```
/* Montgomery 3.24 data */
data mont3_24;
INPUT noise design @@;
datalines;
19 1
20 1
19 1
30 1
8 1
80 2
61 2
73 2
56 2
80 2
47 3
26 3
25 3
35 3
50 3
95 4
46 4
83 4
78 4
97 4
;
run;

/* take a look at the data */
title 'Plot of Circuit Design vs Noise';
proc sgplot data=mont3_24;
scatter y=noise x=design;
run;

 /* run ANOVA analysis */
title 'ANOVA analysis';
proc glm data=mont3_24;
class design;
model noise=design;
output out=diag p=pred r=res;
run;

title 'Normality checking';
proc univariate data=diag normal;
var res;
qqplot res/normal(mu=est sigma=est);
run;

/* outlier checking */
title 'Outlier';
data outlier;
set diag;
stdres=res/sqrt(184.3);
run;

proc print data=outlier;
run;

/* independence checking */
title 'Plot of residuals vs design';
proc sgplot data=diag;
scatter y=res x=design;
refline 0;
run;


 /* constant variance checking */;
title 'constant variance checking';
proc glm data=mont3_24;
class design;
model noise=design;
means design / hovtest=bartlett hovtest=levene;
output out=diag1 p=pred r=res;
run;

proc sgplot data=diag1;
scatter x=pred y=res;
refline 0;
run;


/* test design 1 vs 3 */

data mont3_24a;
INPUT noise design @@;
datalines;
19 1
20 1
19 1
30 1
8 1
47 3
26 3
25 3
35 3
50 3

;
run;

proc ttest data=mont3_24a alpha=0.05;
  class design;
  var noise;
```

3.24. Four different designs for a digital computer circuit are being studied to compare the amount of noise present. The following data have been obtained:

Circuit Design	Noise Observed				
1	19	20	19	30	8
2	80	61	73	56	80
3	47	26	25	35	50
4	95	46	83	78	97

(**a**) Is the same amount of noise present for all four designs? Use $\alpha = 0.05$.

(**b**) Analyze the residuals from this experiment. Are the analysis of variance assumptions satisfied?

(**c**) Which circuit design would you select for use? Low noise is best.

b.) The resudial plots above indicate no significant deviations from normality in the residuals, which is important for the validity of our hypothesis testing on data as-is. Without normality for the residuals, we would need to perform a transformation of the data to achieve normality to do our analyses. Graphically, there appears to be one possible outlier, but an objective approach in the "Outliers" analysis above indicates that observation in question, number 17, has a standardized residual of $|-2.49| < 3$, which indicates it is not a real statistical outlier. Lastly, Levene's test and Bartlett's Test for Homogeneity both indicate that it is acceptable to conclude constant variance (P values 0.273 and 0.297, respectively are significant), therefore all conditions for analysis are met.

c.) Clearly design 1 is the best design as compared to designs 2 and 4. However, since

8

Alex Salce Homework 2 - STAT571B Spring 2023 February 15th, 2023

our ANOVA only test the hypothesis that at least one of the means is different, we do not have a direct comparison from the analysis above of design 1 to design 3.

```
run;
```

$\mu_1 - \mu_3$ t-test of difference in means, $H_0: \mu_1 - \mu_3 = 0$, $H_a: \mu_1 - \mu_3 \neq 0$.

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	8	-2.79	0.0237
Satterthwaite	Unequal	7.0024	-2.79	0.0270

For a significance level of $\alpha = 0.05$, we would reject the null hypothesis that the means are the same based upon P value here of 0.0237, therefore design 1 is objectively the best choice for lowest level of noise.

```
run;
```

```
/* import data */

PROC IMPORT DATAFILE="/home/u63048916/STAT571B/Homework/Homework 2/Q3.28.csv"
    DBMS=CSV
    OUT=mont3_28
    REPLACE;
    GETNAMES=YES;
RUN;


/* take a look at the data */
title 'Plot of Material vs Time';
proc sgplot data=mont3_28;
scatter y=Time x=Material;
run;

 /* run ANOVA analysis */
title 'ANOVA analysis';
proc glm data=mont3_28;
class Material;
model Time=Material;
output out=diag p=pred r=res;
run;

title 'Normality checking';
proc univariate data=diag normal;
var res;
qqplot res/normal(mu=est sigma=est);
run;

/* outlier checking */
title 'Outlier';
data outlier;
set diag;
stdres=res/sqrt(4167043.8);
run;

proc print data=outlier;
run;

/* independence checking */
title 'Plot of residuals vs Material';
proc sgplot data=diag;
scatter y=res x=Material;
refline 0;
run;


 /* constant variance checking */;
title 'constant variance checking';
proc glm data=mont3_28;
class Material;
model Time=Material;
means Material / hovtest=bartlett hovtest=levene;
output out=diag1 p=pred r=res;
run;

proc sgplot data=diag1;
scatter x=pred y=res;
refline 0;
run;


 /* if the constant variance assumption does not hold, use Welch ANOVA, see the code below */;
 title 'Welch';
PROC anova data=mont3_28;
CLASS Material;
MODEL Time=Material;
MEANS Material / WELCH;
RUN;


/* ************************************* */

/* part2: if non-constant variance, do transformation: boxcox */


PROC IMPORT DATAFILE="/home/u63048916/STAT571B/Homework/Homework 2/Q3.28.csv"
    DBMS=CSV
    OUT=two
    REPLACE;
    GETNAMES=YES;
RUN;

proc glm data=two;
class Material;
model Time=Material;
output out=diag2 p=pred r=res;
run;

title 'Residual Plot';
proc sgplot data=diag2;
scatter y=res x=pred;
refline 0;
run;


/* find an appropriate transformation parameter */
proc transreg data=diag2;
model boxcox(Time/convenient lambda=-2.0 to 2.0 by 0.1)=class(Material);
run;
```

3.28. An experiment was performed to investigate the effectiveness of five insulating materials. Four samples of each material were tested at an elevated voltage level to accelerate the time to failure. The failure times (in minutes) are shown below:

Material	Failure Time (minutes)			
1	110	157	194	178
2	1	2	4	18
3	880	1256	5276	4355
4	495	7040	5307	10,050
5	7	5	29	2

(a) Do all five materials have the same effect on mean failure time?

(b) Plot the residuals versus the predicted response. Construct a normal probability plot of the residuals. What information is conveyed by these plots?

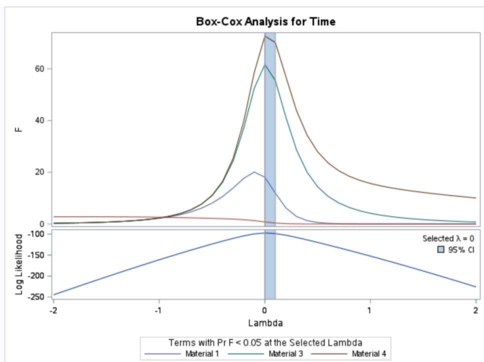
(c) Based on your answer to part (b) conduct another analysis of the failure time data and draw appropriate conclusions.

```
/* do transformation on the Timeonse */
data new;
set two;
newY=log(Time);
run;

/* run a new ANOVA analysis on the transformed data */
proc glm data=new;
class Material;
model newY=Material;
output out=newout p=pred r=res;
run;

/* check the constant variance for the new result */
title 'New: Residual Plot';
proc sgplot data=newout;
scatter y=res x=pred;
refline 0;
run;

title 'New: variance checking';
proc glm data=newout;
class Material;
model newY=Material;
means Material / hovtest=bartlett hovtest=levene:
```



Box-Cox shows $\lambda = 0$, so log is appropriate transformation. Rerunning data with log transform (note transformed data is $newY$).

4	4	8.21400778	1.36487123
5	4	1.90394777	1.10994332

c.) Using the Box Cox transformation method we find the appropriate parameter to be $\lambda = 0$, which implies a $log(Time)$ transformation to be appropriate for the data. We can see that this transformation reconciles the problems with the original data's residuals, and now show qualitative (residual plot) and objective (Levene and Bartlett tests) improvement to homogeneous variance. Therefore the results of this ANOVA can now be correctly interpreted to show that for a significance level of $\alpha = 0.05$, our P value for our computed F Value of 37.66 is $< .0001$ and is significant, so concluding at least one of the Materials are different is now appropriate. -1, need to assess ANOVA assumptions

```
/* import data */

PROC IMPORT DATAFILE="/home/u63048916/STAT571B/Homework/Homework 2/Q3.28.csv"
    DBMS=CSV
    OUT=mont3_28_PR5
    REPLACE;
    GETNAMES=YES;
RUN;

title 'Non-parametric analysis';
proc npar1way data=mont3_28_PR5;
class Material;
var Time;
run;
```

5. **Perform a nonparametric analysis on the dataset Q3.28.csv and draw a conclusion.**

Reference all results from problem 4. **Montgomery 3.28**. Since our residuals for the original data exhibit properties of nonhomogeneous variance, we can use a nonparametric analysis to test the hypothesis

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$H_1: \text{ at least one "=" does not hold}$$

Wilcoxon Scores (Rank Sums) for Variable Time Classified by Variable Material

Material	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	4	42.00	42.0	10.579026	10.5000
2	4	14.50	42.0	10.579026	3.6250
3	4	62.00	42.0	10.579026	15.5000
4	4	70.00	42.0	10.579026	17.5000
5	4	21.50	42.0	10.579026	5.3750

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square	DF	Pr > ChiSq
16.8734	4	0.0020

Taking a significance level of $\alpha = 0.05$, we can see that the Kruskal-Wallis Test has H score of 16.8734 with an associated P value of 0.0020, less than α, therefore we reject H_0 and conclude that at least one Material is different.