

Analyzing adolescent health and risk factors

While there were several factors, I decided to study the dependence of satisfactory communication with mother(H1PF4), satisfactory communication with father (H1PF24) and student's perception of sexual knowledge of close friends (H1PF17) to their chances of getting AIDS(H1GH44) and chances of getting other sexually transmitted diseases (H1GH46). **So, the hypothesis is: Students with better communication with mother and / or father and with friends having better knowledge about birth control rhythm would report lesser chances of getting AIDS and other STD's.**

Step 1:

Tabulating the variables in action and analyzing them.

The FREQ Procedure

| S18Q4 MOM-GOOD COMMUNICATION-W1 | | | | |
|---------------------------------|-----------|---------|----------------------|--------------------|
| H1PF4 | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
| 1 | 2399 | 36.88 | 2399 | 36.88 |
| 2 | 2502 | 38.47 | 4901 | 75.35 |
| 3 | 579 | 8.90 | 5480 | 84.26 |
| 4 | 489 | 7.52 | 5969 | 91.77 |
| 5 | 153 | 2.35 | 6122 | 94.13 |
| 6 | 3 | 0.05 | 6125 | 94.17 |
| 7 | 370 | 5.69 | 6495 | 99.86 |
| 8 | 8 | 0.12 | 6503 | 99.98 |
| 9 | 1 | 0.02 | 6504 | 100.00 |

| S18Q24 DAD-GOOD COMMUNICATION-W1 | | | | |
|----------------------------------|-----------|---------|----------------------|--------------------|
| H1PF24 | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
| 1 | 1521 | 23.39 | 1521 | 23.39 |
| 2 | 1925 | 29.60 | 3446 | 52.98 |
| 3 | 522 | 8.03 | 3968 | 61.01 |
| 4 | 435 | 6.69 | 4403 | 67.70 |
| 5 | 136 | 2.09 | 4539 | 69.79 |
| 6 | 9 | 0.14 | 4548 | 69.93 |
| 7 | 1952 | 30.01 | 6500 | 99.94 |
| 8 | 3 | 0.05 | 6503 | 99.98 |
| 9 | 1 | 0.02 | 6504 | 100.00 |

| S3Q46 CHANGES OF GETTING OTHER STDs-W1 | | | | |
|--|-----------|---------|----------------------|--------------------|
| H1GH46 | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
| 1 | 72 | 1.11 | 72 | 1.11 |
| 2 | 191 | 2.94 | 263 | 4.04 |
| 3 | 878 | 13.50 | 1141 | 17.54 |
| 4 | 1827 | 28.09 | 2968 | 45.63 |
| 5 | 3455 | 53.12 | 6423 | 98.75 |
| 6 | 20 | 0.31 | 6443 | 99.06 |
| 8 | 60 | 0.92 | 6503 | 99.98 |
| 9 | 1 | 0.02 | 6504 | 100.00 |

A random sample of 6,504 adolescents were asked if they were satisfied with the communication with their mother. Of the total number 36.88% chose category 1 and strongly agreed, 38.47% chose category 2 and just agreed to it, 5.69% chose category 7 to skip the question (no resident mother), 0.05% chose category 6 and refused to answer the question whereas 0.12% chose category 8 and did not know the answer to the question.

For the next question, the same students were asked if they were satisfied with the communication with their father. Of the total number 23.39% chose category 1 and strongly agreed to it, 29.60 % chose category 2 and just agreed to it, 30.01% chose category 7 to skip the question (no resident dad), 0.14% chose category 6 and refused to answer the question whereas 0.05% chose category 8 and did not know the answer to the question.

The same set of students were asked about their chances are of getting sexually transmitted diseases other than AIDS, such as gonorrhea or genital herpes according to them. 53.12% chose category 5 saying that they had no chance of getting such diseases, 28.09% chose category 4 saying that they had very little chance of getting such diseases and 0.92% chose category 8 saying that they did not know about it and 0.31% chose category 6 refusing to answer the question.

Step 2: Managing data of the variables in the question.

Factors affecting adolescent risk and health behavior
 GoodCommunicationWithMother, GoodCommunicationWithFather, CloseFriendsKnowBirthControl (1- Yes, 2- Neutral, 3- No)
 ChancesOfGettingAids, ChancesOfGettingOtherSTD (1- High, 2- Low, 3- No chance)
 The FREQ Procedure

| GoodCommunicationWithMother | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|-----------------------------|-----------|---------|----------------------|--------------------|
| 1 | 5283 | 81.23 | 5283 | 81.23 |
| 2 | 579 | 8.90 | 5862 | 90.13 |
| 3 | 642 | 9.87 | 6504 | 100.00 |

| GoodCommunicationWithFather | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|-----------------------------|-----------|---------|----------------------|--------------------|
| 1 | 5411 | 83.19 | 5411 | 83.19 |
| 2 | 522 | 8.03 | 5933 | 91.22 |
| 3 | 571 | 8.78 | 6504 | 100.00 |

| CloseFriendsKnowBirthControl | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|------------------------------|-----------|---------|----------------------|--------------------|
| 1 | 385 | 5.92 | 385 | 5.92 |
| 2 | 1082 | 16.64 | 1467 | 22.56 |
| 3 | 5037 | 77.44 | 6504 | 100.00 |

| ChancesOfGettingAids | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|----------------------|-----------|---------|----------------------|--------------------|
| 1 | 385 | 5.92 | 385 | 5.92 |
| 2 | 3089 | 47.49 | 3474 | 53.41 |
| 3 | 3030 | 46.59 | 6504 | 100.00 |

| ChancesOfGettingOtherSTD | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|--------------------------|-----------|---------|----------------------|--------------------|
| 1 | 344 | 5.29 | 344 | 5.29 |
| 2 | 2705 | 41.59 | 3049 | 46.88 |
| 3 | 3455 | 53.12 | 6504 | 100.00 |

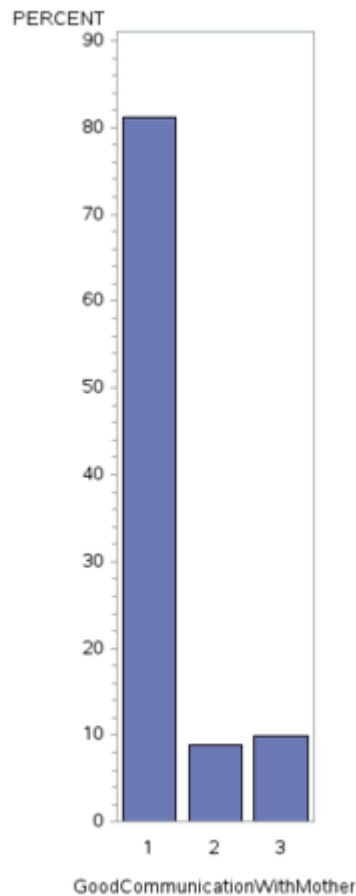
I collapsed the responses for all the variables and divided the valid responses for GoodCommunicationWithMother, GoodCommunicationWithFather and CloseFriendsKnowBirthControl into 3 categories 1 for **YES**, 2 for **NEUTRAL** and 3 for **NO**. In the first two cases majority falls into the 'yes' category and in the case of CloseFriendsKnowBirthControl, a majority of roughly 77% falls into the 'no' category. In the case of variables like ChancesOfGettingAids and ChancesOfGettingOtherSTD I divided the responses into 3 categories namely 1 for **HIGH**, 2 for **LOW**, 3 for **NO CHANCE**. The majority of the students responded with a 'low' chance' ie 47.49% for ChancesOfGettingAids and a 53.12% which means a 'no chance' for ChancesOfGettingOtherSTD.

Step 3: Creating graphs for data variables

Factors affecting adolescent risk and health behavior
 GoodCommunicationWithMother, GoodCommunicationWithFather, CloseFriendsKnowBirthControl (1- Yes, 2- Neutral, 3- No)
 ChancesOfGettingAids, ChancesOfGettingOtherSTD (0- no chance, 1- chance)

The FREQ Procedure

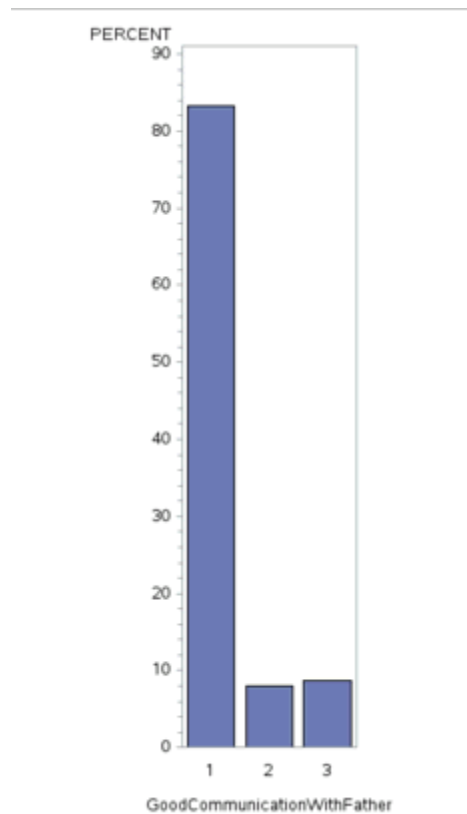
| GoodCommunicationWithMother | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|-----------------------------|-----------|---------|----------------------|--------------------|
| 1 | 5283 | 81.23 | 5283 | 81.23 |
| 2 | 579 | 8.90 | 5862 | 90.13 |
| 3 | 642 | 9.87 | 6504 | 100.00 |



This graph is unimodal with maximum students reporting satisfactory communication with their mother. It also seems skewed to the left as there are higher frequencies to the left .

The FREQ Procedure

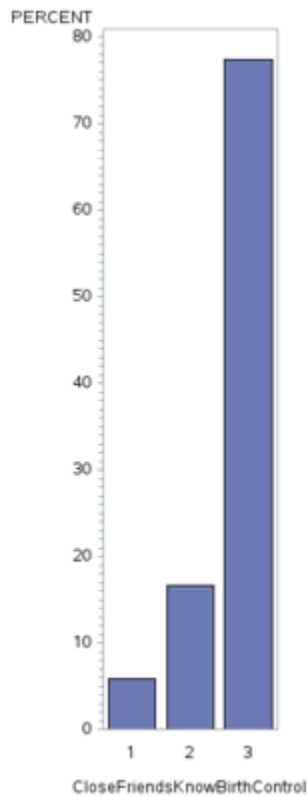
| GoodCommunicationWithFather | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|-----------------------------|-----------|---------|----------------------|--------------------|
| 1 | 5411 | 83.19 | 5411 | 83.19 |
| 2 | 522 | 8.03 | 5933 | 91.22 |
| 3 | 571 | 8.78 | 6504 | 100.00 |



This graph is unimodal with maximum students reporting satisfactory communication with their father. It also seems skewed to the left as there are higher frequencies to the left .

The FREQ Procedure

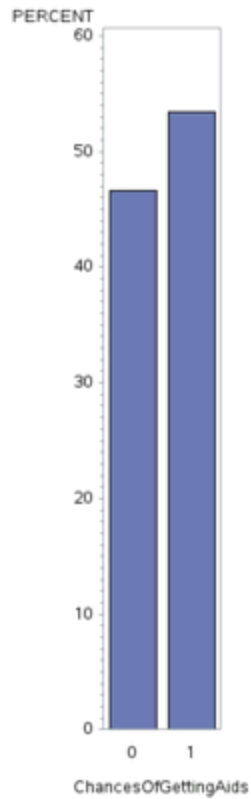
| CloseFriendsKnowBirthControl | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|------------------------------|-----------|---------|----------------------|--------------------|
| 1 | 385 | 5.92 | 385 | 5.92 |
| 2 | 1082 | 16.64 | 1467 | 22.56 |
| 3 | 5037 | 77.44 | 6504 | 100.00 |



This graph is unimodal with maximum students reporting that their friends do not know much about birth control rhythm. It also seems skewed to the right as there are higher frequencies to the right.

The FREQ Procedure

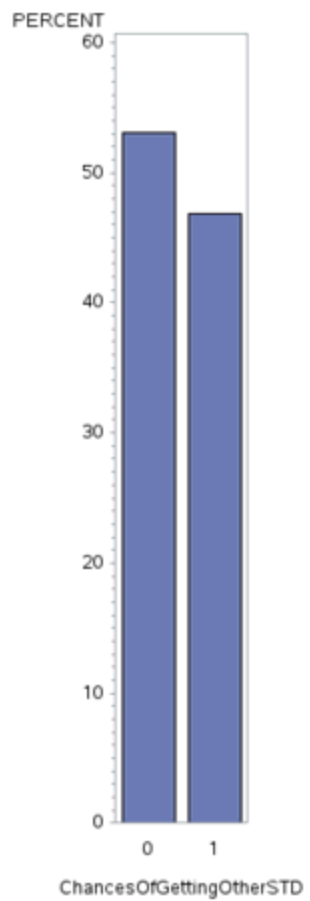
| ChancesOfGettingAids | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|----------------------|-----------|---------|----------------------|--------------------|
| 0 | 3030 | 46.59 | 3030 | 46.59 |
| 1 | 3474 | 53.41 | 6504 | 100.00 |



This graph is unimodal with maximum students reporting that they have chances of getting AIDS. It also seems skewed to the right as there are higher frequencies to the right.

The FREQ Procedure

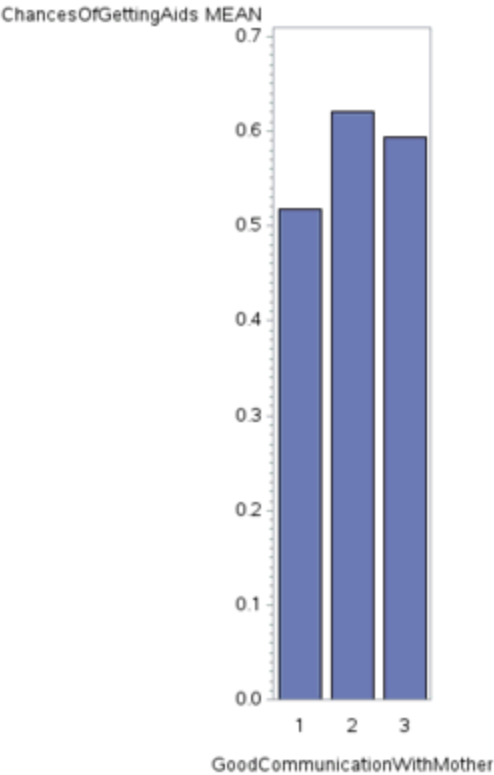
| ChancesOfGettingOther STD | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---------------------------|-----------|---------|----------------------|--------------------|
| 0 | 3455 | 53.12 | 3455 | 53.12 |
| 1 | 3049 | 46.88 | 6504 | 100.00 |

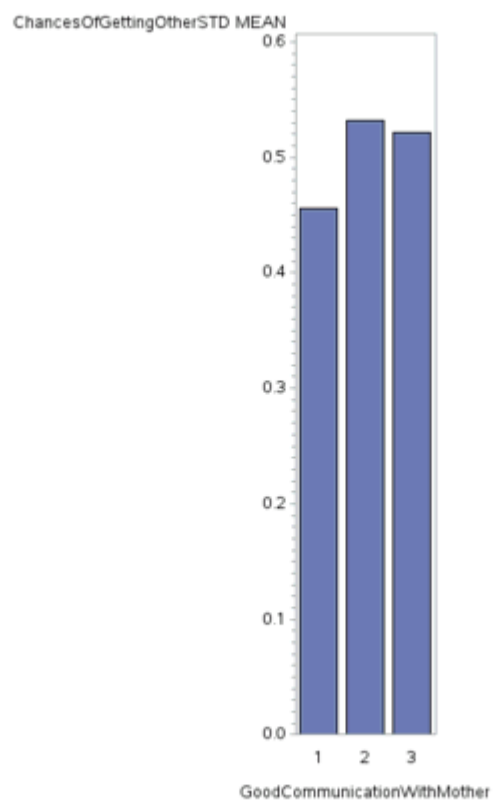


This graph is unimodal with maximum students reporting no chances of getting other STD's. It also seems skewed to the left as there are higher frequencies to the left.

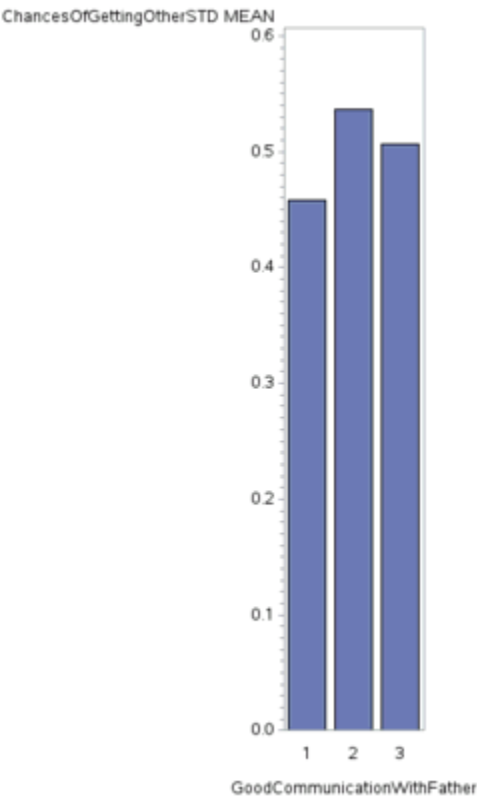
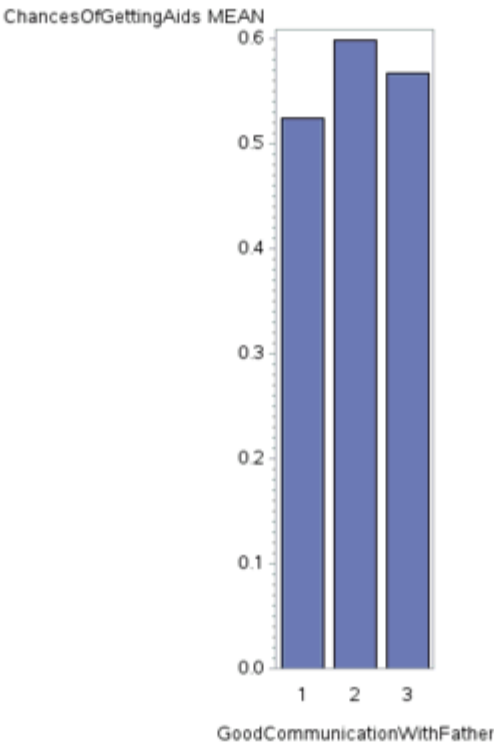
Bivariate Graphs:

GoodCommunicationWithMother vs ChancesOfGettingAids and ChancesOfGettingOtherSTD

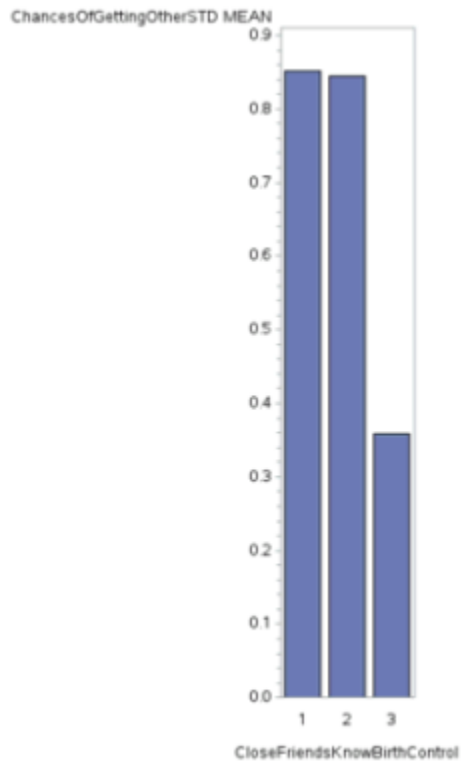
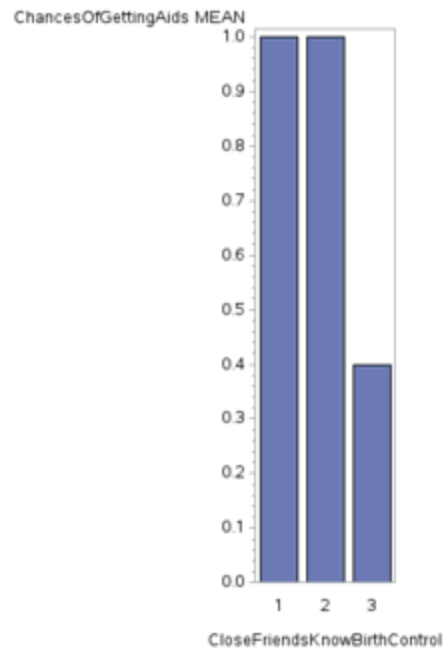




GoodCommunicationWithFather vs ChancesOfGettingAids and ChancesOfGettingOtherSTD



CloseFriendsKnowBirthControl vs ChancesOfGettingAids and ChancesOfGettingOtherSTD



Conclusion:

The graphs show that students who reported satisfactory communication with their mother showed lower tendency to AIDS and other STD's than students who reported neutral or non-satisfactory communication.

The graphs show that students who reported satisfactory communication with their father reported lesser chances of getting AIDS or other STD's than the students who reported neutral or non-satisfactory communication.

Surprisingly, students who reported that their close friends knew the rhythm of birth control and students who were neutral about the issue reported the highest percentages of getting AIDS and others STD's than students who reported that their friends didnt know the rhythm of birth control.

Hence the hypothesis that better communication with mother and/ or father leads to lesser chances of AIDS and other STD's in teenage children is correct. But the hypothesis that having friends with knowledge of birth control rhythm leads to lesser chances of AIDS and other STD's is incorrect.

SAS Code:

```
LIBNAME mydata "/courses/d1406ae5ba27fe300" access=readonly;
DATA new; set mydata.addhealth_pds;
```

```
title 'Factors affecting adoslescent risk and health behavior' ;
title2 'GoodCommunicationWithMother, GoodCommunicationWithFather,
CloseFriendsKnowBirthControl ( 1- Yes, 2- Nuetral, 3- No )';
title3 ' ChancesOfGettingAids, ChancesOfGettingOtherSTD (0- no chance, 1- chance)';
```

```
IF H1PF4 GE 6 then H1PF4 =.;
IF H1PF24 GE 6 then H1PF24=.;
IF H1PF17 GE 6 then H1PF17=.;
IF H1GH44 GE 6 then H1GH44=.;
IF H1GH46 GE 6 then H1GH46=.;
```

```
IF H1PF4 LE 2 THEN GoodCommunicationWithMother=1;
ELSE IF H1PF4 LE 3 THEN GoodCommunicationWithMother=2;
ELSE IF H1PF4 LE 5 THEN GoodCommunicationWithMother=3;
```

```
IF H1PF24 LE 2 THEN GoodCommunicationWithFather=1;
ELSE IF H1PF24 LE 3 THEN GoodCommunicationWithFather=2;
ELSE IF H1PF24 LE 5 THEN GoodCommunicationWithFather=3;
```

```
IF H1PF17 LE 2 THEN CloseFriendsKnowBirthControl=1;
ELSE IF H1PF17 LE 3 THEN CloseFriendsKnowBirthControl=2;
ELSE IF H1PF17 LE 5 THEN CloseFriendsKnowBirthControl=3;
```

```
IF H1GH44 LE 2 THEN CloseFriendsKnowBirthControl=1;
ELSE IF H1GH44 LE 3 THEN CloseFriendsKnowBirthControl=2;
ELSE IF H1GH44 LE 5 THEN CloseFriendsKnowBirthControl=3;
```

```
IF H1GH44 LE 4 THEN ChancesOfGettingAids=1;/*chance*/
ELSE IF H1GH44 LE 5 THEN ChancesOfGettingAids=0;/*no chance*/
```

```

IF H1GH46 LE 4 THEN ChancesOfGettingOtherSTD=1;/*chance*/
ELSE IF H1GH44 LE 5 THEN ChancesOfGettingOtherSTD=0;/*no chance*/

PROC SORT; by AID;

/*Univariate graphs*/
PROC FREQ; TABLES GoodCommunicationWithMother;
PROC GCHART; VBAR GoodCommunicationWithMother /DISCRETE type =PCT Width= 5;
title;

title2;
title3;
PROC FREQ; TABLES GoodCommunicationWithFather;
PROC GCHART; VBAR GoodCommunicationWithFather /DISCRETE type =PCT Width= 5;

PROC FREQ; Tables CloseFriendsKnowBirthControl ;
PROC GCHART; VBAR CloseFriendsKnowBirthControl /DISCRETE type =PCT Width= 5;

PROC FREQ; TABLES ChancesOfGettingAids ;
PROC GCHART; VBAR ChancesOfGettingAids /DISCRETE type =PCT Width= 5;

PROC FREQ; TABLES ChancesOfGettingOtherSTD ;
PROC GCHART; VBAR ChancesOfGettingOtherSTD /DISCRETE type =PCT Width= 5;

/*PROC UNIVARIATE; VAR
GoodCommunicationWithMother GoodCommunicationWithFather CloseFriendsKnowBirthControl
;*/

PROC GCHART ; VBAR GoodCommunicationWithMother / DISCRETE type= mean
sumvar=ChancesOfGettingAids ;
title4 'GoodCommunicationWithMother vs ChancesOfGettingAids and
ChancesOfGettingOtherSTD ' ;

PROC GCHART ; VBAR GoodCommunicationWithMother / DISCRETE type= mean
sumvar=ChancesOfGettingOtherSTD ;

title4;

PROC GCHART ; VBAR GoodCommunicationWithFather / DISCRETE type= mean
sumvar=ChancesOfGettingAids ;
title5 'GoodCommunicationWithFather vs ChancesOfGettingAids and
ChancesOfGettingOtherSTD ' ;
PROC GCHART ; VBAR GoodCommunicationWithFather / DISCRETE type= mean
sumvar=ChancesOfGettingOtherSTD ;

title5;

PROC GCHART ; VBAR CloseFriendsKnowBirthControl / DISCRETE type= mean
sumvar=ChancesOfGettingAids ;
title6 'CloseFriendsKnowBirthControl vs ChancesOfGettingAids and ChancesOfGettingOtherSTD ' ;
PROC GCHART ; VBAR CloseFriendsKnowBirthControl / DISCRETE type= mean
sumvar=ChancesOfGettingOtherSTD ;

title6;

```

run;

Running an analysis of variance

We try to understand the relation between 'AGE AT ONSET OF ALCOHOL DEPENDENCE' and the factor 'BLOOD/NATURAL FATHER EVER AN ALCOHOLIC OR PROBLEM DRINKER'. Here the null hypothesis is 'AGE AT ONSET OF ALCOHOL DEPENDENCE' and the factor 'BLOOD/NATURAL FATHER EVER AN ALCOHOLIC OR PROBLEM DRINKER' are unrelated.

The second set of factors are 'AGE AT ONSET OF ALCOHOL DEPENDENCE' and the factor 'BLOOD/NATURAL MOTHER EVER AN ALCOHOLIC OR PROBLEM DRINKER'. Here the null hypothesis is 'AGE AT ONSET OF ALCOHOL DEPENDENCE' and the factor 'BLOOD/NATURAL MOTHER EVER AN ALCOHOLIC OR PROBLEM DRINKER' are unrelated.

The ANOVA Procedure

Class Level Information

| Class | Levels | Values |
|-------|--------|--------|
| S2DQ1 | 2 | 1 2 |

| | |
|-----------------------------|------|
| Number of Observations Read | 4038 |
|-----------------------------|------|

| | |
|-----------------------------|------|
| Number of Observations Used | 3781 |
|-----------------------------|------|

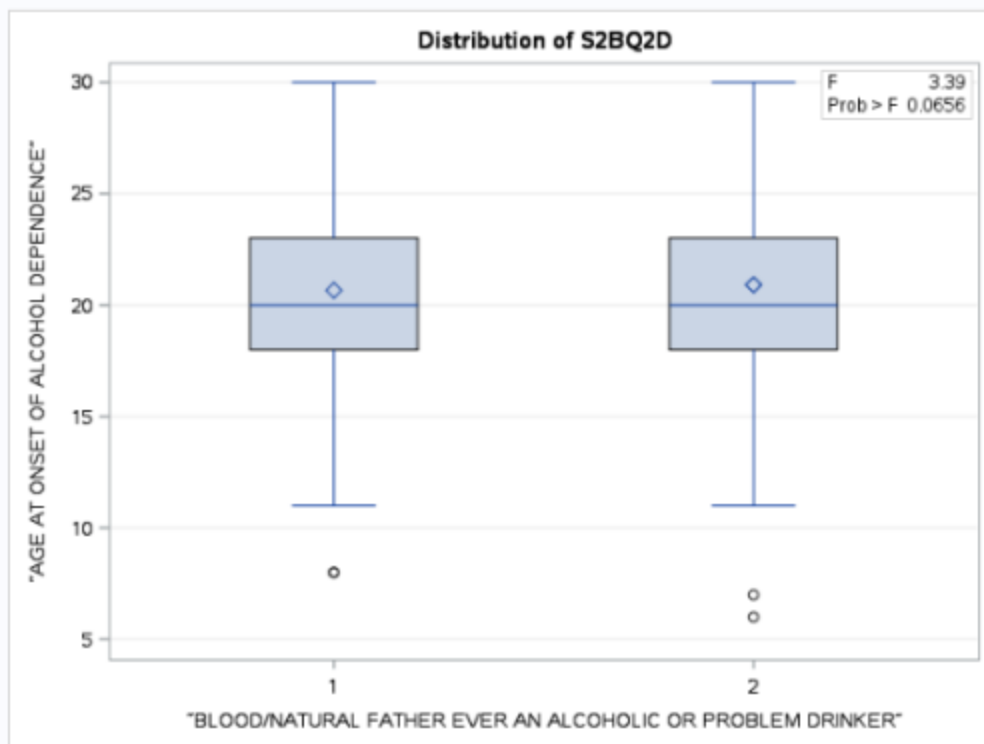
The ANOVA Procedure

Dependent Variable: S2BQ2D "AGE AT ONSET OF ALCOHOL DEPENDENCE"

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|------|----------------|-------------|---------|--------|
| Model | 1 | 52.40170 | 52.40170 | 3.39 | 0.0656 |
| Error | 3779 | 58374.49753 | 15.44708 | | |
| Corrected Total | 3780 | 58426.89923 | | | |

| R-Square | Coeff Var | Root MSE | S2BQ2D Mean |
|----------|-----------|----------|-------------|
| 0.000897 | 18.87847 | 3.930277 | 20.81883 |

| Source | DF | Anova SS | Mean Square | F Value | Pr > F |
|--------|----|-------------|-------------|---------|--------|
| S2DQ1 | 1 | 52.40170471 | 52.40170471 | 3.39 | 0.0656 |



| Level of S2DQ1 | N | S2BQ2D | |
|----------------|------|------------|------------|
| | | Mean | Std Dev |
| 1 | 1402 | 20.8854779 | 4.07268973 |
| 2 | 2379 | 20.9092055 | 3.84390490 |

The ANOVA Procedure

| Class Level Information | | |
|-------------------------|--------|--------|
| Class | Levels | Values |
| S2DQ2 | 2 | 1 2 |

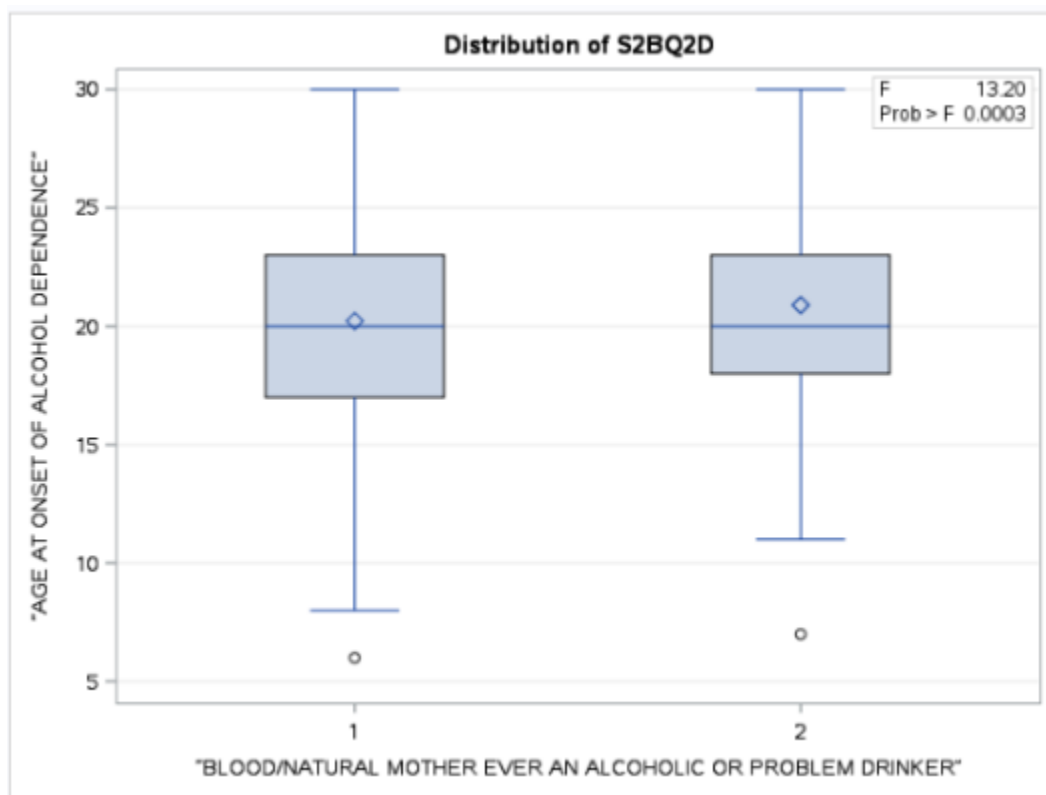
| | |
|-----------------------------|------|
| Number of Observations Read | 4038 |
| Number of Observations Used | 3927 |

The ANOVA Procedure
 Dependent Variable: S2BQ2D "AGE AT ONSET OF ALCOHOL DEPENDENCE"

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|------|----------------|-------------|---------|--------|
| Model | 1 | 204.72170 | 204.72170 | 13.20 | 0.0003 |
| Error | 3925 | 60875.12297 | 15.50959 | | |
| Corrected Total | 3926 | 61079.84467 | | | |

| R-Square | Coeff Var | Root MSE | S2BQ2D Mean |
|----------|-----------|----------|-------------|
| 0.003352 | 18.92671 | 3.938221 | 20.80774 |

| Source | DF | Anova SS | Mean Square | F Value | Pr > F |
|--------|----|-------------|-------------|---------|--------|
| S2DQ2 | 1 | 204.7216957 | 204.7216957 | 13.20 | 0.0003 |



| Level of S2DQ2 | N | S2BQ2D | |
|----------------|------|------------|------------|
| | | Mean | Std Dev |
| 1 | 518 | 20.2220077 | 4.34522551 |
| 2 | 3409 | 20.8987439 | 3.87274294 |

When examining the association between 'age of onset of smoking' (quantitative response) and 'BLOOD/NATURAL FATHER EVER AN ALCOHOLIC OR PROBLEM DRINKER' (categorical explanatory) an Analysis of Variance (ANOVA) revealed that among daily, young adult smokers (below 25 years of age, sample taken), those with alcoholic or a problem drinker father reported a

lesser mean starting drinking age (mean:20.66, sd: 4.07) and those without alcoholic or a problem drinker father reported a higher mean starting drinking age (mean: 20.90, sd : 3.84). $F(1, 52.40) = 3.39$, $p = 0.0656$.

Examining the 'age of onset of smoking' (quantitative response) and 'BLOOD/NATURAL MOTHER EVER AN ALCOHOLIC OR PROBLEM DRINKER' (categorical explanatory) an Analysis of Variance (ANOVA) revealed that among daily, young adult smokers (below 25 years of age, sample taken), those with alcoholic or a problem drinker father reported a lesser mean starting drinking age (mean: 20.22, sd: 4.34) and those without alcoholic or a problem drinker father reported a higher mean starting drinking age (mean: 20.89, sd : 3.87). $F(1, 204.72) = 13.20$, $p = 0.0003$.

Note that the degrees of freedom that I report in parentheses following 'F' can be found in the OLS table as the DF model and DF residuals.

Model Interpretation:

ANOVA revealed that among young adult smokers (my sample), the p factor for this data is 0.0656, > 0.05 . But the value is too low to accept null hypothesis and hence we reject it and accept the alternate hypothesis ('AGE AT ONSET OF ALCOHOL DEPENDENCE' and the factor 'BLOOD/NATURAL FATHER EVER AN ALCOHOLIC OR PROBLEM DRINKER' are related and people with alcoholic father have a earlier age of alcohol onset).

ANOVA revealed that among young adult smokers (my sample), the p factor for this data is 0.0003, < 0.05 . So, we reject null hypothesis and accept the alternate hypothesis ('AGE AT ONSET OF ALCOHOL DEPENDENCE' and the factor 'BLOOD/NATURAL MOTHER EVER AN ALCOHOLIC OR PROBLEM DRINKER' are related and people with alcoholic MOTHER have a earlier age of alcohol onset).

*****Code*****

```
LIBNAME mydata "/courses/d1406ae5ba27fe300 " access=readonly;
DATA new; set mydata.nesarc_pds;
label S2BQ2D = "AGE AT ONSET OF ALCOHOL DEPENDENCE"
      S2DQ1  ="BLOOD/NATURAL FATHER EVER AN ALCOHOLIC OR PROBLEM DRINKER"
      S2DQ2  ="BLOOD/NATURAL MOTHER EVER AN ALCOHOLIC OR PROBLEM DRINKER";

IF S2BQ2D =. THEN S2BQ2D=100;
IF S2DQ1 = 9 THEN S2DQ1=.;
IF S2DQ2 = 9 THEN S2DQ2=.;

IF S2BQ2D LE 30;
PROC sort; by IDNUM;

PROC ANOVA; CLASS S2DQ1;
MODEL S2BQ2D=S2DQ1;
MEANS S2DQ1;

PROC ANOVA; CLASS S2DQ2;
MODEL S2BQ2D=S2DQ2;
MEANS S2DQ2;

RUN;
```