**SAS HOMEWORK 6 (Group Homework)**

***1. Run the following Tobit model (Use PROC QLIM)***

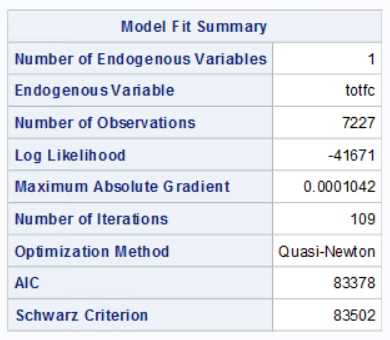
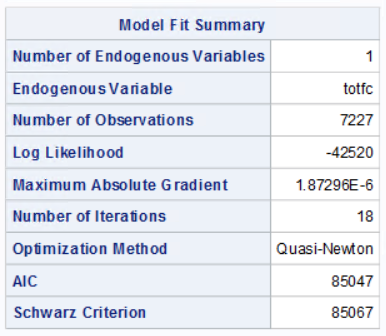
***Model profit = age, totaltrans, rewards, limit, numcard, modes of acquisition, type of card, types of affinity***

***Write a summary of the results. Focus on important effects, interpretation, model fit etc.***

*Assumption: - Duration is 0 for inactive customers (Duration was blank for inactive customers & removing them was leading to only one value of active variable)*



* One-year increase in **age** of the customer is associated with approx. 23 units decrease in Total finance charges paid by the customer over a 3-year period. And this impact is statistically significant.
* 1000 units increase in **Total transaction** amount spent by a customer in 3-year period is associated with approx. 34 units increase in Total finance charges paid by the customer over a 3-year period. And this impact is statistically significant.
* Increase in **number of cards** owned by a customer doesn’t have significant impact on Total finance charges paid by the customer over a 3-year period.
* Total finance charges paid by the customer over a 3-year period are 233 units more if the customer’s mode of acquisition is ‘**Direct Mail**’ as compared to the customer acquired through ‘Internet’. And this impact is statistically significant.
* Total finance charges paid by the customer over a 3-year period are 736 units less if the customer’s mode of acquisition is ‘**Direct Selling**’ as compared to the customer acquired through ‘Internet’. And this impact is statistically significant.
* Total finance charges paid by the customer over a 3-year period are 759 units less if the customer’s mode of acquisition is ‘**Telephone Selling**’ as compared to the customer acquired through ‘Internet’. And this impact is statistically significant.
* There is no significant difference in Total finance charges paid by a customer over a 3-year period between customers with **Standard card & Gold card.**
* Total finance charges paid by the customer over a 3-year period are 751 units more if the customer has a ‘**Platinum card’** as compared to the customer who has a ‘Standard card’. And this impact is statistically significant.
* Total finance charges paid by the customer over a 3-year period are 558 units less if the customer has a ‘**Quantum card**’ as compared to the customer who has a ‘Standard card’. And this impact is statistically significant.
* Total finance charges paid by the customer over a 3-year period are approx. 344 units less if the customer has a **Rewards card** as compared to the customer who doesn’t have a Rewards card. And this impact is statistically significant.
* 1000 units increase in **Credit limit** of customer is associated with approx. 14 units decrease in Total finance charges paid by the customer over a 3-year period. And this impact is statistically significant.
* Total finance charges paid by the customer over a 3-year period are 394 units less for the ones with **Affinity cards affiliated with Sports institutions** as compared to the customers who have non-affinity cards. And this impact is statistically significant.
* Total finance charges paid by the customer over a 3-year period are 234 units less for the ones with **Affinity cards affiliated with Financial institutions** as compared to the customers who have non-affinity cards. And this impact is statistically significant.
* Total finance charges paid by the customer over a 3-year period are 447 units less for the ones with **Affinity cards affiliated with University** as compared to the customers who have non-affinity cards. And this impact is statistically significant.
* Total finance charges paid by the customer over a 3-year period are 631 units less for the ones with **Affinity cards affiliated with Commercial institutions** like Macy’s as compared to the customers who have non-affinity cards. And this impact is statistically significant.



AIC for the model with only one explanatory variable limit is 85047 which significantly reduces to 83378 after adding all the relevant explanatory variables.

***2. Run a selection model (Use PROC QLIM)***

***Model active = age, rewards, limit, numcard, modes of acquition, type of card, types of affinity***

***Model profit = age, totaltrans, rewards, limit, numcard, modes of acquition, type of card, types of affinity***

***Write a summary of the results. Focus on important effects, interpretation, model fit etc.***



Correlation estimate (Rho) is large & significant. This indicates that selection bias is a problem in the estimation of totfc equation.

* One-year increase in **age** of the customer is associated with approx. $27.6 decrease in Total finance charges paid by the customer over a 3-year period.
* $100000 increase in **Total transaction** amount spent by a customer in 3-year period is associated with approx. $39 increase in Total finance charges paid by the customer over a 3-year period.
* Total finance charges paid by the customer over a 3-year period are approx. $403 less if the customer has a **Rewards card** as compared to the customer who doesn’t have a Rewards card.
* $100000 increase in **Credit limit** of customer is associated with approx. $226 increase in Total finance charges paid by the customer over a 3-year period.
* 1 unit increase in **number of cards** owned by a customer is associated with $ 179 increase in Total finance charges paid by the customer over a 3-year period. And this impact is statistically significant.
* Total finance charges paid by the customer over a 3-year period are $295 more if the customer’s mode of acquisition is ‘**Direct Mail**’ as compared to the customer acquired through ‘Internet’.
* Total finance charges paid by the customer over a 3-year period are $916 less if the customer’s mode of acquisition is ‘**Direct Selling**’ as compared to the customer acquired through ‘Internet’.
* Total finance charges paid by the customer over a 3-year period are $942 less if the customer’s mode of acquisition is ‘**Telephone Selling**’ as compared to the customer acquired through ‘Internet’.
* There is no significant difference in Total finance charges paid by a customer over a 3-year period between customers with **Standard card & Gold card.**
* Total finance charges paid by the customer over a 3-year period are $726 more if the customer has a ‘**Platinum card’** as compared to the customer who has a ‘Standard card’.
* Total finance charges paid by the customer over a 3-year period are $793 less if the customer has a ‘**Quantum card**’ as compared to the customer who has a ‘Standard card’.
* Total finance charges paid by the customer over a 3-year period less for the ones with **Affinity cards** as compared to the customers who have non-affinity cards.

$444: Sector C

$274: Sector D

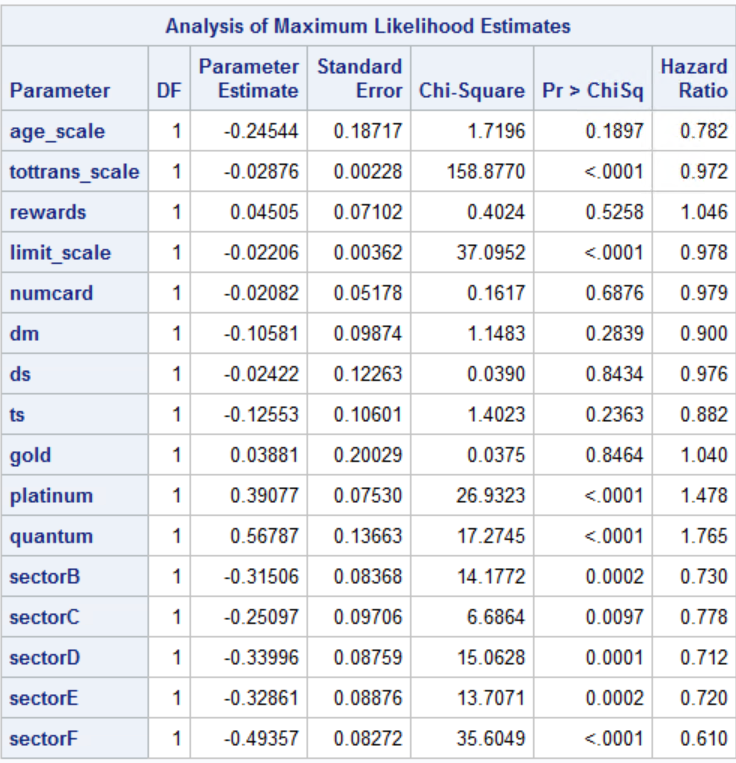
$518: Sector E

$503: Sector F

* The probability of a customer being active decreases with increase in **customer’s age**.
* The probability of a customer being active is less for **Rewards** card customer as compared to non-reward card customer.
* The probability of a customer being active increases with increase in **customer’s credit limit**.
* The probability of a customer being active increases with increase in **number of cards customer has from the firm**.
* The probability of a customer being active is less for customers acquired through **Direct Selling & Telephone selling** as compared to customers acquired through Internet.
* The probability of a customer being active is more for customers having **Platinum card** as compared to customers having standard card.
* The probability of a customer being active is less for customers having **Quantum card** as compared to customers having standard card.
* The probability of customer being active is less for **Affinity card** (Sector C & Sector E) holders as compared to non-affinity card holders.

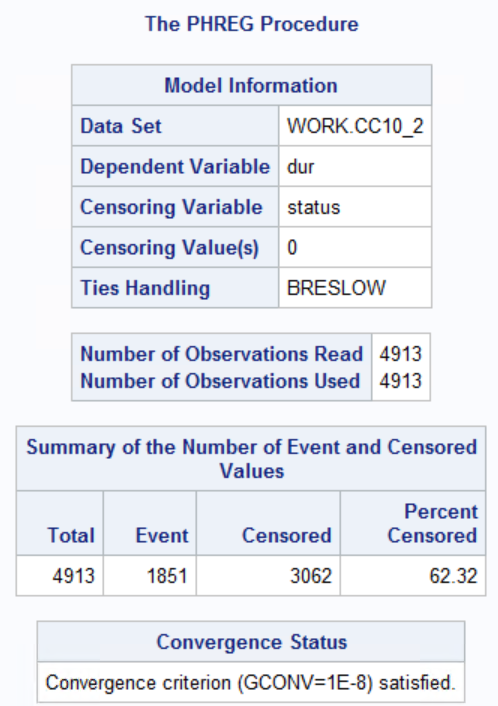
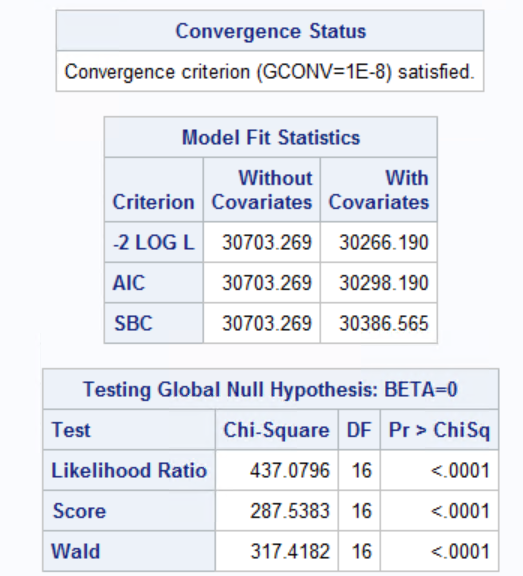
***Q3: Survival analysis***

1. ***Delete all customers who are inactive.***
2. ***Run a proportional hazards model (PROC PHREG)***



*Censoring variable is Status, if status = 0 that means duration is censored otherwise it is not. There are no intercepts in the model as it is a characteristic of partial likelihood.*

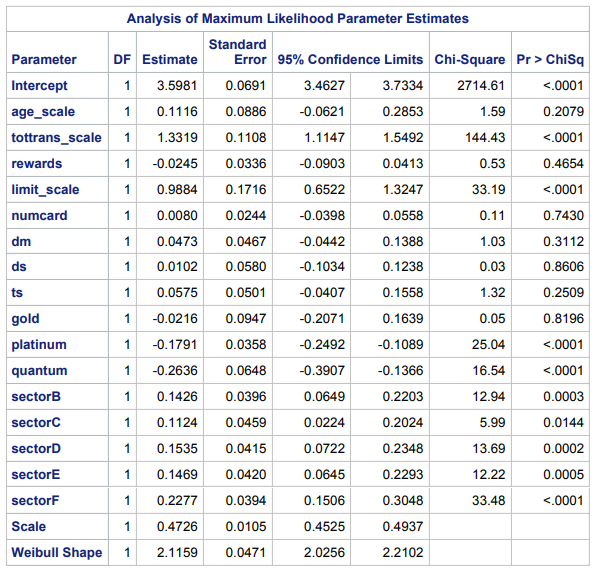
* Every $1000 increase in total transaction by the customer over 3-year period decreases the hazard of leaving the firm by 2.8%.
* Every $1000 increase in credit limit of the customer decreases the hazard of leaving the firm by 2.2%.
* The hazard of a platinum card customer leaving the firm is 47.8% more than the standard card customer.
* The hazard of a quantum card customer leaving the firm is 76.5 % more than the standard card customer.
* The hazard of customers with cards affiliated to professional organisations is 27% less than non-affinity card holders.
* The hazard of customers with cards affiliated to sports organisations is 22.2% less than non-affinity card holders.
* The hazard of customers with cards affiliated to Financial organisations is 28.8% less than non-affinity card holders.
* The hazard of customers with cards affiliated to University organisations is 28% less than non-affinity card holders.
* The hazard of customers with cards affiliated to commercial organisations is 39% less than non-affinity card holders.



Based on all the above three criteria, we can see that the values of the model with intercept and covariates is less than the intercept only criterion. Therefore, the model which we used is a better fit than the intercept only model.

1. ***Run three models using PROC LIFEREG with exponential, Lognormal, and Weibull distributions. Write a summary of the results. Focus on important effects, interpretation, model fit etc.***

We have run the PROC LIFEREG function to predict the time a person has spent with the credit card company. The data is right censored. To test which distribution to use, we have run the model for Exponential, Lognormal and Weibull distributions. The results of Weibull distribution are given below:



From the results displayed above we can conclude with 95% confidence that that following columns are significant in our model:

1. Tottrans\_scale: For every $100000 increase in total transactions (as transactions is has been scaled by dividing it by 100000) there will be a 133.19% increase in the duration a person is customer for the credit card company
2. Limit\_scale: For every $100000 increase in credit card limit (as limit is has been scaled by dividing it by 100000) there will be a 98.84% increase in the duration a person is customer for the credit card company
3. Gold: People with Gold cards are 2.16% less likely to remain as customers as compare to standard cards
4. Platinum: This means that people with Platinum Credit cards are 17.91% less likely to remain as customers as people with standard cards
5. Quantum: This means that people with Quantum Credit cards are 26.36% less likely to remain as customers as people with standard credit cards
6. People with Sector B affiliated cards are 14.26% more likely to remain as customers as people without any affiliations
7. People with Sector C affiliated cards are 11.24% more likely to remain as customers as people without any affiliations
8. People with Sector D affiliated cards are 15.35% more likely to remain as customers as people without any affiliations
9. People with Sector E affiliated cards are 14.69 % more likely to remain as customers as people without any affiliations
10. People with Sector F affiliated cards are 22.77% more likely to remain as customers as people without any affiliations
11. ***Run a regression model with duration as the dependent variable and the same variables above as independent variables. Comment on how the regression results are different from those obtained in survival analysis models****.*

After comparing both the significant beta coefficients from LIFEREG (distribution = Weibull) and Linear Regression model, following shows the results of the coefficients.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Weibull | Exponentials | Lognormal | OLS | % Biased (OLS wrt Weibull) | % Biased (OLS wrt Exponentials) | % Biased (OLS wrt Lognormal |
| tottrans\_scale | 1.33 | 3.24 | 2.03 | 6.55 | 20% | 49% | 31% |
| limit\_scale | 0.98 | 3.17 | 0.83 | 4.39 | 22% | 72% | 19% |
| platinium | -0.18 | -0.48 | -0.24 | -1.38 | 13% | 35% | 17% |
| quantum | -0.26 | -0.76 | -0.45 | -3.35 | 8% | 23% | 13% |
| sectorB | 0.14 | 0.41 | 0.16 | 0.99 | 14% | 41% | 16% |
| sectorD | 0.15 | 0.41 | 0.2 | 1.33 | 11% | 31% | 15% |
| secotrE | 0.15 | 0.4 | 0.2 | 1.46 | 10% | 27% | 14% |
| sectoreF | 0.22 | 0.6 | 0.34 | 1.88 | 12% | 32% | 18% |

We can’t compare different distributions through AIC score however we can compare distribution results with OLS to verify there is a significant bias.

There is a significant bias shown by linear regression model as it increases the coefficients value

