STAT 6800 Final Exam

Augustine Ennin

December 2025

Q1.(a)

$$H_0 : p = 0.06$$
$$H_1 : p \neq 0.06$$

Where p is the true proportion of tires that experience tread separation.

Q1.(b)
Syntax error:

- No semicolon after the line `proc freq data = one`
- The keyword should be `tables` and not `table`
- `tread` is a wrong variable name. It should be `treadsep` instead.

Mathematical error: The `testp` values should be `(0.06, 0.94)` and not `(0.4, 0.6)`. Also, the `z` should be changed to `binomial`.

Q1.(c)

```
data one;
input treadsep $;
datalines;
no
no
no
no
no
no
no
no
no
yes
;
run;

proc print data = one;
    title "Tire Tread Separation Data";
run;

proc freq data = one;
    tables treadsep / binomial(p=0.06);
run;
```

Tire Tread Separation Data

Obs	treadsep
1	no
2	no
3	no
4	no
5	no
6	no
7	no
8	no
9	no
10	yes

Tire Tread Separation Data

The FREQ Procedure

treadsep	Frequency	Percent	Cumulative Frequency	Cumulative Percent
no	9	90.00	9	90.00
yes	1	10.00	10	100.00

Binomial Proportion	
treadsep = no	
Proportion	0.9000
ASE	0.0949
95% Lower Conf Limit	0.7141
95% Upper Conf Limit	1.0000
Exact Conf Limits	
95% Lower Conf Limit	0.5550
95% Upper Conf Limit	0.9975

Test of H0: Proportion = 0.06	
ASE under H0	0.0751
Z	11.1851
One-sided Pr > Z	<.0001
Two-sided Pr > |Z|	<.0001

Sample Size = 10

The p-value for the two-sided test is $< .0001$, less than $\alpha = 0.05$, hence, reject the null hypothesis. Therefore, we can conclude that there is significant evidence to show that the proportion of tires that experience tread separation is not 0.06.

Q2.(a)
This is a paired sample t-test because each citizen's reaction time was measured twice: before and after consuming alcohol. Furthermore, our interest is in how these two metrics differ for each citizen.

Q2.(b)
$H_0 : \mu_B - \mu_A = 0$
$H_a : \mu_B - \mu_A < 0$

Q2.(c)

The p-value is 0.0083. Since this is less than $\alpha = 0.05$, we reject the null hypothesis. We conclude that alcohol consumption significantly increases (slows) reaction times, with an average increase of 0.066 seconds after drinking.

Q3.(a) and Q3(b)

```
data btt;
infile "/home/u63997979/sasuser.v94/Elliott and Morrell/btt.dat.txt";
input
childid 1-4
sex 6
bweight 8-11
gestage 13-14
momage 16-17
parity 19
mdbp 21-23
msbp 25-27
momeduc 29
mmedaid 31
socio 33
dbp5 35-37
sbp5 39-41
ht5 43-47
wt5 49-52
hdl5 54-57
ldl5 59-62
trig5 64-67
smoke5 69
medaid 71
socio5 73;
bmi = wt5 / (ht5)**2;
run;

%macro analyze_btt(data=btt, start=1, end=4);

    %do momeduc = &start %to &end;

        title "Analysis for Mother's Education Level &momeduc";

        %if &momeduc = 1 %then %do;
            proc means data=&data n mean std min max maxdec=2;
                where momeduc = &momeduc;
                var dbp5 sbp5;
                title2 "Descriptive Statistics: Diastolic and Systolic BP at Age 5
                    ";
            run;
        %end;

        %else %if &momeduc = 2 %then %do;
            proc ttest data=&data;
                where momeduc = &momeduc;
                class sex;
                var trig5;
                title2 "T-test: Triglycerides at Age 5 by Child's Sex";
            run;
        %end;

        %else %if &momeduc = 3 %then %do;
```

```
            proc sgplot data=&data;
                where momeduc = &momeduc;
                scatter x=wt5 y=ht5 / markerattrs=(size=7);
                reg x=wt5 y=ht5 / lineattrs=(color=red thickness=2);
                xaxis label="Weight at Age 5 (kg)";
                yaxis label="Height at Age 5 (cm)";
                title2 "Scatter Plot: Height vs Weight at Age 5";
            run;

            proc reg data=&data;
                where momeduc = &momeduc;
                model ht5 = wt5;
                title2 "Regression: Height on Weight at Age 5";
            run;
        %end;

        %else %if &momeduc = 4 %then %do;
            proc sgplot data=&data;
                where momeduc = &momeduc;
                vbox sbp5 / category=sex;
                xaxis label="Sex";
                yaxis label="Systolic Blood Pressure at Age 5";
                title2 "Box Plot: Systolic BP Distribution by Sex";
            run;
        %end;

        %else %do;
            %put Note: No analysis defined for momeduc = &momeduc;
        %end;

    %end;

%mend analyze_btt;

%analyze_btt(data=btt, start=1, end=4);
```

Q3.(c)

Analysis for Mother's Education Level 1
Descriptive Statistics: Diastolic and Systolic BP at Age 5

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
dbp5	28	64.57	6.51	51.00	79.00
sbp5	28	110.75	10.37	98.00	133.00

The average diastolic blood pressure (dbp5) and systolic blood pressure (sbp5) of five-year-old children with the least educated mothers (level 1) are 64.57 and 110.75, respectively. SBP5 showed greater variation (10.37) compared to DBP5 (6.51). The minimum and maximum of dbp5 are 51 and 79, respectively, while those of sbp5 are 98 and 133, respectively.

Analysis for Mother's Education Level 2
T-test: Triglycerides at Age 5 by Child's Sex

The TTEST Procedure

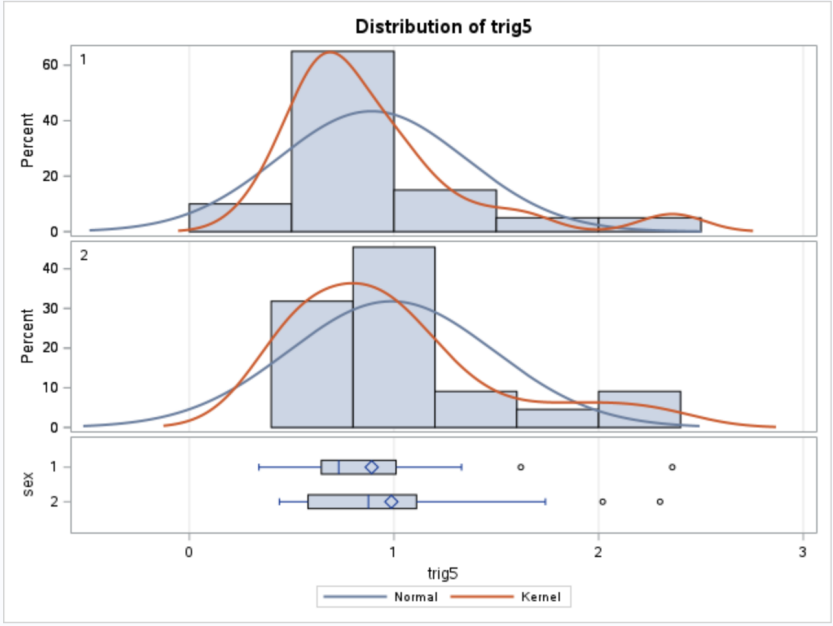
Variable: trig5

sex	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1		20	0.8910	0.4596	0.1028	0.3400	2.3600
2		22	0.9873	0.5019	0.1070	0.4400	2.3000
Diff (1-2)	Pooled		-0.0963	0.4823	0.1490		
Diff (1-2)	Satterthwaite		-0.0963		0.1484		

sex	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		0.8910	0.6759	1.1061	0.4596	0.3496	0.6713
2		0.9873	0.7647	1.2098	0.5019	0.3862	0.7173
Diff (1-2)	Pooled	-0.0963	-0.3974	0.2049	0.4823	0.3960	0.6171
Diff (1-2)	Satterthwaite	-0.0963	-0.3962	0.2036			

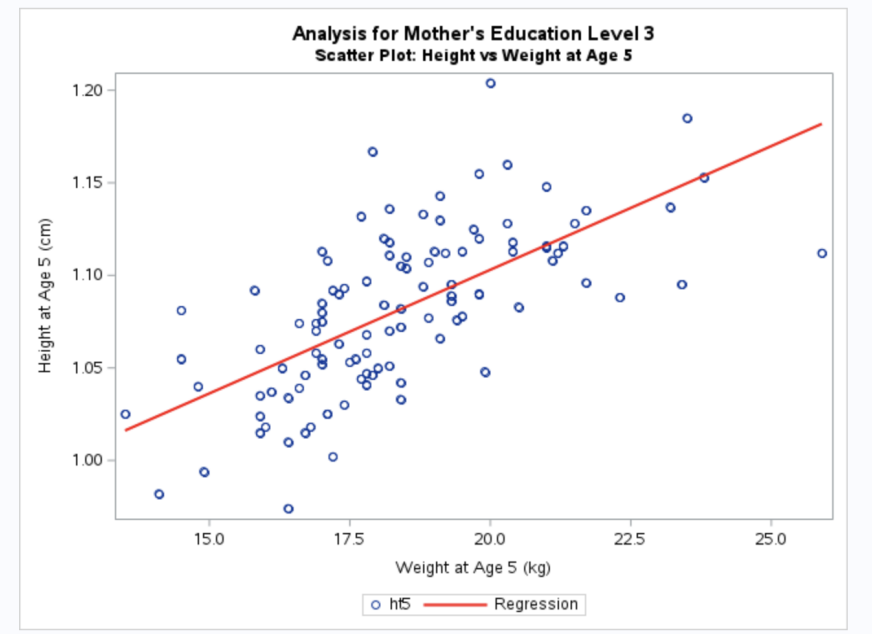
Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	40	-0.65	0.5219
Satterthwaite	Unequal	39.996	-0.65	0.5202

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	21	19	1.19	0.7040



The average female child's triglyceride level at 5 years is higher than that of the male child (0.9873 vs. 0.8910). The Folded F test has a p-value of 0.7040, which is greater than $\alpha = 0.05$. Hence, we do not reject the null hypothesis, which suggests that the assumption of equal variance is valid. Therefore, for this test

we would go with the pooled method, which assumes equal variances. Based on the p-value of the pooled method, we do not reject the null hypothesis and conclude that there is not enough statistical evidence to suggest there is a difference in the triglyceride level between the male and female child at age 5 when the mother's education level is 2. This is supported by the confidence interval (-0.3974, 0.2049), which includes 0.



The plot above shows a positive relationship between the weight and height of children at age 5 when the mother's education level is 3; as height increases, weight also increases.

Analysis for Mother's Education Level 3
Regression: Height on Weight at Age 5

The REG Procedure
Model: MODEL1
Dependent Variable: ht5

Number of Observations Read	108
Number of Observations Used	108

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.08951	0.08951	80.56	<.0001
Error	106	0.11777	0.00111		
Corrected Total	107	0.20728			

Root MSE	0.03333	R-Square	0.4318
Dependent Mean	1.08143	Adj R-Sq	0.4265
Coeff Var	3.08226		

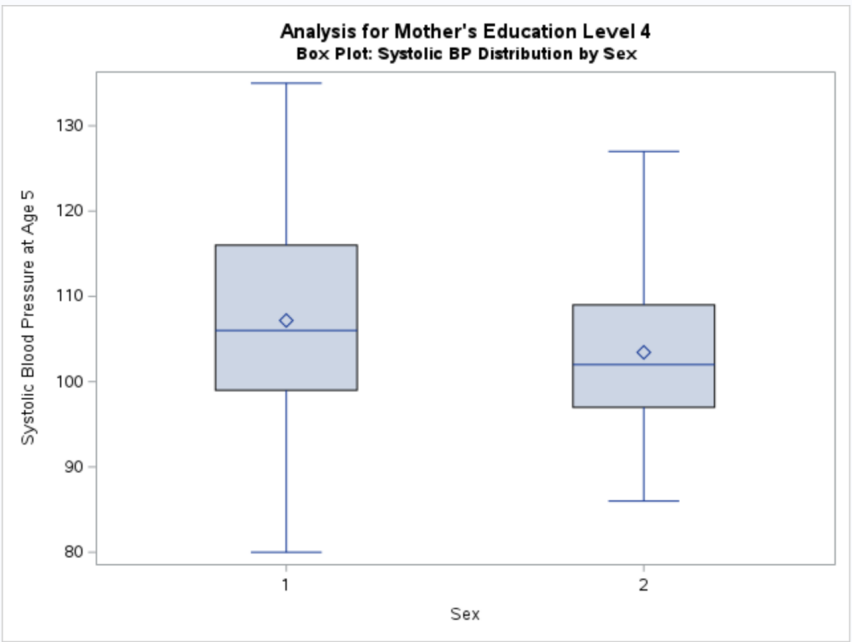
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.83577	0.02756	30.33	<.0001
wt5	1	0.01336	0.00149	8.98	<.0001

The ANOVA table is statistically significant (p-value $< .0001$). The parameter estimates table suggests that for each additional kg increase in weight, height increases statistically significantly (p-value $< .0001$) by 0.01336 m. The regression equation is:

$$Height = 0.83577 + 0.01336 \times Weight$$

An R^2 of 0.4318 suggests that about 43.2% of the variability in height is explained by weight.



Among children whose mothers have education level 4 (at least high school), the male children have a higher median SBP than female children. The male distribution also shows greater variability, as seen in the larger interquartile range and wider spread from min to max. There are no outliers in both classes.

Q4.(a)

```sql
proc sql;
create table Merchants_With_No_Sales as
select distinct
    m.MerchantName,
    m.MerchantID,
    m.Type label = "Merchant Type",
    m.Zip label = "Merchant Zipcode"
from sq.merchant as m
left join sq.transaction as t
    on m.MerchantID = t.MerchantID
where m.zip = 10001
    and t.TransactionID is null;
quit;

proc print data=Merchants_With_No_Sales label;
run;
```

Obs	Merchant Name	Merchant ID	Merchant Type	Merchant Zipcode
1	13th Street Grocery	565543	Groceries	10001
2	Gerivest Services	585048	Retirement	10001
3	Gifting Hands	517039	Charity	10001
4	Good Service Auto Repair	502136	Auto	10001
5	Miasma Mitigation, Inc.	518339	Charity	10001

Q4.(b)

```
proc sql;
title "Potential NYC Customers for Merchant With No Sales";
select distinct
        m.MerchantName,
        m.MerchantID,
        m.Type label = "Merchant Type",
        m.Zip label = "Merchant Zipcode",
        c.CustomerID,
        c.CustomerName,
        c.Address label = "Customer Address"

from Merchants_With_No_Sales as m

        inner join sq.transactionfull as c
        on m.Type = c.Type

where input(scan(c.Address,-1), 5.) = 10001

order by m.MerchantID, c.CustomerID;

quit;
```

Potential NYC Customers for Merchant With No Sales

Merchant Name	Merchant ID	Merchant Type	Merchant Zipcode	CustomerID	CustomerName	Customer Address
Good Service Auto Repair	502136	Auto	10001	1929444655	Pennacchio, Joan Lynn	28 Crest Bend New York NY 10001
Good Service Auto Repair	502136	Auto	10001	1931110664	Comstock, Olga Cathy	653 County Line Court New York NY 10001
Good Service Auto Repair	502136	Auto	10001	1970082571	Alexander, Ruth Helen	14 16th Street New York NY 10001
Good Service Auto Repair	502136	Auto	10001	1978669535	Bowers, Margaret Katie	173 Lakeshore Terrace New York NY 10001
Good Service Auto Repair	502136	Auto	10001	1989612017	Bower, Omar Randy	367 Lake Lane New York NY 10001
Good Service Auto Repair	502136	Auto	10001	1998323808	Sienkiewicz, Gary Roman	301 Wood Drive New York NY 10001
Gifting Hands	517039	Charity	10001	1929444655	Pennacchio, Joan Lynn	28 Crest Bend New York NY 10001
Gifting Hands	517039	Charity	10001	1931110664	Comstock, Olga Cathy	653 County Line Court New York NY 10001
Gifting Hands	517039	Charity	10001	1970082571	Alexander, Ruth Helen	14 16th Street New York NY 10001
Gifting Hands	517039	Charity	10001	1978669535	Bowers, Margaret Katie	173 Lakeshore Terrace New York NY 10001
Gifting Hands	517039	Charity	10001	1989612017	Bower, Omar Randy	367 Lake Lane New York NY 10001
Gifting Hands	517039	Charity	10001	1998323808	Sienkiewicz, Gary Roman	301 Wood Drive New York NY 10001
Miasma Mitigation, Inc.	518339	Charity	10001	1929444655	Pennacchio, Joan Lynn	28 Crest Bend New York NY 10001
Miasma Mitigation, Inc.	518339	Charity	10001	1931110664	Comstock, Olga Cathy	653 County Line Court New York NY 10001
Miasma Mitigation, Inc.	518339	Charity	10001	1970082571	Alexander, Ruth Helen	14 16th Street New York NY 10001
Miasma Mitigation, Inc.	518339	Charity	10001	1978669535	Bowers, Margaret Katie	173 Lakeshore Terrace New York NY 10001
Miasma Mitigation, Inc.	518339	Charity	10001	1989612017	Bower, Omar Randy	367 Lake Lane New York NY 10001
Miasma Mitigation, Inc.	518339	Charity	10001	1998323808	Sienkiewicz, Gary Roman	301 Wood Drive New York NY 10001

Q4.(c)

Miasma Mitigation, Inc. can now market to 6 new customers.

Q5.

Data and Method

This data set is from kaggle, and it provides detailed medical and demographic information for 5,000 patients diagnosed with various types of eye cancer, including Melanoma, Retinoblastoma, and Lymphoma. It includes:

- Patient Demographics: Age (1–90), Gender, Country

- Clinical Details: Cancer type, Laterality (Left/Right/Bilateral), Stage at diagnosis
- Diagnosis and Treatment: Date of diagnosis, Type of treatment (Surgery, Radiation, Chemotherapy), Treatment intensity (e.g., radiation dose in Gy, number of chemo sessions)
- Outcomes: Survival time in months, Outcome status (In Remission, Active, Deceased)
- Genetics and History: Genetic markers (e.g., BRAF mutation), Family history of eye cancer

I started preparing the data by creating indicator variables for some of the categorical variables. Exploratory data analysis followed.

```
proc import datafile="/home/u63997979/eye_cancer_patients.csv"
        out=eye_cancer
        dbms=csv
        replace;
        getnames=yes;
run;

proc print data=eye_cancer (OBS=5);
title "First 5 Observation of the Eye Cancer Dataset";
run;

/* Data for survival analysis */
data ec_cox_analysis;
        set eye_cancer;

        /* Event indicator: 1=Deceased, 0=Censored */
        event = (Outcome_Status = 'Deceased');

        /* Recode key categorical variables */
        braf = (Genetic_Markers = 'BRAF Mutation');
        family_hx = (Family_History = 'True');

        /* Age Categories for clinical relevance */
        if Age < 18 then age_cat = 1;
        else if age < 65 then age_cat = 2;
        else age_cat = 3;

        /* Label variables */
        label event = 'Death Event'
                Survival_Time_Months = 'Survival Time (Months)'
                braf = 'BRAF Mutation'
                familt_hx = 'Family History';
run;

/* Descriptive Statistics */
proc freq data=ec_cox_analysis;
        tables event Cancer_Type Stage_at_Diagnosis Treatment_Type;
run;

proc means data=ec_cox_analysis N mean std min max;
        var Age Survival_Time_Months Radiation_Therapy Chemotherapy;
        class event;
run;
```

First 5 Observation of the Eye Cancer Dataset run

Obs	Patient_ID	Age	Gender	Cancer_Type	Laterality	Date_of_Diagnosis	Stage_at_Diagnosis	Treatment_Type	Surgery_Status	Radiation_Therapy	Chemotherapy	Outcome_Status	Survival_Time_Months	Genetic_Markers	Family_History	Country
1	PID00001	58	F	Retinoblastoma	Left	2019-01-25	Stage IV	Radiation	False	15	3	Deceased	85	None	True	UK
2	PID00002	15	Other	Retinoblastoma	Right	2021-10-21	Stage III	Chemotherapy	True	69	6	In Remission	10	None	True	Japan
3	PID00003	64	M	Retinoblastoma	Bilateral	2021-03-12	Stage IV	Surgery	False	47	6	In Remission	3	BRAF Mutation	False	UK
4	PID00004	33	M	Melanoma	Right	2021-05-10	Stage II	Radiation	True	36	6	Active	40	None	False	Canada
5	PID00005	8	Other	Lymphoma	Left	2019-11-24	Stage I	Chemotherapy	False	14	14	In Remission	26	BRAF Mutation	True	USA

The FREQ Procedure

Death Event				
event	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	3290	65.80	3290	65.80
1	1710	34.20	5000	100.00

Cancer_Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Lymphoma	1637	32.74	1637	32.74
Melanoma	1691	33.82	3328	66.56
Retinoblastoma	1672	33.44	5000	100.00

Stage_at_Diagnosis	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Stage I	1190	23.80	1190	23.80
Stage II	1287	25.74	2477	49.54
Stage III	1281	25.62	3758	75.16
Stage IV	1242	24.84	5000	100.00

Treatment_Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Chemotherapy	1665	33.30	1665	33.30
Radiation	1656	33.12	3321	66.42
Surgery	1679	33.58	5000	100.00

The MEANS Procedure

Death Event	N Obs	Variable	Label	N	Mean	Std Dev	Minimum	Maximum
0	3290	Age		3290	44.8282675	26.0312274	1.0000000	90.0000000
		Survival_Time_Months	Survival Time (Months)	3290	60.3346505	34.6312835	1.0000000	120.0000000
		Radiation_Therapy		3290	35.5699088	20.4870197	0	70.0000000
		Chemotherapy		3290	10.0079027	5.9954831	0	20.0000000
1	1710	Age		1710	45.3538012	25.7511036	1.0000000	90.0000000
		Survival_Time_Months	Survival Time (Months)	1710	61.4497076	34.3906173	1.0000000	120.0000000
		Radiation_Therapy		1710	35.2122807	20.7203784	0	70.0000000
		Chemotherapy		1710	10.1526316	6.0992936	0	20.0000000

There were a total of 5,000 patients diagnosed with eye cancer. Out of this, 1,710 (34.2%) experienced a death event during follow-up. The remaining 3,290 (65.8%) patients were censored. The distribution of

cancer types was well-balanced: melanoma (33.8%), retinoblastoma (33.4%), and lymphoma (32.7%). Cancer stages were also evenly distributed across the cohort: Stage I accounted for 23.8% of patients, Stage II for 25.7%, Stage III for 25.6%, and Stage IV for 24.8%. Treatment allocation was similarly balanced, with surgery administered to 33.6% of patients, radiation to 33.1%, and chemotherapy to 33.3%. The mean age of the cohort was 45.0 years, and this value was similar between patients who died and those who were censored. Treatment intensity showed a mean radiation dose of 35.4 Gray, ranging from 0 to 70 Gy, while chemotherapy averaged 10.1 cycles, ranging from 0 to 20 cycles. The mean overall survival time was 60.7 months, equivalent to approximately 5.1 years. I then fitted cox proportional hazard model with `Survival_Time_Months` as my response variable using `PHREG` procedure in SAS.

Results

```
/* Simple Cox Proportional Hazards Model with Stage only */
proc phreg data=ec_cox_analysis;
    class Stage_at_Diagnosis (ref='Stage I');

    model Survival_Time_Months * event(0) =
          Stage_at_Diagnosis;

    hazardratio Stage_at_Diagnosis;

    title 'Simple Cox Model: Effect of Cancer Stage on Mortality';
run;

/* Multiple Cox Proportional Hazards Model */
proc phreg data=ec_cox_analysis;
        class Cancer_Type (ref='Melanoma')
              stage_at_Diagnosis (ref='Stage I')
              Treatment_Type (ref='Surgery')
              age_cat (ref='1');

        model Survival_Time_Months * event(0) =
                  Age
                  Cancer_Type
                  Stage_at_Diagnosis
                  Treatment_Type
                  Radiation_Therapy
                  Chemotherapy
                  braf
                  family_hx;

        /* Hazard ratios for key variables */
        hazardratio Cancer_Type;
        hazardratio Stage_at_Diagnosis;
        hazardratio Treatment_Type;

        title 'Cox Proportional Hazards Model for Eye Cancer Mortality';
run;
```

Simple Cox Model: Effect of Cancer Stage on Mortality

The PHREG Procedure

Model Information		
Data Set	WORK.EC_COX_ANALYSIS	
Dependent Variable	Survival_Time_Months	Survival Time (Months)
Censoring Variable	event	Death Event
Censoring Value(s)	0	
Ties Handling	BRESLOW	

Number of Observations Read	5000
Number of Observations Used	5000

Class Level Information				
Class	Value	Design Variables		
Stage_at_Diagnosis	Stage I	0	0	0
	Stage II	1	0	0
	Stage III	0	1	0
	Stage IV	0	0	1

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
5000	1710	3290	65.80

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

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	25771.813	25768.078
AIC	25771.813	25774.078
SBC	25771.813	25790.411

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	3.7344	3	0.2916
Score	3.7957	3	0.2844
Wald	3.7922	3	0.2848

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Stage_at_Diagnosis	3	3.7922	0.2848

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Stage_at_Diagnosis	Stage II	1	-0.09066	0.06830	1.7619	0.1844	0.913	Stage_at_Diagnosis Stage II
Stage_at_Diagnosis	Stage III	1	-0.10452	0.06872	2.3132	0.1283	0.901	Stage_at_Diagnosis Stage III
Stage_at_Diagnosis	Stage IV	1	-0.12106	0.06797	3.1724	0.0749	0.886	Stage_at_Diagnosis Stage IV

Hazard Ratios for Stage_at_Diagnosis			
Description	Point Estimate	95% Wald Confidence Limits	
Stage_at_Diagnosis Stage I vs Stage II	1.095	0.958	1.252
Stage_at_Diagnosis Stage I vs Stage III	1.110	0.970	1.270
Stage_at_Diagnosis Stage I vs Stage IV	1.129	0.988	1.290
Stage_at_Diagnosis Stage II vs Stage III	1.014	0.886	1.161
Stage_at_Diagnosis Stage II vs Stage IV	1.031	0.902	1.178
Stage_at_Diagnosis Stage III vs Stage IV	1.017	0.889	1.163

Cox Proportional Hazards Model for Eye Cancer Mortality

The PHREG Procedure

Model Information		
Data Set	WORK.EC_COX_ANALYSIS	
Dependent Variable	Survival_Time_Months	Survival Time (Months)
Censoring Variable	event	Death Event
Censoring Value(s)	0	
Ties Handling	BRESLOW	

Number of Observations Read	5000
Number of Observations Used	5000

Class Level Information				
Class	Value	Design Variables		
Cancer_Type	Lymphoma	1	0	
	Melanoma	0	0	
	Retinoblastoma	0	1	
Stage_at_Diagnosis	Stage I	0	0	0
	Stage II	1	0	0
	Stage III	0	1	0
	Stage IV	0	0	1
Treatment_Type	Chemotherapy	1	0	
	Radiation	0	1	
	Surgery	0	0	

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
5000	1710	3290	65.80

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

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	25771.813	25761.796
AIC	25771.813	25785.796
SBC	25771.813	25851.127

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	10.0170	12	0.6145
Score	10.0904	12	0.6080
Wald	10.0846	12	0.6085

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Age	1	0.0351	0.8514
Cancer_Type	2	0.6693	0.7156
Stage_at_Diagnosis	3	3.6082	0.3070
Treatment_Type	2	2.2185	0.3298
Radiation_Therapy	1	1.7735	0.1829
Chemotherapy	1	0.0694	0.7922
braf	1	1.2787	0.2581
family_hx	1	0.3686	0.5438

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Age		1	0.0001769	0.0009444	0.0351	0.8514	1.000	
Cancer_Type	Lymphoma	1	-0.02132	0.05916	0.1298	0.7186	0.979	Cancer_Type Lymphoma
Cancer_Type	Retinoblastoma	1	-0.04824	0.05904	0.6676	0.4139	0.953	Cancer_Type Retinoblastoma
Stage_at_Diagnosis	Stage II	1	-0.08876	0.06845	1.6815	0.1947	0.915	Stage_at_Diagnosis Stage II
Stage_at_Diagnosis	Stage III	1	-0.10324	0.06882	2.2502	0.1336	0.902	Stage_at_Diagnosis Stage III
Stage_at_Diagnosis	Stage IV	1	-0.11769	0.06810	2.9867	0.0840	0.889	Stage_at_Diagnosis Stage IV
Treatment_Type	Chemotherapy	1	0.07053	0.05884	1.4367	0.2307	1.073	Treatment_Type Chemotherapy
Treatment_Type	Radiation	1	-0.00983	0.05993	0.0269	0.8697	0.990	Treatment_Type Radiation
Radiation_Therapy		1	-0.00156	0.00117	1.7735	0.1829	0.998	
Chemotherapy		1	-0.00105	0.00400	0.0694	0.7922	0.999	
braf		1	-0.05479	0.04845	1.2787	0.2581	0.947	BRAF Mutation
family_hx		1	-0.02941	0.04844	0.3686	0.5438	0.971	

Hazard Ratios for Cancer_Type			
Description	Point Estimate	95% Wald Confidence Limits	
Cancer_Type Lymphoma vs Melanoma	0.979	0.872	1.099
Cancer_Type Lymphoma vs Retinoblastoma	1.027	0.913	1.155
Cancer_Type Melanoma vs Retinoblastoma	1.049	0.935	1.178

Hazard Ratios for Stage_at_Diagnosis			
Description	Point Estimate	95% Wald Confidence Limits	
Stage_at_Diagnosis Stage I vs Stage II	1.093	0.956	1.250
Stage_at_Diagnosis Stage I vs Stage III	1.109	0.969	1.269
Stage_at_Diagnosis Stage I vs Stage IV	1.125	0.984	1.286
Stage_at_Diagnosis Stage II vs Stage III	1.015	0.886	1.161
Stage_at_Diagnosis Stage II vs Stage IV	1.029	0.900	1.177
Stage_at_Diagnosis Stage III vs Stage IV	1.015	0.887	1.161

Hazard Ratios for Treatment_Type			
Description	Point Estimate	95% Wald Confidence Limits	
Treatment_Type Chemotherapy vs Radiation	1.084	0.965	1.217
Treatment_Type Chemotherapy vs Surgery	1.073	0.956	1.204
Treatment_Type Radiation vs Surgery	0.990	0.880	1.114

Univariate Cox Model

From `Testing Global Null Hypothesis` table, Likelihood Ratio Test indicates $\chi^2 = 3.7344$, df = 3, p = 0.2916. This shows that `Stage` alone does not significantly improve model fit.
The hazard ratios are

- Stage II vs Stage I: HR = 0.913 (95% CI: 0.958 - 1.252, p = 0.1844)
- Stage III vs Stage I: HR = 0.901 (95% CI: 0.970 - 1.270, p = 0.1283)
- Stage IV vs Stage I: HR = 0.886 (95% CI: 0.988 - 1.290, p = 0.0749)

All hazard ratios fall within a narrow range near 1.0 (0.886–1.129), and every confidence interval includes 1, indicating no statistically significant associations. Although the point estimates for advanced stages are slightly below 1, these differences are not meaningful from a statistical perspective.

Multivariate Cox Model

Similar to the simple Cox model, all hazard ratios remain close to 1, and their confidence intervals include 1, indicating that none of the parameters are statistically significant.

Conclusion

My analysis of these 5,000 eye cancer patients' data showed that either cancer stage alone or a large set of clinical and demographic variables were not strong predictors of mortality. All hazard ratios were close to 1.0, with non-significant p-values, which indicates no real survival differences according to stage, age,

treatment type, or genetic factors. The data are interpreted to represent either uniformly effective care, unmeasured prognostic variables, or limitations in the dataset itself.