Statistical Analysis of Counts and Proportions Final Exam

Vinny Paris

December 12, 2017

1 Danish Alcohol Consumption

1.1 Major Findings

A study done to collect the count of Danish citizens alcoholic consumption rates was preformed and marriage status, income and location of their home was recorded. It was found that Danes income level in DKK[1] increases the odds of alcohol consumption rates in general. There was a slight drop off though at the top echelon for those who made over 150,000DKK It was also shown that the less urban a citizen is the less likely they are to drink on a regular basis.

2 Methods

2.1 The Working Model

Using the stepAIC method the following model was fit to the data:

```
Count = Alc + Income + Urban + Mar + Alc:Income +
    Alc:Urban + Income:Urban + Alc:Mar +
    Income:Mar + Urban:Mar +
    Alc:Income:Mar
```

Where Alc is one of three levels where an individual has less than one drink daily, 1-2 drinks daily, and more than 2 drinks.

Income is one of 4 defined groups with them being 0-50k, 50k-100k, 100k-150k, 150k+. Marriage status is widowed, married, and never married. It is unclear what divorcées are labeled as. Urban refers to the subjects place of resident which is Copenhagen, Suburbs of Copenhagen, 3 largest towns besides Copenhagen, Other Cities, or Countryside.

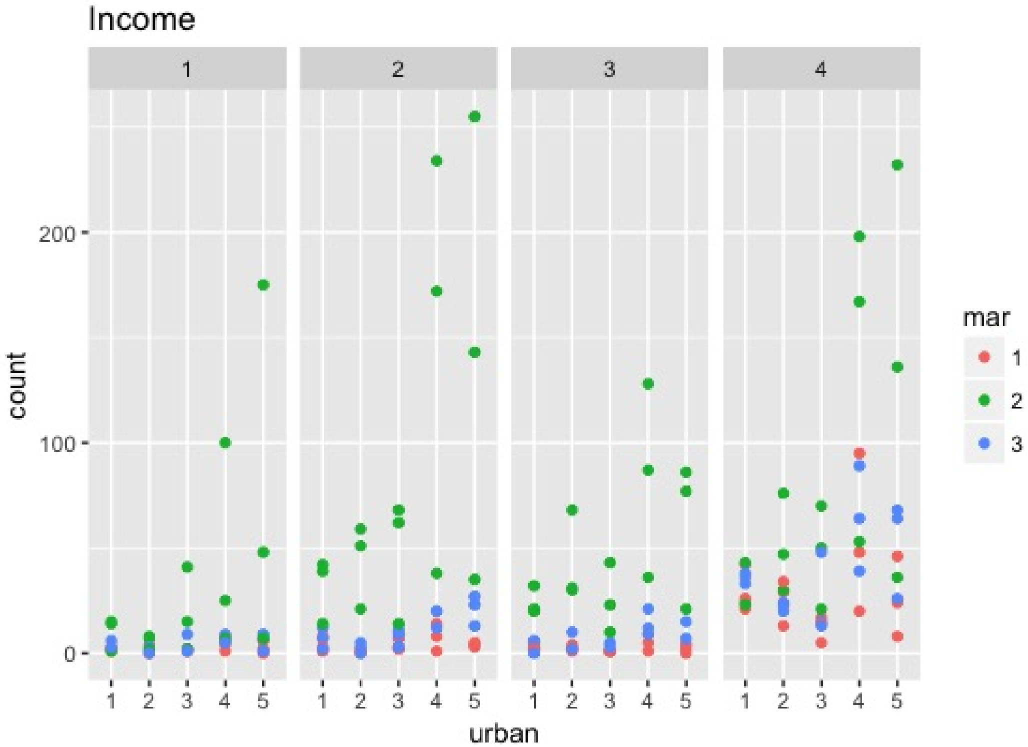
Please note that I am are treating Alc, Income and Urban as nominal variables in this analysis until the near final part of the analysis. The author wanted to allow for the most amount of freedom among the levels as possible and decided to sacrifice the degrees of freedom for it. Also, while the groups were clearly cut in income there could be a lot of lurking variables that went unaccounted for. Part of this was justified in the data graphical exploratory stage of the analysis where a linear trend seems to be poorly justified.

The following graph gives the relative counts along the y-axis with the x-axis being made up of the urban status. The coloration is based on marriage status. Alcohol is not displayed in this graph, it just gives a sense of the number of respondents and the interactions between the other variables. The one I'd like to draw your attention to currently is the overall dispersion of the counts going from one income group to another (the four large, distinct boxes). The elongated upper tails seem to be most prominent in income groups 2 and 4 (50k-100k DKK and 150k+ DKK respectively) while the other two groups (1 with 0-50k DKK and 3 from 100k-150k DKK) seem to be more modest. And then finally the location of the bulk of the data seems to increase rather dramatically in the final income group 4 while the others are more or less constant ignoring the tails.

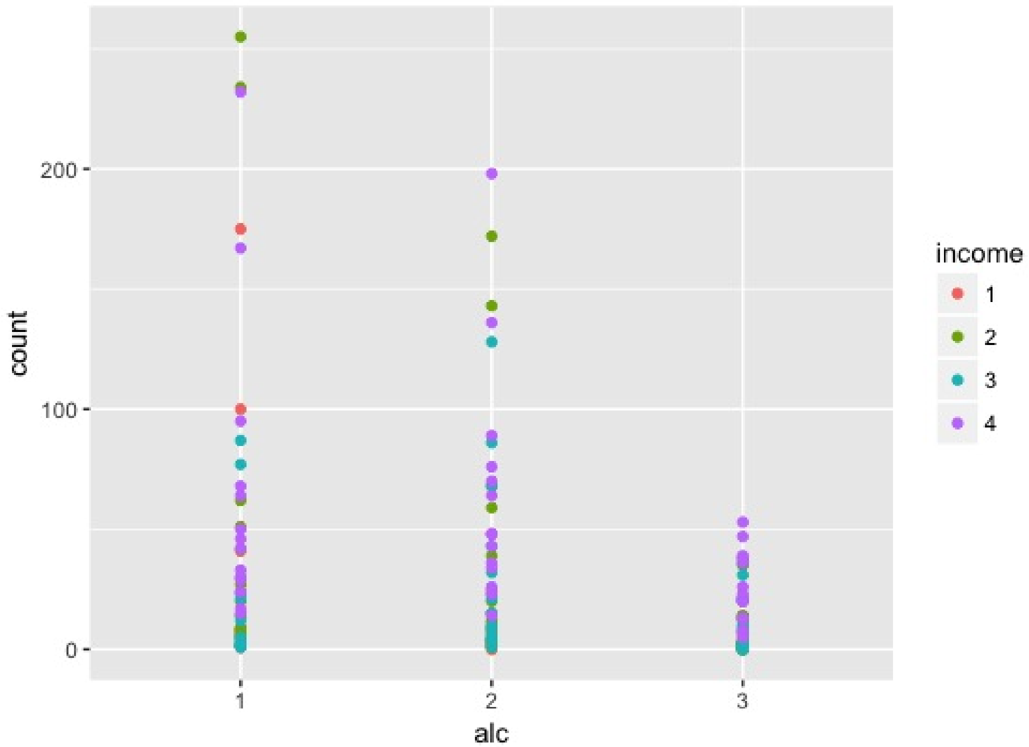
The second thing to note about the graph is that the dispersion seems to increase quiet rapidly the more rural you go. Again, I am still uneasy about thinking of urban as ordinal without better understanding of how it was decided a suburb of Copenhagen qualified as “more city” than the

[1] Danish Krone, currently trading at 6.32 DKK for 1 USD

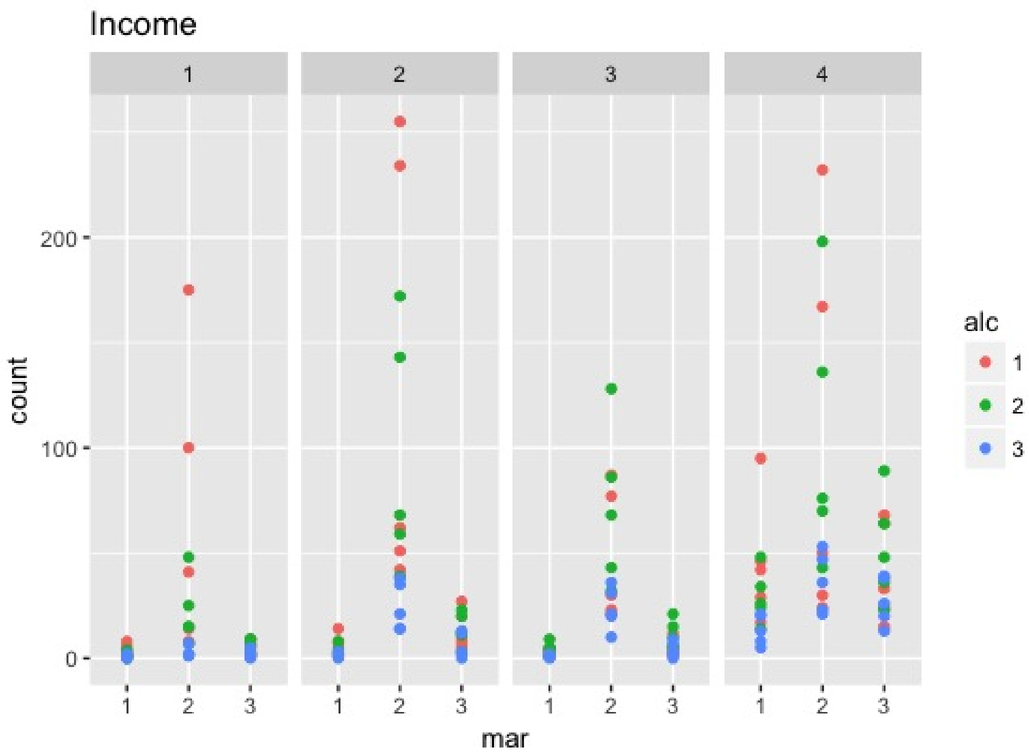
other three largest cities in Denmark. The interesting thing about that though is the final two levels ("other cities" and "countryside") do seem to have a pretty intuitive ordering and it is them that act most strangely. Groupings 1-3 are comparable but 4 becomes more dispersed and 5 even more so. What this means I am not positive. It could simply be that more rural people fill out surveys more often[2]. Also note that being married is by far the largest group. Regardless there is not a linear trend in urban's levels; possibly an exponential curve or maybe even if we combined the first three levels to create a 3 level variable with "urban", "small city", "rural" we could do better for a linear fit. The second graph presented shows the count for the three alcohol groups. Again, there isn't so much a linear fit as maybe a curved fit. So with these graphs in mind, all variables were treated as nominal.



[2] I grew up in a town of 112 people, you have time on your hands to fill out a survey



The final graph that is presented in this section is honestly rather boring. It suggests that most Danes in the target population are in fact married and has strong effects at that between married and the other two (single, widowed) with again ambiguous grouping for the divorced.



Further analysis showed that the model suffered from an overdisperison of about 40% The author then fit a quasipoisson model to account for that variation. After that the working model was tested with an anova ran using the χ^2-test and it was discovered that the three term interaction was non-signficant so it was removed. The resulting and final model then was strictly main effects and all two way interactions. Again, a quasipoisson was fit and the overdispersion parameter was found to be 1.57701 so about a 60% increase in the amount of variation would exist in a regular poisson model without this correction. The final model then is as follows:

```
Count = Alc + Income + Urban + Mar + Alc:Income +
    Alc:Urban + Income:Urban + Alc:Mar +
    Income:Mar + Urban:Mar
```

With "Count" being considered a quasipoisson variable.

A Goodman-Kruskal Gamma will still be ran to test the relative agreement between alcohol (which has three distinct ordinal levels), income (four distinct ordinal levels) and urban (which does **not** have as clear an ordinal relationship but I decided to play with it some like it did). The results will be presented in part three.

The last major step in the analysis was building a logit model out of the log linear form. R's "set first level to 0" method allowed for an intuitive approach of comparing alc level 1 vs 2 and alc level 1 vs 3. The baseline of level 1 which corresponds to less than one drink a night seems to the author to be the most common category and just screams natural baseline.

3 Results

The following is an ANOVA using the χ^2 test as well as a goodness of fit test using deviance and Pearson's tests. This should be taken with some salt though due to the very limited counts the low income, high alc groups, some of which are outright zero. Of the two goodness of fit tests Deviance will be the more robust.

```
 Pearson test =  195.5
  Degrees of freedom =  124
     p-value =  4e-05

 Deviance test =  189.85
   df =  124
     p-value =  0.00013
```

```
Model: quasipoisson, link: log

Response: count

Terms added sequentially (first to last)


             Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                           179     8268.4
alc           2  1014.22       177     7254.2 < 2.2e-16 ***
income        3  1289.60       174     5964.6 < 2.2e-16 ***
urban         4  1399.79       170     4564.8 < 2.2e-16 ***
mar           2  3061.08       168     1503.8 < 2.2e-16 ***
alc:income    6   224.42       162     1279.3 < 2.2e-16 ***
alc:urban     8   198.53       154     1080.8 < 2.2e-16 ***
income:urban 12   100.43       142      980.4 4.773e-09 ***
alc:mar       4    86.40       138      894.0 3.598e-11 ***
income:mar    6   570.86       132      323.1 < 2.2e-16 ***
urban:mar     8   133.26       124      189.9 6.037e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Then our model is still statistically worse than a fully parameterized option. This is not good but the only model I could find that wasn't the full model was and exhausted all three way interactions model with an effect hierarchy principle used and even then the goodness of fit tests are still weak and not robust enough and the model is severely punished by the AIC.

The parameter estimates in log-linear form and their standard errors are reported on the back page as well as a confidence interval on the linear scale. You may look through the estimates on your own. The major highlights are that income group 4 does have a major boost in the number of counts. Due to all the interactions it is hard to get a hold of what is happening in the data though, why I chose to include the graphs above. It allows for a much cleaner reading. A point of interest is that marriage and widow both will end up in the favoring the upper two alc levels over the baseline single and in alcohol group 1.

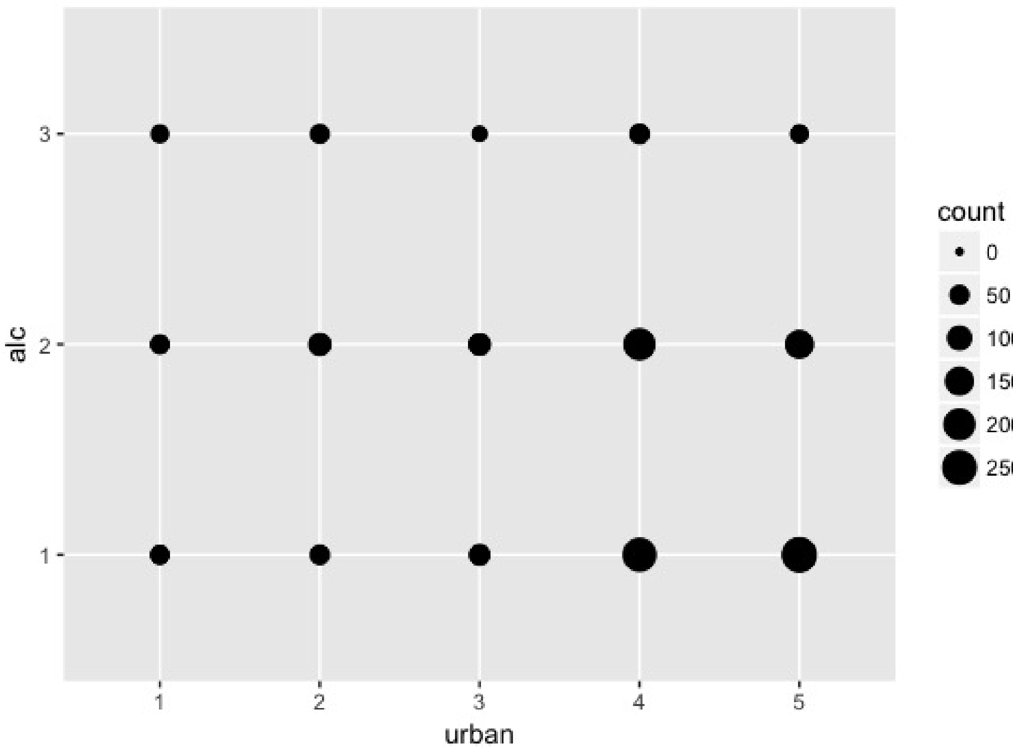
A small test using the Goodman-Kruskal Gamma was used to test the ordinal agreement between income and urban, income and alcohol, and urban and alcohol. The following were the results:

Urban-Income = -0.1311951

Income-Alcohol = 0.2056987

Urban-Alcohol = -0.2328338

The most interesting one to me is the Urban and Alcohol result saying that alcohol and urban work in opposite directions. This is surprising to me personally because of how often I hear of drugs ravaging rural America on the news (I'm thinking alcohol would be roughly the same and that Denmark and the US's effects of urbanization would be similar. This could be very wrong thinking but it was my thought process. A look at the parameter estimates on the back page will support the Gamma's conclusion and was as the following graph:



The other thing to note is the postive relationship between Income and Alcohol. It appears to be a modest relationship between income growth and alcohol consumption.

The following is the logit models. Again, alcohol group 1 was the base line from which the other two levels were studied allowing for the difference in the log odds of being in category (2 or 3) compared being in group 1. The following table is for 1vs2 and secondly 1vs3

Notice as well that the more rural one becomes the lower the odds of being in a alcohol group larger than 1. Also please note that income does have that strange curvature with a max peak in group 3. I could hazard a few guesses as to why but they would be mostly shots in the dark. The major advantages of using these logit models is that it actually lets us see how different variables effect alcohol. Up until now all our model was doing was talking about relative associations and independence. Now we can actually say something about how different variables directly influence a citizens drinking habit (or at least the correlation). This also allows us to strip away the main effects of variables not alcohol from the model thanks to the difference in the log odds

Table 1: Logit 1 vs 2

	Parameter	Std. Error	Lower CI	Upper CI
int	-1.13	0.20	0.22	0.48
income2	0.61	0.14	1.40	2.44
income3	1.32	0.16	2.75	5.09
income4	0.86	0.14	1.82	3.13
urban2	0.28	0.17	0.94	1.86
urban3	0.18	0.17	0.86	1.66
urban4	-0.19	0.14	0.63	1.09
urban5	-0.53	0.14	0.44	0.78
mar2	0.42	0.13	1.18	1.98
mar3	0.79	0.15	1.63	2.97

Table 2: Logit 1 vs 3

	Parameter	Std. Error	Lower CI	Upper CI
int	-1.97	0.30	0.08	0.25
income2	0.81	0.25	1.40	3.78
income3	1.62	0.26	3.06	8.62
income4	1.31	0.24	2.35	6.11
urban2	0.13	0.20	0.76	1.70
urban3	-0.54	0.22	0.37	0.90
urban4	-0.79	0.18	0.32	0.64
urban5	-1.18	0.18	0.21	0.44
mar2	0.24	0.18	0.90	1.82
mar3	0.97	0.20	1.80	3.89

(or log odds of the ratio which is the same thing). I do not believe there are major draw backs to presenting in this format. The log odds ratio is easier to understand and the contrasts are already built in to allow this to occur. It is a small paradigm shift that more closely answers the question this survey was designed to study. Also, and while it shouldn't bother me I am still having a hard time building a model without a clear response variable; it's just a strange shift it all.

4 Right Heart Catheterization Data

4.1 Major Findings

Among the major findings is that what the underlying disease is has a strong pull on whether you will not get the surgery or whether you might get the surgery, as will be shown later. Cancer levels decreased the odds of the catherterization being preformed; with a shot in the dark being perhaps the individual wasn't healthy enough to survive the surgery if cancer is already ravaging their body. In better news who provides your insurance doesn't push one way or the other for the surgery, or at least not statistically significantly. Interestingly "Do Not Resuscitate" actually had little effect on the choice of surgery within the first 24 hours of hospitalization but did hold some interactions.

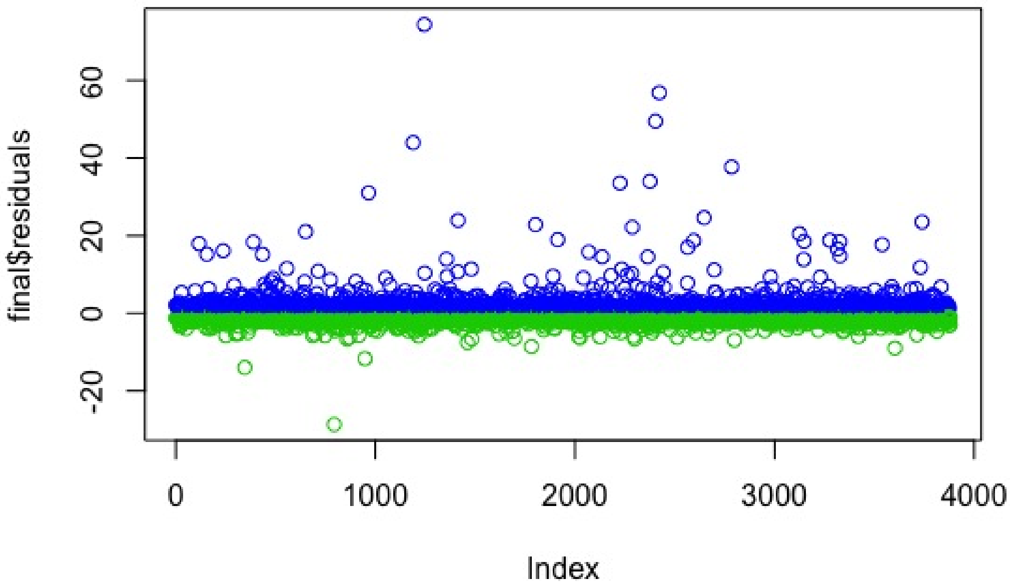
5 Methods

The data had some contaminations that had to be dealt with including values simply plugged in as 0 when not recorded. Either that or there is a striking large amount of people who's weight 0 lbs. While these data points were not deleted (as per instructions) it did make analysis difficult and somewhat untrustworthy. Beyond that for a few of the graphs I will present at the end there was some necessary removal of points simply to see the real underlying structure (or lack thereof) in what does have numbers recorded.

The first two major steps in the analysis was turning the variables that were nominal categories into nominal categories instead of just as numeric of how they came in. The second thing I did which I'm kind of proud of was that I reduced the number of categories for admission diagnosis. Actually, reduced isn't correct, I had a mild transformation on them. I ran a simple logistic regression on all the parameters and invented a new one labeled "admissions" which was the summation of all admission diagnosis's outside of cardio (which given the subject matter I decided to keep separate). The model was then fit. For all categories of admission diagnosis that were not significant at the the .05 level were thrown out/simply left folded into "admissions". It is worth noting that the admissions diagnosis's are not mutually exclusive. The highest level is had four different ones. Other variables that had non-significant main effects were also removed to allow for more reasonable computation time.

With the reduced variable collection then a stepAIC was ran with a limit on looking only at second order interactions and main effects. This took some time but eventually a model was fit that the computer was almost happy. The interesting thing about the model is that it had so many dummy variables and categorical variables that some were starting to become linear combinations of others leading to singularities which you can see on the parameter outputs with the NA's.

Once the model was fit a simple look was made to see how the residuals looked. The following graph is the then colored according to whether you were given the surgery or not given the surgery. Notice the extreme outliers. This led me to try and fit a quasibinomial to see if there was some extra dispersion that could be happening but the overdispersion parameter came out to be $\tilde{1}.04$ which is not a really signficant difference so the quasibinomial was rejected in favor of the regular binomial.



A Shapiro-Wilks Test for Normality was ran on the residuals and the predictable result of them being rejected as normal was found. Partly because we had so many data points and partly because of those major outliers. That being said the residual graph does not actually concern me too much. While it does not have the "traditional" look of a logisitic residual graph with a bulge in the middle and outliers on the ends it still seems to suggest approximately equal variance along the whole way which is good since we have a lot of categorical variables that only exist in the extremes in our model (for example if you were let in with the diagnosis admission "trauma" then you had a larger probability of having the surgery, This allowed for more accurate predictions near the extremes). More so since we have so many explanatory variables there isn't one line that everything is being based off of and most continuous variables are interacting with categorical. So, don't worry too much about the residual plot unfortunately.

After the initial analysis more or less ended some small graphs will be displayed showing strange behavior that I cannot explain.

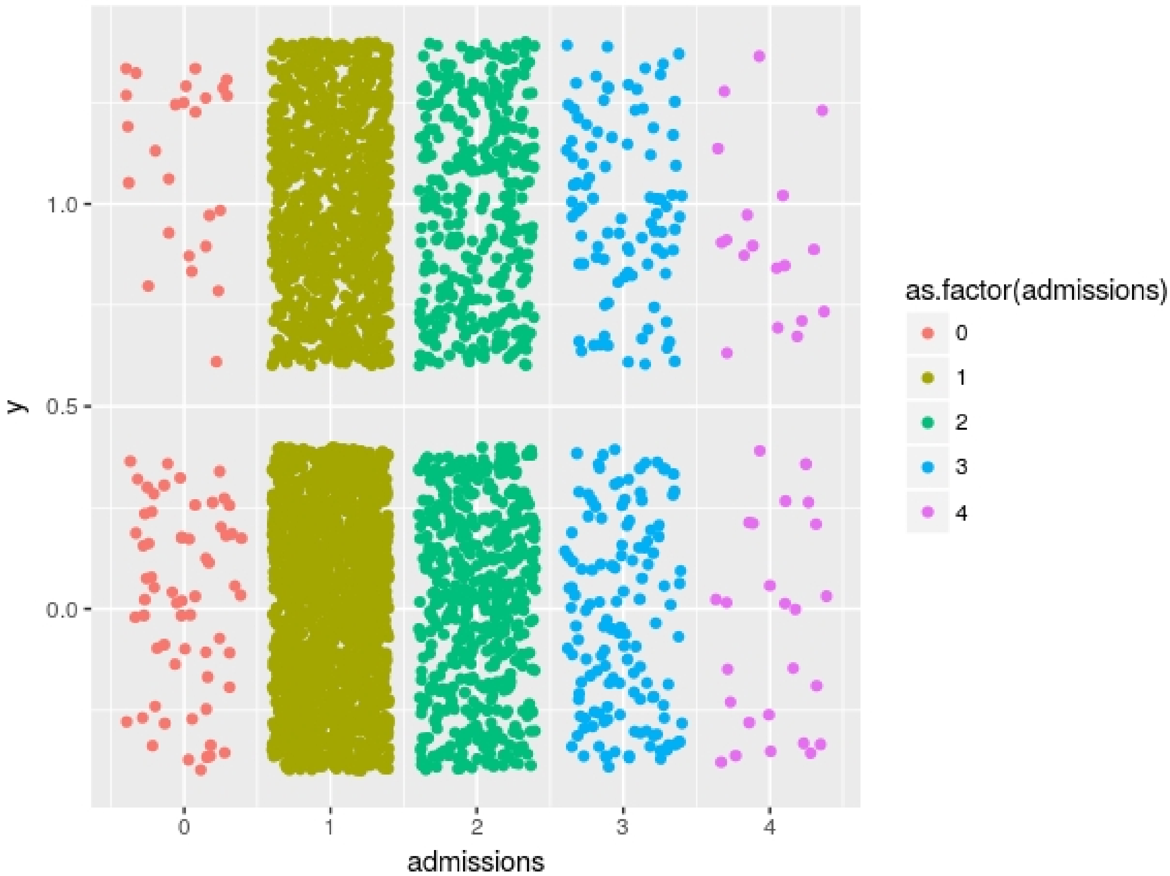
6 Results

The following binomial model was fit to the data (4021.559 as the AIC) and it's a dozy:

```
y ~ edu + insur + disease + dnr + cancer + aps +
    weight + rrate + hrt + pafi + paco2 + pH + hemat + sod +
    pot + card + gastr + seps + trauma + admissions + cancer:pH +
    disease:aps + disease:pafi + disease:pot + hemat:gastr +
    disease:gastr + insur:pH + card:gastr + disease:trauma +
    pafi:trauma + dnr:rrate + edu:hemat + pafi:paco2 + rrate:pot +
    aps:rrate + rrate:hrt + sod:gastr + disease:seps + disease:hemat +
    aps:hemat + hemat:trauma + pot:gastr + dnr:seps + paco2:hemat +
    disease:sod + aps:card + pH:hemat + paco2:admissions + cancer:pot +
    pot:admissions + dnr:weight + dnr:paco2 + hrt:gastr + weight:paco2 +
    paco2:trauma + rrate:trauma + weight:sod + insur:weight +
    aps:seps
```

The parameter estimates for the final model are presented in the last section of the paper. Notice that disease3:trauma has NA for all values (but disease4:trauma and disease2:trauma are both estimable). This is because it is some linear combination of the other parameters. While not statistically very important its a first for me to see something like this. TThis model is not very

easy to read and understand what's happening unfortunately. Notice "admissions" appear on the list kept in the model and in the massive parameter tables you'll find that admissions main effect is none significant but that the interaction of it and "Paco2" and "Pot" are significant at the 10% level. The following is a a jittered scatter plot for the values of admission vs surgery status. It appears
that across the board the no surgery group is just a bit more common with no real changes apparent.



The goodness of fit tests though strongly suggest that our model fits excellently.

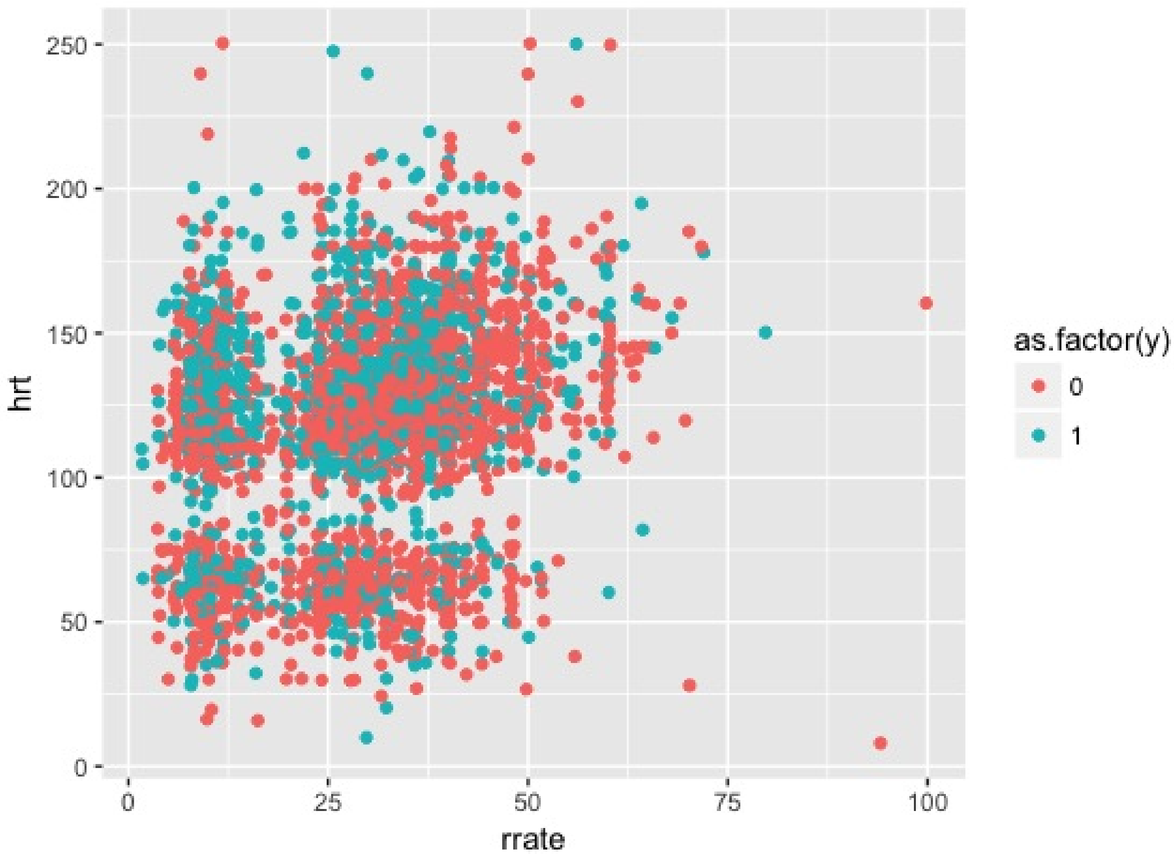
```
Pearson test =  2561.97
 Degrees of freedom =  3789
 p-value =  1

Deviance test =  2077.05
  Degrees of freedom =  3789
  p-value =  1
```

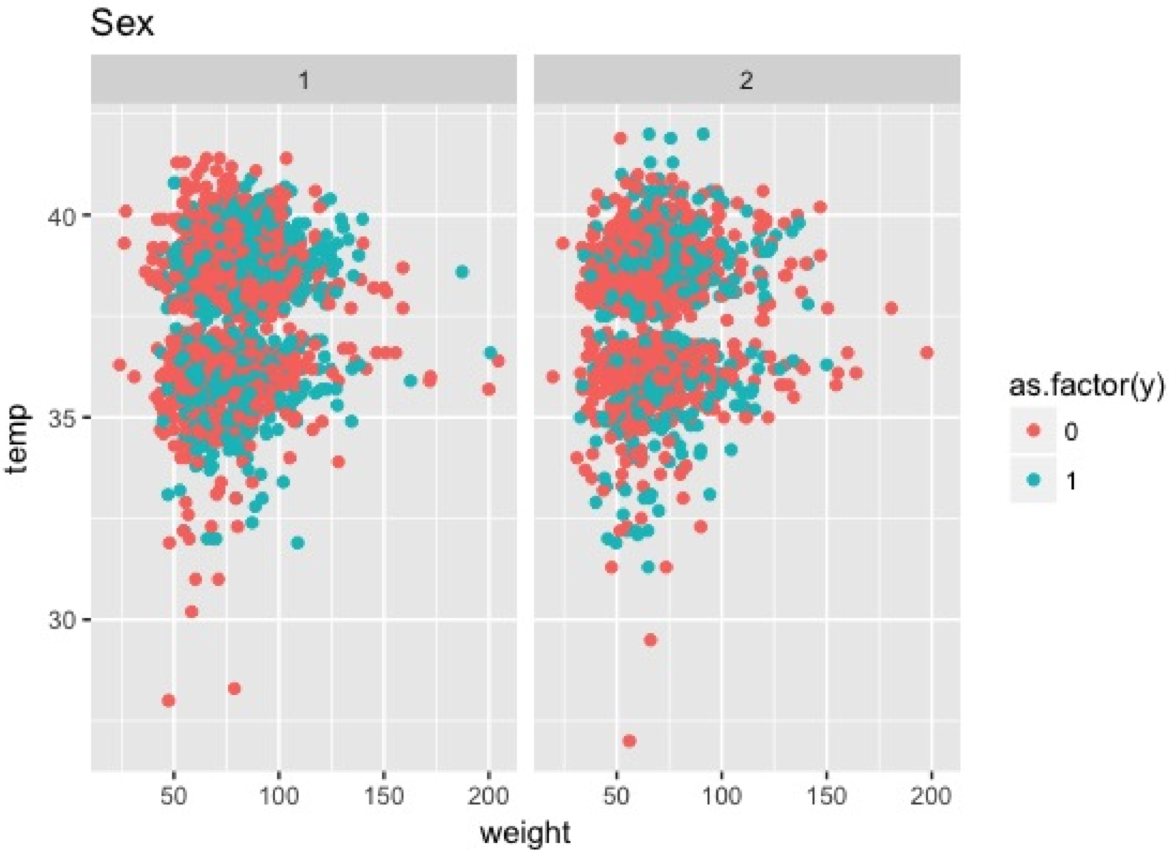
Having p-values that high are actually suspicious that we might have over fit. We are so not different from the full-parameterization that it is statistically strange.

The mean absolute deviation of the true observations against our fitted values came out to be 0.3285063 which is strange given that our model was suppose to be so close to fully parameterized one.

I am going to showcasesome of the strange happenings in this data since there appears to be a lot of missing measurements.



Please note the distinctive breaks in the above data around heart beat rate "hrt" = 80 and respitory rate ("rrate") = 20. I have found no reason for such a strange and distinctive split since both lines appear to go roughly through the middle of the data I would think more people would qualify. Also notice please that whetherthey got the surgery or not doesn't seem to be that indicative of what "quadrant" the patient was in.



Again there is this weird split down the middle for temperature that doesn't seem to effect either sex or the whether the patient had the surgery. An important note is that the above both have had values labeled as 0 in a continuous variable removed only for plotting purposes. Otherwise the structure of the graphs is lost on the elongated scales.

The following table gives the disease vs surgery status for individuals who have Acute Respiratory Failure ("ARF"), MultiOrgan System Failure ("MOSF"), Congestive Heart Failure ("CHF") or some other disease. Probably the most interesting thing off of it is that MOSF actually ends up justifying the surgery more often than the others (ignoring the multitude of other variables) and CHF is approaching 50:50. The remaining diseases seem to not be as critical in getting the surgery.

Table 3: For the Four Types of Disease and Surgery Status

	ARF	MOSF	CHF	Other
No Surgery	1085	525	167	654
Surgery	615	554	133	148

The following is the table for "Do Not Resuscitate" vs surgery status. The major highlights are that people who where in the "No" group for DNR had a much higher probability of getting the surgery than the "Yes" group. At first glance I thought this seemed counterintuitive but it is actually caused by a strange double negative. The "No" group DOES want to be resuscitated while the "Yes" group does not want to be. I spent longer trying to figure out that out than I should have. So then the question becomes what does this imply? I would argue that one of two things are happening. Firstly, the patient is still able to give or not give their consent to the surgery and (having already decided they are prepared for the end over extensive surgeries) opt to not have it preformed. The second possible case is that the patients are unable to give consent and so their stated wishes (DNR status) is basis of the decision. Note that this analysis gets much more complicated in the first scenario. We will effectively end up with a double analysis of when doctors want to give someone a surgery plus when the patient is willing. This would be similar to the assignment with the milk study where we had two populations pooled. In this case the one where the doctors choice is followed and one where the patient overrides. As such, DNR is a very important variable in this analysis as it corrects (somewhat) for this strange behavior.

	No	Yes
0	2106	325
1	1343	107

The last thing I am going to present is the table looking at the surgery vs cancer status. The most interesting this is that people without any form of cancer seem to be getting the surgery more often which is understandable given that usually cancer patients bodies are heavily worn and their immune systems are compromised. Any non-critical surgeries should be put on hold.

	Localized	Metastatic	None
No Surgery	447	175	1809
Surgery	218	85	1147

Ultimately I found that the model fit poorly I think. The mean absolute deviation from the prediction line is quiet high even for a model that is so heavily parameterized. Partly because of the data's strange behavior including the recorded 0's and the strange missing middle parts of the data. It's a working model but it is neither simple to interpret due to all the parameters nor extremely accurate but still the best according to my program and the goodness of fit tests.

Table 4: Right Heart Catheterization Surgery Parameter Estimates

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	85.4070	19.4741	4.39	0.0000
edu	-0.0813	0.0556	-1.46	0.1435
insur2	-16.3663	12.2187	-1.34	0.1805
insur3	-4.3816	12.5926	-0.35	0.7279
insur4	19.7362	13.6233	1.45	0.1475
insur5	-1.0916	16.8476	-0.06	0.9483
insur6	-19.2954	12.1886	-1.58	0.1135
disease2	0.1941	1.8065	0.11	0.9144
disease3	8.5814	3.9288	2.18	0.0290
disease4	-2.6273	2.6448	-0.99	0.3206
dnrYes	-0.6336	0.7386	-0.86	0.3910
cancer2	-27.7272	15.0445	-1.84	0.0654
cancer3	-36.9292	9.1806	-4.02	0.0001
aps	-0.0241	0.0123	-1.96	0.0504
weight	0.0558	0.0297	1.88	0.0604
rrate	0.0007	0.0168	0.04	0.9671
hrt	0.0081	0.0020	4.07	0.0000
pafi	-0.0096	0.0015	-6.60	0.0000
paco2	-0.0827	0.0216	-3.83	0.0001
pH	-11.0170	2.5462	-4.33	0.0000
hemat	-1.0262	0.4416	-2.32	0.0202
sod	0.0032	0.0162	0.20	0.8428
pot	0.3747	0.1749	2.14	0.0323
card	1.5405	0.3073	5.01	0.0000
gastr	-6.3906	2.1430	-2.98	0.0029
seps	0.5178	0.4387	1.18	0.2380
trauma	-153.5766	3440.0948	-0.04	0.9644
admissions	-0.1885	0.2978	-0.63	0.5268
cancer2:pH	3.9673	2.0108	1.97	0.0486
cancer3:pH	5.2200	1.2300	4.24	0.0000
disease2:aps	-0.0048	0.0058	-0.83	0.4092
disease3:aps	0.0133	0.0113	1.18	0.2395
disease4:aps	0.0306	0.0075	4.10	0.0000
disease2:pafi	0.0008	0.0010	0.78	0.4354
disease3:pafi	0.0063	0.0016	3.94	0.0001
disease4:pafi	0.0029	0.0012	2.35	0.0186
disease2:pot	0.1597	0.0955	1.67	0.0946
disease3:pot	-0.5074	0.1889	-2.69	0.0073
disease4:pot	-0.0423	0.1305	-0.32	0.7458
hemat:gastr	0.0656	0.0166	3.94	0.0001
disease2:gastr	-0.0310	0.2520	-0.12	0.9021
disease3:gastr	-1.2664	1.0500	-1.21	0.2278
disease4:gastr	-1.1908	0.3286	-3.62	0.0003
insur2:pH	2.1428	1.6545	1.30	0.1953
insur3:pH	0.6210	1.7047	0.36	0.7157
insur4:pH	-2.7348	1.8481	-1.48	0.1390
insur5:pH	0.0064	2.2825	0.00	0.9978
insur6:pH	2.6011	1.6501	1.58	0.1150
card:gastr	-0.7930	0.3910	-2.03	0.0426
disease2:trauma	168.0036	3216.7055	0.05	0.9583
disease4:trauma	603.8850	14062.2559	0.04	0.9657

Table 5: Right Heart Catheterization Surgery Parameter Estimates Continued

	Estimate	Std. Error	t value	Pr(>|t|)
pafi:trauma	-2.6446	39.5572	-0.07	0.9467
dnrYes:rrate	0.0228	0.0107	2.13	0.0336
edu:hemat	0.0039	0.0017	2.26	0.0241
pafi:paco2	0.0001	0.0000	2.50	0.0126
rrate:pot	-0.0089	0.0031	-2.89	0.0038
aps:rrate	0.0004	0.0002	2.29	0.0219
rrate:hrt	-0.0002	0.0001	-2.42	0.0155
sod:gastr	0.0369	0.0140	2.63	0.0085
disease2:seps	-0.5545	0.2361	-2.35	0.0189
disease3:seps	19.8311	7047.2473	0.00	0.9978
disease4:seps	0.0695	0.4980	0.14	0.8891
disease2:hemat	-0.0019	0.0141	-0.13	0.8937
disease3:hemat	0.0730	0.0232	3.15	0.0016
disease4:hemat	-0.0189	0.0164	-1.15	0.2487
aps:hemat	0.0011	0.0003	3.02	0.0026
hemat:trauma	11.1557	179.4253	0.06	0.9504
pot:gastr	0.2458	0.1123	2.19	0.0286
dnrYes:seps	-0.8144	0.3655	-2.23	0.0259
paco2:hemat	0.0014	0.0005	2.95	0.0032
disease2:sod	0.0002	0.0120	0.01	0.9887
disease3:sod	-0.0794	0.0272	-2.92	0.0036
disease4:sod	0.0068	0.0179	0.38	0.7015
aps:card	-0.0081	0.0052	-1.57	0.1155
pH:hemat	0.1135	0.0573	1.98	0.0478
paco2:admissions	0.0088	0.0049	1.81	0.0701
cancer2:pot	-0.3848	0.2077	-1.85	0.0640
cancer3:pot	-0.2650	0.1150	-2.30	0.0212
pot:admissions	-0.0968	0.0550	-1.76	0.0785
dnrYes:weight	0.0090	0.0056	1.61	0.1080
dnrYes:paco2	-0.0272	0.0158	-1.72	0.0848
hrt:gastr	-0.0050	0.0031	-1.60	0.1093
weight:paco2	-0.0002	0.0001	-1.74	0.0825
paco2:trauma	9.7382	151.9104	0.06	0.9489
rrate:trauma	-3.6071	56.9551	-0.06	0.9495
weight:sod	-0.0003	0.0002	-1.52	0.1278
insur2:weight	0.0066	0.0069	0.96	0.3390
insur3:weight	-0.0033	0.0066	-0.50	0.6162
insur4:weight	-0.0005	0.0075	-0.06	0.9492
insur5:weight	0.0171	0.0089	1.92	0.0549
insur6:weight	0.0027	0.0063	0.43	0.6638
aps:seps	0.0088	0.0064	1.38	0.1670

Table 6: Parameter Estimates for Log-Linear Model in Part 1

	Parameter	Std. Error	Lower CI	Upper CI
(Intercept)	1.51	0.29	2.47	7.80
alc2	-1.13	0.20	0.22	0.48
alc3	-1.97	0.30	0.08	0.25
income2	0.30	0.33	0.72	2.60
income3	-0.42	0.36	0.33	1.34
income4	2.16	0.29	5.04	15.54
urban2	-1.20	0.37	0.14	0.61
urban3	-0.45	0.32	0.34	1.20
urban4	0.71	0.26	1.24	3.43
urban5	0.56	0.27	1.04	2.99
mar2	1.44	0.26	2.56	7.22
mar3	-0.09	0.31	0.50	1.69
alc2:income2	0.61	0.14	1.40	2.44
alc3:income2	0.81	0.25	1.40	3.78
alc2:income3	1.32	0.16	2.75	5.09
alc3:income3	1.62	0.26	3.06	8.62
alc2:income4	0.86	0.14	1.82	3.13
alc3:income4	1.31	0.24	2.35	6.11
alc2:urban2	0.28	0.17	0.94	1.86
alc3:urban2	0.13	0.20	0.76	1.70
alc2:urban3	0.18	0.17	0.86	1.66
alc3:urban3	-0.54	0.22	0.37	0.90
alc2:urban4	-0.19	0.14	0.63	1.09
alc3:urban4	-0.79	0.18	0.32	0.64
alc2:urban5	-0.53	0.14	0.44	0.78
alc3:urban5	-1.18	0.18	0.21	0.44
income2:urban2	0.67	0.35	1.00	3.96
income3:urban2	1.05	0.36	1.43	5.86
income4:urban2	0.86	0.34	1.24	4.67
income2:urban3	-0.14	0.28	0.50	1.49
income3:urban3	-0.50	0.31	0.33	1.11
income4:urban3	-0.30	0.27	0.44	1.25
income2:urban4	0.18	0.24	0.73	1.91
income3:urban4	0.07	0.26	0.64	1.78
income4:urban4	0.09	0.23	0.68	1.71
income2:urban5	-0.24	0.24	0.49	1.25
income3:urban5	-0.62	0.26	0.32	0.89
income4:urban5	-0.36	0.23	0.44	1.08
alc2:mar2	0.42	0.13	1.18	1.98
alc3:mar2	0.24	0.18	0.90	1.82
alc2:mar3	0.79	0.15	1.63	2.97
alc3:mar3	0.97	0.20	1.80	3.89
income2:mar2	0.44	0.28	0.89	2.66
income3:mar2	0.32	0.30	0.76	2.51
income4:mar2	-1.54	0.24	0.13	0.33
income2:mar3	0.33	0.33	0.73	2.64
income3:mar3	0.17	0.36	0.59	2.40
income4:mar3	-0.39	0.28	0.39	1.16
urban2:mar2	0.66	0.21	1.28	2.94
urban3:mar2	1.07	0.24	1.85	4.70
urban4:mar2	0.85	0.18	1.64	3.29
urban5:mar2	1.59	0.20	3.32	7.25
urban2:mar3	-0.26	0.24	0.48	1.25
urban3:mar3	0.56	0.26	1.06	2.93
urban4:mar3	0.20	0.20	0.83	1.79
urban5:mar3	0.85	0.22	1.53	3.59