STAT 6800 HW5

Augustine Ennin

November 2025

Q1

```
Title "Created on %qsysfunc(today(), weekdate.)";
proc sql number;
select
substr(FirstName, 1, 1) label = "Initial",
LastName label = "Last Name",
CreditScore label = "Credit Score",
case
when DOB between '01Jan1928'd and '31DEC1945'd then "Silent"
when DOB between '01Jan1946'd and '31DEC1964'd then "Boomer"
when DOB between '01Jan1965'd and '31DEC1979'd then "GenX"
when DOB between '01Jan1980'd and '31DEC1996'd then "Millennial"
when DOB >= '01JAN1997'd then "Post-Millennial"
else "Unknown"
end as Generation
from sq.customer
where CreditScore is not missing and state = "VT"
order by Generation, CreditScore desc;
quit;
title;
```

Created on Saturday, November 29, 2025

Row	Initial	Last Name	Credit Score	Generation
1	G	Reece	784	Boomer
2	C	Hebert	749	Boomer
3	J	Hockema	734	Boomer
4	M	Santone	722	Boomer
5	C	Smitley	717	Boomer
6	R	Angst	703	Boomer
7	L	Parks	702	Boomer
8	K	Armistead	693	Boomer
9	C	Wiseley	690	Boomer
10	D	Pickle	689	Boomer
11	J	Mackie	684	Boomer
12	C	Madonia	680	Boomer
13	R	Espinoza	665	Boomer
14	A	Barnes	660	Boomer
15	C	Jenkins	650	Boomer
16	S	Renn	636	Boomer
17	H	Lukasik	634	Boomer
18	N	Bosque	783	GenX
19	J	Gallagher	732	GenX
20	J	Nash	698	GenX
21	G	Kretsinger	696	GenX
22	S	Carlson	662	GenX
23	N	Underwood	660	GenX
24	K	Jackson	640	GenX
25	E	Hatala	620	GenX
26	L	Betts	611	GenX
27	A	Kufeldt	752	Millennial
28	J	Shipley	736	Millennial

29	M	Smith	688	Millennial
30	S	Mattox	673	Millennial
31	W	Mangone	650	Millennial
32	F	Davis	642	Millennial
33	C	Gonzalez	630	Millennial
34	J	Willard	622	Millennial
35	A	Farrell	603	Millennial
36	L	Gray	709	Post-Millennial
37	M	Hitt	670	Post-Millennial
38	A	Dodson	668	Post-Millennial
39	M	Miller	654	Post-Millennial
40	D	Adams	652	Post-Millennial
41	L	Smith	651	Post-Millennial
42	K	Mazurowski	628	Post-Millennial
43	K	Bowden	581	Post-Millennial
44	L	Dodson	579	Post-Millennial
45	D	Ptaszynski	577	Post-Millennial
46	R	Martinez	541	Post-Millennial

Q2

```
Title "Customer Marital Status Count By Primary Bank";
proc sql;
select
c.BankID,
m.MaritalStatus,
b.Name,
put(Count(*), comma9.) as Count
from
sq.customer as c
left join
sq.maritalcode as m
on c.married = m.MaritalCode
left join
sq.bank as b
on c.BankID = b.BankID
where c.BankID is not null
group by c.BankID, m.MaritalStatus, b.Name
order by Count desc;
```

```
Quit;
title;
```

Customer Marital Status Count By Primary Bank

BankID	MaritalStatus	Name	Count
101010101	Married	Biggest Bank, Inc.	22,933
202020202	Married	Sailors Credit Union	16,997
303030303	Married	Wheatberry Bank, Inc.	14,258
101010101	Single	Biggest Bank, Inc.	8,793
202020202	Single	Sailors Credit Union	6,589
303030303	Single	Wheatberry Bank, Inc.	5,458
101010101	Divorced	Biggest Bank, Inc.	4,896
202020202	Divorced	Sailors Credit Union	3,644
303030303	Divorced	Wheatberry Bank, Inc.	3,054
101010101		Biggest Bank, Inc.	2,814
202020202		Sailors Credit Union	2,162
303030303		Wheatberry Bank, Inc.	1,802
101010101	Widowed	Biggest Bank, Inc.	739
202020202	Widowed	Sailors Credit Union	549
303030303	Widowed	Wheatberry Bank, Inc.	362

Q2(c). Widowed customers at Wheatberry Bank, Inc., with a count of 362

Q3

```
libname mylib "/home/u63997979/Survival Analysis";
data worcester;
set mylib.worcester;
run;

proc print data = worcester (obs=10);
title "First 10 Observation of the Worcester Dataset";
run;
```

First 10 Observation of the Worcester Dataset

Obs	ID	Age	Gender	Cardiac_Enzyme	Cardio_Compl	Heart_Compl	MI_Order	MI_Type	DepOnset	Length_Follow	Status	MI_Type1
1	1	62	1	485	1	1	0	1	-1	1	1	1
2	2	78	1	910	0	1	1	1	-1	1	1	1
3	3	81	1	320	1	1	0	1	-1	1	1	1
4	4	79	1	3290	1	1	1	1	-1	1	1	1
5	5	60	1	2500	1	1	1	1	-1	2	1	1
6	6	72	0	99	0	0	0	1	-1	2	1	1
7	7	60	1	1200	0	0	0	1	-1	2	1	1
8	8	83	1	160	0	0	0	1	-1	3	1	1
9	9	78	0	66	0	1	1	1	-1	3	1	1
10	10	72	1	99	0	0	0	1	-1	5586	0	1

```
/* Cox proportional hazards model with exact method for ties; */;
proc phreg data = worcester;
class gender (ref='0') cardio_compl (ref='0') heart_compl (ref='0') mi_order (ref
   ='0') mi_type1 (ref='0') / param=ref;
model length_follow*status(0) = age gender cardiac_enzyme cardio_compl heart_compl
    mi_order mi_type1 / ties=exact;
/* Profile-likelihood confidence limits for each categorical predictor */
hazardratio age / cl=pl;
hazardratio gender / cl=pl;
hazardratio cardiac_enzyme / cl=pl;
hazardratio cardio_compl / cl=pl;
hazardratio heart_compl / cl=pl;
hazardratio mi_order / cl=pl;
hazardratio mi_type1 / cl=pl;
title "Cox PH model (ties=EXACT) for long-term survival after AMI";
run;
```

Cox PH model (ties=EXACT) for long-term survival after AMI

The PHREG Procedure

Model Information	
Data Set	WORK.WORCESTER
Dependent Variable	Length_Follow
Censoring Variable	Status
Censoring Value(s)	0
Ties Handling	EXACT

Number of Observations Read	481
Number of Observations Used	481

Class Level Information		
Class	Value	Design Variables
Gender	0	0
	1	1
Cardio_Compl	0	0
	1	1
Heart_Compl	0	0
	1	1
MI_Order	0	0
	1	1
MI_Type1	0	0
	1	1

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
481	249	232	48.23

Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	2645.732	2466.814
AIC	2645.732	2480.814
SBC	2645.732	2505.436

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	178.9179	7	<.0001
Score	310.5805	7	<.0001
Wald	213.4985	7	<.0001

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Age	1	34.0755	<.0001
Gender	1	0.0896	0.7646
Cardiac_Enzyme	1	1.0410	0.3076
Cardio_Compl	1	74.2020	<.0001
Heart_Compl	1	16.1331	<.0001
MI_Order	1	5.2924	0.0214
MI_Type1	1	2.4081	0.1207

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Age		1	0.03464	0.00593	34.0755	<.0001	1.035	
Gender	1	1	0.04022	0.13434	0.0896	0.7646	1.041	Gender 1
Cardiac_Enzyme		1	0.0000686	0.0000672	1.0410	0.3076	1.000	
Cardio_Compl	1	1	1.80273	0.20928	74.2020	<.0001	6.066	Cardio_Compl 1
Heart_Compl	1	1	0.57378	0.14285	16.1331	<.0001	1.775	Heart_Compl 1
MI_Order	1	1	0.30679	0.13336	5.2924	0.0214	1.359	MI_Order 1
MI_Type1	1	1	0.21395	0.13787	2.4081	0.1207	1.239	MI_Type1 1

Hazard Ratios for Age			
Description	Point Estimate	95% Profile Likelihood Confidence Limits	
Age Unit=1	1.035	1.023	1.047

Hazard Ratios for Gender			
Description	Point Estimate	95% Profile Likelihood Confidence Limits	
Gender 0 vs 1	0.961	0.739	1.251

Hazard Ratios for Cardiac_Enzyme			
Description	Point Estimate	95% Profile Likelihood Confidence Limits	
Cardiac_Enzyme Unit=1	1.000	1.000	1.000

Hazard Ratios for Cardio_Compl			
Description	Point Estimate	95% Profile Likelihood Confidence Limits	
Cardio_Compl 0 vs 1	0.165	0.110	0.251

Hazard Ratios for Heart_Compl			
Description	Point Estimate	95% Profile Likelihood Confidence Limits	
Heart_Compl 0 vs 1	0.563	0.426	0.745

Hazard Ratios for MI_Order			
Description	Point Estimate	95% Profile Likelihood Confidence Limits	
MI_Order 0 vs 1	0.736	0.567	0.957

Hazard Ratios for MI_Type1			
Description	Point Estimate	95% Profile Likelihood Confidence Limits	
MI_Type1 0 vs 1	0.807	0.615	1.056

```
/* Descriptive Statistics for predictor variables */;
proc means data=worcester n mean std min p25 median p75 max;
var age cardiac_enzyme;
title "Descriptive Statistics for Continuous Predictors";
run;
```

Descriptive Statistics for Continuous Predictors

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	25th Pctl	Median	75th Pctl	Maximum
Age	481	67.4844075	12.6805406	24.0000000	59.0000000	68.0000000	77.0000000	98.0000000
Cardiac_Enzyme	481	941.5426195	1132.08	10.0000000	270.0000000	587.0000000	1146.00	9000.00

```
proc freq data=worcester;
tables gender cardio_compl heart_compl mi_order mi_type1;
title "Frequency Distributions for Categorical Predictors";
run;
```

Frequency Distributions for Categorical Predictors

The FREQ Procedure

Gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	287	59.67	287	59.67
1	194	40.33	481	100.00

Cardio_Compl	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	443	92.10	443	92.10
1	38	7.90	481	100.00

Heart_Compl	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	285	59.25	285	59.25
1	196	40.75	481	100.00

MI_Order	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	308	64.03	308	64.03
1	173	35.97	481	100.00

MI_Type1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	201	41.79	201	41.79
1	280	58.21	481	100.00

Q3.(a)

The statistic to use for the test is the ‘likelihood ratio test‘ from the ”Testing Global Null Hypothesis: BETA=0” table.

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$
$$H_1 : \text{At least one } \beta_i \neq 0 \text{ for } i = 1, 2, \ldots, 7$$

Since the p-value (<.0001) is less than the significance level of 0.05, we reject the null hypothesis that all regression coefficients are zero. We conclude that there is statistically significant evidence that at least one of the predictor variables (age, gender, cardiac enzyme, cardiogenic shock complications, heart complications, MI order, or MI type) is associated with survival time following acute myocardial infarction.

Q3.(b)

The parameter estimate for Cardio_Compl is 1.80273, which represents the log hazard ratio comparing patients with cardiogenic shock complications to those without such complications, after adjusting for all other predictors in the model. This positive coefficient indicates that patients who experienced cardiogenic shock have a log hazard that is 1.80273 units higher than patients without these complications, reflecting a substantially increased hazard (risk) of death at any point during the follow-up period.

Q3.(c)

The hazard ratio of 0.165 indicates that patients without cardiogenic shock complications have only 16.5% of the hazard of death—an 83.5% lower risk—compared to patients with cardiogenic shock, after adjusting for all other predictors. The 95% profile likelihood confidence interval (0.110, 0.251) shows that the true hazard ratio in the population lies between 0.110 and 0.251. Because the entire interval is well below 1, it confirms that cardiogenic shock complications are a strong and statistically significant risk factor for increased mortality.