Multinomial Logit Models - Overview

Richard Williams, University of Notre Dame, https://www3.nd.edu/~rwilliam/
Last revised January 17, 2022

This is adapted heavily from Menard's Applied Logistic Regression analysis; also, Borooah's Logit and Probit: Ordered and Multinomial Models; Also, Hamilton's Statistics with Stata, Updated for Version 7.

When categories are unordered, Multinomial Logistic regression is one often-used strategy. Mlogit models are a straightforward extension of logistic models.

Suppose a DV has M categories. One value (typically the first, the last, or the value with the most frequent outcome of the DV) is designated as the reference category. (Stata's `mlogit` defaults to the most frequent outcome, which I personally do not like because different subsample analyses may use different baseline categories). The probability of membership in other categories is compared to the probability of membership in the reference category.

For a DV with M categories, this requires the calculation of M-1 equations, one for each category relative to the reference category, to describe the relationship between the DV and the IVs.

Hence, if the first category is the reference, then, for m = 2, …, M,

$$\ln \frac{P(Yi = m)}{P(Yi = 1)} = \alpha_m + \sum_{k=1}^{K} \beta_{mk} X_{ik} = Z_{mi}$$

Hence, for each case, there will be M-1 predicted log odds, one for each category relative to the reference category. (Note that when m = 1 you get ln(1) = 0 = Z_{11}, and exp(0) = 1.)

When there are more than 2 groups, computing probabilities is a little more complicated than it was in logistic regression. For m = 2, …, M,

$$P(Y_i = m) = \frac{\exp(Z_{mi})}{1 + \sum_{h=2}^{M} \exp(Z_{hi})}$$

For the reference category,

$$P(Y_i = 1) = \frac{1}{1 + \sum_{h=2}^{M} \exp(Z_{hi})}$$

In other words, you take each of the M-1 log odds you computed and exponentiate it. Once you have done that the calculation of the probabilities is straightforward.

Note that, when M = 2, the mlogit and logistic regression models (and for that matter the ordered logit model) become one and the same.

We'll redo our Challenger example, this time using Stata's `mlogit` routine. In Stata, the most frequent category is the default reference group, but we can change that with the basecategory option, abbreviated b:

```
. mlogit  distress date temp, b(1)

Iteration 0:   log likelihood = -24.955257
Iteration 1:   log likelihood = -19.232647
Iteration 2:   log likelihood = -18.163998
Iteration 3:   log likelihood = -17.912395
Iteration 4:   log likelihood = -17.884218
Iteration 5:   log likelihood = -17.883654
Iteration 6:   log likelihood = -17.883653

Multinomial logistic regression                   Number of obs   =         23
                                                  LR chi2(4)      =      14.14
                                                  Prob > chi2     =     0.0069
Log likelihood = -17.883653                       Pseudo R2       =     0.2834

------------------------------------------------------------------------------
    distress |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1 or 2       |
        date |   .0017686   .0014431     1.23   0.220    -.0010599     .004597
        temp |  -.1054113   .1343361    -0.78   0.433    -.3687052    .1578826
       _cons |  -8.405851   10.47099    -0.80   0.422    -28.92862    12.11692
-------------+----------------------------------------------------------------
3 plus       |
        date |   .0067752   .0033931     2.00   0.046     .0001248    .0134256
        temp |  -.2964675   .1568354    -1.89   0.059    -.6038594    .0109243
       _cons |  -40.43276   25.17892    -1.61   0.108    -89.78254    8.917024
------------------------------------------------------------------------------
(Outcome distress==none is the comparison group)
```

For group 2 (one or two distress incidents), the coefficients tell us that lower temperatures and higher dates increase the likelihood that you will have one or two distress incidents as opposed to none. We see the same thing in group 3, but the effects are even larger.

To have Stata compute the Z values and the predicted probabilities of being in each group:

```
. predict z2, xb outcome(2)
. predict z3, xb outcome(3)
. * You could predict z1 - but it would be 0 for every case!
. predict mnone monetwo mthreeplus, p
```

```
. list flight temp date distress z2 z3 mnone monetwo mthreeplus
```

	flight	temp	date	distress	z2	z3	mnone	monetwo	mthree~s
1.	STS-1	66	7772	none	-1.6178	-7.342882	.8340411	.1654192	.0005398
2.	STS-2	70	7986	1 or 2	-1.660975	-7.078863	.8397741	.1595182	.0007077
3.	STS-3	69	8116	none	-1.325651	-5.901621	.7884166	.209427	.0021563
4.	STS-4	80	8213	.	-2.313626	-8.505571	.9098317	.0899842	.0001841
5.	STS-5	68	8350	none	-.8063986	-4.019761	.6828641	.3048736	.0122624
6.	STS-6	67	8494	1 or 2	-.4463157	-2.747666	.5868342	.3755631	.0376027
7.	STS-7	72	8569	none	-.8407306	-3.721865	.6870095	.2963726	.0166179
8.	STS-8	73	8642	none	-.8170375	-3.523744	.6797047	.3002516	.0200437
9.	STS-9	70	8732	none	-.3416339	-2.024575	.5426942	.385643	.0716627
10.	STS_41-B	57	8799	1 or 2	1.147206	2.28344	.0716345	.2256043	.7027612
11.	STS_41-C	63	8862	3 plus	.6261569	.9314718	.184889	.345818	.469293
12.	STS_41-D	70	9008	3 plus	.1464868	-.154624	.3317303	.384064	.2842057
13.	**STS_41-G**	**78**	**9044**	**none**	**-.6331355**	**-2.282458**	**.6123857**	**.3251306**	**.0624836**
14.	STS_51-A	67	9078	none	.5865193	1.209041	.1626547	.2924077	.5449376
15.	STS_51-C	53	9155	3 plus	2.198456	5.881276	.0027153	.0244682	.9728165
16.	STS_51-D	67	9233	3 plus	.8606451	2.259195	.0772794	.1827414	.7399792
17.	STS_51-B	75	9250	3 plus	.0474203	.0026329	.32774	.3436559	.3286041
18.	STS_51-G	70	9299	3 plus	.6611357	1.816955	.11001	.2130884	.6769016
19.	STS_51-F	81	9341	1 or 2	-.424109	-1.159631	.5081418	.3325039	.1593543
20.	STS_51-I	76	9370	1 or 2	.1542354	.5191875	.259914	.3032586	.4368274
21.	STS_51-J	79	9407	none	-.096562	-.1195333	.3577449	.3248158	.3174394
22.	STS_61-A	75	9434	3 plus	.3728341	1.249267	.1683607	.2444334	.5872059
23.	STS_61-B	76	9461	1 or 2	.3151737	1.135729	.1823506	.249911	.5677384
24.	STS_61-C	58	9508	3 plus	2.295699	6.790579	.0011107	.0110305	.9878589
25.	**STS_51-L**	**31**	**9524**	**.**	**5.1701**	**14.90361**	**3.37e-07**	**.0000593**	**.9999404**

To verify that Stata got it right, note that

$$Z_{2i} = -8.4059 -.10541*Temp + .001769*Date$$

$$Z_{3i} = -40.433 -.29647*Temp + .006775*Date.$$

Hence, for flight 13, where Temp = 78 and Date = 9044, we get

$$Z_2 = -8.4059 -.10541*78 + .001769*9044 = -.629$$

$$Z_3 = -40.433 -.29647*78 + .006775*9044 = -2.2846$$

In each case, the negative numbers tell us flight 13 was more likely to fall in the reference category. From these numbers, we can compute that, for Flight 13,

$$P(Y_i = 1) = \frac{1}{1 + \sum_{h=2}^{M} \exp(Z_{hi})} = \frac{1}{1 + \exp(-.629) + \exp(-2.2846)} = .6116$$

$$P(Y_i = 2) = \frac{\exp(Z_{1i})}{1 + \sum_{h=2}^{M} \exp(Z_{hi})} = \frac{\exp(-.629)}{1 + \exp(-.629) + \exp(-2.2846)} = .326$$

$$P(Y_i = 3) = \frac{\exp(Z_{2i})}{1 + \sum_{h=2}^{M} \exp(Z_{hi})} = \frac{\exp(-2.2846)}{1 + \exp(-.629) + \exp(-2.2846)} = .0623$$

These numbers are similar to what we got with the ordinal regression. If we do similar calculations for Challenger, we get P(Y = 1) = .0005367, P(Y = 2) = .0000593, P(Y = 3) = .9999404.

So, in this case, both the multinomial and ordinal regression approaches produce virtually identical results, but the ordinal regression model is somewhat simpler and requires the estimation of fewer parameters. Note too that in the Ordered Logit model the effects of both Date and Time were statistically significant, but this was not true for all the groups in the Mlogit analysis; this probably reflects the greater efficiency of the Ordered Logit approach. Particularly in a model with more X variables and/or categories of Y, the ordinal regression approach would be simpler and hence preferable, provided its assumptions are met.

In short, the models get more complicated when you have more than 2 categories, and you get a lot more parameter estimates, but the logic is a straightforward extension of logistic regression.

Closing Comments. A few other things you may want to consider:

- You may want to combine some categories of the DV, partly to make the analysis simpler, and partly because the number of cases in some categories may be very small. Remember, the more categories you have, the more parameters you will estimate, and the more difficult it may be to get significant results. It is simplest, of course, to only have two categories, but you'll have to decide whether or not that is justified for your particular problem.

- Make sure you understand what the reference category is, since different programs do it differently. You may need to recode the variable if there is no other way of changing the reference category. However, in Stata, you can just use the b option; b is short for baseoutcome. I usually choose b(1).

- If the DV is ordinal, other techniques may be appropriate and more parsimonious.

Appendix A: Adjusted Predictions and Marginal Effects for Multinomial Logit Models

We can use the exact same commands that we used for `ologit` (substituting `mlogit` for `ologit` of course). Since there is nothing new here I will simply give the commands and output. Make sure you understand what is happening at each step. If you compare with the earlier `ologit` handout, you'll see that results are not identical but (at least for this example) are pretty similar.

```
. * Appendix A: Adjusted predictions & Marginal effects
. * Requires Stata 14+
. webuse nhanes2f, clear

. keep if !missing(diabetes, black, female, age)
(2 observations deleted)

. label define black 0 "nonBlack" 1 "black"
. label define female 0 "male" 1 "female"
. label values black black
. label values female female
. mlogit health i.female i.black c.age, nolog b(1)

Multinomial logistic regression                 Number of obs     =     10,335
                                                LR chi2(12)       =    1821.98
                                                Prob > chi2       =     0.0000
Log likelihood = -14853.408                     Pseudo R2         =     0.0578

------------------------------------------------------------------------------
      health |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
poor         |  (base outcome)
-------------+----------------------------------------------------------------
fair         |
      female |
     female  |   .3712131   .0894146     4.15   0.000     .1959637    .5464626
             |
       black |
      black  |  -.4491975   .1173988    -3.83   0.000    -.6792949      -.2191
         age |  -.0208594   .0034329    -6.08   0.000    -.0275878   -.0141309
       _cons |   1.927039   .2153915     8.95   0.000     1.504879    2.349198
-------------+----------------------------------------------------------------
average      |
      female |
     female  |    .276952   .0844963     3.28   0.001     .1113424    .4425616
             |
       black |
      black  |  -.7897314   .1129536    -6.99   0.000    -1.011116   -.5683463
         age |  -.0505401    .003225   -15.67   0.000     -.056861   -.0442191
       _cons |   4.160382   .2008492    20.71   0.000     3.766724    4.554039
-------------+----------------------------------------------------------------
good         |
      female |
     female  |   .2296885   .0871759     2.63   0.008     .0588268    .4005502
             |
       black |
      black  |  -1.425797   .1260638   -11.31   0.000    -1.672878   -1.178716
         age |  -.0715066   .0032844   -21.77   0.000    -.0779439   -.0650693
       _cons |   5.093431   .2019058    25.23   0.000     4.697703    5.489159
-------------+----------------------------------------------------------------
```

```
excellent    |
      female |
     female  |   .0204885   .0889547     0.23   0.818    -.1538596    .1948365
             |
       black |
      black  |  -1.721134   .1348555   -12.76   0.000    -1.985446   -1.456822
         age |  -.0842692   .0033392   -25.24   0.000     -.090814   -.0777245
       _cons |   5.679135   .2028395    28.00   0.000     5.281577    6.076693
------------------------------------------------------------------------------

. * AAPs using margins
. margins black

Predictive margins                              Number of obs     =     10,335
Model VCE    : OIM

1._predict   : Pr(health==poor), predict(pr outcome(1))
2._predict   : Pr(health==fair), predict(pr outcome(2))
3._predict   : Pr(health==average), predict(pr outcome(3))
4._predict   : Pr(health==good), predict(pr outcome(4))
5._predict   : Pr(health==excellent), predict(pr outcome(5))

--------------------------------------------------------------------------------
               |            Delta-method
               |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
_predict#black |
   1#nonBlack  |   .0627775   .0024596    25.52   0.000     .0579567    .0675982
      1#black  |   .1406454   .0104604    13.45   0.000     .1201435    .1611474
   2#nonBlack  |   .1535468   .0036354    42.24   0.000     .1464216    .1606721
      2#black  |   .2307221     .01267    18.21   0.000     .2058895    .2555548
   3#nonBlack  |   .2785696   .0046427    60.00   0.000       .26947    .2876692
      3#black  |   .3275166   .0141872    23.09   0.000     .2997103     .355323
   4#nonBlack  |   .2595737   .0045198    57.43   0.000      .250715    .2684324
      4#black  |   .1736632   .0111181    15.62   0.000     .1518721    .1954544
   5#nonBlack  |   .2455324   .0043418    56.55   0.000     .2370226    .2540421
      5#black  |   .1274526    .009619    13.25   0.000     .1085997    .1463054
--------------------------------------------------------------------------------

. *spost13
. mtable, at(black = (0 1))

Expression: Pr(health), predict(outcome())

           |    black      poor      fair   average      good  excellent
 ----------+------------------------------------------------------------
         1 |        0     0.063     0.154     0.279     0.260      0.246
         2 |        1     0.141     0.231     0.328     0.174      0.127

Specified values where .n indicates no values specified with at()

           |  No at()
 ----------+---------
   Current |       .n
```

```
. * AMEs using margins
. margins, dydx(black)

Average marginal effects                          Number of obs     =     10,335
Model VCE    : OIM

dy/dx w.r.t. : 1.black
1._predict   : Pr(health==poor), predict(pr outcome(1))
2._predict   : Pr(health==fair), predict(pr outcome(2))
3._predict   : Pr(health==average), predict(pr outcome(3))
4._predict   : Pr(health==good), predict(pr outcome(4))
5._predict   : Pr(health==excellent), predict(pr outcome(5))

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1.black      |
    _predict |
          1  |    .077868    .010746     7.25   0.000     .0568062    .0989297
          2  |   .0771753   .0131821     5.85   0.000     .0513389    .1030118
          3  |    .048947   .0149289     3.28   0.001     .0196868    .0782072
          4  |  -.0859105   .0120031    -7.16   0.000    -.1094361   -.0623849
          5  |  -.1180798   .0105546   -11.19   0.000    -.1387665   -.0973931
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

. mtable, dydx(black)

Expression: Marginal effect of Pr(health), predict(outcome())

    poor      fair   average      good  excellent
-------------------------------------------------
   0.078     0.077     0.049    -0.086     -0.118


. * mtable
. mtable, at (black = (0 1) age = 20 ) at (black = (0 1) age = 47 ) at (black = (0 1) age = 74 ) dec(4)

Expression: Pr(health), predict(outcome())

           |    black       age      poor      fair   average      good  excellent
 ----------+------------------------------------------------------------------------
         1 |        0        20    0.0076    0.0417    0.2039    0.3321     0.4147
         2 |        1        20    0.0270    0.0947    0.3294    0.2842     0.2647
         3 |        0        47    0.0435    0.1361    0.2988    0.2764     0.2452
         4 |        1        47    0.1159    0.2306    0.3603    0.1765     0.1167
         5 |        0        74    0.1660    0.2948    0.2905    0.1526     0.0960
         6 |        1        74    0.3072    0.3487    0.2443    0.0679     0.0318

Specified values where .n indicates no values specified with at()

           |  No at()
 ----------+---------
   Current |       .n

. quietly mtable, at (black = 0 age = 20 ) rown(20 year old white) dec(4)
. quietly mtable, at (black = 1 age = 20 ) rown(20 year old black) dec(4) below
. quietly mtable, at (black = 0 age = 47 ) rown(47 year old white) dec(4) below
. quietly mtable, at (black = 1 age = 47 ) rown(47 year old black) dec(4) below
. quietly mtable, at (black = 0 age = 74 ) rown(74 year old white) dec(4) below
. mtable, at (black = 1 age = 74 ) rown(74 year old black) dec(4) below
```

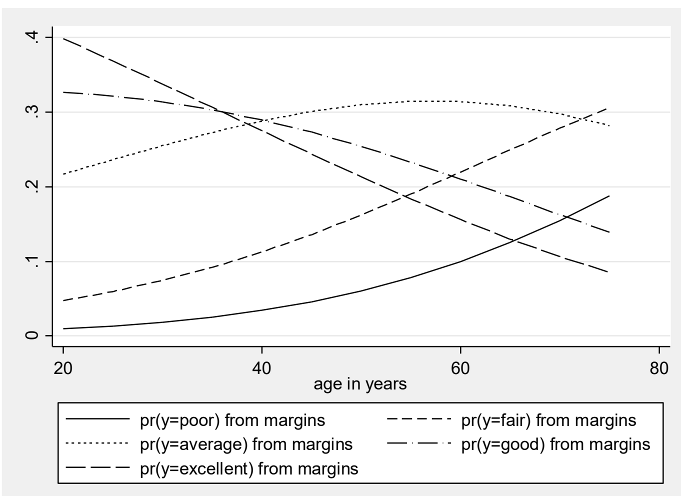
```
Expression: Pr(health), predict(outcome())

                    |     poor      fair   average      good  excellent
 -------------------+--------------------------------------------------
  20 year old white |   0.0076    0.0417    0.2039    0.3321     0.4147
  20 year old black |   0.0270    0.0947    0.3294    0.2842     0.2647
  47 year old white |   0.0435    0.1361    0.2988    0.2764     0.2452
  47 year old black |   0.1159    0.2306    0.3603    0.1765     0.1167
  74 year old white |   0.1660    0.2948    0.2905    0.1526     0.0960
  74 year old black |   0.3072    0.3487    0.2443    0.0679     0.0318

Specified values of covariates

           |    black       age
 ----------+-------------------
     Set 1 |        0        20
     Set 2 |        1        20
     Set 3 |        0        47
     Set 4 |        1        47
     Set 5 |        0        74
   Current |        1        74
```

```
* Graphics using mgen
* mgen for all groups pooled together
mgen, at(age = (20(5)75)) stub(all)
list  allpr1 allpr2 allpr3 allpr4 allpr5 allage in 1/15
line allpr1 allpr2 allpr3 allpr4 allpr5 allage, scheme(sj) name(pooled)
```




```
* mgen for groups
drop allpr1 - allCpr5
mgen, at(age = (20(5)75) black  = 0) stub(wh) predn(whpr)
mgen, at(age = (20(5)75) black  = 1) stub(bl) predn(blpr)
line whwhpr1 blblpr1 whwhpr5 blblpr5 whage, scheme(sj) name(byrace)
```


```
. * mchange
. mchange black female age, stats(change start end) dec(5) delta(10)

mlogit: Changes in Pr(y) | Number of obs = 10335

Expression: Pr(health), predict(outcome())

                   |      poor       fair    average       good  excellent
-------------------+------------------------------------------------------
black              |
 black vs nonBlack |   0.07787    0.07718    0.04895   -0.08591   -0.11808
              From |   0.06278    0.15355    0.27857    0.25957    0.24553
                To |   0.14065    0.23072    0.32752    0.17366    0.12745
female             |
    female vs male |  -0.01537    0.02542    0.02077    0.00868   -0.03951
              From |   0.07869    0.14817    0.27340    0.24619    0.25355
                To |   0.06333    0.17360    0.29417    0.25487    0.21404
age                |
                +1 |   0.00337    0.00469    0.00099   -0.00342   -0.00562
              From |   0.07054    0.16159    0.28428    0.25070    0.23290
                To |   0.07390    0.16627    0.28527    0.24728    0.22728
            +delta |   0.03889    0.04812    0.00359   -0.03660   -0.05399
              From |   0.07054    0.16159    0.28428    0.25070    0.23290
                To |   0.10943    0.20970    0.28787    0.21410    0.17890
          Marginal |   0.00331    0.00466    0.00106   -0.00339   -0.00564
              From |        .z         .z         .z         .z         .z
                To |        .z         .z         .z         .z         .z

Average predictions

             |      poor       fair    average       good  excellent
-------------+------------------------------------------------------
  Pr(y|base) |   0.07054    0.16159    0.28428    0.25070    0.23290

1: Delta equals 10.
```

If you are condemned to using Stata 13 or earlier you can similarly adapt the code that was given earlier for ologit.

Appendix B: Using SPSS NOMREG for Multinomial Logistic Regression

Note: I have not used SPSS in years, but this code did work in the past and may or may not still work now.

```
NOMREG
  distress  (base = first) WITH temp date
  /CRITERIA = CIN(95) DELTA(0) MXITER(100) MXSTEP(5) CHKSEP(20) LCONVERGE(0)
  PCONVERGE(1.0E-6) SINGULAR(1.0E-8)
  /MODEL
  /INTERCEPT = INCLUDE
  /PRINT = PARAMETER SUMMARY LRT
  /Save = ESTPROB (MLog) .
```

Nominal Regression

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	49.911			
Final	35.767	14.143	4	.007

Pseudo R-Square

Cox and Snell	.459
Nagelkerke	.519
McFadden	.283

Likelihood Ratio Tests

Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	40.714	4.946	2	.084
TEMP	42.739	6.972	2	.031
DATE	47.243	11.475	2	.003

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

Parameter Estimates

DISTRESS thermal distress incidents[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
2 1 or 2	Intercept	-8.4059	10.471	.644	1	.422			
	TEMP	-.10541	.134	.616	1	.433	.900	.692	1.171
	DATE	.001769	.001	1.502	1	.220	1.002	.999	1.005
3 3 plus	Intercept	-40.433	25.179	2.579	1	.108			
	TEMP	-.29647	.157	3.573	1	.059	.743	.547	1.011
	DATE	.006775	.003	3.987	1	.046	1.007	1.000	1.014

a. The reference category is: 1 none.

Because we included the parameter /Save = ESTPROB (MLog), we can also get the estimated probabilities for each case of falling into each of the three groups (again with the exception of the case we really want, case 25).

```
Formats mlog1_1 mlog2_1 mlog3_1 (f8.4).
List flight temp date distress mlog1_1 mlog2_1 mlog3_1 .
```

List

FLIGHT	TEMP	DATE	DISTRESS	MLOG1_1	MLOG2_1	MLOG3_1
1	66	7772	1	.8340	.1654	.0005
2	70	7986	2	.8398	.1595	.0007
3	69	8116	1	.7884	.2094	.0022
4	80	8213	.	.	.	.
5	68	8350	1	.6829	.3049	.0123
6	67	8494	2	.5868	.3756	.0376
7	72	8569	1	.6870	.2964	.0166
8	73	8642	1	.6797	.3003	.0200
9	70	8732	1	.5427	.3856	.0717
10	57	8799	2	.0716	.2256	.7028
11	63	8862	3	.1849	.3458	.4693
12	70	9008	3	.3317	.3841	.2842
13	**78**	**9044**	**1**	**.6124**	**.3251**	**.0625**
14	67	9078	1	.1627	.2924	.5449
15	53	9155	3	.0027	.0245	.9728
16	67	9233	3	.0773	.1827	.7400
17	75	9250	3	.3277	.3437	.3286
18	70	9299	3	.1100	.2131	.6769
19	81	9341	2	.5081	.3325	.1594
20	76	9370	2	.2599	.3033	.4368
21	79	9407	1	.3577	.3248	.3174
22	75	9434	3	.1684	.2444	.5872
23	76	9461	2	.1824	.2499	.5677
24	58	9508	3	.0011	.0110	.9879
25	**31**	**9524**	.	.	.	.

```
Number of cases read:  25    Number of cases listed:  25
```