What is the distribution of gender, vehicle size, and vehicle class?

**PROC** **IMPORT** OUT= WORK.CI

        DATAFILE= "H:\data\car\_insurance\_19.csv"

        DBMS=CSV REPLACE;

GETNAMES=YES;

DATAROW=**2**;

**RUN**;

**PROC** **CONTENTS**; **RUN**;

TITLE 'Distribution of Gender';

**PROC** **FREQ** DATA = WORK.CI;

TABLES Gender; **RUN**;

**PROC** **GCHART** DATA = WORK.CI; HBAR Gender; **RUN**;

TITLE 'Distribution of Vehicle Size';

**PROC** **FREQ** DATA = WORK.CI;

TABLES Vehicle\_Size; **RUN**;

**PROC** **GCHART** DATA = WORK.CI; HBAR Vehicle\_Size; **RUN**;

TITLE 'Distribution of Vehicle Class';

**PROC** **FREQ** DATA = WORK.CI;

TABLES Vehicle\_Class; **RUN**;

**PROC** **GCHART** DATA = WORK.CI; HBAR Vehicle\_Class; **RUN**;

What is the average customer life time value of each level of gender, vehicle size, and vehicle class?

TITLE 'Average Customer Lifetime Value of each Gender';

**PROC** **MEANS**; VAR Customer\_Lifetime\_Value; CLASS Gender; **RUN**;

TITLE 'Average Customer Lifetime Value for different Vehicle Sizes';

**PROC** **MEANS**; VAR Customer\_Lifetime\_Value; CLASS Vehicle\_Size; **RUN**;

TITLE 'Average Customer Lifetime Value for different Vehicle Classes';

**PROC** **MEANS**; VAR Customer\_Lifetime\_Value; CLASS Vehicle\_Class; **RUN**;

Do Large cars have a higher lifetime value than medsize cars. Do a ttest and report on your findings.

Q3 data a2;set CI;

if Vehicle\_Size='Large' or Vehicle\_Size='Medsize';

proc ttest; var Customer\_Lifetime\_Value;class Vehicle\_Size;run;

Is there a significant difference between men and women in customer life time value?

PROC ttest; VAR Customer\_Lifetime\_Value;CLASS Gender;run;

Use ANOVA to test whether there is difference in customer lifetime value across different sales channels. Which sales channel generates the highest lifetime value?

**PROC SORT** DATA = WORK.CI;BY Sales\_Channel; **run**;

**PROC ANOVA**; CLASS Sales\_Channel;

MODEL Customer\_Lifetime\_Value = Sales\_Channel; **run**;

**PROC MEANS**; VAR Customer\_Lifetime\_Value; BY Sales\_Channel; **run**;

Is there a relationship between renew\_offer\_type and response (use Chi-sq test)? Which offer type generates the highest response rate?

ods graphics on;

proc freq data = WORK.CI;

tables (Renew\_Offer\_Type)\*(Response) /chisq

plots=(freqplot(twoway=grouphorizontal

scale=percent));

run;

ods graphics off;

Do different renew\_offer\_types have different lifetime values? Which offer type is the best?

TITLE 'Renew offer types and corresponding Customer Lifetime';

**PROC** **ANOVA**; CLASS Renew\_Offer\_Type;

MODEL Customer\_Lifetime\_Value = Renew\_Offer\_Type; **RUN**;

**PROC** **MEANS**; VAR Customer\_Lifetime\_Value; CLASS Renew\_Offer\_Type; **RUN**;

Is the effectiveness of renew\_offer\_type different across different states with respect to lifetime value?

**PROC** **MEANS**; VAR Customer\_Lifetime\_Value; CLASS State Renew\_Offer\_Type; **RUN**;