STAT 410 REVIEW FOR FIRST MIDTERM EXAM

Problem 1. A clinical trial is conducted to compare the efficacy of two drugs (A and B) in reduction of excess body weight. Drug (A or B), age, and gender were recorded at the baseline. The percent excess body weight loss (EWL) was recorded 6 months into the study. Study the included R script and output to answer the questions below.

(a) Discuss normality of the response variable. Write down the fitted model. Specify all parameters.
(b) Discuss the overall model fit and significance of regression slopes. Interpret only the significant estimated regression coefficients.
(c) Use the fitted model to predict the percent decrease in excess body weight loss for a 30-year-old female who is taking drug B.

Problem 2. Investigators at a large medical center conducted a quality improvement (QI) study which consisted of a six-month-long series of seminars and practical instructional tools on how to improve quality assurance for future projects at this center. Data were collected on participants' designation (nurse/doctor/staff), years of work at the center, whether had a prior experience with QI projects, and the score on the knowledge test taken at the end of the study. The score was measured on a one-hundred point scale with larger values indicating better knowledge. Study the included SAS code and output to answer the questions that follow.

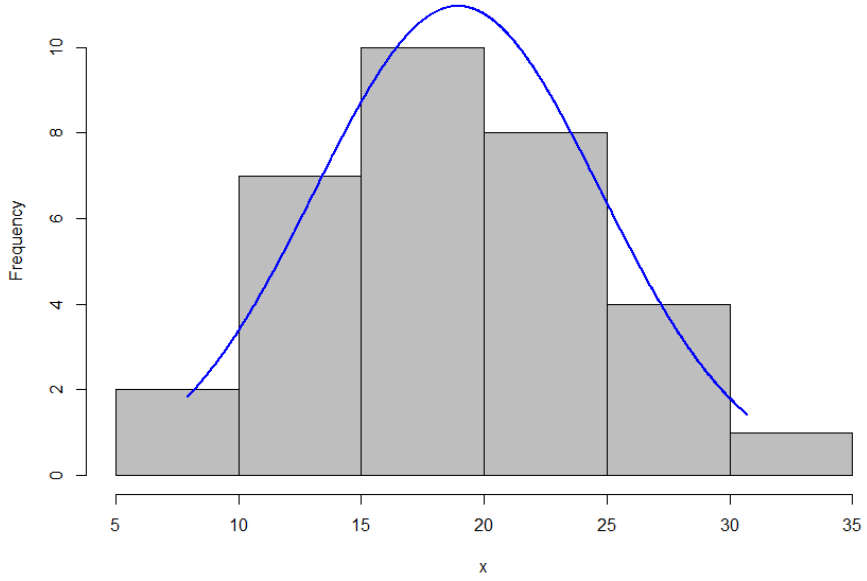
(a) What transformation was used to model the data? Give the name. Write down the explicit formula for the transformation. Explain how the formula was obtained.
(b) Write down the fitted model. What predictors are significant at the 5% level?
(c) How good is the model fit?
(d) Interpret the significant estimated regression coefficients.
(e) Predict the score for a doctor who has worked at the center for 15 years and who had previously been a co-PI on a grant that involved quality assurance component. Compute by hand and compare to the SAS output.
(f) What other model was used for the analysis? Give the name. Write down the fitted regression model. Estimate all parameters.
(g) Discuss significance of the regression coefficients at the 5% level. Discuss the overall goodness of fit of the model.
(h) Interpret the estimated regression coefficients only for significant predictors.
(i) Predict the score for a doctor who has worked at the center for 15 years and who had previously been a co-PI on a grant that involved quality assurance component. Compute by hand and compare to the SAS output.

REVIEW FOR FIRST MIDTERM EXAM

CODES AND OUTPUTS

Problem 1.

```
> weightloss.data<- read.csv(file="./pr1data.csv", header=TRUE,
sep = ",")

> library(rcompanion)
> plotNormalHistogram(weightloss.data$EWL)
```

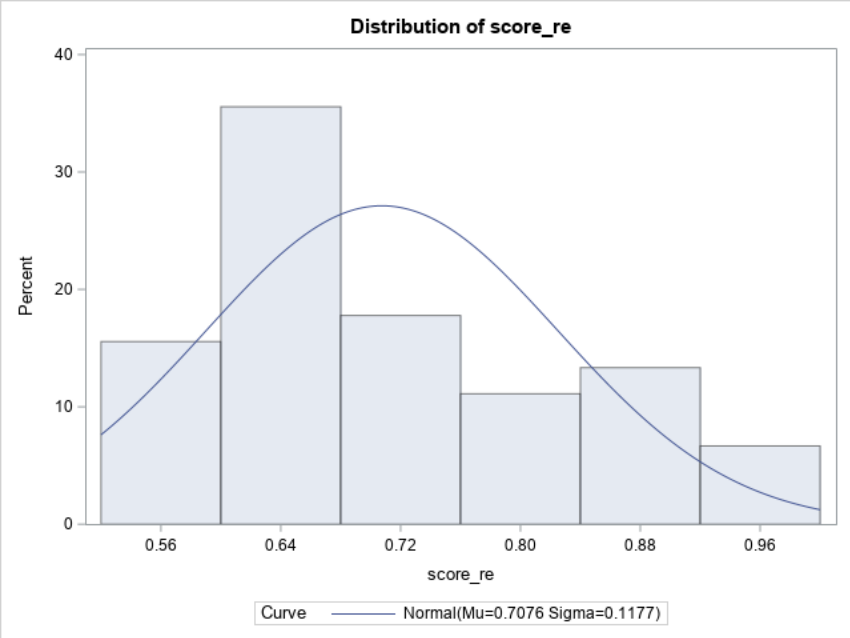



```
> shapiro.test(weightloss.data$EWL)

        Shapiro-Wilk normality test

data:  weightloss.data$EWL
W = 0.97608, p-value = 0.6803

> drug.rel<- relevel(weightloss.data$drug, ref="A")
> gender.rel<- relevel(weightloss.data$gender, ref="M")

> summary(fitted.model<- glm(EWL ~ drug.rel + age + gender.rel,
data = weightloss.data, family=gaussian(link=identity)))
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.1899     5.6573   1.801   0.0825
drug.relB     4.2034     1.9472   2.159   0.0396
age           0.1162     0.1111   1.046   0.3046
gender.relF   2.5593     1.9437   1.317   0.1986

> sigma(fitted.model)

[1] 5.462401

> null.model<- glm(EWL~1, data=weightloss.data, family=gaussian(
link=identity))
> print(deviance<- -2*(logLik(null.model)-logLik(fitted.model)))

'log Lik.' 7.341607

> print(p.value<- pchisq(deviance, df=3, lower.tail=FALSE))

'log Lik.' 0.061771

> print(predict(fitted.model,data.frame(drug.rel="B", age=30, ge
nder.rel="F")))

20.43717
```

Problem 2.

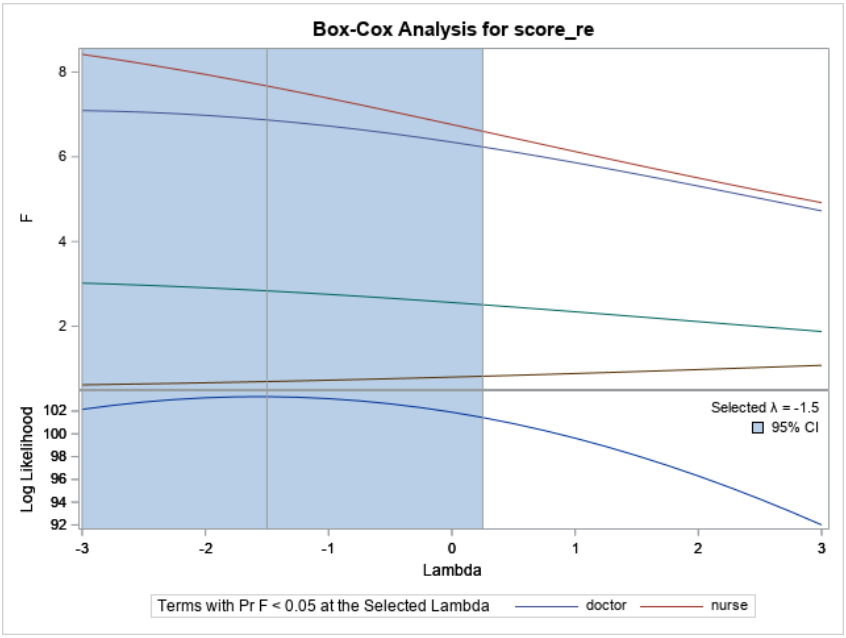
```
proc import out=QIscore datafile='./pr2data.csv' dbms=csv
replace;
run;

data QIscore;
set QIscore;
 score_re=score/100;
run;

proc univariate data=QIscore;
 var score_re;
  histogram/normal;
run;
```



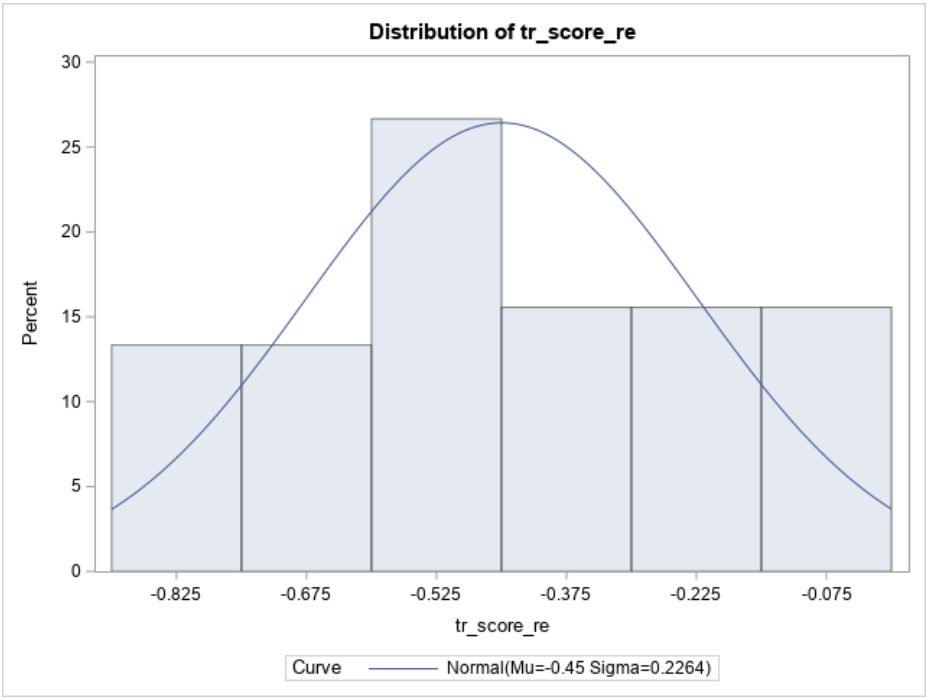
Goodness-of-Fit Tests for Normal Distribution

Test		Statistic		p Value
Kolmogorov-Smirnov	**D**	0.15929952	**Pr > D**	<0.010
Cramer-von Mises	**W-Sq**	0.20886687	**Pr > W-Sq**	<0.005
Anderson-Darling	**A-Sq**	1.19454731	**Pr > A-Sq**	<0.005

```
data QIscore;
set QIscore;
doctor=(desgn='doctor');
nurse=(desgn='nurse');
priorQIyes=(priorQI='yes');
run;

proc transreg;
model BoxCox(score_re)=identity(doctor nurse priorQIyes wrkyrs);
run;
```




```
data QIscore;
 set QIscore;
  tr_score_re=1-(1/score_re);
run;

proc univariate;
 var tr_score_re;
  histogram /normal;
  run;
```



Goodness-of-Fit Tests for Normal Distribution

Test		Statistic		p Value
Kolmogorov-Smirnov	**D**	0.09928641	**Pr > D**	>0.150
Cramer-von Mises	**W-Sq**	0.08343724	**Pr > W-Sq**	0.189
Anderson-Darling	**A-Sq**	0.53550113	**Pr > A-Sq**	0.168

```
data predict;
input desgn$ priorQI$ wrkyrs;
cards;
doctor yes 15
;

data QIscore;
set QIscore predict;
run;

proc genmod data=QIscore;
class desgn priorQI (ref='no');
 model tr_score_re=desgn priorQI wrkyrs/dist=normal
link=identity;
  output out=outdata p=ptrscore_re;
run;
```

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.6437	0.0886	-0.8174	-0.4701	52.79	<.0001
desgn	**doctor**	1	0.2369	0.0861	0.0681	0.4057	7.57	0.0059
desgn	**nurse**	1	0.2364	0.0820	0.0756	0.3972	8.30	0.0040
desgn	**staff**	0	0.0000	0.0000	0.0000	0.0000	.	.
priorQI	**yes**	1	0.1132	0.0643	-0.0129	0.2392	3.10	0.0784
priorQI	**no**	0	0.0000	0.0000	0.0000	0.0000	.	.
wrkyrs		1	-0.0035	0.0038	-0.0110	0.0041	0.82	0.3657
Scale		1	0.1955	0.0206	0.1590	0.2404		

```
data outdata;
 set outdata;
```

```
 pred_score=100/(1-ptrscore_re);
run;

proc print data=outdata (firstobs=46) noobs;
var pred_score;
run;
```

pred_score
74.3040

```
proc genmod data=QIscore;
 model tr_score_re=/dist=normal link=identity;
run;

data deviance_test;
deviance=-2*(3.4961-(9.5903));
pvalue=1-probchi(deviance,4);
run;

proc print data=deviance_test;
run;
```

deviance	pvalue
12.1884	**0.016004**

```
proc genmod data=QIscore;
class desgn priorQI (ref='no');
 model score_re=desgn priorQI wrkyrs/dist=gamma link=log;
  output out=outdata p=pscore_re;
run;
```

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.4840	0.0635	-0.6084	-0.3596	58.16	<.0001
desgn	**doctor**	1	0.1714	0.0620	0.0499	0.2929	7.65	0.0057
desgn	**nurse**	1	0.1698	0.0591	0.0540	0.2857	8.26	0.0041
desgn	**staff**	0	0.0000	0.0000	0.0000	0.0000	.	.
priorQI	**yes**	1	0.0757	0.0466	-0.0156	0.1670	2.64	0.1042
priorQI	**no**	0	0.0000	0.0000	0.0000	0.0000	.	.
wrkyrs		1	-0.0028	0.0027	-0.0081	0.0026	1.04	0.3067

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Scale	1	50.4777	10.6067	33.4380	76.2008		

```
data outdata;
set outdata;
pred_score=100*pscore_re;
run;

proc print data=outdata (firstobs=46) noobs;
 var pred_score;
run;
```

pred_score
75.6695

```
proc genmod data=QIscore;
 model score_re=/dist=gamma link=log;
run;

data deviance_test;
deviance=-2*(34.3723-(40.3845));
pvalue=1-probchi(deviance,4);
run;

proc print data=deviance_test noobs;
run;
```

deviance	pvalue
12.0244	0.017171