BAN 5733 Exercise 7 (10 points)

Data Set Description:

A pharmaceutical company has developed a new drug that has shown amazing regenerative capabilities in cartilage regrowth in lab tests. It has also passed phase I clinical trials and proven to not be harmful to human health. Now the company has conducted a phase II clinical trial to determine if patients with arthritis show improvement in pain after treatment. Patients were assigned to either a treatment group of the new arthritis drug or a placebo. Researchers also believe that there may be some interaction with gender because the drug works on certain biologic pathways. The pharmaceutical researchers have asked your analytics company to examine the data to see if you can explain and predict whether someone will have a successful treatment outcome or not. This information will be used to pass FDA approval processes and develop marketing campaigns for the drug.

Your answers will be looked at by the managers at the pharmaceutical company, the researchers and the marketing analysts. So, write your report with a short section in a manager friendly way with a separate detailed technical section for the researchers and marketing analysts. The managers are interested in primarily non-technical explanation of your final model and recommendations based on that model. The researchers and marketing analysts are interested in knowing all the technical details including but not limited to testing of assumptions of your model. The Researchers/Analysts do not want you to transform any variable for this modeling exercise.

Your analysis may be conducted in JMP or SAS (use effect coding for categorical IVs). Your report should be restricted to 7-pages maximum. Please back up your assertions with appropriate statistics and graphs as needed.

Variable Descriptions:

- ID is identification variable
- Improve: A binary dependent variable indicating whether a patient shows marked improvement (1) or no improvement (0) in arthritis pain.
- Gender: 0 = Male, 1 = Female
- Treatment: 0= Placebo, 1 = Active

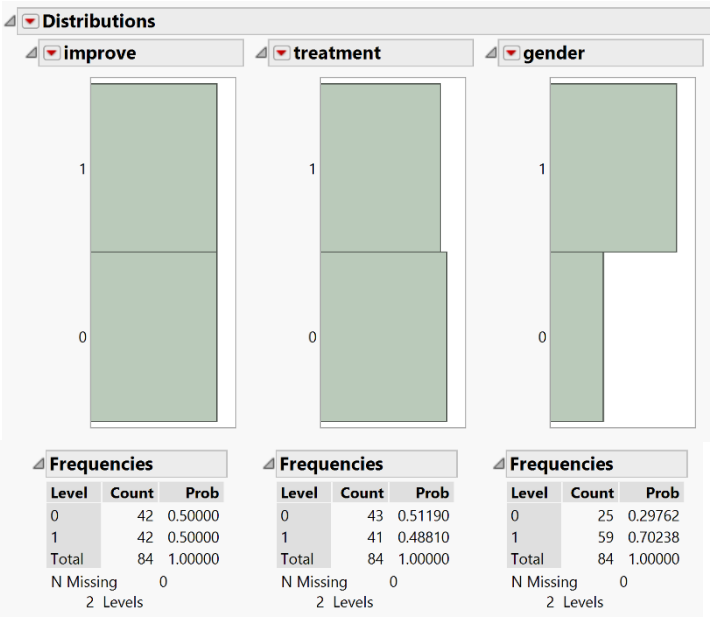
Manager Friendly Report

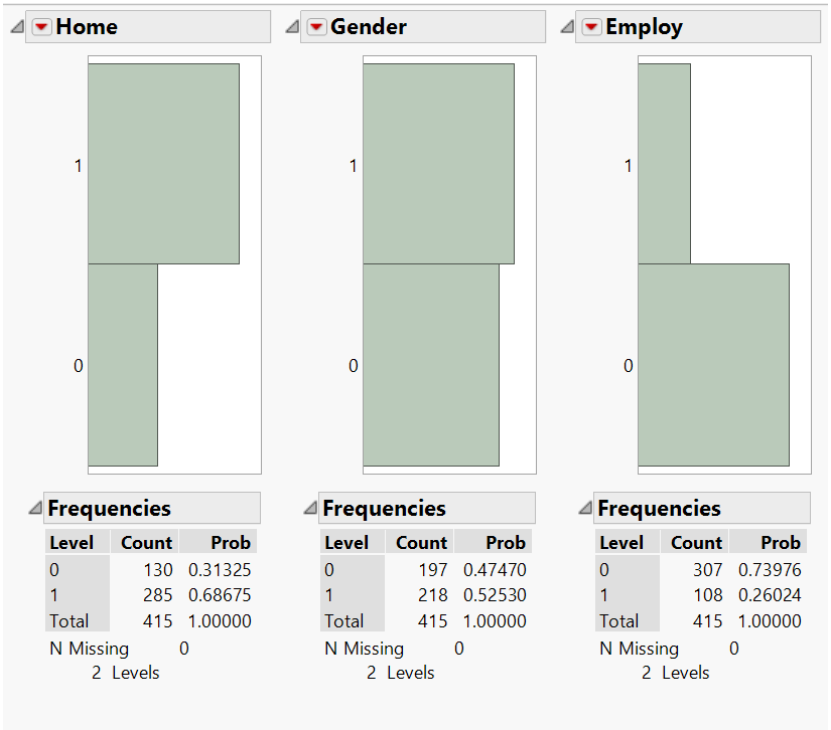
We have explored your data and built several models to explain and predict whether a person is satisfied with the web-based service using all the available variables in the data set. Our overall model was statistically significant, resulted in both (treatment and gender) variables being statistically significant, and had a 32% misclassification rate. The resulting model accurately predicted whether a person's arthritis pain would improve or not 68% of the time. This is better than a 50/50 guess at improvement. Below you will find highlights of the results. For detailed results, please refer to the accompanying technical report that is attached.

- Gender is important for explaining improvement in arthritis pain.
- There was a distinct difference in treatment (68.3%) over placebo (31.7%) group for reported improvement in arthritis pain.
- The odds of arthritis pain improving were 5.9 times greater for those receiving the treatment versus the placebo.
- The odds of a woman experiencing improvement in pain were 4.3 times higher than for males.
- There was not an interaction between gender and treatment indicating the treatment effect does not alter directions when gender is taken into account.

Technical Report

A quick look at the summary statistics indicates all three variables are nominal in nature. Further, the categorical values do not seem to have issues with low cell count or extremely unbalanced data. None of the variables have missing values. There are no major indications for transforming variables at this time but we did need to change modeling types for gender and treatment as they were listed as continuous.





A series of variable selection models using stepwise (forward, backward, mixed) regression using 5% level of significance identified the following variables as statistically significant. The same set of variables were identified in the mixed and backward

Method	Variables identified by Method	Generalized R-Sqr	Misclassification	Comments
Forward	Gender, Treatment	15.9%	32.1%	Interaction between Gender and Treatment non-sig.
Backward	Gender, Treatment	15.9%	32.1%	Interaction between Gender and Treatment non-sig.
Stepwise	Gender, Treatment	15.9%	32.1%	Interaction between Gender and Treatment non-sig.

Running a nominal logistic regression model with improve (1) being the event of interest using the 2 variables found significant above, we find the overall model to be statistically significant and explained about 15.9% variance in Improve. The model used all 84 observations.

Nominal Logistic Fit for improve

Effect Summary

Source	LogWorth	PValue
treatment	3.631	0.00023
gender	2.153	0.00703

Remove Add Edit FDR

Converged in Gradient, 4 iterations

Iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	9.113590	2	18.22718	0.0001*
Full	49.110774			
Reduced	58.224363			

RSquare (U)	0.1565
AICc	104.522
BIC	111.514
Observations (or Sum Wgts)	84

Fit Details

Measure	Training	Definition
Entropy RSquare	0.1565	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.2601	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.5847	∑ -Log(ρ[j])/n
RMSE	0.4477	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.4003	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.3214	∑ (ρ[j]≠ρMax)/n
N	84	n

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	1	0.138782	0.277564
Saturated	3	48.971992	Prob>ChiSq
Fitted	2	49.110774	0.5983

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-0.278544	0.2726509	1.04	0.3070
treatment[Treatment]	0.89084011	0.2593814	11.80	0.0006*
gender[Male]	-0.7343271	0.2878287	6.51	0.0107*

For log odds of 1/0

The estimates for coefficients show both independent variables are statistically significant. Being female has a positive effect on Improvement independent of the treatment while receiving the treatment also shows a positive effect on improvement.

Examining the odds ratios, we can see the degree to which the variables are associated with IMPROVE.

Odds Ratios

For improve odds of 1 versus 0

Odds Ratios for treatment

Level1	/Level2	Odds Ratio	Prob>Chisq	Lower 95%	Upper 95%
Placebo	Treatment	0.168355	0.0006*	0.0609052	0.465369
Treatment	Placebo	5.9398282	0.0006*	2.148831	16.41895

Odds Ratios for gender

Level1	/Level2	Odds Ratio	Prob>Chisq	Lower 95%	Upper 95%
Female	Male	4.3433856	0.0107*	1.4054906	13.422358
Male	Female	0.2302351	0.0107*	0.0745026	0.7114953

Because GENDER and TREATMENT are categorical variables, effect coding was used to analyze these variables in relation to IMPROVE. We can see from the odds ratio table that going from placebo to treatment, the odds of improving the arthritis pain were 5.9 times higher for the treatment group over the placebo.

With the 2 variable model, we can see…

1. that we correctly identified 67.9% of the records ((21+36)/84);

2. there were 32% misclassified;
3. of those who actually improved, 50% were correctly classified (sensitivity); and
4. 85.7% of the customers who did not improve were correctly classified (specificity).

That indicates that our model is better at explaining those people who did not improve but still does better than a random guess for those who have arthritis pain.

Most Likely improve

improve	Count Total % Col % Row %	0	1	Total
	0	36 42.86 63.16 85.71	6 7.14 22.22 14.29	42 50.00
	1	21 25.00 36.84 50.00	21 25.00 77.78 50.00	42 50.00
	Total	57 67.86	27 32.14	84

.

Assumptions for the use of the nominal logistic regression were examined.

1. The dependent variable was measured on a 0/1 scale – not improved/ improved as seen in the descriptive statistics at the beginning of this report.
2. Observations were independent from each other as the clinical trial was conducted among different participants and persons did not repeat the survey according to information given.
3. The two independent variables were not associated according to chi-square analysis.

While this model has room for improvement by adding other variables, it does predict improvement better than a random guess. The pharmaceutical company should continue research into this new drug but it does seem very promising in the treatment of arthritis.

SAS code:

```
proc glm data = work.arthrit;
  class gender treatment;
  model improve = gender treatment;
run;
```