Summary Statistics Page 1

Sunday, December 6, 2020 4:39 PM

Variable Name	Type
x1	Continuous
x2	Categorical
x3	Continuous
x4	Continuous
x5	Continuous
x6	Categorical
x7	Continuous
x8	Categorical
x9	Categorical
x10	Categorical
x11	Continuous
x12	Categorical
x13	Continuous
x14	Categorical

Continuous Variables

```
. summ x1 x3 x4 x5 x7 x11 x13
```

Variable	Obs	Mean	Std. Dev.	Min	Max
x1	151	11.66225	3.257581	6	21
x3	151	493.5762	186.7685	120	985
x4	151	8.900662	3.526575	2	23
x5	151	48.09934	11.19866	29	90
x7	151	.1986755	.5035854	0	3
x11	151	.2980132	.7098798	0	4
x13	151	81.90728	3.988903	65	86

Categorical Variables

```
. tab x2
```

x2	Freq.	Percent	Cum.
0	136	90.07	90.07
1	15	9.93	100.00
Total	151	100.00	

```
. tab x6
```

x6	Freq.	Percent	Cum.
0	44	29.14	29.14
1	107	70.86	100.00
Total	151	100.00	

```
. tab x8
```

x8	Freq.	Percent	Cum.
1	11	7.28	7.28
2	63	41.72	49.01
3	54	35.76	84.77
4	10	6.62	91.39
5	13	8.61	100.00
Total	151	100.00	

Summary Statistics Page 2

Sunday, December 6, 2020 10:31 PM

Categorical Variables Cont.

```
. tab x9
```

x9	Freq.	Percent	Cum.
0	66	43.71	43.71
1	85	56.29	100.00
Total	151	100.00	

```
. tab x10
```

x10	Freq.	Percent	Cum.
0	64	42.38	42.38
1	87	57.62	100.00
Total	151	100.00	

```
. tab x12
```

x12	Freq.	Percent	Cum.
1	55	36.42	36.42
2	52	34.44	70.86
3	28	18.54	89.40
4	8	5.30	94.70
5	8	5.30	100.00
Total	151	100.00	

```
. tab x14
```

x14	Freq.	Percent	Cum.
0	75	49.67	49.67
1	76	50.33	100.00
Total	151	100.00	

Logistic Regression w/ Odds Ratio Page 1

Sunday, December 6, 2020 7:44 PM

Model 1

```
. logit x2 x1 x3 x4 x7 x9 x10 x11 i.x12

note: 4.x12 != 0 predicts failure perfectly
      4.x12 dropped and 8 obs not used

Iteration 0:   log likelihood = -48.006155
Iteration 1:   log likelihood = -28.207472
Iteration 2:   log likelihood = -9.5303157
Iteration 3:   log likelihood = -5.6013261
Iteration 4:   log likelihood =  -4.161876
Iteration 5:   log likelihood = -4.1183138
Iteration 6:   log likelihood = -4.1157896
Iteration 7:   log likelihood = -4.1156154
Iteration 8:   log likelihood = -4.1156142
Iteration 9:   log likelihood = -4.1156142

Logistic regression                             Number of obs     =        143
                                                LR chi2(10)       =      87.78
                                                Prob > chi2       =     0.0000
Log likelihood = -4.1156142                     Pseudo R2         =     0.9143

------------------------------------------------------------------------------
          x2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |    .513863   .3922332     1.31   0.190    -.2548999    1.282626
          x3 |   -.060554   .0313958    -1.93   0.054    -.1220885    .0009806
          x4 |  -2.637786   1.689333    -1.56   0.118    -5.948819    .6732465
          x7 |  -3.194342   45.99427    -0.07   0.945    -93.34146    86.95278
          x9 |  -1.695295   3.299289    -0.51   0.607    -8.161782    4.771193
         x10 |   3.244547   4.537504     0.72   0.475    -5.648798    12.13789
         x11 |  -1.099486   11.69203    -0.09   0.925    -24.01545    21.81648
             |
         x12 |
          2  |   .8463053   4.075242     0.21   0.835    -7.141022    8.833633
          3  |  -1.140228    3.94459    -0.29   0.773    -8.871483    6.591027
          4  |           0  (empty)
          5  |   9.037978   61.78213     0.15   0.884    -112.0528    130.1287
             |
       _cons |   28.79344   19.34202     1.49   0.137    -9.116231    66.70311
------------------------------------------------------------------------------
Note: 60 failures and 2 successes completely determined.

. logit, or

note: 4.x12 != 0 predicts failure perfectly
      4.x12 dropped and 8 obs not used

Logistic regression                             Number of obs     =        143
                                                LR chi2(10)       =      87.78
                                                Prob > chi2       =     0.0000
Log likelihood = -4.1156142                     Pseudo R2         =     0.9143

------------------------------------------------------------------------------
          x2 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.671737   .6557105     1.31   0.190     .7749941    3.606096
          x3 |    .941243    .029551    -1.93   0.054       .88507    1.000981
          x4 |  .0715194   .1208202    -1.56   0.118     .0026089    1.960592
          x7 |  .0409935   1.885466    -0.07   0.945     2.90e-41    5.80e+37
          x9 |  .1835451   .6055683    -0.51   0.607     .0002854      118.06
         x10 |   25.6501   116.3874     0.72   0.475     .0035217    186818.4
         x11 |  .3330421    3.89394    -0.09   0.925     3.72e-11    2.98e+09
             |
         x12 |
          2  |  2.331019   9.499464     0.21   0.835     .0007919    6861.165
          3  |  .3197461   1.261267    -0.29   0.773     .0001403    728.5286
          4  |          1  (empty)
          5  |  8416.738     520004     0.15   0.884     2.17e-49    3.27e+56
             |
       _cons |   3.20e+12   6.18e+13     1.49   0.137     .0001099    9.31e+28
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
Note: 60 failures and 2 successes completely determined.
```

Model 2

```
. logit x2 x1 x3 x4 x7 x9 x10 x11

Iteration 0:   log likelihood = -48.867438
Iteration 1:   log likelihood = -29.001628
Iteration 2:   log likelihood = -7.6920668
Iteration 3:   log likelihood = -5.0095908
Iteration 4:   log likelihood = -4.2762379
Iteration 5:   log likelihood = -4.2417221
Iteration 6:   log likelihood = -4.2413143
Iteration 7:   log likelihood =  -4.241314

Logistic regression                             Number of obs     =        151
                                                LR chi2(7)        =      89.25
                                                Prob > chi2       =     0.0000
Log likelihood =  -4.241314                     Pseudo R2         =     0.9132

------------------------------------------------------------------------------
          x2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .5918773   .3527541     1.68   0.093     -.099508    1.283263
          x3 |  -.0586078   .0262244    -2.23   0.025    -.1100068   -.0072089
          x4 |  -2.914245    1.48098    -1.97   0.049    -5.816912   -.0115773
          x7 |    1.16314   6.854251     0.17   0.865    -12.27094    14.59722
          x9 |  -1.265774    2.59441    -0.49   0.626    -6.350724    3.819176
         x10 |   2.831577   3.649284     0.78   0.438    -4.320888    9.984042
         x11 |  -1.166463   17.65575    -0.07   0.947    -35.77109    33.43817
       _cons |    29.1867   16.55568     1.76   0.078    -3.261837    61.63524
------------------------------------------------------------------------------
Note: 59 failures and 0 successes completely determined.

. logit, or

Logistic regression                             Number of obs     =        151
                                                LR chi2(7)        =      89.25
                                                Prob > chi2       =     0.0000
Log likelihood =  -4.241314                     Pseudo R2         =     0.9132

------------------------------------------------------------------------------
          x2 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.807378    .63756     1.68   0.093     .9052828    3.608393
          x3 |   .9430765   .0247317    -2.23   0.025     .8958281    .9928171
          x4 |    .054245   .0803357    -1.97   0.049     .0029768    .9884894
          x7 |   3.199967   21.93337     0.17   0.865     4.69e-06     2185215
          x9 |    .282021   .7316781    -0.49   0.626     .0017455    45.56666
         x10 |   16.97221    61.9364     0.78   0.438     .0132881    21677.75
         x11 |   .3114666   5.499176    -0.07   0.947     2.92e-16    3.33e+14
       _cons |   4.74e+12   7.84e+13     1.76   0.078      .038318    5.86e+26
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
Note: 59 failures and 0 successes completely determined.
```

Models Interpretations Page 1

Sunday, December 6, 2020 10:13 PM

Model #1

Coefficients

- x1 has a coefficient of 0.51
  - for every one minute that travel time increases, the driver is more likely to take the freeway instead of the rural road by 0.51
- x3 has a coefficient of -0.61
  - for every one vehicle per hour increase, the driver is less likely to take the freeway to the rural road by -0.61
- x12.4 has no coefficient
  - this variable is collinear and therefore violates one of the assumptions and must be omitted for a valid model
- this model has a constant of 28.79
  - given all other factors are zero, the driver is more likely to take the freeway over the rural road by 28.79

Odds Ratios

- x1 has an odds ratio of 1.67
  - for every one minute that the travel time increases, the driver is 1.67 times more likely to take the freeway over the rural road
- x3 has an odds ratio of 0.94
  - for every one vehicle per hour increase, the driver is 0.94 times more likely to take the freeway over the rural road. This ratio is less than one which means the chances of the driver taking the freeway is decreasing as the rate of cars per hour increases.
- x12.4 has an odds ratio of 1.00
  - this variable is collinear and therefore, this variables is omitted from the final calculation
- this model has a constant of $3.20 * 10^{12}$
  - If all other factors are ignored, the driver is $3.20 * 10^{12}$ times more likely to take the freeway over the rural road
  - because this number is so high, not having any other information makes this model basically unusable

Model #2

Main Difference Relative to Model #1

This model excludes the x12 variable. By removing the x12 variable, variables x3 and x4 become more reliable to a realistic level with new P-Values under 0.05.

Both Models

- Neither model had more than two, if any variables under 0.05 which makes most of the data unreliable.
- If a logistic model only has x2 and one other variable, the model would work
- If all variables are included, the model breaks
- If any variable is added that isn't included in these two models, the model breaks
- by "breaks" I mean the model is too perfect and any data other than coefficients or odds ratios cannot be determined.

Stata Code Page 1

Sunday, December 6, 2020 11:05 PM

```
import excel "C:\Users\lmm56\Documents\School\Poly 20 - 21\QMB 3200\Final Exam\Data_FinalTakeHome.xlsx",
sheet("Sheet1") firstrow
summ x1 x3 x4 x5 x7 x11 x13
summ x1 x3 x4 x5 x7 x11 x13
gen route1 = cond(x2 == 1, 1, 0)
gen route2 = cond(x2 == 2, 1, 0)
gen seatBeltYes = cond(x6 == 1, 1, 0)
gen seatBeltNo = cond(x6 == 0, 1, 0)
gen driverAge18_23 = cond(x8 == 1, 1, 0)
gen driverAge24_29 = cond(x8 == 2, 1, 0)
gen driverAge30_39 = cond(x8 == 3, 1, 0)
gen driverAge40_49 = cond(x8 == 4, 1, 0)
gen driverAge50_Up = cond(x8 == 5, 1, 0)
gen genderMale = cond(x9 == 1, 1, 0)
gen genderFemale = cond(x9 == 0, 1, 0)
gen marriedYes = cond(x10 == 0, 1, 0)
gen marriedNo = cond(x10 == 1, 1, 0)
gen annIncomeLess_200k = cond(x12 == 1, 1, 0)
gen annIncome200k_299k = cond(x12 == 2, 1, 0)
gen annIncome300k_399k = cond(x12 == 3, 1, 0)
gen annIncome400k_499k = cond(x12 == 4, 1, 0)
gen annIncome500k_Up = cond(x12 == 5, 1, 0)
gen carDomestic = cond(x14 == 1, 1, 0)
gen carForeign = cond(x14 == 0, 1, 0)
tab route1 route2 seatBeltYes seatBeltNo driverAge18_23 driverAge24_29 driverAge30_39 driverAge40_49 driverAge50
_Up genderMale genderFemale marriedYes marriedNo annIncomeLess_200k annIncome200k_299k annIncome300k_
399k annIncome400k_499k annIncome500k_Up carDomestic carForeign
tab x12
tab x2
tab x6
tab x8
tab x9
tab x10
tab x14
logit route1 x1 x3 x4 x5 i.x6 x7 i.x8 i.x9 i.x10 x11 i.x12 x13 i.x14
logit route1 x1 x3 x4 x5 i.x6 x7 i.x8 i.x9 i.x10 x11 i.x12 x13 i.x14
logit route1 x1 x3 x4 x5 i.x6 x7 i.x8 i.x9 i.x10 x11 x13 i.x14
logit x2 x1 x3 x4 x5 i.x6 x7 i.x8 i.x9 i.x10 x11 i.x12 x13 i.x14
logit route1 x1 x3 x4 x5 i.x6 x7 i.x8 i.x9 i.x10 x11 i.x12 x13 i.x14
logit route2 x1 x3 x4 x5 i.x6 x7 i.x8 i.x9 i.x10 x11 i.x12 x13 i.x14
logit route1 x1 x3 x4 x5 x6 x7 i.x8 x9 x10 x11 i.x12 x13 x14
logit route1 x1 x3 x4 x5 x6 x7 i.x8 x9 x10 x11 i.x12 x13 x14
logit route1 x1 x3 x4 x5 x6 x7 i.x8 x9 x10 x11 x13 x14
logit route1 x1
logit route1 x1 x3
logit route1 x1 x3 x4
logit route1 x1 x3 x4 x5
```

Stata Code Page 2

Sunday, December 6, 2020 11:08 PM

```
anova route1 x1 x3 x4 x5 x7 x11 x13
loneway route1 x1 x3 x4 x5 x7 x11 x13
logit route1 x1 x3 x4 x6
logit route1 x1 x3 x4 i.x6
logit route1 x1 x3 x4
logit route1 x1 x3 x4 x7
logit route1 x1 x3 x4 x7 i.x8
logit route1 x1 x3 x4 x7 x9
logit route1 x1 x3 x4 x7 x9 x10
logit route1 x1 x3 x4 x7 x9 x10 x11
logit route1 x1 x3 x4 x7 x9 x10 x11 x13
logit route1 x1 x3 x4 x7 x9 x10 x11 x14
logit route1 x1 x3 x4 x7 x9 x10 x11
logit, or
logit route1 x8
logit route1 i.x8
logit route1 x3 x4 i.x8
replace x2 = x2-1
logit x2 x1 x3 x4 x5 x7 i.x9 i.x10 x11 x12
logit x2 x3 x4
logit x2 x1 x3 x4
logit x2 x3 x4 x5
logit route1 x3 x4
logit route2 x1 x3 x4 x7 x9 x10 x11
logit route2 x3 x4
logit, or
logit route1 x3 x4
logit, or
logit x2 x1
logit x2 x3
logit x2 x4
logit x2 x5
logit x2 x6
logit x2 i.x6
logit x2 x7
logit x2 x8
logit x2 i.x8
logit x2 x9
logit x2 x10
logit x2 x11
logit x2 i.x12
tab x12
logit x2 x13
logit x2 x14
logit x2 x1 x3 x4 x6 x7 x8 x9 x10 x11 x12 x13 x14
logit route1 x1 x3 x4 x6 x7 x8 x9 x10 x11 x12 x13 x14
logit x2 x1 x3 x4 x6 x7 i.x8 x9 x10 x11 i.x12 x13 x14
logit x2 x1 x3 x4 x7 i.x9 i.x10 x11 x12
```

Stata Code Page 3

Sunday, December 6, 2020 11:09 PM

```
logit x2 x1 x3 x4 x7 x9 i.x10 x11 x12
logit x2 x1 x3 x4 x7 x9 x10 x11 i.x12
logit, or
logit x2 x3 x4
logit, or
logit x2 x1 x3 x4 x5 x6 x7 i.x8 x9 x10 x11 i.x12 x13 x14
logit x2 x1 x3 x4 x6 x7 i.x8 x9 x10 x11 i.x12 x13 x14
logit x2 x1 x3 x4 x6 x7 x9 x10 x11 i.x12 x13 x14
logit x2 x1 x3 x4 x6 x7 x9 x10 x11 i.x12 x13
logit x2 x1 x3 x4 x6 x7 x9 x10 x11 i.x12
logit x2 x1 x3 x4 x7 x9 x10 x11 i.x12
logit x2 x1 x3 x4 x7 x9 x10 x11
anova x2 x1 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 x14
logit x2 x1 x3 x4 x7 x9 x10 x11
logit x2 x3 x4
logit x2 x1 x3 x4
logit x2 x3 x4 i.x8
logit x2 x1 x3 x4 x7 x9 x10 x11 i.x12
logit, or
logit x2 x3 x4
logit, or
summ x1 x3 x4 x5 x7 x11 x13
tab x2
tab x6
tab x8
tab x9
tab x10
tab x12
tab x14
logit x2 x1 x3 x4 x7 x9 x10 x11 i.x12
logit, or
logit route x1 x3 x4 x7 x9 x10 x11 i.x12
logit route2 x1 x3 x4 x7 x9 x10 x11 i.x12
logit, or
logit x2 x1 x3 x4 x7 x9 x10 x11
logit, or

log close
clear
```

My stata code is three pages because I was basically brute forcing the models until I found one that worked :(
p.s. check out the times when all of these pages were made.