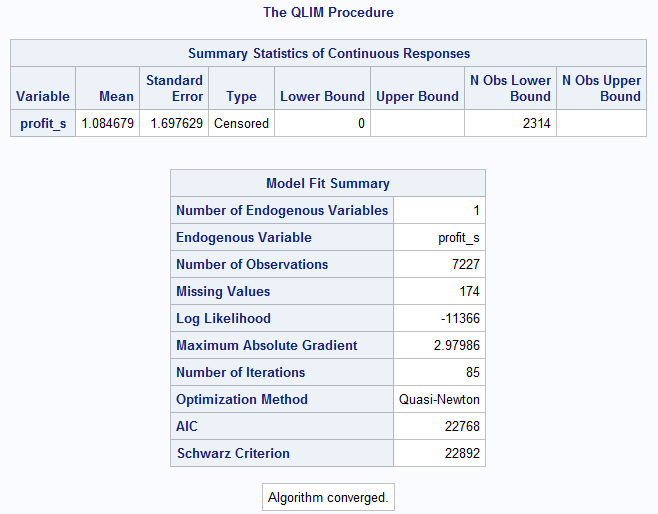
We created the profit and the active variables. We standardized the limit, tottrans and profit variables by dividing them by 1000, and the age variable is standardized by dividing it by 100.

**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

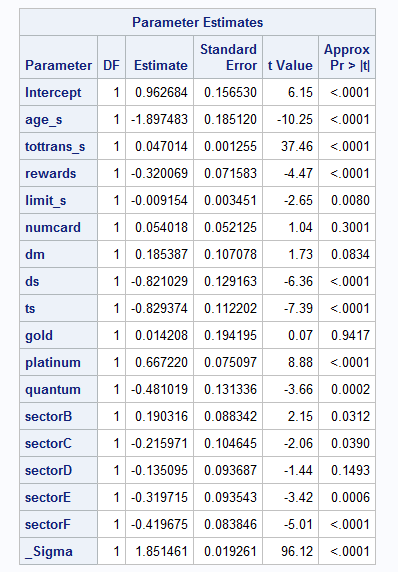
Write a summary of the results. Focus on important effects, interpretation, model fit etc.



**Inference:**

The above tables provide a summary of the number of censored values at Lower bound (where lower bound= 0 for the profit variable) 2314.

The Model Fit Summary table includes information on the number of observations which is 7227 out of which it has 174 missing values, the log likelihood for the model is -11366 which can be used to compare nested models, and the AIC and Schwarz Criterion (also known as the BIC) for the model are 22768 and 22892 respectively, which can be used to compare nested and non-nested models. The selected mode has only one Endogenous variable



**Inference:**

The above table displays the coefficients, their standard errors, the t-statistics, and associated p-values of the variables. The coefficients for age\_s (i.e. Age), tottrans\_s (i.e. total transactions), rewards are statistically significant at alpha = 5%, whereas numcard (i.e. the number of cards) is not even statistically significant at alpha = 10%.

The modes of acquisition dm (Direct mail) is statistically significant at alpha = 10%, ds (Direct Selling) and ts (Telephone Selling) are statistically significant at alpha = 5% with NET (internet) as the reference group for all the free mode of acquisition.

The type of card, gold is not statistically significant at alpha = 10%, platinum and quantum are statistically significant at alpha = 5% with Standard as the reference group for all type of cards.

The type of affinity cards, sectorD is not statistically significant at alpha = 10%, sectorB, sectorC, sectorE and sectorF are statistically significant at alpha = 5% with sectorA as the reference group for all type of affinity cards.

* AGE:

A 100-year increase in age is associated with a $189.75 decrease in the predicted value of profit.

* TOTTRANS:

A $1000 increase in Tottrans is associated with a $47.01 increase in the predicted value of profit.

* REWARDS:

A Customer with rewards card is expected to have profit $320 less than those with no rewards card.

* LIMIT:

A $1000 increase in Limit is associated with a $9.15 decrease in the predicted value of profit.

* Number of Cards:

Not Significant as p-value is 0.30.

* Mode of acquisition:

The predicted value of profit is $185.4 higher for customer acquired through dm (Direct Mail) than for customers acquired through NET(Internet).

The predicted value of profit is $821 less for customer acquired through ds (Direct Selling) than for customers acquired through NET(Internet).

The predicted value of profit is $829.4 less for customer acquired through ts (Telephone Selling) than for customers acquired through NET(Internet).

* Type of Card:

The predicted value of profit is $14.2 higher for customer having gold card than for customers having standard card.

The predicted value of profit is $667.2 higher for customer having platinum card than for customers having standard card.

The predicted value of profit is $481.01 less for customer having quantum card than for customers having standard card.

* Type of Affinity Card:

The predicted value of profit is $190.3 higher for customer having sectorB affinity card than for customers having sectorA affinity card.

The predicted value of profit is $215.97 less for customer having sectorC affinity card than for customers having sectorA affinity card.

SectorD is Not Significant at alpha = 10% as p-value 0.14

The predicted value of profit is $319.71 less for customer having sectorE affinity card than for customers having sectorA affinity card.

The predicted value of profit is $419.67 less for customer having sectorF affinity card than for customers having sectorA affinity card.

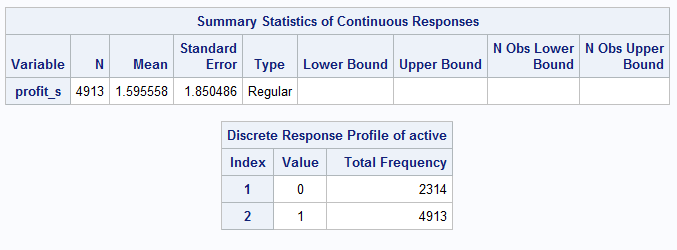
* The ancillary statistic \_sigma is equivalent to the square root of the residual variance in OLS regression. The value of 1.85 can be compared to the standard deviation of profit which was 1.69. That \_sigma is statistically significant at alpha = 5% which means that the estimated coefficient (1.85) is statistically significantly different from 0.

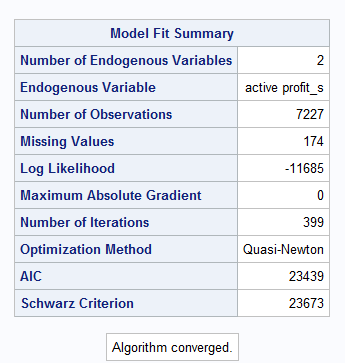
**2. Run a selection model (Use PROC QLIM)**

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.

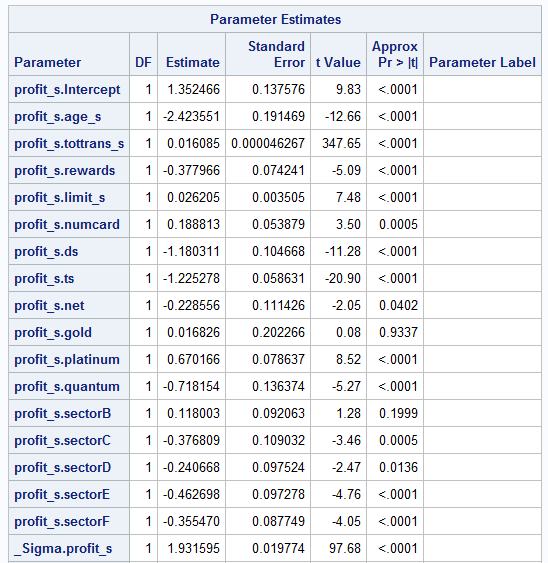


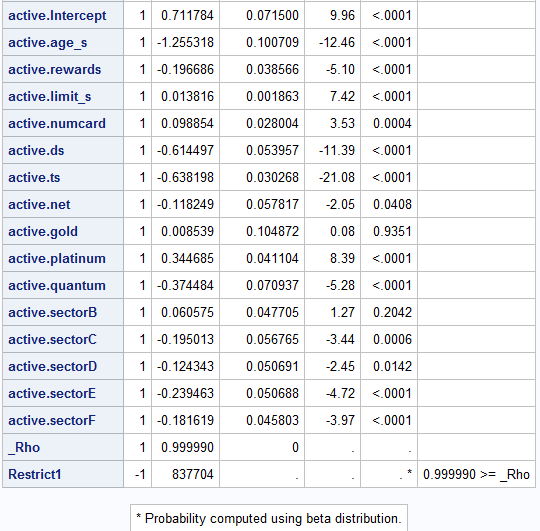


**Inference:**

According to the above Model Fit Summary:

* Log Likelihood -11685, AIC 23439 & SC(BIC) 23673 which when compared to TOBIT model it is much high, suggesting that model is not an improved model.
* SAS took 399 Iterations to converge the model
* Selected model has 2 Endogenous variables active





**Inference:**

Parameter Estimates:

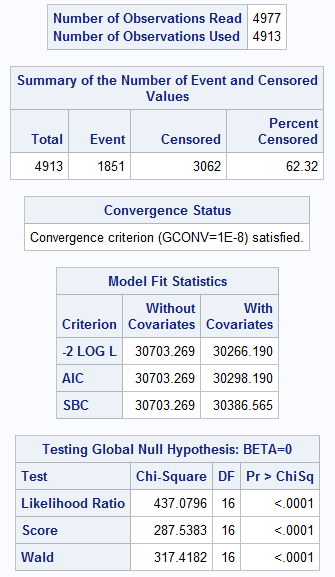
* The Model gave us Probit and Regression Parameters.
* The Model gave us Rho, which is used to detect significance of selection bias in the model. If \_Rho is statistically significant, then it indicates that there is significant impact of selection on model and the two models should not be run separately i.e. the model should be run together but if it is not significant then there is no selection bias in the model.
* When \_Rho is positive, this indicates that an unobservable are positively correlated with one another and when rho is negative, this indicates that unobservable are negatively correlated with one another.
* \_Sigma.profit\_s: Is the square root of the residual variance in Ordinary least squares regression.The value of 1.93 can be compared to the standard deviation of profit which was 1.85. It is statistically significant at alpha = 5% means that the estimated coefficient (1.93) is statistically significant and different from 0.

**3. Survival analysis**

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

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

Note that duration is censored if its value is 37 as we have only 37 months of data.



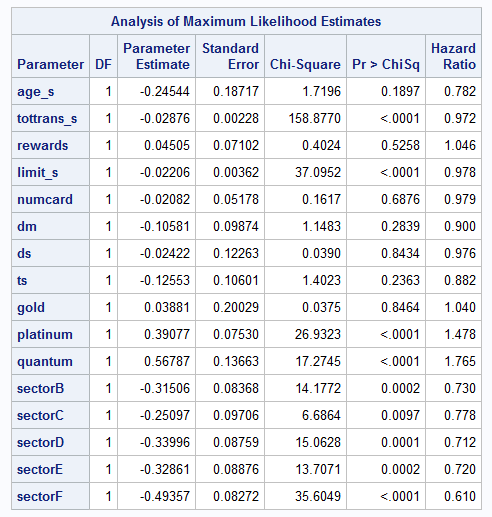
**Inference:**

Model Fit Statistics:

We see here that adding parameters as predictors improves the fit of the model, as all three statistics decreases when compared to model with no parameters.

Hypothesis test that at least one coefficient is different than 0.

All three Likelihood Ratio, Score, Wald are significant at alpha = 5% and agrees that at least one coefficient is different from 0.



**Inference:**

Parameter Estimated: [100(exp(β)-1)], exp(β)=Hazard Ratio

* AGE:

Not Significant at alpha = 10%

* Total Transaction:

With $1000 increase in total transaction Hazard rate decreases by 2.8% (Hazard Ratio=0.972)

* REWARDS:

Not Significant at alpha = 10%

* LIMIT:

With $1000 increase in Limit Hazard rate decreases by 2.2% (Hazard Ratio=0.978)

* Numcards:

Not Significant at alpha = 10%

* Mode of Acquisition:

It appears that all the modes of acquisition (i.e. DS, DM and TS) are Not Significant at alpha = 10%

* Type of Card:

Affect by customer having Gold Card cannot be estimate as its Not Significant at alpha = 10%.

Customer with Quantum card have 76.5% (Hazard Ratio = 1.765) increase in Hazard Rate compared to customer with Standard Card.

Customer with Platinum card have 47.8% (Hazard Ratio = 1.478) increase in Hazard Rate compared to customer with Standard Card.

* Types of Affinity Card:

SectorB affinity card have 31% (Hazard Ratio = 0.730) decrease in Hazard Rate compared to SectorA affinity Card.

SectorC affinity card have 22.8% (Hazard Ratio = 0.778) decrease in Hazard Rate compared to SectorA affinity Card.

SectorD affinity card have 28.8% (Hazard Ratio = 0.712) decrease in Hazard Rate compared to SectorA affinity Card.

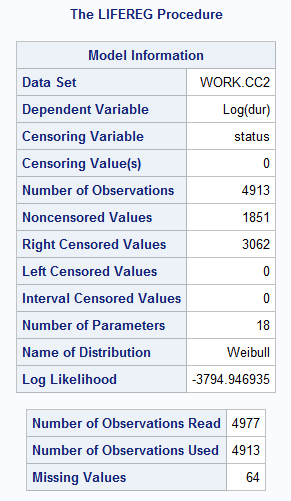
SectorE affinity card have 28% (Hazard Ratio = 0.720) decrease in Hazard Rate compared to SectorA affinity Card.

SectorF affinity card have 39% (Hazard Ratio = 0.610) decrease in Hazard Rate compared to SectorA affinity Card.

* There is no intercept. In Cox regression, the intercept is absorbed into the baseline hazard function, which is left unspecified.

**4. Run a model using PROC LIFEREG with Weibull distribution.**

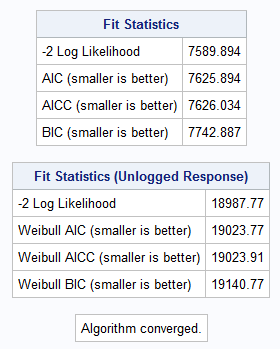
Write a summary of the results. Focus on important effects, interpretation, model fit etc.



There are 1851 uncensored observations and 3062 right-censored observations. The log likelihood for the Weibull distribution is -3794.946935.

The log-likelihood value can be used to compare the goodness of fit for nested models with different

covariates, but with the same distribution.



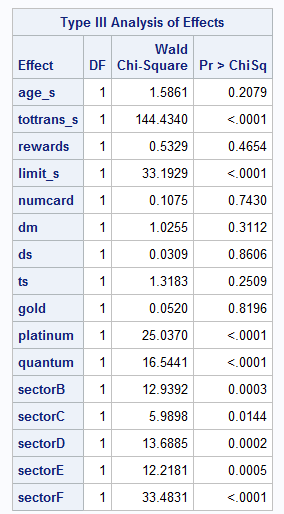
**Inference:**

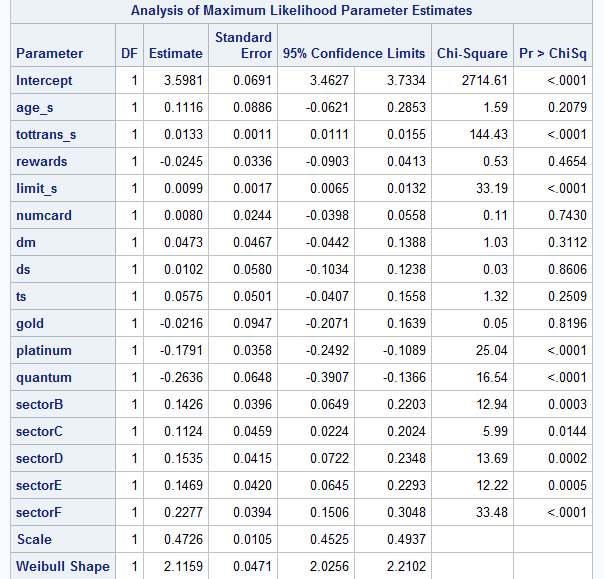
MODEL FIT SUMMARY:

There are 4977 observations out of which 3062 are right censored,1851 observations are noncensored and remaining 64 are missing values. The log likelihood is -3794.94.

Using the Weibull distribution, the 2 log likelihood = 18987.77, Weibull AIC = 19023.77, Weibull BIC = 19140.77 and Weibull AICC = 19023.9.

The “Fit Statistics” table displays statistics based on the maximum extreme-value log likelihood fit by using log(Duration)/ as the response. These statistics are useful in comparing the fit of a different model when the fit criteria from the model that you compare is also based on the log likelihood using log(duration)/ as the response. The “Fit Statistics (Unlogged Response)” table is based on the maximum Weibull log likelihood using Minutes as the response. The AIC, BIC, and AICC statistics in this table can be used to compare models with different covariates, in addition to models with different distributions, as long as the fit statistics for the models that you compare use Minutes as the response.





**Inference:**

Maximum Likelihood Parameter Estimates:

* exp(β) gives the ratio of the expected survival times for the two groups if the X variables is binary.
* for a continuous variable, 100(exp(β)-1) gives the percent increase in expected survival time for each unit increase in the variable.

Parameter Interpretations:

The interpretation of the Estimates of Weibull distribution model is:

* AGE: Not Significant at alpha = 10%
* tottrans: A $1000 increase in Total transactions by a customer the expected duration of customer being active increases by 1.34%.
* Rewards: Not Significant at alpha = 10%
* Limit: For a $1000 increase in limit of customer card the expected duration of customer being active increases by 0.99%.
* Number of Cards: Not Significant at alpha = 10%
* Mode of Acquisition: All three variables are insignificant at alpha = 10%
* Type of Card:

For Gold Card: Not Significant at alpha = 10%

For Platinum: A platinum card is expected to stay 0.17% less duration than standard Card

For Quantum Card: A Quantum card is expected to stay 0.26% less duration than standard card

* Type of affinity card:

For SectorB: B type affinity card is expected to stay 0.14% more duration with customer than customer with no affinity card.

SectorC: C type affinity card is expected to stay 0.11% more duration with than customer with customer with no affinity card.

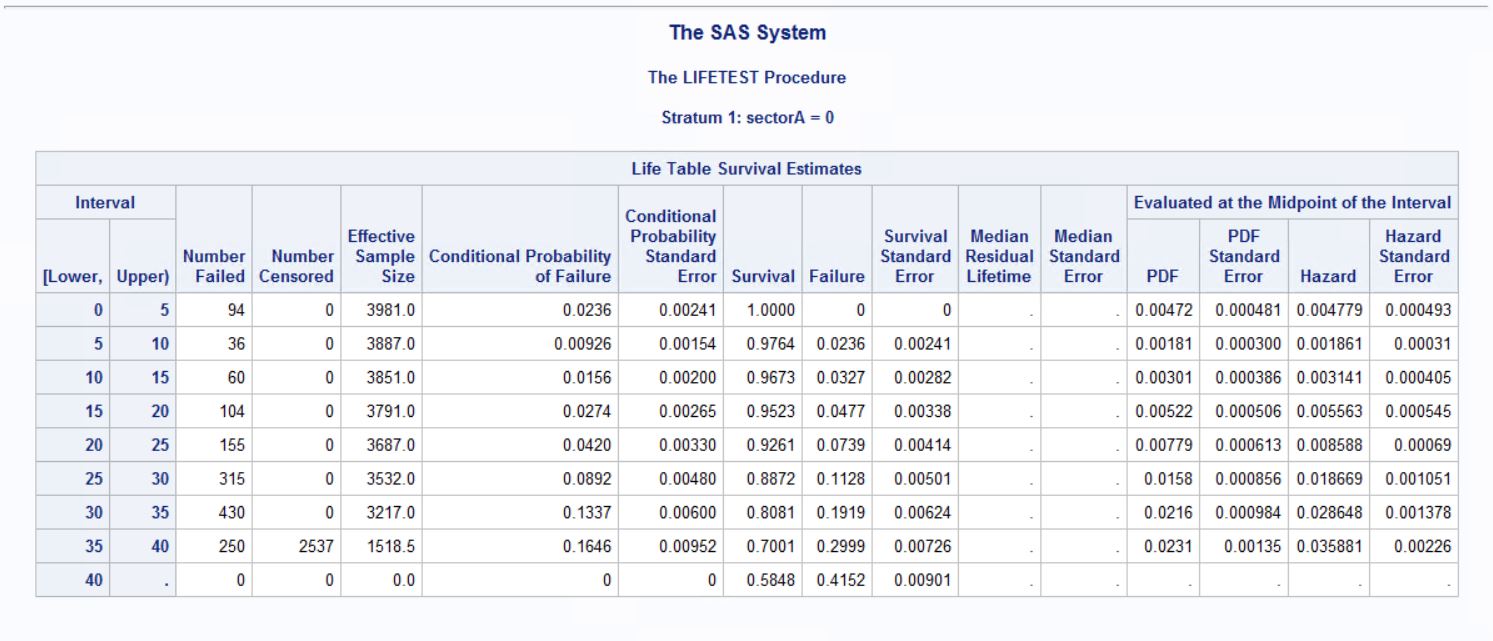
SectorD: D type affinity card is expected to stay 0.15% more duration with than customer with customer with no affinity card.

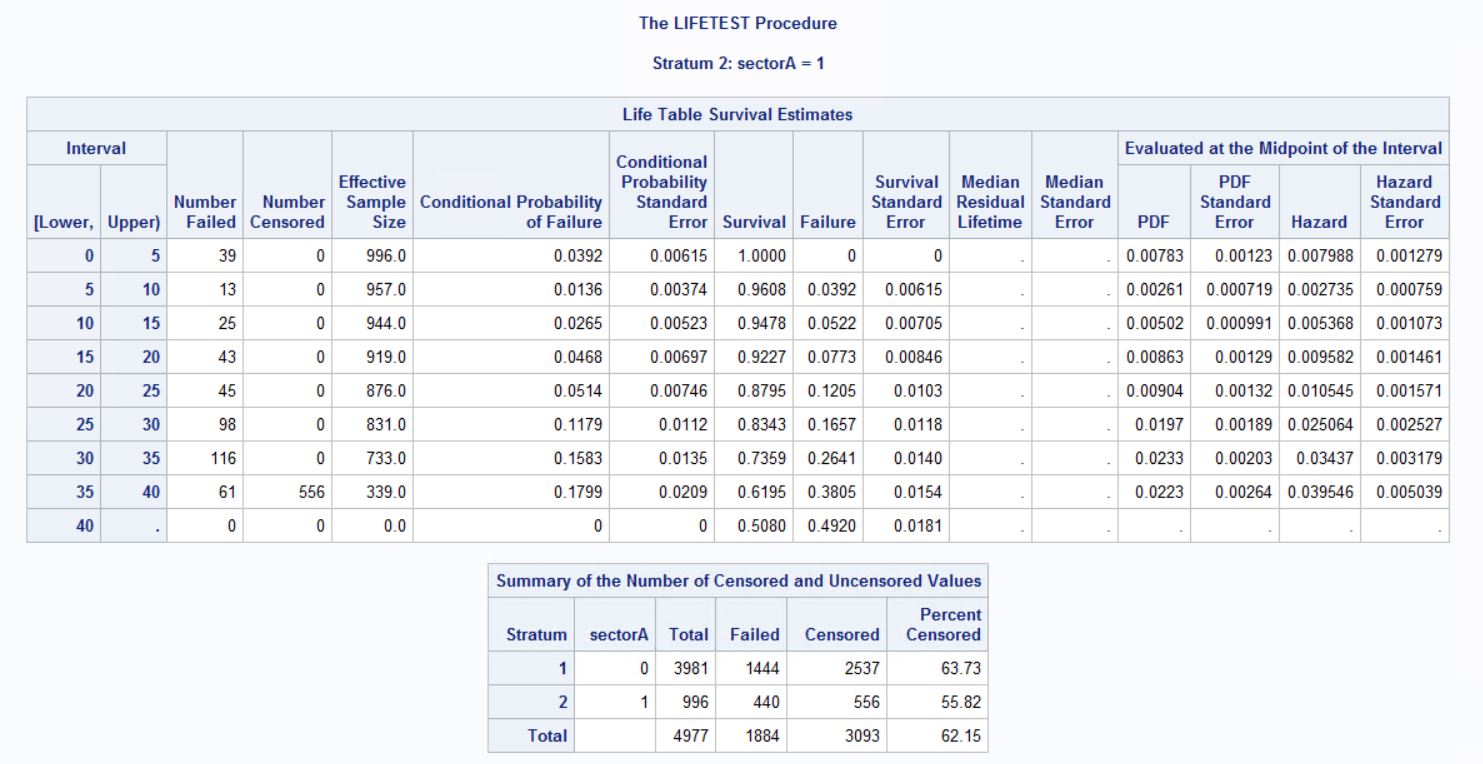
SectorE: E type affinity card is expected to stay 0.14% more duration with than customer with customer with no affinity card.

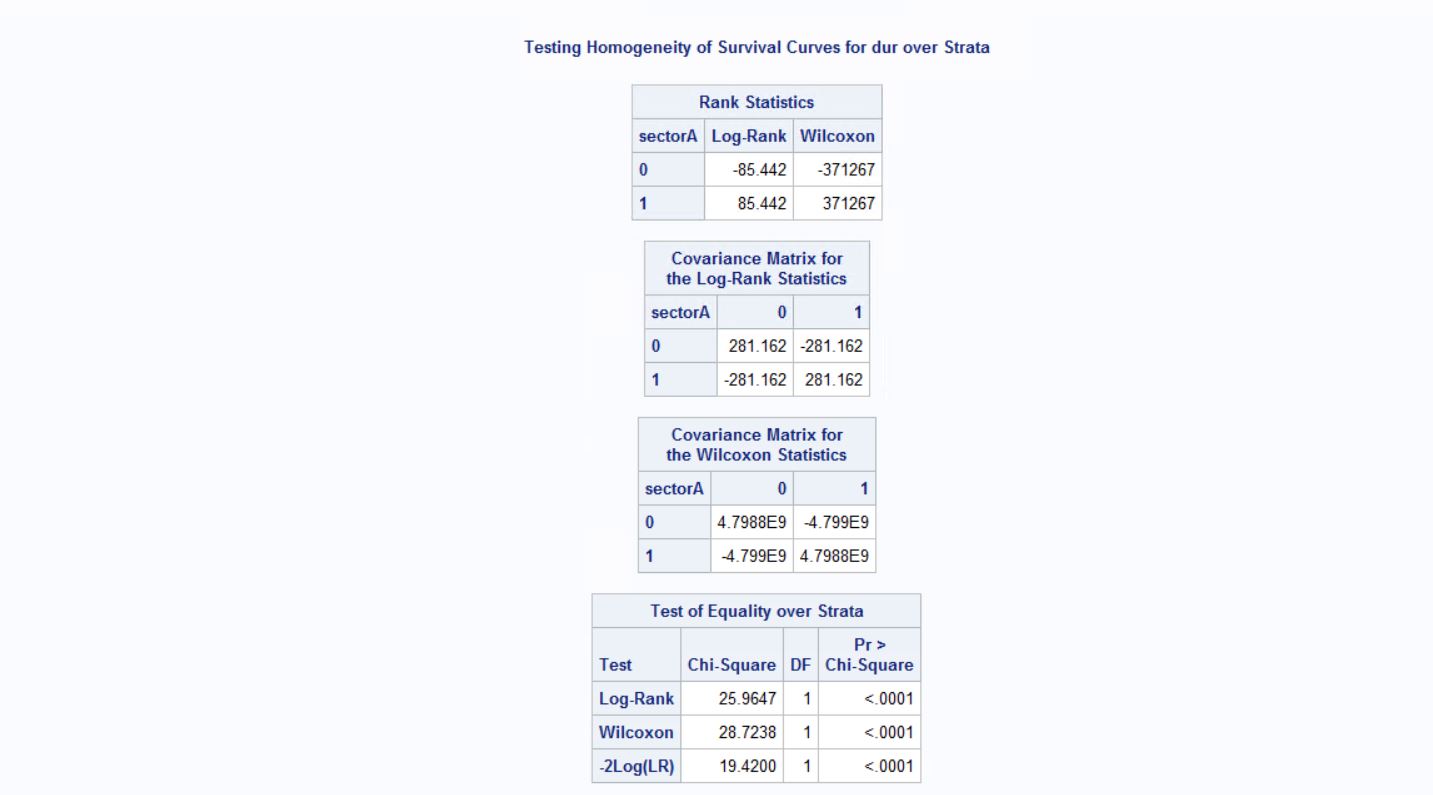
SectorF: F type affinity card is expected to stay 0.22% more duration with than customer with customer with no affinity card.

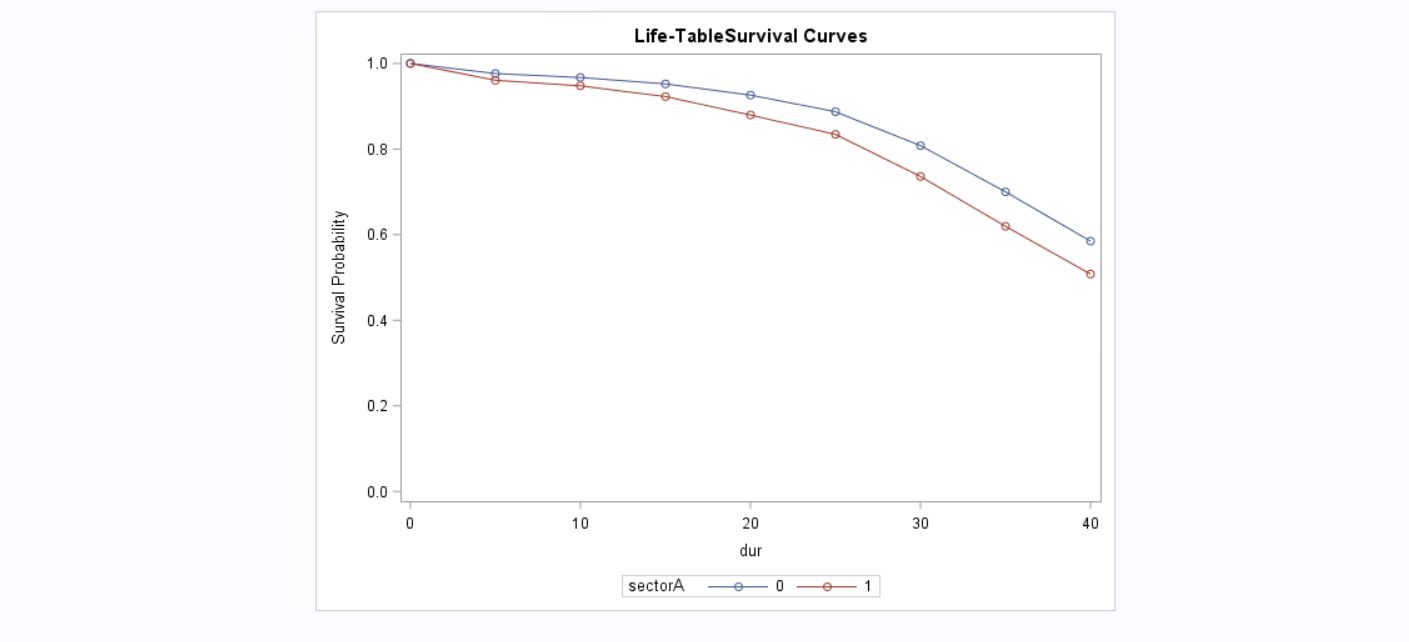
**5. Use PROC LIFETEST to test whether survivor function of affinity groups are significantly different from that of non-affinity groups.**

What do you conclude?









**Inference:**

We can conclude on the basis of statistics obtained via Test of Equality over strata:

Ho: Two groups have exactly the same Survivor function

H1: Two groups have different Survivor function

On considering Log-rank and Wilcoxon p-values which is (<.0001) and (<.0001) respectively

Considering Alpha=0.05(5%) rejection region, since p-value<aplha i.e. (0.001<0.05)

We have enough statistical evidence to reject the H0 and conclude that affinity groups are significantly different from that of non-affinity groups