/\* Question 1, these data are cross-sectional data \*/

proc print data=class.creditcard(obs=10); run;

proc contents data=class.creditcard; run;

proc sql;

create table aa as

select count(distinct resp\_id) AS identifiervariable

from class.creditcard;

;quit;

proc print data=aa; run;

/\* Question 2 \*/

proc means data=class.creditcard mean std max min median q1 q3 p5 p10 p25 p50 p75 p90 p95;

var age;

run;

proc univariate data=class.creditcard;

var age;

run;

/\* Question 3 \*/

/\* The outlier is age of 200 \*/

proc sgplot data=class.creditcard;

histogram age;

run;

data class.newdataset;

set class.creditcard;

where age < 100;

run;

proc print data=class.newdataset(obs=5); run;

proc sgplot data=class.newdataset;

histogram age;

run;

/\* Question 4 \*/

/\* Tabulate age. Using your results, fill in the blanks with the appropriate number of observations: Of \*/

/\* the 1,603 respondents, \_\_3\_\_ did not answer the question about age, \_\_98\_\_\_are under 21 years old, \*/

/\* \_\_83\_\_are 24 years old (the mode), 777\_\_\_are over 30, and \_\_27\_\_\_are 65 years old or older. (Again, use \*/

/\* the original age variable as these summary values are not sensitive to outliers.) \*/

proc freq data=class.creditcard;

table age;

run;

/\* Questoin 5 \*/

/\* The shape of the histogram is right skewed \*/

proc sgplot data=class.newdataset;

histogram age;

run;

/\* Question 6 \*/

/\* The shape of historgram is right skewed \*/

data class.logdataset;

set class.creditcard;

where age < 100;

log\_age = log(age);

run;

/\* Step 2: Generate the histogram \*/

proc sgplot data=class.logdataset;

histogram log\_age / binwidth=0.5;

xaxis label="Natural Log of Age";

yaxis label="Frequency";

title "Histogram of Natural Log of Age";

run;

/\* Question 7 \*/

/\* We can conclude that the second dominant card is unfixed APR card because this card has been chosen \*/

/\* by more than nearly 50% of the customers who do not choose the dominant card. The business should promote \*/

/\* more the unfixed APR card to their customers\*/

/\* Dominant card:48.85 \*/

/\* High fee card:10.04 \*/

/\* Unfixed APR card: 24.77 \*/

/\* High APR card:16.34 \*/

proc freq data=class.creditcard;

table chosen\_terms;

run;

/\* Question 8\*/

/\* We can conclude that even though people have credit card, they will still likely to choose\*/

/\* the wrong card card as the percentage of people who chose wrong card is 51.15%, from the 74.42% of people \*/

/\* who have credit card, it is reasonable more than half of the people will still make a wrong decision. \*/

proc freq data=class.creditcard;

table chosedom havecard;

run;

/\* Question 9 \*/

/\* Female make more right decisions compares to male as the percentage on female who made the right decision is more than male \*/

/\* Female who made the right decisions = 380 / 753 = 50.46%, male who made the right decisions = 403/850 = 47.41% \*/

proc freq data=class.creditcard;

table chosedom\*male / norow nocol;

run;

/\* Question 10 \*/

/\* Lets say I classify income levels as follows: Under $25000 - $49999 as low, $50000 - $99999 as medium, \*/

/\* and $100000 - $150000 or over as high. The total percentages for high are 100.03%, 98.16% for medium, \*/

/\* and 97.17% for low. Since the total percentage for high is 100.03% to choose the dominant card \*/

/\* it is reasonable to conlude higher income respondent tend to make better choices. \*/

proc freq data=class.creditcard;

table hh\_inc \* chosedom /nopercent nocol nofreq;

run;

/\* Question 11 \*/

/\* There is no clear pattern showing that higher educated respondents consistently make better choices \*/

/\* as high school respondents have a higher percentage of choosing the dominant card comapres to \*/

/\* professional degree respondents meanwhile doctoral degree respondents have the highest percentage \*/

/\* of making the right decision. \*/

proc freq data=class.creditcard;

table highest\_ed \* chosedom / nopercent nocol nofreq;

run;

/\* Question 12 \*/

/\* Based on the graph results, lets say the business's goal is to maximize profit.\*/

/\* The percentages of people who chose the dominant card, the business should avertise the credit \*/

/\* cards with the superfluous taglines and using the baseline video because this strategies \*/

/\* will mislead customers and the chances on selecting the dominant card is 36.802% \*/

proc means data=class.creditcard mean noprint;

class video tagline;

var chosedom;

output out=dominant\_card\_summary mean=choice\_share;

run;

/\* Print the results \*/

proc print data=dominant\_card\_summary;

var video tagline choice\_share;

run;

/\* Create a bar chart to visualize the results \*/

proc sgplot data=dominant\_card\_summary;

vbar video / response=choice\_share group=tagline groupdisplay=cluster datalabel;

yaxis label="Choice share of dominant card" grid;

xaxis label="Video";

keylegend / location=inside position=topright across=1;

run;

/\* Question 13 \*/

/\* The difference between 0.65 and 0.51 means the difference in the percentage of choice share of dominant card \*/

/\* between no tagline and superfluous tagline given the implemental video is launched \*/

/\* Question 14 \*/

/\* It shows that the results in figure 6 are different and have little variations by sex, although there are slight differences, there is \*/

/\* no strong evidence to conclude that results are varied by sex because the discrepancies are really small. \*/

/\* Among female respondents \_23.71\_\_percent chose the dominant card. Among male respondents \_25.14\_\_percent chose the dominant card. \*/

/\* Among female respondents that saw the implemental video and no superfluous taglines 26.58\_\_\_percent chose the dominant card. \*/

/\* Among male respondents that saw the implemental video and no superfluous taglines \_26.30\_\_percent chose the dominant card. \*/

/\* Among female respondents that saw the baseline video and superfluous taglines \_18.95\_\_percent chose the dominant card. \*/

/\* Among male respondents that saw the baseline video and superfluous taglines \_18.11\_\_percent chose the dominant card. \*/

proc freq data=class.creditcard;

table male\*chosedom;

run;

proc freq data=class.creditcard;

table male\*chosedom\*video\*tagline/ nofreq;

run;

/\* Question 15 \*/

/\* Using your cross-tabulation, fill in the blanks with the appropriate percent (rounded to the nearest first decimal place). \*/

/\* 6.1 percent of respondents quite strongly disagreed (2 or lower) with both the easiness and confidence questions.Among those respondents that quite strongly disagreed (2 or lower) with the easiness question, \*/

/\* 4.3\_\_percent quite strongly disagreed (2 or lower) with the confidence question. Among those respondents that quite strongly disagreed \*/

/\* (2 or lower) with the confidence question, 5.3 percent quite strongly disagreed (2 or lower) with the easiness question. Among those respondents that were neutral (4) on the confidence question,\_9.18 percent agreed (5 or higher) with the easiness question. \*/

proc freq data=class.creditcard;

table easy\_choice\*confidence;

run;

/\* Question 16 \*/

/\* Step 1: Random sampling of 1000 participants without replacement \*/

proc surveyselect data=class.creditcard out=class.samples method=srs sampsize=1000 seed=42;

run;

/\* Step 2: Test the results using class.samples\*/

/\* Q15 After comparing the results, I can confirm that the results are generally close to question 16. \*/

/\* Based on this finding, I will suggest the business to reduce the population size of survey to 1000 \*/

/\* to save costs as the results are relatively closed for Q15\*/

proc freq data=class.samples;

table easy\_choice\*confidence / nofreq;

run;

/\* Q15 \*/

proc freq data=class.samples;

table easy\_choice\*confidence;

run;

/\* Q14 \*/

proc freq data=class.samples;

table male\*chosedom;

run;

proc freq data=class.samples;

table male\*chosedom\*video\*tagline/ nofreq;

run;

/\* Q13 \*/

/\* The difference between 0.64 and 0.54 means the difference in the percentage of choice share of dominant card \*/

/\* between no tagline and superfluous tagline given the implemental video is launched \*/

/\* Q12 \*/

proc means data=class.samples mean noprint;

class video tagline;

var chosedom;

output out=dominant\_card\_samples mean=choice\_share;

run;

/\* Print the results \*/

proc print data=dominant\_card\_samples(obs=10);

var video tagline choice\_share;

run;

/\* Create a bar chart to visualize the results \*/

proc sgplot data=dominant\_card\_samples;

vbar video / response=choice\_share group=tagline groupdisplay=cluster datalabel;

yaxis label="Choice share of dominant card" grid;

xaxis label="Video";

keylegend / location=inside position=topright across=1;

run;

/\* Q11 \*/

proc freq data=class.samples;

table highest\_ed \* chosedom / nopercent nocol nofreq;

run;

/\* Q10 \*/

proc freq data=class.samples;

table hh\_inc \* chosedom /nopercent nocol nofreq;

run;

/\* Q9 \*/

proc freq data=class.samples;

table chosedom\*male / norow nocol;

run;

/\* Q8 \*/

proc freq data=class.samples;

table chosedom havecard;

run;

/\* Q7 \*/

proc freq data=class.samples;

table chosen\_terms;

run;

/\* Q6 \*/

data class.logdatasetsamples;

set class.samples;

where age < 100;

log\_age = log(age);

run;

/\* Step 2: Generate the histogram \*/

proc sgplot data=class.logdataset;

histogram log\_age / binwidth=0.5;

xaxis label="Natural Log of Age";

yaxis label="Frequency";

title "Histogram of Natural Log of Age";

run;

/\* Q5 \*/

proc sgplot data=class.samples;

histogram age;

run;

/\* Q4 \*/

proc freq data=class.samples;

table age;

run;

/\* Q3 \*/

proc sgplot data=class.samples;

histogram age;

run;

data class.samples1;

set class.creditcard;

where age < 100;

run;

proc print data=class.samples1(obs=5); run;

proc sgplot data=class.samples1;

histogram age;

run;

/\* Q2 \*/

proc means data=class.samples mean std max min median q1 q3 p5 p10 p25 p50 p75 p90 p95;

var age;

run;

proc univariate data=class.samples;

var age;

run;

/\* Q1 \*/

proc print data=class.samples(obs=10); run;

proc contents data=class.samples; run;

/\* After comparing the results, I can confirm that the results are generally close to between the . \*/

/\* oringinal dataset and samples dataset Based on this finding, I will suggest the business to reduce the population size of survey to 1000 \*/

/\* to save costs as the results are relatively close to the population size