YINGQIAO GOU
STAT430
HW3

```
DATA HW3;
INFILE '/home/u50368724/my_shared_file_links/schimiak/OldClassData.csv'
delimiter=',' dsd;
/* Use this stmt to deal with csv file */
INPUT Subject $
        Gender $
        Phone $
        Campus $
        Grade $
        Car $
        Optimist
        Math
        Siblings
        Pets
        Credit_Hours
        Social_Media
        Extra_Curricular
        Height
        HS_GPA
        Exercise
        Time_To_Get_Ready
        Distance
        LETTER_GRADE $;

Q1:
PROC FORMAT;
VALUE $New_Gender ‘M’=‘Male’
                  ‘F’ =‘Female’;
RUN;

PROC FORMAT;
VALUE NEW_MATH 1 ="I really like math."
               2="I somewhat like math."
               3="I could take math or leave it."
               4="I really don't like math."
               5="I'd rather have a root canal.";
RUN;

PROC PRINT;
FORMAT Gender $New_Gender.;
FORMAT Math NEW_MATH.;
RUN;
```

Q2:

```
IF HS_GPA GT 4.0 THEN LETTER_GRADE = 'A';
ELSE IF HS_GPA GE 3.0 AND HS_GPA LT 4.0 THEN LETTER_GRADE = 'B';
ELSE IF HS_GPA GE 2.0 AND HS_GPA LT 3.0 THEN LETTER_GRADE = 'C';
ELSE IF HS_GPA GE 1.0 AND HS_GPA LT 2.0 THEN LETTER_GRADE = 'D';
ELSE IF HS_GPA LT 1.0 THEN LETTER_GRADE = 'F';
ELSE IF HS_GPA EQ . THEN LETTER_GRADE = '';
```

Q3:

```
PROC GCHART DATA = HW3;
VBAR LETTER_GRADE;
RUN;
QUIT;
```

Q4:

```
PROC TTEST ALPHA=0.05;
CLASS Gender;
VAR HS_GPA;
RUN;
PROC NPAR1WAY WILCOXON;
CLASS Gender;
VAR HS_GPA;
RUN;
```

Null Hypothesis (H0): GPA(males) = GPA(females).
Alternative Hypothesis (Ha): GPA(males) ≠ GPA(females).

Since there is an outlier in the sample, we cannot use a traditional T-test. We need to use a Wilcoxon for our case. The p-value in Wilcoxon is 0.2288, which is greater than 0.05. This means that we don't reject the null, and there is not enough evidence to support the alternative hypothesis.

Q5;

```
PROC TTEST ALPHA=0.05 H0=3.5;
VAR Exercise;
RUN;
PROC TTEST ALPHA=0.05 H0=3.5 SIDES=U;
VAR Exercise;
RUN;
```

Null Hypothesis (H0): Students exercise 3.5 times a week.
Alternative Hypothesis (Ha): Students exercise over 3.5 times a week.

Since there is no outlier in the sample, we can use a traditional T-test. The p-value in the T-test is 0.0184, which is smaller than 0.05. This means that we reject the null, and there is enough evidence to support the alternative hypothesis.

Q6;
PROC MEANS CLM ALPHA=0.05 MAXDEC=2;
VAR TIME_TO_GET_READY;
RUN;

The average time that people take to get ready is between 38.77 and 55.14.
(95% confidence interval)