AlcoholEdu survey results
Survey 1 = Pre-survey (149 items)
Survey 2 = Module effectiveness (22 items)
Survey 3 = Post-survey (138 items)

N = 333 (100%)
116 males (35%)
217 females (65%)
309 Caucasian (93%)
326 age 18-19 (98%)
321 live on-campus (96%)
17 live in substance-free hall (5%)
3 transfer students (1%)
72 from Iowa (22%)
165 from Illinois (50%)

163 volunteer/com service (49%)
50 student religious group (15%)
141 athletics (42%)
162 intramural (49%)
15 health education group (5%)
15 media organization (5%)
51 music/performing arts (15%)

88 never started drinking (26%)
35 started drinking before age 15 (11%)
199 started drinking before age 18 (60%)

122 never got drunk (37%)
21 got drunk before age 15 (6%)
167 got drunk before age 18 (50%)

Pre: 238 consumed alcohol in the past year (71%)
Post: 254 consumed alcohol in the past year (76%)

Pre: 111 see no need to change the way they drink (33%)
Post: 156 see no need to change the way they drink (47%)

==========

Q1: Does age of first drink predict the likelihood of use or more frequent use?

-- First, I want to see how old students were when they started drinking and when they first got drunk. These variables will be our predictors.

Q145. How old were you when you first started drinking?	Q146. Personal Characteristics - How old were you when you first got drunk?					
	Never did	11 years	12-14 yea	15-17 yea	18-20 yea	Total
Never did this	86	0	0	0	1	87
11 years or younger	0	1	1	0	0	2
12-14 years old	0	0	16	17	0	33
15-17 years old	17	0	2	127	18	164
18-20 years old	18	0	1	2	22	43
Total	121	1	20	146	41	329

-- Next, I want to see the indicators of "frequency of use." The first measure will be question #5 from the pre-survey.

Q5. Frequency of High-Risk Drinking

in the past 2 weeks	Freq.	Percent	Cum.
Never	67	20.12	20.12
Once	34	10.21	30.33
Twice	27	8.11	38.44
Three or more times	5	1.50	39.94
(no response)	200	60.06	100.00
Total	333	100.00	

-- This might be a problem - 200 students did not respond to this item. For these 133 students, let's look at their age of first use.

Q145. How old were you when you first started drinking?	Q146. Personal Characteristics - How old were you when you first got drunk?					
	Never did	11 years	12-14 yea	15-17 yea	18-20 yea	Total
Never did this	0	0	0	0	1	1
11 years or younger	0	1	1	0	0	2
12-14 years old	0	0	11	13	0	24
15-17 years old	2	0	0	78	8	88
18-20 years old	6	0	1	1	10	18
Total	8	1	13	92	19	133

92 of the 133 respondents first got drunk between the ages of 15-17. Since so many students are in this single group, question 146 won't predict much of anything. We might get a good prediction from question 145.

Q5. Frequency of High-Risk Drinking in the past 2 weeks	Q145. Personal Characteristics - How old were you when you first started drinking?					
	Never did	11 years	12-14 yea	15-17 yea	18-20 yea	Total
Never	1	1	7	45	13	67
Once	0	1	10	19	4	34
Twice	0	0	5	21	1	27
Three or more times	0	0	2	3	0	5
Total	1	2	24	88	18	133

From this table, there doesn't appear to be much of a relationship (since most students did not drink in the past 2 weeks). If anything, it looks as though students who began drinking between the ages of 15-17 more often engage in high-risk drinking. I'll go ahead and run a linear and ordinal logistic regression to see if we get a good prediction.

Linear regression

Number of obs	133
F(1, 131)	3.03
Prob > F	0.0842
R-squared	0.0230
Root MSE	.89429

oneQ5	Coef.	Robust Std. Err.	t	P>|t|	Beta
oneQ145	-.2066945	.1187887	-1.74	0.084	-.1516571
_cons	1.374316	.3657241	3.76	0.000	.

Ordered logistic regression

Number of obs	133
LR chi2(1)	3.27

```
                                                Prob > chi2     =     0.0705
Log likelihood = -150.13532                     Pseudo R2       =     0.0108

------------------------------------------------------------------------------
       oneQ5 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     oneQ145 |   .6399635   .1608285    -1.78   0.076     .3910576    1.047297
-------------+----------------------------------------------------------------
       /cut1 |  -1.276212   .7466991                     -2.739716    .1872913
       /cut2 |  -.1221908   .7404325                     -1.573412     1.32903
       /cut3 |    1.97886   .8403015                      .3318993    3.625821
------------------------------------------------------------------------------
```

Neither model provides a good prediction. The logistic regression shows, if anything, students who begin drinking later in life (slightly) less often engage in high-risk drinking. But, really, there's no significant relationship between age of first drink and frequency of high-risk drinking (on the pre-survey). Running the same analyses on the post-survey finds virtually the same (non) results.

Let's move on to the other measures of "frequency of use" (questions 6-8). Here are the responses to questions 6-7 on the pre-survey:

```
      Q6. |
    Heavy |
  Episodic |   Q7. Problematic
  Drinkers |       Drinkers
           |        No        Yes |     Total
-----------+----------------------+----------
        No |       241          0 |       241
       Yes |        65         27 |        92
-----------+----------------------+----------
     Total |       306         27 |       333
```

I'll first see if age of first drink (and age of first time drunk) predict if students are more likely to be heavy episodic drinkers. The one model that provides an interesting conclusion is:

```
Logistic regression                             Number of obs   =        330
                                                LR chi2(1)      =      65.69
                                                Prob > chi2     =     0.0000
Log likelihood = -162.45186                     Pseudo R2       =     0.1682

------------------------------------------------------------------------------
       oneQ6 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     oneQ146 |   2.212869   .2678854     6.56   0.000     1.745465    2.805437
       _cons |   .0586177    .021161    -7.86   0.000     .0288898    .1189359
------------------------------------------------------------------------------
```

I'm going to overstate the conclusion here, but it could be interpreted that: The odds of becoming a heavy episodic drinker double for students who first get drunk later in life. Let's see if I can state that more clearly: Students who get drunk earlier in life are half as likely to report becoming heavy episodic drinkers. This conclusion also holds on the post-survey.

Let's see if a similar conclusion can be made about students reporting to be problematic drinkers.

```
Logistic regression                             Number of obs   =        330
                                                LR chi2(1)      =      12.22
                                                Prob > chi2     =     0.0005
Log likelihood = -87.344434                     Pseudo R2       =     0.0654

------------------------------------------------------------------------------
```

```
      oneQ7 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+----------------------------------------------------------------
     oneQ146 |    1.71078   .3032158     3.03   0.002     1.208728    2.421362
       _cons |   .0241354   .0130988    -6.86   0.000     .0083309    .0699229
------------------------------------------------------------------------------
```

Yep, students who first get drunk later in life are more likely to report becoming problematic drinkers. This conclusion also holds on the post-survey.

The final way to address this question would be to look at question #9 from the pre-survey.

```
   Q8. Self-Reported Drinking Rates and |
    Behaviors - Students Categorized by |
                                reporte |      Freq.     Percent        Cum.
----------------------------------------+-----------------------------------
Non-drinkers (0 drinks in previous two  |        203       60.96       60.96
Light/Moderate drinkers (1-3 drinks for |         38       11.41       72.37
High Risk drinkers (4 or more drinks fo |         92       27.63      100.00
----------------------------------------+-----------------------------------
                                  Total |        333      100.00
```

Let's see if age of first drink predicts these categories of drinking.

```
Ordered logistic regression                     Number of obs   =        329
                                                LR chi2(2)      =      91.61
                                                Prob > chi2     =     0.0000
Log likelihood = -253.49562                     Pseudo R2       =     0.1530

------------------------------------------------------------------------------
      oneQ8 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+----------------------------------------------------------------
     oneQ145 |   1.354526   .1833219     2.24   0.025     1.038928    1.765994
     oneQ146 |   1.852475   .2022613     5.65   0.000     1.495596    2.294511
-------------+----------------------------------------------------------------
       /cut1 |   2.535096   .3633494                      1.822944    3.247248
       /cut2 |   3.180742   .3739365                       2.44784    3.913644
------------------------------------------------------------------------------
```

Here we get the same conclusion: For students who begin drinking later in life, the odds of becoming a high-risk drinker (versus a combined non-drinker or moderate drinker category) are 1.35 times higher. Likewise, for students who are older when they first get drunk, the odds of becoming a high-risk drinker (versus a non-drinker or moderate drinker) are 1.85 times higher. This result held for the post-survey.

==========

Q2: Does age of first use predict the use of protective behaviors?

Protective behaviors are measured by items 76 and 78-94. Since they all measure protective behaviors and they're all measured on the same 7-point scale, I'm going to create a single variable called "protective" that's simply the sum of all these item responses. So students can score from 0 (no protective behaviors) to 126 (7 x 18 items = 126). Let's look at the distribution of this variable:

```
   2* | 049
   3* | 9
   4* | 0135666778999
   5* | 01123344555556777778899
   6* | 000000111122222334446666777788889
   7* | 00001112222222223344444445566678889999
   8* | 000111111112222233444445555666777888888999
   9* | 00011133444444455566677777888899
  10* | 002223345577899
```

```
  11* | 0002223344577
  12* | 01566
```

We ended up with 221 students who answered all 18 items related to protective behaviors. There's a pretty nice distribution here, with only two students (those scoring 126) who reported using all 18 protective behaviors all the time.

I don't know much about AlcoholEdu, but I would (naively) expect students to report using protective behaviors more often on the post-survey. Let's look at the distribution for the post-survey:

```
   1* | 88
   2* | 1
   3* | 03456688
   4* | 2599
   5* | 011222333444444445568889999
   6* | 00122222233334444445555566667778888888
   7* | 000000112222222222222223333444445555666677788888889
   8* | 000111111111223333455568888
   9* | 000112233334444567777999
  10* | 000002222444456677888889
  11* | 00123444567779
  12* | 0233344556666666
```

Following the AlcoholEdu program, the number of students using all protective behaviors all the time increased to 7. Let's run a quick dependent-samples t-test to see if we have a significant change in the use of protective behaviors:

```
Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
presurvey|     197    78.38071     1.49768    21.02094    75.42707    81.33435
postsurv |     197     79.6599    1.647589    23.12502    76.41062    82.90918
---------+--------------------------------------------------------------------
    diff |     197   -1.279188    1.421626    19.95347   -4.082835    1.524459
------------------------------------------------------------------------------
     mean(diff) = mean(protective - postprotective)                t =  -0.8998
 Ho: mean(diff) = 0                                     degrees of freedom =      196

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.1847         Pr(|T| > |t|) = 0.3693          Pr(T > t) = 0.8153
```

The average "protective behaviors" score for students at the beginning of the program was 78. By the end of the program, that average increased slightly to 80. It wasn't a statistically significant difference -- we cannot conclude the AlcoholEdu program was associated with an increase in the reported use of protective behaviors.

The following stem plot shows the change in "protective behaviors" following the AlcoholEdu program:

```
 -7* | 0
 -6* | 4
 -5* | 3
 -4* | 75210
 -3* | 6533
 -2* | 9988544332110
 -1* | 999888654333322000
 -0* | 99988777777666665554443333222222222211111
  0* | 000000001111111122222233344455555666666678888999999999
  1* | 011112222233333334445555666677899
  2* | 00011224445667
  3* | 1236
  4* | 00124
```

```
  5* | 7
  6* | 03
```

Once again, it shows that just as many (if not more) students reporting using protective behaviors LESS frequently following the program.

Ok, back to the original question: Does age of first use predict the use of protective behaviors?

```
      Source |       SS       df       MS              Number of obs =     221
-------------+------------------------------           F(  2,   218) =   19.66
       Model |  14112.5556     2  7056.27778           Prob > F      =  0.0000
    Residual |  78250.6662   218   358.94801           R-squared     =  0.1528
-------------+------------------------------           Adj R-squared =  0.1450
       Total |  92363.2217   220  419.832826           Root MSE      =  18.946

------------------------------------------------------------------------------
  protective |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     oneQ145 |   6.212621   1.849309     3.36   0.001     2.567807    9.857435
     oneQ146 |  -5.450851   1.026691    -5.31   0.000    -7.474362    -3.42734
       _cons |   74.18919   6.259451    11.85   0.000      61.8524    86.52598
------------------------------------------------------------------------------
```

This seems to indicate:
a) Students who begin drinking at a later age are more likely to use protective behaviors
b) Students who first get drunk at a later age are less likely to use protective behaviors

Let's look at some conditional means (and ANOVA tables) to see what's really going on:

```
         oneQ145 |         N      mean        sd
-----------------+------------------------------
  Never did this |         3  80.33333  54.50076
11 years or youn |         2      78.5  9.192388
 12-14 years old |        30      68.1  20.10893
 15-17 years old |       146  78.20548  18.97436
 18-20 years old |        40      87.3  20.28161
-----------------+------------------------------
           Total |       221  78.51131  20.48982
-------------------------------------------------
```

```
                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      6365.11944      4   1591.27986      4.00     0.0038
 Within groups      85998.1023    216   398.139362
------------------------------------------------------------------------
    Total           92363.2217    220   419.832826

Bartlett's test for equal variances:  chi2(4) =   9.2007  Prob>chi2 = 0.056
```

```
Row Mean-|
Col Mean |   Never di   11 years   12-14 ye   15-17 ye
---------+--------------------------------------------
11 years |   -1.83333
         |      1.000
         |
12-14 ye |   -12.2333      -10.4
         |      1.000      1.000
         |
15-17 ye |   -2.12785   -.294521    10.1055
         |      1.000      1.000      0.122
         |
18-20 ye |    6.96667        8.8       19.2    9.09452
```

```
         |      1.000      1.000      0.001      0.113
```

```
        oneQ146 |         N      mean        sd
----------------+----------------------------
  Never did this |        35  98.94286  20.73352
11 years or youn |         1        85         .
 12-14 years old |        17  64.82353  19.21664
 15-17 years old |       130  74.17692  17.70143
 18-20 years old |        38  80.47368  17.23563
----------------+----------------------------
           Total |       221  78.51131  20.48982
-------------------------------------------------
```

```
                      Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups       20426.461      4   5106.61524     15.33     0.0000
 Within groups      71936.7608    216   333.040559
------------------------------------------------------------------------
    Total           92363.2217    220   419.832826

Bartlett's test for equal variances:  chi2(3) =   1.7406  Prob>chi2 = 0.628
```

```
Row Mean-|
Col Mean |   Never di   11 years   12-14 ye   15-17 ye
---------+--------------------------------------------
11 years |   -13.9429
         |      1.000
         |
12-14 ye |   -34.1193   -20.1765
         |      0.000      1.000
         |
15-17 ye |   -24.7659   -10.8231    9.35339
         |      0.000      1.000      0.482
         |
18-20 ye |   -18.4692   -4.52632    15.6502    6.29676
         |      0.000      1.000      0.037      0.627
```

These tables seem to indicate:
a) Students who began drinking between the ages of 18-20 are significantly more likely to use protective behaviors (compared to ages 12-14)
b) Students who have never gotten drunk are significantly more likely to use protective behaviors.
c) Students who got drunk for the first time between the ages of 18-20 are significantly more likely to use protective behaviors (compared to ages 12-14).

When I put all these factors into an AxB ANOVA, the effects disappear. All these conclusions hold for the post-survey as well.

==========

Q3: Does sex predict use patterns?

This is easier to see in a graph, but here's a t-test:

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    Male |      81    72.33333    2.178529    19.60676    67.99792    76.66874
  Female |     140    82.08571    1.707383    20.20203    78.70991    85.46151
```

```
---------+--------------------------------------------------------------------
combined |     221    78.51131    1.378295    20.48982    75.79496    81.22766
---------+--------------------------------------------------------------------
    diff |           -9.752381    2.790161               -15.25139   -4.253377
------------------------------------------------------------------------------
    diff = mean(Male) - mean(Female)                              t =  -3.4953
Ho: diff = 0                                     degrees of freedom =      219

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0003         Pr(|T| > |t|) = 0.0006          Pr(T > t) = 0.9997
```

Females reported using protective behaviors more frequently than males. Looking at the protective behaviors individually, these stick out:

Females are more likely to use…
78 - Pace your drinks to 1 or fewer per hour
80 - Alternate non-alcoholic drinks
85 - Have friends let you know when you've had enough
88 - Avoid drinking games
89 - Know where your drink is at all times
91 - Put extra ice in your drink

==========

Q4: If students report expecting to get in trouble, did they tend to report less use?

The two items related to "getting in trouble" were items 43-44. Let's look at those distributions:

Q43 Likelihood of getting in trouble with authorities	Q44. Likelihood of getting in trouble with your parents (1) V Unl.	(2)	(3)	(4)	(5)	(6)	(7)	Total
(1) Very unlikely	43	14	15	13	4	3	8	100
(2)	9	27	14	10	5	4	3	72
(3)	4	5	8	10	3	4	5	39
(4)	0	4	6	12	4	0	9	35
(5)	3	0	4	2	6	6	8	29
(6)	0	0	2	1	1	10	5	19
(7) Very Likely	0	0	2	1	1	1	24	29
Total	59	50	51	49	24	28	62	323

211 students (65%) think it's unlikely that they will get in trouble with authorities if they had 3-4 drinks. 160 students (50%) think it's unlikely that they will get in trouble with their parents if they had 3-4 drinks.

I can compare this to item #5 (frequency of use) to see if a relationship exists. Rather than run a test, I think this table shows a relationship:

Q43 Likelihood of getting in trouble with authorities	Q5. Self-Reported Drinking Rates and Behaviors - Frequency of High-Risk Drinking Never	Once	Twice	Three or	Total
(1) Very unlikely	19	12	12	1	44
(2)	21	13	7	4	45
(3)	6	6	2	0	14
(4)	8	1	3	0	12
(5)	5	1	1	0	7
(6)	4	0	0	0	4
(7) Very Likely	4	1	2	0	7

Total	67	34	27	5	133

Of the 67 students who did not engage in high-risk drinking, the majority thought it was unlikely that they would get in trouble with authorities if they had 3-4 drinks. 19% of these students, however, did think it was likely. I'll summarize the table another way to show the result:

Students who never engage in high-risk drinking:
69% think it is unlikely they will get in trouble with authorities
19% think it is likely

Students who engaged in high-risk drinking one time in 2 weeks:
91% think it is unlikely they will get in trouble with authorities
6% think it is likely

Students who engaged in high-risk drinking 2 times in 2 weeks:
78% think it is unlikely they will get in trouble with authorities
11% think it is likely

Students who engaged in high-risk drinking 3+ times in 2 weeks:
100% think it is unlikely they will get in trouble with authorities
0% think it is likely

This shows that students who drink more are less likely to think they will get in trouble with authorities if they have 3-4 drinks. Let's try the same analysis focused on getting in trouble with parents.

Q44 Likelihood of getting in trouble with parents	Q5. Self-Reported Drinking Rates and Behaviors - Frequency of High-Risk Drinking				
	Never	Once	Twice	Three or	Total
(1) Very unlikely	17	7	8	1	33
(2)	12	12	7	1	32
(3)	9	5	8	1	23
(4)	12	5	4	1	22
(5)	7	3	0	1	11
(6)	3	0	0	0	3
(7) Very Likely	7	2	0	0	9
Total	67	34	27	5	133

Students who never engage in high-risk drinking:
57% think it is unlikely they will get in trouble with authorities
25% think it is likely

Students who engaged in high-risk drinking one time in 2 weeks:
71% think it is unlikely they will get in trouble with authorities
15% think it is likely

Students who engaged in high-risk drinking 2 times in 2 weeks:
85% think it is unlikely they will get in trouble with authorities
0% think it is likely

Students who engaged in high-risk drinking 3+ times in 2 weeks:
60% think it is unlikely they will get in trouble with authorities
20% think it is likely

This shows a similar trend: Students who drink more are less likely to think they will get in trouble with their parents if they have 3-4 drinks.

These results held for the post-survey.

==========

Q5: Do scores on AlcoholEdu indicate use patterns (more use; lower score)?

The dataset doesn't actually contain their test scores (for items 134-135). Instead, every student just received a "1" for the first test, "2" for the second test, and so on.