BAS 320 - Assignment 8 - Categorical Predictors

Jack Guptill

## Analysis of Customer Tenure by Monthly Charges and the Option of Paperless Billing

I’m using a multiple regression model to predict how long someone might be the customer for a phone company in regards to their phone plan. For this analysis I have built two different models that take in different predictors. For the first model I am utilizing the amount of money a customer spends on their plan and whether or not they utilize paperless billing. For the second model I am diving further into how the exact payment method used interacts with the tenure and the monthly amount that a given customer spends.

I made these models because I was genuinely curious to how long a customer stays with a company by the way in which they pay for the service. As we all know, the frequency to which we look at our bills tends to change how long we tend to stick with a service or product for that matter. I was wondering how drastic this was with cell phone plans.

The data I’m using comes from the BAS 320 library and it contains a total of 7032 rows with 20 columns. There are a total of 3 quantitative and 17 categorical columns.

## Task 1 - Investigation of the Relationship Between Tenure, Monthly Charges and the Option of Paperless Billing

For the model that I built both the indicator variable with and without the interaction is statistically significant. Due to it being statistically significant with the the interaction the full regression equation is as follows:

Customer Tenure = 22.14035 + .195\*MonthlyCharges - 8.45PaperlessBilling + .061\*MonthlyCharges\*PaperlessBilling

Implicit Equation for Customer Tenure that don’t include Paperless Billing (No Level):

Customer Tenure = 22.14035 + .195\*MonthlyCharges

Implicit Equation for Cellphone Tenure that do include Paperless Billing (Yes Level):

Customer Tenure = 13.690 + .256\*MonthlyCharges

Customers that have paperless billing have a stronger relationship between their tenure and the monthly charges than does the relationship for those that do not have paperless billing. We can infer that this might be due to an underlying increase in overall tolerance because of the feature of being paperless. The difference in the average value of the monthly payment shrinks between having paperless billing and not having paperless billing.

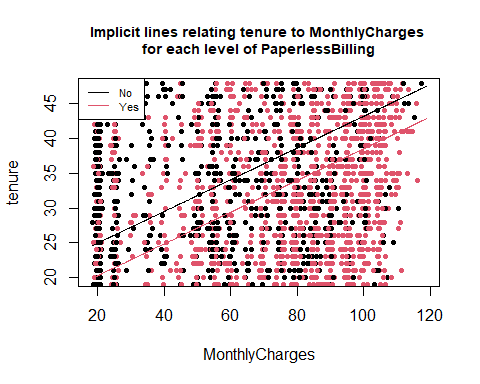
M1 = lm(tenure~MonthlyCharges+PaperlessBilling,data=TELCO) #Fit a regression predicting y from x1 (numeric) and x2 (categorical) that EXCLUDES the interaction  
  
summary(M1)

##   
## Call:  
## lm(formula = tenure ~ MonthlyCharges + PaperlessBilling, data = TELCO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -42.05 -21.45 -3.38 21.57 51.77   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 20.40203 0.68443 29.809 < 2e-16 \*\*\*  
## MonthlyCharges 0.22829 0.01003 22.754 < 2e-16 \*\*\*  
## PaperlessBillingYes -4.67837 0.61432 -7.616 2.97e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 23.69 on 7029 degrees of freedom  
## Multiple R-squared: 0.06863, Adjusted R-squared: 0.06836   
## F-statistic: 259 on 2 and 7029 DF, p-value: < 2.2e-16

drop1(M1,test="F")

## Single term deletions  
##   
## Model:  
## tenure ~ MonthlyCharges + PaperlessBilling  
## Df Sum of Sq RSS AIC F value Pr(>F)   
## <none> 3945270 44517   
## MonthlyCharges 1 290597 4235866 45015 517.735 < 2.2e-16 \*\*\*  
## PaperlessBilling 1 32553 3977822 44573 57.997 2.969e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

visualize\_model(M1)

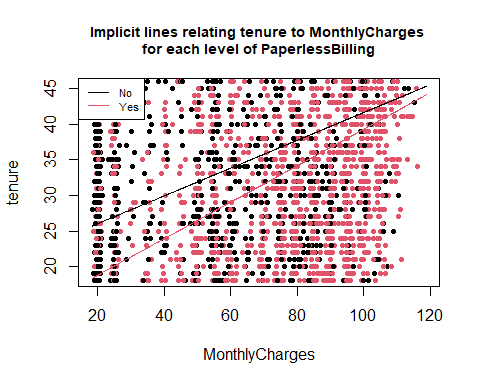


##   
## Effect test for PaperlessBilling has p-value 2.969e-14

#summary(TELCO$tenure) #32.42 as the mean of Tenure  
  
M2 = lm(tenure~MonthlyCharges\*PaperlessBilling,data=TELCO) #Fit a regression predicting y from x1 (numeric) and x2 (categorical) that INCLUDES the interaction  
  
summary(M2)

##   
## Call:  
## lm(formula = tenure ~ MonthlyCharges \* PaperlessBilling, data = TELCO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -40.177 -21.657 -3.406 21.566 53.257   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 22.14035 0.89242 24.809 < 2e-16 \*\*\*  
## MonthlyCharges 0.19488 0.01490 13.082 < 2e-16 \*\*\*  
## PaperlessBillingYes -8.45318 1.38782 -6.091 1.18e-09 \*\*\*  
## MonthlyCharges:PaperlessBillingYes 0.06109 0.02014 3.033 0.00243 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 23.68 on 7028 degrees of freedom  
## Multiple R-squared: 0.06984, Adjusted R-squared: 0.06945   
## F-statistic: 175.9 on 3 and 7028 DF, p-value: < 2.2e-16

visualize\_model(M2)



##   
## Effect test for interaction with PaperlessBilling has p-value 0.002431

## Task 2 - Investigation of the relationship between Customer Tenure by Monthly Charges and Payment Type

(You can use the same numeric predictor X1 from Task 1 if you’d like or a different one. Your categorical predictor X4 must have between 3 and 12 levels)

(Fit two models, one without an interaction between the two predictors and one with an interaction. Produce a summary and drop1 for both, and include visualize\_model for both.)

(A requirement for this part is that at least one of the following must be true: 1) the categorical predictor is statistically significant in the model without an interaction or 2) the categorical predictor is statistically significant in the model with an interaction. Check drop1 to determine if this is true or not. What you’ll write depends on whether the interaction is or is not statistically significant.)

For both models, including and not including the interaction of the payment type, the variable is statistically significant.

For the model that includes the interaction, the overall equation is as follows:

Customer Tenure = 33.400 + .153\*MonthlyCharges - 3.598\*PaymentMethodCreditCard - 30.636\*PaymentMethodElectronicCheck - 17.206\*PaymentMethodMailedCheck + .049\*MonthlyCharges\*PaymentMethodCreditCard + .140\*MonthlyCharges\*PaymentMethodElectronicCheck - .023\*MonthlyCharges\*PaymentMethodMailedCheck

The equation for the reference level:

Customer Tenure = 33.400 + .153\*MonthlyCharges

The implicit equation for Electronic Check:

Customer Tenure = 2.764 + .293\*MonthlyCharges

The implicit equation for Mailed Check:

Customer Tenure = 16.194 + .130\*MonthlyCharges

The implicit equation for Credit Card:

Customer Tenure = 29.802 + .202\*MonthlyCharges

For the sake of this Analysis I would like to split the four levels of payment types into group A and group B. Group A consists of those that pay by bank transfer and credit cards Automatically. Group B consists of mailed check and electronic check. In general those that are in Group A tend to stay with their phone plan for longer than those in group B. This might be due to the convenience factor of their payment method. If we inspect group A more closely, it seems that at low levels of monthly payment, those that use bank transfers tend to stay with their phone plan longer. However, at large monthly rates, this flips where those that utilize automatic credit credit card payments tend to have a longer tenure. If we look at group B we see a similar relationship with e-checks and mailed in checks. At low levels of monthly charges, those that utilize mailed in checks, tend to have a longer tenure. However, when a customer has a large monthly payment, those that electronic checks tend to have a longer tenure.

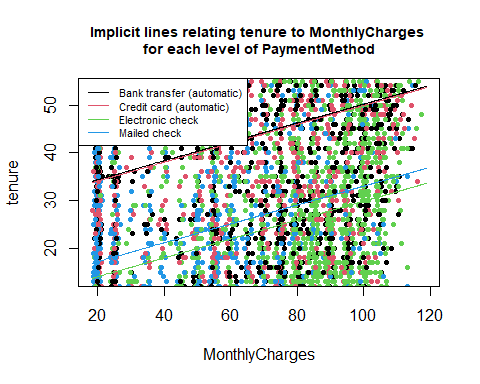
From this we can assume that convenience of method of payment has a drastic change in how long the customer tends to stay on their phone plan. Where automatic type payment, such as Group A, in general tend to stay with their phone plan longer than manual methods such as Group B. We can confirm this, by looking at the various strengths of relationships. Automatic credit card payments have a stronger relationship then does automatic bank transfers. Electronic Check has a much stronger relationship than does the more inconvenient method of the mail in check.

What is interesting from the output is that the difference in average tenure tends to shrink in larger amounts of monthly payment between the electronic check and the bank transfer method. For further analysis I would like to explore how often they check their banking statements and what type of visibility they might have over their banking accounts. Do they get their statements in the mail once a month? Can they go on an app at anypoint to see? This might also have drastic effects on how they view their cell phone plan as well.

M3 <- lm(tenure~MonthlyCharges+PaymentMethod, data=TELCO) #Fit a regression predicting y from x3 (numeric) and x4 (categorical) that EXCLUDES the interaction  
drop1(M3,test="F")

## Single term deletions  
##   
## Model:  
## tenure ~ MonthlyCharges + PaymentMethod  
## Df Sum of Sq RSS AIC F value Pr(>F)   
## <none> 3349759 43371   
## MonthlyCharges 1 209207 3558966 43795 438.87 < 2.2e-16 \*\*\*  
## PaymentMethod 3 628063 3977822 44573 439.18 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

visualize\_model(M3)

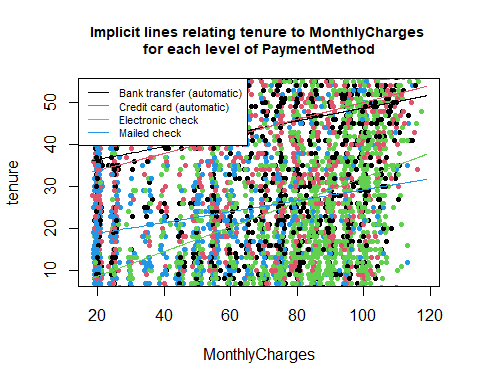


##   
## Effect test for PaymentMethod has p-value 1.607e-261

M4 <- lm(tenure~MonthlyCharges\*PaymentMethod, data=TELCO) #Fit a regression predicting y from x3 (numeric) and x4 (categorical) that INCLUDES the interaction  
drop1(M4,test="F")

## Single term deletions  
##   
## Model:  
## tenure ~ MonthlyCharges \* PaymentMethod  
## Df Sum of Sq RSS AIC F value Pr(>F)   
## <none> 3329263 43333   
## MonthlyCharges:PaymentMethod 3 20496 3349759 43371 14.414 2.329e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

visualize\_model(M4)



##   
## Effect test for interaction with PaymentMethod has p-value 2.329e-09

summary(M4)

##   
## Call:  
## lm(formula = tenure ~ MonthlyCharges \* PaymentMethod, data = TELCO)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -49.14 -17.82 -2.54 18.52 63.26   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) 33.39952 1.33929  
## MonthlyCharges 0.15347 0.01814  
## PaymentMethodCredit card (automatic) -3.59823 1.89126  
## PaymentMethodElectronic check -30.63596 2.00231  
## PaymentMethodMailed check -17.20621 1.70707  
## MonthlyCharges:PaymentMethodCredit card (automatic) 0.04943 0.02572  
## MonthlyCharges:PaymentMethodElectronic check 0.14043 0.02599  
## MonthlyCharges:PaymentMethodMailed check -0.02276 0.02749  
## t value Pr(>|t|)   
## (Intercept) 24.938 < 2e-16 \*\*\*  
## MonthlyCharges 8.460 < 2e-16 \*\*\*  
## PaymentMethodCredit card (automatic) -1.903 0.0571 .   
## PaymentMethodElectronic check -15.300 < 2e-16 \*\*\*  
## PaymentMethodMailed check -10.079 < 2e-16 \*\*\*  
## MonthlyCharges:PaymentMethodCredit card (automatic) 1.922 0.0547 .   
## MonthlyCharges:PaymentMethodElectronic check 5.403 6.79e-08 \*\*\*  
## MonthlyCharges:PaymentMethodMailed check -0.828 0.4078   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 21.77 on 7024 degrees of freedom  
## Multiple R-squared: 0.214, Adjusted R-squared: 0.2133   
## F-statistic: 273.3 on 7 and 7024 DF, p-value: < 2.2e-16