```sas
/* Step 1: Assign a libref */
libname grp6 '/home/u60908848/stats project 2024';

/* Step 2: Print the dataset */
proc print data=grp6.group_6_train;
run;

proc means data=grp6.group_6_train;
run;


DATA grp6.group_6_train;
SET LIBRARY.group_1_train
(RENAME=
  (_1=Student_Age
_2 = Sex
_3 = Grad_High_School_Type
_4 = Scholarship_Type
_5 = Additional_Mark
_6 = Artistic_Sports_Activity
_7 = Partner
_8 = Salary
_9 = Transportation
_10 = Accommodation
_11 = Mother_Edu
_12 = Father_Edu
_13 = No_Siblings
_14 = Parental_Stat
_15 = Mothers_Occu
_16 = Fathers_Occu
_17 = Weekly_Stu_Hour
_18 = Read_Freq_Non_Scien
_19 = Read_Freq_Scien
_20 = Seminar_Attend
_21 = Impact_Proj_Stud_Succ
_22 = Attendance
_23 = Prep_Mid_Term_Exam_1
_24 = Prep_Mid_Term_Exam_2
_25 = Taking_Notes_Class
_26 = Listening_Class
_27 = Discussions
_28 = Flip_Class
_29 = Cumal_GPA_Last_Sem
_30 = Expected_GPA_Grad));
DROP _dataobs_;
RUN;

proc format;
value GRADEfmt
0 = 'Fail'
1 = 'DD'
2 = 'DC'
3 = 'CC'
4 = 'CB'
5 = 'BB'
6 = 'BA'
7 = 'AA';
value Student_Agefmt
1 = 'Young Adult'
2 = 'Mature Adult'
3 = 'Mature Adult';
value sexfmt
1 = 'Female'
2 = 'Male';
value Grad_High_School_Typefmt
1 = 'Private'
2 = 'Non-Private'
3 = 'Non-Private';
value Scholarship_Typefmt
1 = 'Low to Moderate Scholarship'
2 = 'Low to Moderate Scholarship'
3 = 'Low to Moderate Scholarship'
4 = 'High Scholarship'
5 = 'High Scholarship';
value Additional_Workfmt
1 = 'Yes'
2 = 'No';
value Artistic_Sports_Activityfmt
1 = 'Yes'
```

```
2 = 'No';
value Partnerfmt
1 = 'Yes'
2 = 'No';
value Salaryfmt
1 = 'Low to Moderate Salary'
2 = 'Low to Moderate Salary'
3 = 'Low to Moderate Salary'
4 = 'High Salary'
5 = 'High Salary';
value Transportationfmt
1 = 'Public Transport'
2 = 'Private Transport'
3 = 'Public Transport'
4 = 'Other';
value Accommodationfmt
1 = 'Rental'
2 = 'Dormitory'
3 = 'Family/Other'
4 = 'Family/Other';
value Mother_Edufmt
1 = 'Low Education'
2 = 'Low Education'
3 = 'Medium Education'
4 = 'Medium Education'
5 = 'Higher Education'
6 = 'Higher Education';
value Father_Edufmt
1 = 'Low Education'
2 = 'Low Education'
3 = 'Medium Education'
4 = 'Medium Education'
5 = 'Higher Education'
6 = 'Higher Education';
value No_Siblingsfmt
1 = 'Few Siblings'
2 = 'Few Siblings'
3 = 'Moderate Siblings'
4 = 'Moderate Siblings'
5 = 'Many Siblings';
value Parental_Statfmt
1 = 'Stable'
2 = 'Unstable'
3 = 'Unstable';
value Mothers_Occufmt
1 = 'Not Employed'
2 = 'Not Employed'
3 = 'Employed'
4 = 'Employed'
5 = 'Employed'
6 = 'Employed';
value Fathers_Occufmt
1 = 'Not Employed'
2 = 'Not Employed'
3 = 'Employed'
4 = 'Employed'
5 = 'Employed'
6 = 'Employed';
value Weekly_Stu_Hourfmt
1 = 'Low Study'
2 = 'Low Study'
3 = 'Moderate Study'
4 = 'Moderate Study'
5 = 'High Study';
value Read_Freq_Non_Scienfmt
1 = 'Low Frequency'
2 = 'Low Frequency'
3 = 'High Frequency';
value Read_Freq_Scienfmt
1 = 'Low Frequency'
2 = 'Low Frequency'
3 = 'High Frequency';
value Seminar_Attendfmt
1 = 'Yes'
2 = 'No';
value Attendancefmt
1 = 'High Attendance'
2 = 'Low Attendance'
3 = 'Low Attendance';
value Prep_Mid_Term_Exam_1fmt
1 = 'Individual Study'
```

```
2 = 'Group Study'
3 = 'Individual Study';
value Prep_Mid_Term_Exam_2fmt
1 = 'Inconsistent Prep'
2 = 'Consistent Prep'
3 = 'Inconsistent Prep';
value Taking_Notes_Classfmt
1 = 'Infrequent Notetaking'
2 = 'Infrequent Notetaking'
3 = 'Consistent Notetaking';
value Listening_Classfmt
1 = 'Infrequent Listening'
2 = 'Infrequent Listening'
3 = 'Consistent Listening';
value Discussionsfmt
1 = 'Infrequent Discussions'
2 = 'Infrequent Discussions'
3 = 'Frequent Discussions';
value Flip_Classfmt
1 = 'Not Useful'
2 = 'Useful'
3 = 'Not Applicable';
value Cumal_GPA_Last_Semfmt
1 = 'Low Performance'
2 = 'Low Performance'
3 = 'High Performance'
4 = 'High Performance'
5 = 'High Performance';
value Expected_GPA_Gradfmt
1 = 'Low Performance'
2 = 'Low Performance'
3 = 'High Performance'
4 = 'High Performance'
5 = 'High Performance';
run;


/*This allows one to print the new dataset that we have created from the previous piece of code that was used by th
ODS HTML;
PROC PRINT DATA grp6.group_6_train;
RUN;
ODS HTML CLOSE;

/* Step 1: Check normality for Attendance */
proc univariate data=grp6.group_6_train normal;
   var Attendance;
   histogram Attendance / normal;
   probplot Attendance / normal(mu=est sigma=est);
   inset mean std skewness kurtosis / position=ne;
run;

proc univariate data=grp6.group_6_train normal;
   var Grade;
   histogram Grade / normal;
   probplot Grade / normal(mu=est sigma=est);
   inset mean std skewness kurtosis / position=ne;
run;

/* Scatter plot for Attendance vs Expected_GPA_Grad */
proc sgplot data=grp6.group_6_train;
   scatter x=Attendance y=Expected_GPA_Grad / markerattrs=(symbol=circlefilled);
   xaxis label="Attendance";
   yaxis label="Expected GPA at Graduation";
run;

proc sgplot data=grp6.group_6_train;
   vbox Attendance / category=Sex;
   xaxis label="Sex";
   yaxis label="Attendance";
run;

proc sgplot data=grp6.group_6_train;
   vbox Grade / category=Sex;
   xaxis label="Sex";
   yaxis label="Expected GPA at Graduation";
run;


proc freq data=grp6.group_6_train;
   tables Grade / nocum;
run;
```

```
proc freq data=grp6.group_6_train;
   tables Attendance * Grade / chisq;
run;


ods graphics on;
proc sgplot data=grp6.group_6_train;
   scatter x=Attendance y=Grade / markerattrs=(symbol=circlefilled);
   xaxis label="Attendance";
   yaxis label="Eat Graduation";
run;


proc logistic data=grp6.group_6_train plots=roc;
    class Attendance;
    model Grade = attendance;
run;


PROC CORR DATA=GRP6.GROUP_6_TRAIN;
    VAR attendance grade;
RUN;




PROC LOGISTIC DATA=grp6.group_6_train DESCENDING;
   CLASS Attendance (PARAM=REF)
         Sex (PARAM=REF)
         Grad_High_School_Type (PARAM=REF)
         Scholarship_Type (PARAM=REF)
         Artistic_Sports_Activity (PARAM=REF)
         Partner (PARAM=REF)
         Salary (PARAM=REF)
         Transportation (PARAM=REF)
         Accommodation (PARAM=REF)
         Mother_Edu (PARAM=REF)
         Father_Edu (PARAM=REF)
         No_Siblings (PARAM=REF)
         Parental_Stat (PARAM=REF)
         Mothers_Occu (PARAM=REF)
         Fathers_Occu (PARAM=REF)
         Read_Freq_Non_Scien (PARAM=REF)
         Read_Freq_Scien (PARAM=REF)
         Seminar_Attend (PARAM=REF)
         Impact_Proj_Stud_Succ (PARAM=REF)
         Prep_Mid_Term_Exam_1 (PARAM=REF)
         Prep_Mid_Term_Exam_2 (PARAM=REF)
         Listening_Class (PARAM=REF)
         Discussions (PARAM=REF)
         Flip_Class (PARAM=REF)
         attendance (PARAM=REF);

   MODEL grade(event='7') =
         Student_Age
         Sex
         Grad_High_School_Type
         Scholarship_Type
         Additional_Mark
         Artistic_Sports_Activity
         Partner
         Salary
         Transportation
         Accommodation
         Mother_Edu
         Father_Edu
         No_Siblings
         Parental_Stat
         Mothers_Occu
         Fathers_Occu
         Weekly_Stu_Hour
         Read_Freq_Non_Scien
         Read_Freq_Scien
         Seminar_Attend
         Impact_Proj_Stud_Succ
         Prep_Mid_Term_Exam_1
         Prep_Mid_Term_Exam_2
         Taking_Notes_Class
         Listening_Class
         Discussions
```

```
         Flip_Class
         Cumal_GPA_Last_Sem
         attendance
         expected_gpa_grad;
RUN;


proc logistic data=grp6.group_6_train;
    model grade(event='7') =
        Student_Age
        Sex
        Grad_High_School_Type
        Scholarship_Type
        Additional_Mark

        Salary
        Transportation
        Accommodation
        Mother_Edu

        No_Siblings

        Fathers_Occu
        Weekly_Stu_Hour
        Read_Freq_Non_Scien

        Seminar_Attend
        Impact_Proj_Stud_Succ
        Prep_Mid_Term_Exam_1
        Prep_Mid_Term_Exam_2

        Listening_Class

        Flip_Class
        Cumal_GPA_Last_Sem


        / selection=forward slentry=0.05;
run;



ods html;
proc logistic data=grp6.group_6_train descending;
    class grade (param=ref ref='4');
    model grade(event='7')= attendance / link=glogit expb;
run;
ods html close;

proc logistic data=grp6.group_6_train;
    model grade(event='7') =
         Student_Age
        Sex
        Grad_High_School_Type
        Scholarship_Type
        Additional_Mark

        Salary
        Transportation
        Accommodation
        Mother_Edu

        No_Siblings

        Fathers_Occu
        Weekly_Stu_Hour
        Read_Freq_Non_Scien

        Seminar_Attend
        Impact_Proj_Stud_Succ
        Prep_Mid_Term_Exam_1
        Prep_Mid_Term_Exam_2

        Listening_Class

        Flip_Class
        Cumal_GPA_Last_Sem

        / selection=backward slstay=0.05;
run;
```

```
proc logistic data=grp6.group_6_train;
    model grade(event='7') =
         Student_Age Sex Grad_High_School_Type  Scholarship_Type Additional_Mark
        Salary Transportation Accommodation  Mother_Edu No_Siblings Fathers_Occu
        Weekly_Stu_Hour Read_Freq_Non_Scien Seminar_Attend Impact_Proj_Stud_Succ
        Prep_Mid_Term_Exam_1 Prep_Mid_Term_Exam_2 Listening_Class Flip_Class
        Cumal_GPA_Last_Sem
        / selection=stepwise slstay=0.05;
run;

ods html;
```

```
proc corr data=grp6.group_6_train PLOTS (ONLY) = SCATTER;

var Expected_GPA_Grad Attendance grade ;

run;

ods html close
```