = sum( Mi^2 *(1-mi/Mi)*(sqi/mi) )
        return ((1-(n/N))*(bigsum/(n-1))*(1/(n*Mbar^2)) + (1/(n*N*Mbar^2))*samllsum)
 }
# varEstimatedYBarRatioDown <- function(dataset, n,N,Mi,mi,Mbar) {
#   ybari = tapply(dataset$downtime,dataset$plant,mean)
#   ybarhatr = sum(Mi*ybari)/sum(Mi)
#   sqi = tapply(dataset$downtime,dataset$plant,var)
#   bigsum = sum(Mi^2 * (ybari - ybarhatr)^2)
#   samllsum = sum( Mi^2 *(1-mi/Mi)*(sqi/mi) )
#
#   return ((1-(n/N))*(bigsum/(n-1))*(1/(n*Mbar^2)) + (1/(n*N*Mbar^2))*samllsum)
# }
varEstimatedYBarUnb <- function(testdata,n,N,Mi,mi,Mbar) {
 ybari = tapply(testdata$cash,testdata$cluster,mean)
 ybarhatUnb = sum(Mi*ybari)/(n*Mbar)
 sqi = tapply(testdata$cash,testdata$cluster,var)
 sb2 = sum((Mi*ybari - Mbar*ybarhatUnb)^2)/(n-1)
 samllsumUnb = sum( Mi^2 *(1-(mi/Mi))*(sqi/mi) )
 return (((1-(n/N))*(sb2) *(1/(n*Mbar^2))+ (1/(n*N*Mbar^2))*samllsumUnb))
}
#
# varEstimatedYBarUnbDown <- function(testdata,n,N,Mi,mi,Mbar) {
#   ybari = tapply(testdata$downtime,testdata$plant,mean)
#   ybarhatUnb = sum(Mi*ybari)/(n*Mbar)
#   sqi = tapply(testdata$downtime,testdata$plant,var)
#   sb2 = sum((Mi*ybari - Mbar*ybarhatUnb)^2)/(n-1)
#   samllsumUnb = sum( Mi^2 *(1-(mi/Mi))*(sqi/mi) )
#   return (((1-(n/N))*(sb2) *(1/(n*Mbar^2))+ (1/(n*N*Mbar^2))*samllsumUnb))
# }
#=================================================================================
=====================
# estimating adjusted Rsqr after deletion of the 2000 outlier
summary(lm(dataNoOut$cash~factor(dataNoOut$cluster)))
anova(lm(dataNoOut$cash~factor(dataNoOut$cluster)))
summary(lm(data$cash~factor(data$cluster)))
anova(lm(dataNoOut$cash~factor(data$cluster)))
# Residual standard error: 75.55 on 156 degrees of freedom
# Multiple R-squared:  0.09218,  Adjusted R-squared:  0.07472
# F-statistic:  5.28 on 3 and 156 DF,  p-value: 0.001705
```

```
#============================================================================
=====================
# ratio estimating average cash carrying for male
#sta302: 13 female, 6 male        male ratio: 32% estimated from the sample
#sta304: 50 female, 26 male       male ratio: 30%
#mat334: 6 female 14 male         male ratio: 70%
#robart: female 21, male 23   male ratio: 52%
maledata = subset(dataNoOut, gender == "male")
MiMale = c(93*0.7,100*0.52,180*0.32,150*0.3) #based on sample proportion of gender, Mi is the total
cluster size for male
miMale=tapply(maledata$gender, maledata$cluster, length) #total male students we have sampled in the
cluster
ybariMale = tapply(maledata$cash,maledata$cluster,mean)
ybarhatrMale = sum(MiMale*ybariMale)/sum(MiMale)
varYbarHatrForMale = varEstimatedYBarRatio(maledata,n,N,MiMale,miMale,Mbar/2)
CIyBarHatRmale = c(ybarhatrMale-
1.96*(varYbarHatrForMale^0.5),ybarhatrMale+1.96*(varYbarHatrForMale^0.5)) #the 95% confidence
interval for the population mean
ybarhatrMale
varYbarHatrForMale
CIyBarHatRmale
# > ybarhatrMale
# [1] 62.37986
# > varYbarHatrForMale
# [1] 546.7696
# > CIyBarHatRmale
# [1]  16.54897 108.21074
#==========================unbiasedestimateforMale=========================
ybarhatUnbMale = sum(MiMale*ybariMale)/(n*(Mbar/2))
varYbarHatUnbMale = varEstimatedYBarUnb(maledata,n,N,MiMale,miMale,Mbar/2)
CIyBarUnbMale = c(ybarhatUnbMale-
1.96*(varYbarHatUnbMale^0.5),ybarhatUnbMale+1.96*(varYbarHatUnbMale^0.5)) #the 95% confidence
interval for the population mean
ybarhatUnbMale
varYbarHatUnbMale
CIyBarUnbMale
# > ybarhatUnbMale
# [1] 67.84582
# > varYbarHatUnbMale
# [1] 697.2034
# > CIyBarUnbMale
# [1]  16.09278 119.59885
#============================================================================
=====================
# ratio estimating average cash carrying for female
#sta302: 13 female, 6 male        male ratio: 32% estimated from the sample
#sta304: 50 female, 26 male       male ratio: 30%
#mat334: 6 female 14 male         male ratio: 70%
#robart: female 21, male 23   male ratio: 52%
femaledata = subset(dataNoOut, gender == "female")
MiFemale = c(93*0.3,100*0.48,180*0.68,150*0.7)  #based on sample proportion of gender, Mi is the total
cluster size for female
miFemale=tapply(femaledata$gender, femaledata$cluster, length) #how many female students we have
sampled in the cluster
ybarifemale = tapply(femaledata$cash,femaledata$cluster,mean)
ybarhatrFemale = sum(MiFemale*ybarifemale)/sum(MiFemale)
```

```
varYbarHatrForFemale = varEstimatedYBarRatio(femaledata,n,N,MiFemale,miFemale,Mbar/2)
CIyBarHatRfemale= c(ybarhatrFemale-
1.96*(varYbarHatrForFemale^0.5),ybarhatrFemale+1.96*(varYbarHatrForFemale^0.5)) #the 95%
confidence interval for the population mean
ybarhatrFemale
varYbarHatrForFemale
CIyBarHatRfemale
# > ybarhatrFemale
# [1] 52.28142
# > varYbarHatrForFemale
# [1] 362.1823
# > CIyBarHatRfemale
# [1] 14.98049 89.58235
#=========================unbiasedestimateforFemale=========================
#ybariUnbMale = tapply(maledata$cash,maledata$cluster,mean)
ybarhatUnbFemale = sum(MiFemale*ybarifemale)/(n*(Mbar/2))
#mi=tapply(femaledata$gender, femaledata$cluster, length)
varYbarHatUnbFemale = varEstimatedYBarUnb(femaledata,n,N,MiFemale,miFemale,Mbar/2)
CIyBarUnbFemale = c(ybarhatUnbFemale-
1.96*(varYbarHatUnbFemale^0.5),ybarhatUnbFemale+1.96*(varYbarHatUnbFemale^0.5)) #the 95%
confidence interval for the population mean
ybarhatUnbFemale
varYbarHatUnbFemale
CIyBarUnbFemale
# > ybarhatUnbFemale
# [1] 78.49978
# > varYbarHatUnbFemale
# [1] 252.6164
# > CIyBarUnbFemale
# [1]  47.34772 109.65184
#==============================================================================
======================
# estimating the confidence interval for mean difference between male and female
CLdifference = c((ybarhatrMale - ybarhatrFemale)-
1.96*(varYbarHatrForFemale+varYbarHatrForMale)^0.5,(ybarhatrMale -
ybarhatrFemale)+1.96*(varYbarHatrForFemale+varYbarHatrForMale)^0.5)
CLdifference
#[1] -48.99327  69.19014
#==============================================================================
======================
# estimating average cash carrying for all using ratio estimate
mi=tapply(dataNoOut$cash, dataNoOut$cluster, length)
ybariRatio = tapply(dataNoOut$cash,dataNoOut$cluster,mean)
ybarhatr = sum(Mi*ybariRatio)/sum(Mi)
varYbarHatr = varEstimatedYBarRatio(dataNoOut,n,N,Mi,mi,Mbar)
CIyBarHatR = c(ybarhatr-1.96*(varYbarHatr^0.5),ybarhatr+1.96*(varYbarHatr^0.5)) #the 95% confidence
interval for the population mean
ybarhatr
varYbarHatr
CIyBarHatR
#
# > ybarhatr
# [1] 56.44764
# > varYbarHatr
# [1] 461.1091
# > CIyBarHatR
```

```
# [1] 14.35967 98.53561
#===============================================================================
=====================
# estimating average cash carrying for all using unbiased estimate
ybariUnb = tapply(dataNoOut$cash,dataNoOut$cluster,mean)
ybarhatUnb = sum(Mi*ybariUnb)/(n*Mbar)
mi=tapply(dataNoOut$gender, dataNoOut$cluster, length)
varYbarHatUnb = varEstimatedYBarUnb(dataNoOut,n,N,Mi,mi,Mbar)
CIyBarUnb = c(ybarhatUnb-1.96*(varYbarHatUnb^0.5),ybarhatUnb+1.96*(varYbarHatUnb^0.5)) #the 95%
confidence interval for the population mean
ybarhatUnb
varYbarHatUnb
CIyBarUnb
# > ybarhatUnb
# [1] 67.48567
# > varYbarHatUnb
# [1] 266.5387
# > CIyBarUnb
# [1] 35.48668 99.48466
#===============================================================================
=====================
# estimating the confidence interval for unbaised mean difference between male and female
CLdifferenceUnb = c((ybarhatUnbMale - ybarhatUnbFemale)-
1.96*(varYbarHatUnbMale+varYbarHatUnbFemale)^0.5,(ybarhatUnbMale -
ybarhatUnbFemale)+1.96*(varYbarHatUnbMale+varYbarHatUnbFemale)^0.5)
CLdifferenceUnb
#[1] -71.05949  49.75156

dataNoOutCredit =subset(clusterdata.complete, cash != "NA" & cash < 2000 & credit <4)
#model with outlier
model = lm(cash ~ credit, data=dataNoOutCredit)
summary(model)
plot(model)
#model.res = resid(model)
plot(dataNoOutCredit$cash~dataNoOutCredit$credit,ylab="Cash", xlab="Credit", main="Cash Vs Credit
With Outliers")
abline(model)
```

Data

cash	credit	gender	cluster
300	2	male	robart
2	1	female	robart
30	1	male	robart
10	2	female	robart
75	1	male	robart
20	7	male	robart
18	4	male	robart
0	2	male	robart
5	1	male	robart
20	1	female	robart
45	2	female	robart
10	1	male	robart
4	2	male	robart
30	1	male	robart
25	2	male	robart
20	2	female	robart
20	2	male	robart
0	0	female	robart
300	2	male	robart
100	4	male	robart
2	3	female	robart
30	2	female	robart
0.25	0	male	robart
4	4	female	robart
40	1	male	robart
0	1	male	robart

3.2	3	female	robart
30	1	female	robart
55	3	male	robart
200	4	female	robart
40	1	female	robart
190.25	2	female	robart
30	3	male	robart
50	2	female	robart
5	2	male	robart
170.4	1	male	robart
120	1	female	robart
125.9	1	female	robart
0	2	female	robart
1	1	female	robart
20	1	male	robart
50	3	female	robart
40	1	female	robart
200	1	male	robart
0	1	male	mat334
200	1	female	mat334
50	1	female	mat334
270	3	male	mat334
20.3	0	male	mat334
100	2	male	mat334
70	3	male	mat334
15	1	female	mat334
65	3	male	mat334
30	1	male	mat334
150	2	male	mat334

100	1	male	mat334
420	3	male	mat334
340	3	male	mat334
5	1	female	mat334
40	4	male	mat334
20	2	male	mat334
0	1	male	mat334
50	3	male	mat334
170	5	female	mat334
330	3	female	mat334
10	3	female	sta302
40	3	female	sta302
5	1	male	sta302
60	2	female	sta302
10	1	male	sta302
5	2	female	sta302
0	3	female	sta302
5	2	male	sta302
21	5	male	sta302
20	4	female	sta302
56	1	female	sta302
85	2	female	sta302
3	2	male	sta302
60	2	male	sta302
50	2	female	sta302
21.4	5	female	sta302
11	2	female	sta302
50	2	female	sta302
0	1	female	sta302

5.35	1	male	sta304
40	4	female	sta304
100	3	female	sta304
93	1	female	sta304
2000	3	male	sta304
35	2	female	sta304
100	4	female	sta304
0	2	female	sta304
80	2	male	sta304
20	1	female	sta304
37.2	2	female	sta304
60	1	male	sta304
80	1	female	sta304
72	2	female	sta304
12	1	male	sta304
57	4	female	sta304
83	11	female	sta304
180	2	female	sta304
0	1	male	sta304
0	4	male	sta304
100	1	female	sta304
0	1	male	sta304
60	0	female	sta304
44.25	2	male	sta304
50	2	female	sta304
35	0	male	sta304
130	1	male	sta304
290	4	female	sta304
100	2	male	sta304

25	4	female	sta304
35	4	female	sta304
45	1	male	sta304
150	2	male	sta304
0.6	2	female	sta304
60	2	female	sta304
1.9	4	female	sta304
25	3	female	sta304
0	3	female	sta304
240	2	female	sta304
90	4	female	sta304
25	1	male	sta304
20	3	male	sta304
48.75	1	male	sta304
40	1	female	sta304
0	1	female	sta304
0	2	male	sta304
70	3	female	sta304
0	3	female	sta304
25	4	female	sta304
20	1	male	sta304
35	4	female	sta304
0	1	male	sta304
29	3	female	sta304
5	2	male	sta304
62	3	female	sta304
6.15	4	male	sta304
0	1	female	sta304
190	3	female	sta304

165	2	female	sta304
60	3	male	sta304
40	1	male	sta304
45	1	female	sta304
10	6	male	sta304
20	1	female	sta304
7.25	1	female	sta304
35	2	female	sta304
1.25	1	female	sta304
20	3	female	sta304
0.35	0	female	sta304
45	2	female	sta304
30	3	female	sta304
80	2	male	sta304
20	3	female	sta304
30	3	female	sta304
200	3	female	sta304
0	2	female	sta304
260	2	male	sta304
NA	NA	female	sta304
NA	NA	female	sta304
NA	NA	female	sta304