Program Summary - Take-Home Test.sas

Execution Environment

Author:	chwang10
File:	/home/chwang10/Take-Home Test.sas
SAS Platform:	Linux LIN X64 3.10.0-1062.9.1.el7.x86_64
SAS Host:	ODAWS04-USW2.ODA.SAS.COM
SAS Version:	9.04.01M6P11072018
SAS Locale:	en_US
Submission Time:	11/23/2020, 7:53:11 PM
Browser Host:	ASTOUND-66-234-210-119.CA.ASTOUND.NET
User Agent:	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_6) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/86.0.4240.198 Safari/537.36
Application Server:	ODAMID01-USW2.ODA.SAS.COM

Code: Take-Home Test.sas

```
* Programmed by Charles Hwang *
* Coded in SAS OnDemand       *
* Monday, November 23, 2020   *
* Course: STAT 403            *
* Title: Take-Home Test       *;

/* 1a */ Proc Format;
Value Status 0='On-Time' 1='Delayed';
Run;

/* 1b */ Data Delays;
/* 1b(i) */ Length Airline$ 6 City1 - City10$ 3;
Infile "/home/chwang10/delays.txt";
Input City1 - City10$;
Input Status1 - Status10;
Input Count1 - Count10;
/* 1b(ii) */ Array a[*] City1 - City10; * Array placeholder variables must be unique ;
Array b[*] Status1 - Status10;
Array c[*] Count1 - Count10;
Do i = 1 to dim(a);
If Count1=497 or Count=62 or Count=221 or Count=12 or Count=212 or Count=20 or Count=503 or Count=102 or Count=1841 or Count=1
else Airline='AMBEST'; * how else to add Airline variable? ;
City=a[i];
Status=b[i];
Count=c[i];
Output;
End;
Drop i City1 - City10 Status1 - Status10 Count1 - Count10;
Run;

/* 1c */ Data DelaysRe;
Set Delays;
Format Status Status.;
Run;

/* 1d */ Axis label=(a=90 "Number of Flights");
Proc Gchart data=DelaysRe;
Title "1d. Flight Status by Airline";
Vbar Status /sumvar=Count group=Airline axis=axis discrete;
Run; * America's Best have far more on-time flights than TWB, but also have more delayed flights than
TWB because they have more flights overall. ;

/* 1e(i) */ * H0: There is no relationship between airline and flight status.
HA: There is a relationship between airline and flight status. ;
/* 1e(ii) */ Proc Freq data=DelaysRe;
Title "1e(ii). Chi-Squared Test of Independence on Status by Airline";
Table Status*Airline /chisq oddsratio;
Weight Count;
Run;
/* 1e(iii) */ * χ = 13.5717, p = 0.0002 ;
/* 1e(iv) */ * We reject H0 at the α = .05 level. There is sufficient evidence that there is a
relationship between airline and flight status. ;
/* 1e(v) */ * The calculated odds ratio is 1.2518. This means it is approximately 25.18 percent more
likely to have a delayed flight on TWB than America's Best. We are 95 percent confident the odds ratio
is between 1.1106 and 1.4109. The odds ratio is significant because its confidence interval does not
include 1, indicating the odds are only moving in a single direction. ;
```

```
/* 1f(i) */ Proc Sort data=DelaysRe;
By City;
/* 1f(ii) */ Proc Freq data=DelaysRe;
Title "1f(ii). Chi-Squared Test of Independence on Status by Airline, Grouped by Airport";
Table Status*Airline /chisq;
By City;
Weight Count;
Output out=p pchi;
Run;
/* 1f(iii) */ * The results do not appear to completely agree with the results in problem 1e. Airports
from San Francisco (p = 0.0277235077), Seattle (p = 0.0000040883999), and San Diego (p = 0.0001132093)
are significant at the α = .05 level, while airports from Los Angeles (p = 0.0718151204) and
Phoenix (p = 0.1255718036) are not. America's Best had more flights at airports in Los
Angeles (811 vs. 559), Phoenix (5255 vs. 233), and San Francisco (448 vs. 232), while TWB had more
flights at airports in San Diego (262 vs. 2146) and Seattle (449 vs. 605). Although the sample size
assumption for chi-squared tests was met in all cells, some tests may be less robust in certain cells
if there is a large difference in flights between the two airlines at an airport. ;
/* 1f(iv) */ * Although America's Best had a greater proportion of delayed flights at each of the
airports in the data, TWB had the greater proportion of delayed flights overall, and the difference
between that proportion and the proportion of delayed America's Best flights was found to be
significant. The large difference in flights between the two airlines at certain airports causes the
proportions to be weighted differently, but because the difference in the overall number of flights
between the two airports is not large, this weighting disappears in the test on the full dataset. ;

/* 2a */ Data SoggyChips;
Infile "/home/chwang10/SoggyChips2020.csv" dlm=',' firstobs=2; * Skipping header row, starting on row 2 ;
Input Time Moisture;
Run;

/* 2b */ Proc Corr data=SoggyChips nosimple; * Descriptive statistics not needed ;
Title "2b. Correlation between Moisture Content and Frying Time";
Var Moisture Time;
Run;
/* 2b(i) */ * The correlation coefficient between moisture content and frying time is -0.77149. The
relationship does appear to be strong. ;
/* 2b(ii) */ * H0: ρ = 0, HA: ρ ≠ 0 ;
/* 2b(iii) */ * p < .0001, We reject H0 at the α = .05 level. There is sufficient evidence that the
correlation is a nonzero value. ;

/* 2c */ Symbol value=circle;
Proc Gplot data=SoggyChips;
Title "2c. Moisture Content vs. Frying Time";
Plot Moisture*Time;
Run; * The relationship is moderately strong and clearly nonlinear. ;

/* 2d */ Proc Reg data=SoggyChips;
Title "2d. Linear Regression of Soggy Chips Data";
Model Moisture=Time; * Linear model: Moisture = 13.34756 - 0.25880*Time ;
Run; * TSS = 805.95958 ;
/* 2d(i) */ * r^2 = 0.5768, Approximately 57.68 percent of the variation in the data is explained by
the linear model. ;
/* 2d(ii) */ * The linearity and constant variance assumptions are violated. From the scatterplot in
problem 2c, the data are clearly not linear, and the residual plot and studentized residual plot
indicate a clear lack of homoscedasticity. ;

/* 2e */ Data SoggyChipsPower;
Set SoggyChips;
MoisturePower=log(Moisture);
TimePower=log(Time);
Proc Reg data=SoggyChipsPower;
Title "2e. Power Regression of Soggy Chips Data";
Model MoisturePower=TimePower;
Output out=SoggyChipsPowerO predicted=MoisturePowerPred;
Run; * TSS = 18.21285 ;
/* 2e(i) */ * r^2 = 0.9406 ;
/* 2e(ii) */ * According to the residual plot and studentized residual plot, there may be a slight
violation of the homoscedasticity assumption. ;
/* 2e(iii) */ * Power model: Moisture = e^(4.88770 - 1.12055*ln(Time)) ;
/* 2e(iv) */ Data SoggyChipsPowerUn;
Set SoggyChipsPowerO;
MoisturePowerUn=exp(MoisturePowerPred);
Run;

/* 2f */ Proc Transreg data=SoggyChips;
Title "2f. Box-Cox Transformation of Soggy Chips Data";
Model Boxcox(Moisture)=identity(Time);
```

```
Run;
/* 2f(i) */ * λ = -0.75 ;
/* 2f(ii) */ Data SoggyChipsBC;
Set SoggyChips;
MoistureBC=Moisture**-0.75;
Proc Reg data=SoggyChipsBC;
Title "2f(ii). Power Regression of Soggy Chips Data with Box-Cox Transformation";
Model MoistureBC=Time;
Output out=SoggyChipsBCO predicted=MoistureBCPred;
Run; * TSS = 1.3932 ;
/* 2f(iii) */ * r^2 = 0.9666 ;
/* 2f(iv) */ * No, none of the assumptions appear to be violated. ;
/* 2f(v) */ * Power model with Box-Cox transformation: Moisture = (0.75444 - 1.80577*Time)^(-4/3) ;
/* 2f(vi) */ Data SoggyChipsBCUn;
Set SoggyChipsBCO;
MoistureBCUn=MoistureBCPred**(-4/3);
Run;

/* 2g(i) */ Data SoggyChipsM;
Merge SoggyChipsPowerUn SoggyChipsBCUn;
By Time;
Run;
/* 2g(ii) */ Proc Sgplot data=SoggyChipsM;
Title "2g(ii). Scatterplot of Power and Box-Cox Models of Soggy Chips Data";
Keylegend "Power" "Box-Cox" /location=inside position=topright across=1;
Yaxis label="Moisture";
Scatter X=Time Y=Moisture;
Series X=Time Y=MoisturePowerUn /name="Power" legendlabel="Power Model" lineattrs=(color=orange);
Series X=Time Y=MoistureBCUn /name="Box-Cox" legendlabel="Power Model with Box-Cox Transformation" lineattrs=(color=maroon);
Run;
/* 2g(iii) */ * I believe the power model with the Box-Cox transformation is the best of the three. The adjusted-r^2
is the highest of the three models and the total sum of squares (TSS) is by far the lowest. The Box-Cox
Transformation also assures we have the best value of λ. ;
/* 2g(iv) */ * Although the r^2 is very strong in the Box-Cox transformation model, it appears to be
overfitting the data. Because Time is the only independent variable and every three of its values are
the same, the model will predict the same Moisture value for them despite the true value being recorded
differently. There may be additional independent variables not in the dataset. ;

/* 3a */ Proc Format;
Value $Color B='Brown' P='Pink' U='Purple' R='Red';
Run;

/* 3b */ Proc Import out=Lipstick file="/home/chwang10/lipsticklead.xlsx" dbms=xlsx;
Run;

/* 3c */ Data LipstickSep;
Set Lipstick;
/* 3c(i) */ Color=substr(ColorPrice, 1, 1);
Price=substr(ColorPrice, 2, 1);
/* 3c(ii) */ Format Color $Color.;
Run;

/* 3d */ Proc Gchart data=LipstickSep;
Title "3d. Lead Content of Lipstick by Color and Price Category";
Vbar Price /group=Color sumvar=Pb type=mean;
Run;
/* 3d(i) */ * It appears that price category 2 has more lead than price categories 1 and 3. However,
it is difficult to visually discern whether a specific color has more lead in its lipstick than the
other colors. ;

/* 3e */ Proc GLM data=LipstickSep;
Title '3e. ANOVA of Lipstick Data With Interaction Term';
Class Color Price;
Model Pb=Color Price Color*Price;
Means Color Price Color*Price /tukey;
Run;
/* 3e(i) */ * H0(1): μ_B = μ_P = μ_U = μ_R , HA(1): At least one of the means is different
H0(2): μ_1 = μ_2 = μ_3 , HA(2): At least one of the means is different
H0(3): There is no interaction between color and price category , HA(3): There is an interaction between color and price categ
/* 3e(ii) */ * Color: F = 1.95 (not significant at the α = .05 level)
Price category: F = 6.99 (significant at the α = .05 level)
Interaction: F = 0.70 (not significant at the α = .05 level);
/* 3e(iii) */ Proc GLM data=LipstickSep;
Title '3e(iii). ANOVA of Lipstick Data Without Interaction Term';
Class Color Price;
Model Pb=Color Price;
Means Color Price /tukey;
```

```
Run; * According to Tukey's HSD test, price categories 1 and 2 and price categories 2 and 3 have
statistically different means. ;
/* 3f */ * I would recommend avoiding lipsticks in price category 2, as it has a significantly higher
lead content than price categories 1 and 3. This difference was shown to be statistically significant
at the α = .05 level. Additionally, red lipsticks have the lowest mean, median, interquartile
range (IQR), and upper outlier bound (Q3+1.5*IQR) of the four colors. ;

/* 4a */ Proc Import out=DrugCon file='/home/chwang10/drugconcentrationB.xls' dbms=xls;
Run;
/* 4a(i) */ Proc Contents data=DrugCon noprint; * Placeholder code indicating that I opened and read the code on my PC ;
/* 4a(ii) */ Data DrugUni;
Set DrugCon;
Array x[*] conc_1 - conc_49;
Do i = 1 to dim(x);
If i = 1 then Time=1/60; * The initial reading is t = 1 minute ;
else Time=1/2*(i-1); * Expressing time in hours ;
Concentration=x[i];
Output;
End;
Drop i conc_1 - conc_49;
Run;
/* 4a(iii) */ Symbol value=circle;
Proc Gplot data=DrugUni;
Title '4a(iii). Drug Concentration vs. Time';
Plot Concentration*Time;
Run;

/* 4b */ Proc Nlin data=DrugUni plots=fit;
Title '4b. Nonlinear Model of Drug Concentration Data';
Parameters α=1 β=64.8 θ1=0.3 θ2=1.5; * Arbitrary choosing unknown α = 1 and median choice of θ_1 and θ_2 parameters ;
If Time < 8 then Model Concentration=β*(exp(-θ1*Time)-exp(-θ2*Time))/(θ2-θ1);
else Model Concentration=β/(θ2-θ1)*(exp(-θ1*Time)-exp(-θ2*Time)+α*exp(-θ1*(Time-8))-α*exp(-θ2*(Time-8)));
Output out=DrugUniO predicted=ConcPred;
Run;
/* 4b(i) */ * α = 1.1903, θ_1 = 0.2988, θ_2 = 1.5491, The patients received an increase in the dosage
at t = 8 hours. The increase was approximately 19.03 percent. ;
/* 4b(ii) */ Proc Sort data=DrugUniO;
By Time;
Proc Sgplot data=DrugUniO;
Title "4b(ii). Scatterplot of Concentration vs. Time with Predicted Curve Overlay";
Keylegend "P" /location=inside position=topright across=1;
Scatter X=Time Y=Concentration;
Series X=Time Y=ConcPred /name="P" legendlabel="Predicted Curve" lineattrs=(color=green);
```

Log: Take-Home Test.sas

Notes (90)

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
70
71         * Programmed by Charles Hwang *
72         * Coded in SAS OnDemand       *
73         * Monday, November 23, 2020   *
74         * Course: STAT 403            *
75         * Title: Take-Home Test       *;
76
77         /* 1a */
77       !          Proc Format;
78         Value Status 0='On-Time' 1='Delayed';
NOTE: Format STATUS is already on the library WORK.FORMATS.
NOTE: Format STATUS has been output.
79         Run;

NOTE: PROCEDURE FORMAT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              300.81k
      OS Memory           37540.00k
      Timestamp           11/24/2020 03:53:06 AM
      Step Count                        412  Switch Count  0
      Page Faults                       0
      Page Reclaims                     14
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           32


80
```

```
81          /* 1b */
81        !          Data Delays;
82          /* 1b(i) */ Length Airline$ 6 City1 - City10$ 3;
83          Infile "/home/chwang10/delays.txt";
84          Input City1 - City10$;
85          Input Status1 - Status10;
86          Input Count1 - Count10;
87          /* 1b(ii) */ Array a[*] City1 - City10; * Array placeholder variables must be unique ;
88          Array b[*] Status1 - Status10;
89          Array c[*] Count1 - Count10;
90          Do i = 1 to dim(a);
91          If Count1=497 or Count=62 or Count=221 or Count=12 or Count=212 or Count=20 or Count=503 or Count=102 or Count=1841 or
91        ! Count=305 then Airline='TWB';
92          else Airline='AMBEST'; * how else to add Airline variable? ;
93          City=a[i];
94          Status=b[i];
95          Count=c[i];
96          Output;
97          End;
98          Drop i City1 - City10 Status1 - Status10 Count1 - Count10;
99          Run;

NOTE: The infile "/home/chwang10/delays.txt" is:
      Filename=/home/chwang10/delays.txt,
      Owner Name=chwang10,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=18Nov2020:21:57:09,
      File Size (bytes)=242

NOTE: 6 records were read from the infile "/home/chwang10/delays.txt".
      The minimum record length was 38.
      The maximum record length was 40.
NOTE: The data set WORK.DELAYS has 20 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.10 seconds
      user cpu time       0.01 seconds
      system cpu time     0.01 seconds
      memory              1034.68k
      OS Memory           37800.00k
      Timestamp           11/24/2020 03:53:06 AM
      Step Count                        413  Switch Count  2
      Page Faults                       0
      Page Reclaims                     94
      Page Swaps                        0
      Voluntary Context Switches        19
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272


100
101         /* 1c */
101       !          Data DelaysRe;
102         Set Delays;
103         Format Status Status.;
104         Run;

NOTE: There were 20 observations read from the data set WORK.DELAYS.
NOTE: The data set WORK.DELAYSRE has 20 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1056.56k
      OS Memory           38060.00k
      Timestamp           11/24/2020 03:53:06 AM
      Step Count                        414  Switch Count  2
      Page Faults                       0
      Page Reclaims                     126
      Page Swaps                        0
      Voluntary Context Switches        11
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


105
106         /* 1d */ Axis label=(a=90 "Number of Flights");
107         Proc Gchart data=DelaysRe;
108         Title "1d. Flight Status by Airline";
109         Vbar Status /sumvar=Count group=Airline axis=axis discrete;
110         Run;

110       !      * America's Best have far more on-time flights than TWB, but also have more delayed flights than
111         TWB because they have more flights overall. ;
112
113         /* 1e(i) */ * H0: There is no relationship between airline and flight status.
114         HA: There is a relationship between airline and flight status. ;
115         /* 1e(ii) */

NOTE: There were 20 observations read from the data set WORK.DELAYSRE.
NOTE: PROCEDURE GCHART used (Total process time):
```

```
      real time           0.17 seconds
      user cpu time       0.17 seconds
      system cpu time     0.00 seconds
      memory              7749.93k
      OS Memory           40728.00k
      Timestamp           11/24/2020 03:53:06 AM
      Step Count                        415  Switch Count  1
      Page Faults                       0
      Page Reclaims                     819
      Page Swaps                        0
      Voluntary Context Switches        6
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           296


115      !             Proc Freq data=DelaysRe;
116        Title "1e(ii). Chi-Squared Test of Independence on Status by Airline";
117        Table Status*Airline /chisq oddsratio;
118        Weight Count;
119        Run;

NOTE: There were 20 observations read from the data set WORK.DELAYSRE.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.06 seconds
      user cpu time       0.07 seconds
      system cpu time     0.01 seconds
      memory              2905.00k
      OS Memory           39344.00k
      Timestamp           11/24/2020 03:53:06 AM
      Step Count                        416  Switch Count  5
      Page Faults                       0
      Page Reclaims                     192
      Page Swaps                        0
      Voluntary Context Switches        32
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           528


120        /* 1e(iii) */ * χ = 13.5717, p = 0.0002 ;
121        /* 1e(iv) */ * We reject H0 at the α = .05 level. There is sufficient evidence that there is a
122        relationship between airline and flight status. ;
123        /* 1e(v) */ * The calculated odds ratio is 1.2518. This means it is approximately 25.18 percent more
124        likely to have a delayed flight on TWB than America's Best. We are 95 percent confident the odds ratio
125        is between 1.1106 and 1.4109. The odds ratio is significant because its confidence interval does not
126        include 1, indicating the odds are only moving in a single direction. ;
127
128        /* 1f(i) */
128      !             Proc Sort data=DelaysRe;
129        By City;
130        /* 1f(ii) */

NOTE: There were 20 observations read from the data set WORK.DELAYSRE.
NOTE: The data set WORK.DELAYSRE has 20 observations and 4 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              943.59k
      OS Memory           39084.00k
      Timestamp           11/24/2020 03:53:06 AM
      Step Count                        417  Switch Count  2
      Page Faults                       0
      Page Reclaims                     113
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


130      !             Proc Freq data=DelaysRe;
131        Title "1f(ii). Chi-Squared Test of Independence on Status by Airline, Grouped by Airport";
132        Table Status*Airline /chisq;
133        By City;
134        Weight Count;
135        Output out=p pchi;
136        Run;

NOTE: There were 20 observations read from the data set WORK.DELAYSRE.
NOTE: The data set WORK.P has 5 observations and 4 variables.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.21 seconds
      user cpu time       0.22 seconds
      system cpu time     0.00 seconds
      memory              2055.06k
      OS Memory           39864.00k
      Timestamp           11/24/2020 03:53:07 AM
      Step Count                        418  Switch Count  11
      Page Faults                       0
```

```
      Page Reclaims                     319
      Page Swaps                        0
      Voluntary Context Switches        67
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           1152


137         /* 1f(iii) */ * The results do not appear to completely agree with the results in problem 1e. Airports
138         from San Francisco (p = 0.0277235077), Seattle (p = 0.0000040883999), and San Diego (p = 0.0001132093)
139         are significant at the α = .05 level, while airports from Los Angeles (p = 0.0718151204) and
140         Phoenix (p = 0.1255718036) are not. America's Best had more flights at airports in Los
141         Angeles (811 vs. 559), Phoenix (5255 vs. 233), and San Francisco (448 vs. 232), while TWB had more
142         flights at airports in San Diego (262 vs. 2146) and Seattle (449 vs. 605). Although the sample size
143         assumption for chi-squared tests was met in all cells, some tests may be less robust in certain cells
144         if there is a large difference in flights between the two airlines at an airport. ;
145         /* 1f(iv) */ * Although America's Best had a greater proportion of delayed flights at each of the
146         airports in the data, TWB had the greater proportion of delayed flights overall, and the difference
147         between that proportion and the proportion of delayed America's Best flights was found to be
148         significant. The large difference in flights between the two airlines at certain airports causes the
149         proportions to be weighted differently, but because the difference in the overall number of flights
150         between the two airports is not large, this weighting disappears in the test on the full dataset. ;
151
152         /* 2a */
152       !         Data SoggyChips;
153         Infile "/home/chwang10/SoggyChips2020.csv" dlm=',' firstobs=2; * Skipping header row, starting on row 2 ;
154         Input Time Moisture;
155         Run;

NOTE: The infile "/home/chwang10/SoggyChips2020.csv" is:
      Filename=/home/chwang10/SoggyChips2020.csv,
      Owner Name=chwang10,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=18Nov2020:21:57:09,
      File Size (bytes)=207

NOTE: 24 records were read from the infile "/home/chwang10/SoggyChips2020.csv".
      The minimum record length was 4.
      The maximum record length was 7.
NOTE: The data set WORK.SOGGYCHIPS has 24 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              762.18k
      OS Memory           38824.00k
      Timestamp           11/24/2020 03:53:07 AM
      Step Count                        419  Switch Count  2
      Page Faults                       0
      Page Reclaims                     96
      Page Swaps                        0
      Voluntary Context Switches        18
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


156
157         /* 2b */
157       !         Proc Corr data=SoggyChips nosimple; * Descriptive statistics not needed ;
158         Title "2b. Correlation between Moisture Content and Frying Time";
159         Var Moisture Time;
160         Run;

NOTE: PROCEDURE CORR used (Total process time):
      real time           0.02 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              756.75k
      OS Memory           38824.00k
      Timestamp           11/24/2020 03:53:07 AM
      Step Count                        420  Switch Count  1
      Page Faults                       0
      Page Reclaims                     51
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           8


161         /* 2b(i) */ * The correlation coefficient between moisture content and frying time is -0.77149. The
162         relationship does appear to be strong. ;
163         /* 2b(ii) */ * H0: ϱ = 0, HA: ϱ ≠ 0 ;
164         /* 2b(iii) */ * p < .0001, We reject H0 at the α = .05 level. There is sufficient evidence that the
165         correlation is a nonzero value. ;
166
167         /* 2c */ Symbol value=circle;
168         Proc Gplot data=SoggyChips;
169         Title "2c. Moisture Content vs. Frying Time";
170         Plot Moisture*Time;
```

```
171        Run;

171      !      * The relationship is moderately strong and clearly nonlinear. ;
172
173        /* 2d */

NOTE: There were 24 observations read from the data set WORK.SOGGYCHIPS.
NOTE: PROCEDURE GPLOT used (Total process time):
      real time           0.16 seconds
      user cpu time       0.15 seconds
      system cpu time     0.01 seconds
      memory              7061.93k
      OS Memory           42776.00k
      Timestamp           11/24/2020 03:53:07 AM
      Step Count                        421  Switch Count  1
      Page Faults                       0
      Page Reclaims                     1007
      Page Swaps                        0
      Voluntary Context Switches        10
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           184


173      !          Proc Reg data=SoggyChips;
174        Title "2d. Linear Regression of Soggy Chips Data";
175        Model Moisture=Time; * Linear model: Moisture = 13.34756 - 0.25880*Time ;
176        Run;

176      !      * TSS = 805.95958 ;
177        /* 2d(i) */ * r^2 = 0.5768, Approximately 57.68 percent of the variation in the data is explained by
178        the linear model. ;
179        /* 2d(ii) */ * The linearity and constant variance assumptions are violated. From the scatterplot in
180        problem 2c, the data are clearly not linear, and the residual plot and studentized residual plot
181        indicate a clear lack of homoscedasticity. ;
182
183        /* 2e */

NOTE: PROCEDURE REG used (Total process time):
      real time           0.60 seconds
      user cpu time       0.25 seconds
      system cpu time     0.05 seconds
      memory              20571.46k
      OS Memory           55112.00k
      Timestamp           11/24/2020 03:53:07 AM
      Step Count                        422  Switch Count  23
      Page Faults                       0
      Page Reclaims                     15467
      Page Swaps                        0
      Voluntary Context Switches        1009
      Involuntary Context Switches      4
      Block Input Operations            0
      Block Output Operations           1512


183      !          Data SoggyChipsPower;
184        Set SoggyChips;
185        MoisturePower=log(Moisture);
186        TimePower=log(Time);

NOTE: There were 24 observations read from the data set WORK.SOGGYCHIPS.
NOTE: The data set WORK.SOGGYCHIPSPOWER has 24 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              974.15k
      OS Memory           52652.00k
      Timestamp           11/24/2020 03:53:07 AM
      Step Count                        423  Switch Count  2
      Page Faults                       0
      Page Reclaims                     129
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


187        Proc Reg data=SoggyChipsPower;
188        Title "2e. Power Regression of Soggy Chips Data";
189        Model MoisturePower=TimePower;
190        Output out=SoggyChipsPowerO predicted=MoisturePowerPred;
191        Run;

191      !      * TSS = 18.21285 ;
192        /* 2e(i) */ * r^2 = 0.9406 ;
193        /* 2e(ii) */ * According to the residual plot and studentized residual plot, there may be a slight
194        violation of the homoscedasticity assumption. ;
195        /* 2e(iii) */ * Power model: Moisture = e^(4.88770 - 1.12055*ln(Time)) ;
196        /* 2e(iv) */
```

```
NOTE: The data set WORK.SOGGYCHIPSPOWERO has 24 observations and 5 variables.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.51 seconds
      user cpu time       0.22 seconds
      system cpu time     0.04 seconds
      memory              11192.03k
      OS Memory           60388.00k
      Timestamp           11/24/2020 03:53:08 AM
      Step Count                        424  Switch Count  25
      Page Faults                       0
      Page Reclaims                     12309
      Page Swaps                        0
      Voluntary Context Switches        1011
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           1432


196     !             Data SoggyChipsPowerUn;
197       Set SoggyChipsPowerO;
198       MoisturePowerUn=exp(MoisturePowerPred);
199       Run;

NOTE: There were 24 observations read from the data set WORK.SOGGYCHIPSPOWERO.
NOTE: The data set WORK.SOGGYCHIPSPOWERUN has 24 observations and 6 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1088.75k
      OS Memory           53164.00k
      Timestamp           11/24/2020 03:53:08 AM
      Step Count                        425  Switch Count  2
      Page Faults                       0
      Page Reclaims                     124
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


200
201       /* 2f */
201     !          Proc Transreg data=SoggyChips;
202       Title "2f. Box-Cox Transformation of Soggy Chips Data";
203       Model Boxcox(Moisture)=identity(Time);
204       Run;

NOTE: There were 24 observations read from the data set WORK.SOGGYCHIPS.
NOTE: PROCEDURE TRANSREG used (Total process time):
      real time           0.14 seconds
      user cpu time       0.06 seconds
      system cpu time     0.01 seconds
      memory              3067.71k
      OS Memory           54068.00k
      Timestamp           11/24/2020 03:53:08 AM
      Step Count                        426  Switch Count  1
      Page Faults                       0
      Page Reclaims                     332
      Page Swaps                        0
      Voluntary Context Switches        353
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           496


205       /* 2f(i) */ * λ = -0.75 ;
206       /* 2f(ii) */
206     !             Data SoggyChipsBC;
207       Set SoggyChips;
208       MoistureBC=Moisture**-0.75;

NOTE: There were 24 observations read from the data set WORK.SOGGYCHIPS.
NOTE: The data set WORK.SOGGYCHIPSBC has 24 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1085.56k
      OS Memory           53164.00k
      Timestamp           11/24/2020 03:53:08 AM
      Step Count                        427  Switch Count  2
      Page Faults                       0
      Page Reclaims                     124
      Page Swaps                        0
      Voluntary Context Switches        15
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           272
```

```
209         Proc Reg data=SoggyChipsBC;
210         Title "2f(ii). Power Regression of Soggy Chips Data with Box-Cox Transformation";
211         Model MoistureBC=Time;
212         Output out=SoggyChipsBCO predicted=MoistureBCPred;
213         Run;

213       !      * TSS = 1.3932 ;
214         /* 2f(iii) */ * r^2 = 0.9666 ;
215         /* 2f(iv) */ * No, none of the assumptions appear to be violated. ;
216         /* 2f(v) */ * Power model with Box-Cox transformation: Moisture = (0.75444 - 1.80577*Time)^(-4/3) ;
217         /* 2f(vi) */

NOTE: The data set WORK.SOGGYCHIPSBCO has 24 observations and 4 variables.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.49 seconds
      user cpu time       0.22 seconds
      system cpu time     0.05 seconds
      memory              11295.00k
      OS Memory           60644.00k
      Timestamp           11/24/2020 03:53:09 AM
      Step Count                        428  Switch Count  25
      Page Faults                       0
      Page Reclaims                     12188
      Page Swaps                        0
      Voluntary Context Switches        1007
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           1416


217       !             Data SoggyChipsBCUn;
218         Set SoggyChipsBCO;
219         MoistureBCUn=MoistureBCPred**(-4/3);
220         Run;

NOTE: There were 24 observations read from the data set WORK.SOGGYCHIPSBCO.
NOTE: The data set WORK.SOGGYCHIPSBCUN has 24 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1077.43k
      OS Memory           53676.00k
      Timestamp           11/24/2020 03:53:09 AM
      Step Count                        429  Switch Count  2
      Page Faults                       0
      Page Reclaims                     124
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


221
222         /* 2g(i) */
222       !            Data SoggyChipsM;
223         Merge SoggyChipsPowerUn SoggyChipsBCUn;
224         By Time;
225         Run;

NOTE: MERGE statement has more than one data set with repeats of BY values.
NOTE: There were 24 observations read from the data set WORK.SOGGYCHIPSPOWERUN.
NOTE: There were 24 observations read from the data set WORK.SOGGYCHIPSBCUN.
NOTE: The data set WORK.SOGGYCHIPSM has 24 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1527.84k
      OS Memory           53936.00k
      Timestamp           11/24/2020 03:53:09 AM
      Step Count                        430  Switch Count  2
      Page Faults                       0
      Page Reclaims                     158
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


226         /* 2g(ii) */
226       !             Proc Sgplot data=SoggyChipsM;
227         Title "2g(ii). Scatterplot of Power and Box-Cox Models of Soggy Chips Data";
228         Keylegend "Power" "Box-Cox" /location=inside position=topright across=1;
229         Yaxis label="Moisture";
230         Scatter X=Time Y=Moisture;
231         Series X=Time Y=MoisturePowerUn /name="Power" legendlabel="Power Model" lineattrs=(color=orange);
232         Series X=Time Y=MoistureBCUn /name="Box-Cox" legendlabel="Power Model with Box-Cox Transformation"
232       ! lineattrs=(color=maroon);
```

```
233        Run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.10 seconds
      user cpu time       0.04 seconds
      system cpu time     0.00 seconds
      memory              1840.68k
      OS Memory           54320.00k
      Timestamp           11/24/2020 03:53:09 AM
      Step Count                        431  Switch Count  2
      Page Faults                       0
      Page Reclaims                     290
      Page Swaps                        0
      Voluntary Context Switches        216
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           472

NOTE: There were 24 observations read from the data set WORK.SOGGYCHIPSM.

234        /* 2g(iii) */ * I believe the power model with the Box-Cox transformation is the best of the three. The adjusted-r^2
235        is the highest of the three models and the total sum of squares (TSS) is by far the lowest. The Box-Cox
236        Transformation also assures we have the best value of λ. ;
237        /* 2g(iv) */ * Although the r^2 is very strong in the Box-Cox transformation model, it appears to be
238        overfitting the data. Because Time is the only independent variable and every three of its values are
239        the same, the model will predict the same Moisture value for them despite the true value being recorded
240        differently. There may be additional independent variables not in the dataset. ;
241
242        /* 3a */
242      !         Proc Format;
243        Value $Color B='Brown' P='Pink' U='Purple' R='Red';
NOTE: Format $COLOR is already on the library WORK.FORMATS.
NOTE: Format $COLOR has been output.
244        Run;

NOTE: PROCEDURE FORMAT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              210.81k
      OS Memory           53156.00k
      Timestamp           11/24/2020 03:53:09 AM
      Step Count                        432  Switch Count  0
      Page Faults                       0
      Page Reclaims                     14
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0


245
246        /* 3b */
246      !         Proc Import out=Lipstick file="/home/chwang10/lipsticklead.xlsx" dbms=xlsx;
247        Run;

NOTE: Import cancelled.  Output dataset WORK.LIPSTICK already exists.  Specify REPLACE option to overwrite it.
NOTE: The SAS System stopped processing this step because of errors.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              822.28k
      OS Memory           53652.00k
      Timestamp           11/24/2020 03:53:09 AM
      Step Count                        433  Switch Count  0
      Page Faults                       0
      Page Reclaims                     138
      Page Swaps                        0
      Voluntary Context Switches        1
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0

248
249        /* 3c */


249      !         Data LipstickSep;
250        Set Lipstick;
251        /* 3c(i) */ Color=substr(ColorPrice, 1, 1);
252        Price=substr(ColorPrice, 2, 1);
253        /* 3c(ii) */ Format Color $Color.;
254        Run;

NOTE: There were 223 observations read from the data set WORK.LIPSTICK.
NOTE: The data set WORK.LIPSTICKSEP has 223 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
```

```
      system cpu time     0.00 seconds
      memory              866.81k
      OS Memory           53676.00k
      Timestamp           11/24/2020 03:53:09 AM
      Step Count                        434  Switch Count  2
      Page Faults                       0
      Page Reclaims                     121
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


255
256        /* 3d */
256      !          Proc Gchart data=LipstickSep;
257        Title "3d. Lead Content of Lipstick by Color and Price Category";
258        Vbar Price /group=Color sumvar=Pb type=mean;
259        Run;

260        /* 3d(i) */ * It appears that price category 2 has more lead than price categories 1 and 3. However,
261        it is difficult to visually discern whether a specific color has more lead in its lipstick than the
262        other colors. ;
263
264        /* 3e */

NOTE: There were 223 observations read from the data set WORK.LIPSTICKSEP.
NOTE: PROCEDURE GCHART used (Total process time):
      real time           0.16 seconds
      user cpu time       0.16 seconds
      system cpu time     0.01 seconds
      memory              6273.43k
      OS Memory           55832.00k
      Timestamp           11/24/2020 03:53:09 AM
      Step Count                        435  Switch Count  1
      Page Faults                       0
      Page Reclaims                     599
      Page Swaps                        0
      Voluntary Context Switches        11
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           160


264      !          Proc GLM data=LipstickSep;
265        Title '3e. ANOVA of Lipstick Data With Interaction Term';
266        Class Color Price;
267        Model Pb=Color Price Color*Price;
268        Means Color Price Color*Price /tukey;
269        Run;

NOTE: Means from the MEANS statement are not adjusted for other terms in the model.  For adjusted means, use the LSMEANS statement.
270        /* 3e(i) */ * H0(1): μ_B = μ_P = μ_U = μ_R , HA(1): At least one of the means is different
271        H0(2): μ_1 = μ_2 = μ_3 , HA(2): At least one of the means is different
272        H0(3): There is no interaction between color and price category , HA(3): There is an interaction between color and price
272      ! category ;
273        /* 3e(ii) */ * Color: F = 1.95 (not significant at the α = .05 level)
274        Price category: F = 6.99 (significant at the α = .05 level)
275        Interaction: F = 0.70 (not significant at the α = .05 level);
276        /* 3e(iii) */

NOTE: PROCEDURE GLM used (Total process time):
      real time           0.56 seconds
      user cpu time       0.36 seconds
      system cpu time     0.02 seconds
      memory              5096.09k
      OS Memory           55104.00k
      Timestamp           11/24/2020 03:53:09 AM
      Step Count                        436  Switch Count  3
      Page Faults                       0
      Page Reclaims                     1149
      Page Swaps                        0
      Voluntary Context Switches        3545
      Involuntary Context Switches      5
      Block Input Operations            0
      Block Output Operations           1768


276      !               Proc GLM data=LipstickSep;
277        Title '3e(iii). ANOVA of Lipstick Data Without Interaction Term';
278        Class Color Price;
279        Model Pb=Color Price;
280        Means Color Price /tukey;
281        Run;

NOTE: Means from the MEANS statement are not adjusted for other terms in the model.  For adjusted means, use the LSMEANS statement.
281      !      * According to Tukey's HSD test, price categories 1 and 2 and price categories 2 and 3 have
282        statistically different means. ;
283        /* 3f */ * I would recommend avoiding lipsticks in price category 2, as it has a significantly higher
284        lead content than price categories 1 and 3. This difference was shown to be statistically significant
```

```
285         at the α = .05 level. Additionally, red lipsticks have the lowest mean, median, interquartile
286         range (IQR), and upper outlier bound (Q3+1.5*IQR) of the four colors. ;
287
288         /* 4a */

NOTE: PROCEDURE GLM used (Total process time):
      real time           0.43 seconds
      user cpu time       0.26 seconds
      system cpu time     0.03 seconds
      memory              4613.06k
      OS Memory           55104.00k
      Timestamp           11/24/2020 03:53:10 AM
      Step Count                        437  Switch Count  3
      Page Faults                       0
      Page Reclaims                     926
      Page Swaps                        0
      Voluntary Context Switches        2607
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           1160


288       !          Proc Import out=DrugCon file='/home/chwang10/drugconcentrationB.xls' dbms=xls;
289         Run;

NOTE: Import cancelled.  Output dataset WORK.DRUGCON already exists.  Specify REPLACE option to overwrite it.
NOTE: The SAS System stopped processing this step because of errors.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              822.34k
      OS Memory           53396.00k
      Timestamp           11/24/2020 03:53:10 AM
      Step Count                        438  Switch Count  0
      Page Faults                       0
      Page Reclaims                     138
      Page Swaps                        0
      Voluntary Context Switches        1
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           0

290         /* 4a(i) */


290       !             Proc Contents data=DrugCon noprint; * Placeholder code indicating that I opened and read the code on my PC ;
291         /* 4a(ii) */

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.01 seconds
      memory              683.53k
      OS Memory           53420.00k
      Timestamp           11/24/2020 03:53:10 AM
      Step Count                        439  Switch Count  0
      Page Faults                       0
      Page Reclaims                     85
      Page Swaps                        0
      Voluntary Context Switches        0
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           8


291       !              Data DrugUni;
292         Set DrugCon;
293         Array x[*] conc_1 - conc_49;
294         Do i = 1 to dim(x);
295         If i = 1 then Time=1/60; * The initial reading is t = 1 minute ;
296         else Time=1/2*(i-1); * Expressing time in hours ;
297         Concentration=x[i];
298         Output;
299         End;
300         Drop i conc_1 - conc_49;
301         Run;

NOTE: There were 6 observations read from the data set WORK.DRUGCON.
NOTE: The data set WORK.DRUGUNI has 294 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1162.21k
      OS Memory           53420.00k
      Timestamp           11/24/2020 03:53:10 AM
      Step Count                        440  Switch Count  2
      Page Faults                       0
      Page Reclaims                     124
      Page Swaps                        0
```

```
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264


302         /* 4a(iii) */ Symbol value=circle;
303         Proc Gplot data=DrugUni;
304         Title '4a(iii). Drug Concentration vs. Time';
305         Plot Concentration*Time;
306         Run;

307
308         /* 4b */

NOTE: There were 294 observations read from the data set WORK.DRUGUNI.
NOTE: PROCEDURE GPLOT used (Total process time):
      real time           0.15 seconds
      user cpu time       0.15 seconds
      system cpu time     0.01 seconds
      memory              6979.03k
      OS Memory           56856.00k
      Timestamp           11/24/2020 03:53:10 AM
      Step Count                        441  Switch Count  1
      Page Faults                       0
      Page Reclaims                     858
      Page Swaps                        0
      Voluntary Context Switches        8
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           248


308       !          Proc Nlin data=DrugUni plots=fit;
309         Title '4b. Nonlinear Model of Drug Concentration Data';
310         Parameters α=1 β=64.8 θ1=0.3 θ2=1.5; * Arbitrary choosing unknown α = 1 and median choice of θ_1 and θ_2
310       ! parameters ;
311         If Time < 8 then Model Concentration=β*(exp(-θ1*Time)-exp(-θ2*Time))/(θ2-θ1);
312         else Model Concentration=β/(θ2-θ1)*(exp(-θ1*Time)-exp(-θ2*Time)+α*exp(-θ1*(Time-8))-α*exp(-θ2*(Time-8)));
313         Output out=DrugUniO predicted=ConcPred;
314         Run;

NOTE: DER.α not initialized or missing. It will be computed automatically.
NOTE: DER.β not initialized or missing. It will be computed automatically.
NOTE: DER.θ1 not initialized or missing. It will be computed automatically.
NOTE: DER.θ2 not initialized or missing. It will be computed automatically.
NOTE: PROC NLIN grid search time was  0: 0: 0.
NOTE: Convergence criterion met.
NOTE: The data set WORK.DRUGUNIO has 294 observations and 3 variables.
NOTE: PROCEDURE NLIN used (Total process time):
      real time           0.26 seconds
      user cpu time       0.16 seconds
      system cpu time     0.00 seconds
      memory              4117.06k
      OS Memory           55348.00k
      Timestamp           11/24/2020 03:53:10 AM
      Step Count                        442  Switch Count  3
      Page Faults                       0
      Page Reclaims                     748
      Page Swaps                        0
      Voluntary Context Switches        299
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           1200


315         /* 4b(i) */ * α = 1.1903, θ_1 = 0.2988, θ_2 = 1.5491, The patients received an increase in the dosage
316         at t = 8 hours. The increase was approximately 19.03 percent. ;
317         /* 4b(ii) */
317       !              Proc Sort data=DrugUniO;
318         By Time;

NOTE: There were 294 observations read from the data set WORK.DRUGUNIO.
NOTE: The data set WORK.DRUGUNIO has 294 observations and 3 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              962.28k
      OS Memory           54188.00k
      Timestamp           11/24/2020 03:53:10 AM
      Step Count                        443  Switch Count  2
      Page Faults                       0
      Page Reclaims                     114
      Page Swaps                        0
      Voluntary Context Switches        16
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           264
```

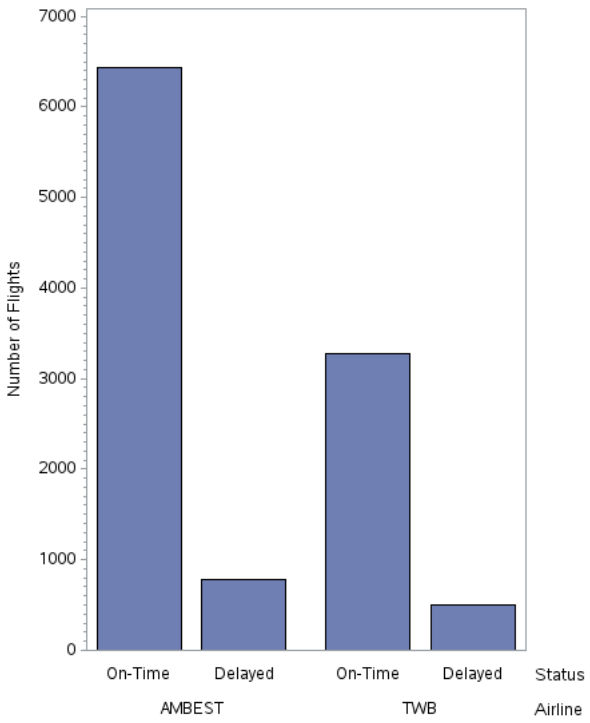
```
319         Proc Sgplot data=DrugUniO;
320         Title "4b(ii). Scatterplot of Concentration vs. Time with Predicted Curve Overlay";
321         Keylegend "P" /location=inside position=topright across=1;
322         Scatter X=Time Y=Concentration;
323         Series X=Time Y=ConcPred /name="P" legendlabel="Predicted Curve" lineattrs=(color=green);
324         Run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.12 seconds
      user cpu time       0.05 seconds
      system cpu time     0.01 seconds
      memory              1535.25k
      OS Memory           54832.00k
      Timestamp           11/24/2020 03:53:10 AM
      Step Count                        444  Switch Count  2
      Page Faults                       0
      Page Reclaims                     291
      Page Swaps                        0
      Voluntary Context Switches        210
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           552

NOTE: There were 294 observations read from the data set WORK.DRUGUNIO.

325
326         OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
337
```

Results: Take-Home Test.sas



1e(ii). Chi-Squared Test of Independence on Status by Airline

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Status by Airline		
	Airline		
Status	AMBEST	TWB	Total
On-Time	6438 58.53 66.29 89.11	3274 29.76 33.71 86.73	9712 88.29
Delayed	787 7.15 61.10 10.89	501 4.55 38.90 13.27	1288 11.71
Total	7225 65.68	3775 34.32	11000 100.00

Statistics for Table of Status by Airline

Statistic	DF	Value	Prob
Chi-Square	1	13.5717	0.0002
Likelihood Ratio Chi-Square	1	13.3469	0.0003
Continuity Adj. Chi-Square	1	13.3426	0.0003
Mantel-Haenszel Chi-Square	1	13.5705	0.0002
Phi Coefficient		0.0351	
Contingency Coefficient		0.0351	
Cramer's V		0.0351	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	6438
Left-sided Pr <= F	0.9999
Right-sided Pr >= F	0.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	0.0003

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.2518	1.1106	1.4109
Relative Risk (Column 1)	1.0849	1.0363	1.1358
Relative Risk (Column 2)	0.8667	0.8049	0.9331

Sample Size = 11000

1f(ii). Chi-Squared Test of Independence on Status by Airline, Grouped by Airport

The FREQ Procedure

City=LA

Frequency Percent Row Pct Col Pct

Table of Status by Airline			
	Airline		
Status	AMBEST	TWB	Total
On-Time	694 50.66 58.27 85.57	497 36.28 41.73 88.91	1191 86.93
Delayed	117 8.54 65.36 14.43	62 4.53 34.64 11.09	179 13.07
Total	811 59.20	559 40.80	1370 100.00

Statistics for Table of Status by Airline

Statistic	DF	Value	Prob
Chi-Square	1	3.2410	0.0718
Likelihood Ratio Chi-Square	1	3.2931	0.0696
Continuity Adj. Chi-Square	1	2.9540	0.0857
Mantel-Haenszel Chi-Square	1	3.2387	0.0719
Phi Coefficient		-0.0486	
Contingency Coefficient		0.0486	
Cramer's V		-0.0486	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	694
Left-sided Pr <= F	0.0420
Right-sided Pr >= F	0.9708
Table Probability (P)	0.0129
Two-sided Pr <= P	0.0733

Sample Size = 1370

1f(ii). Chi-Squared Test of Independence on Status by Airline, Grouped by Airport

The FREQ Procedure

City=PHO

Frequency Percent Row Pct Col Pct

Table of Status by Airline			
	Airline		
Status	AMBEST	TWB	Total
On-Time	4840 88.19 95.63 92.10	221 4.03 4.37 94.85	5061 92.22
Delayed	415 7.56 97.19 7.90	12 0.22 2.81 5.15	427 7.78
Total	5255 95.75	233 4.25	5488 100.00

Statistics for Table of Status by Airline

Statistic	DF	Value	Prob
Chi-Square	1	2.3464	0.1256
Likelihood Ratio Chi-Square	1	2.6277	0.1050
Continuity Adj. Chi-Square	1	1.9792	0.1595
Mantel-Haenszel Chi-Square	1	2.3460	0.1256
Phi Coefficient		-0.0207	
Contingency Coefficient		0.0207	
Cramer's V		-0.0207	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	4840
Left-sided Pr <= F	0.0741
Right-sided Pr >= F	0.9579
Table Probability (P)	0.0320
Two-sided Pr <= P	0.1345

Sample Size = 5488

1f(ii). Chi-Squared Test of Independence on Status by Airline, Grouped by Airport

The FREQ Procedure

City=SD

Frequency Percent Row Pct Col Pct

Table of Status by Airline			
	Airline		
Status	AMBEST	TWB	Total
On-Time	201 8.35 9.84 76.72	1841 76.45 90.16 85.79	2042 84.80
Delayed	61 2.53 16.67 23.28	305 12.67 83.33 14.21	366 15.20
Total	262 10.88	2146 89.12	2408 100.00

Statistics for Table of Status by Airline

Statistic	DF	Value	Prob
Chi-Square	1	14.9026	0.0001
Likelihood Ratio Chi-Square	1	13.4084	0.0003
Continuity Adj. Chi-Square	1	14.2072	0.0002
Mantel-Haenszel Chi-Square	1	14.8964	0.0001
Phi Coefficient		-0.0787	
Contingency Coefficient		0.0784	
Cramer's V		-0.0787	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	201
Left-sided Pr <= F	0.0002
Right-sided Pr >= F	0.9999
Table Probability (P)	<.0001
Two-sided Pr <= P	0.0002

Sample Size = 2408

1f(ii). Chi-Squared Test of Independence on Status by Airline, Grouped by Airport

The FREQ Procedure

City=SEA

Frequency Percent Row Pct Col Pct

Table of Status by Airline			
	Airline		
Status	AMBEST	TWB	Total
On-Time	320 30.36 38.88 71.27	503 47.72 61.12 83.14	823 78.08
Delayed	129 12.24 55.84 28.73	102 9.68 44.16 16.86	231 21.92
Total	449 42.60	605 57.40	1054 100.00

Statistics for Table of Status by Airline

Statistic	DF	Value	Prob
Chi-Square	1	21.2229	<.0001
Likelihood Ratio Chi-Square	1	21.0223	<.0001
Continuity Adj. Chi-Square	1	20.5349	<.0001
Mantel-Haenszel Chi-Square	1	21.2028	<.0001
Phi Coefficient		-0.1419	
Contingency Coefficient		0.1405	
Cramer's V		-0.1419	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	320
Left-sided Pr <= F	<.0001
Right-sided Pr >= F	1.0000
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 1054

1f(ii). Chi-Squared Test of Independence on Status by Airline, Grouped by Airport

The FREQ Procedure

City=SF

Frequency Percent Row Pct Col Pct

Table of Status by Airline			
	Airline		
Status	AMBEST	TWB	Total
On-Time	383 56.32 64.37 85.49	212 31.18 35.63 91.38	595 87.50
Delayed	65 9.56 76.47 14.51	20 2.94 23.53 8.62	85 12.50
Total	448 65.88	232 34.12	680 100.00

Statistics for Table of Status by Airline

Statistic	DF	Value	Prob
Chi-Square	1	4.8452	0.0277
Likelihood Ratio Chi-Square	1	5.1136	0.0237
Continuity Adj. Chi-Square	1	4.3218	0.0376
Mantel-Haenszel Chi-Square	1	4.8381	0.0278
Phi Coefficient		-0.0844	
Contingency Coefficient		0.0841	
Cramer's V		-0.0844	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	383
Left-sided Pr <= F	0.0170
Right-sided Pr >= F	0.9913
Table Probability (P)	0.0083
Two-sided Pr <= P	0.0280

Sample Size = 680

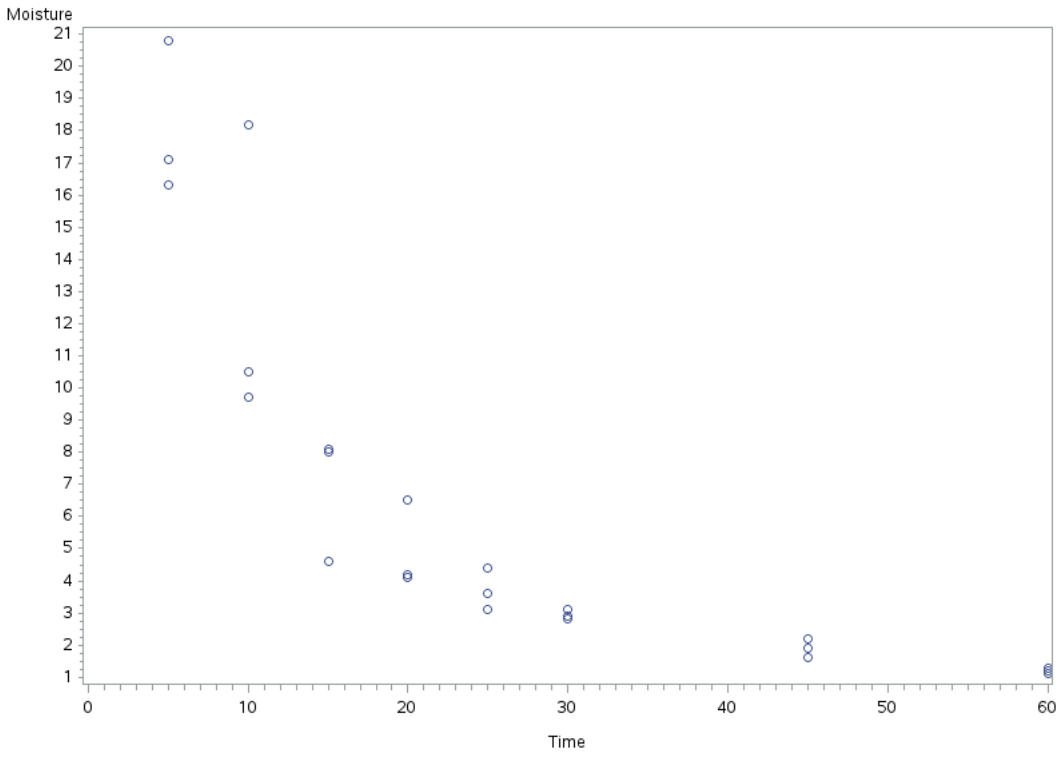
2b. Correlation between Moisture Content and Frying Time

The CORR Procedure

2 Variables:	Moisture Time

Pearson Correlation Coefficients, N = 24 Prob > |r| under H0: Rho=0		
	Moisture	Time
Moisture	1.00000	-0.77149 <.0001
Time	-0.77149 <.0001	1.00000

2c. Moisture Content vs. Frying Time



2d. Linear Regression of Soggy Chips Data

The REG Procedure
Model: MODEL1
Dependent Variable: Moisture

Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	479.71039	479.71039	32.35	<.0001
Error	22	326.24920	14.82951		
Corrected Total	23	805.95958			

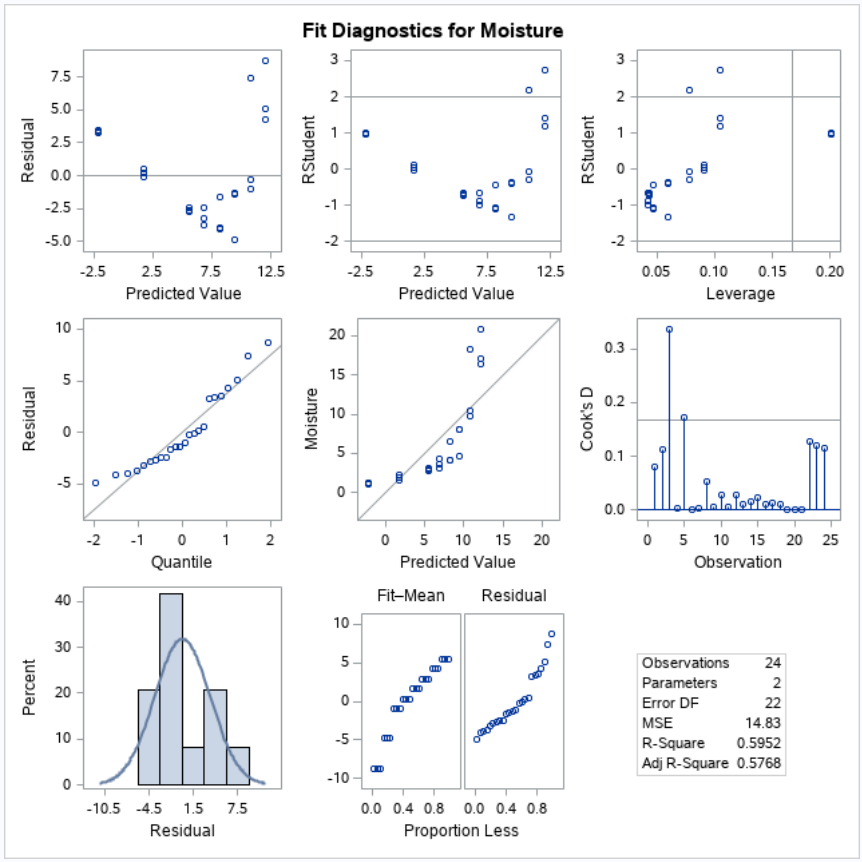
Root MSE	3.85091	R-Square	0.5952
Dependent Mean	6.55417	Adj R-Sq	0.5768
Coeff Var	58.75515		

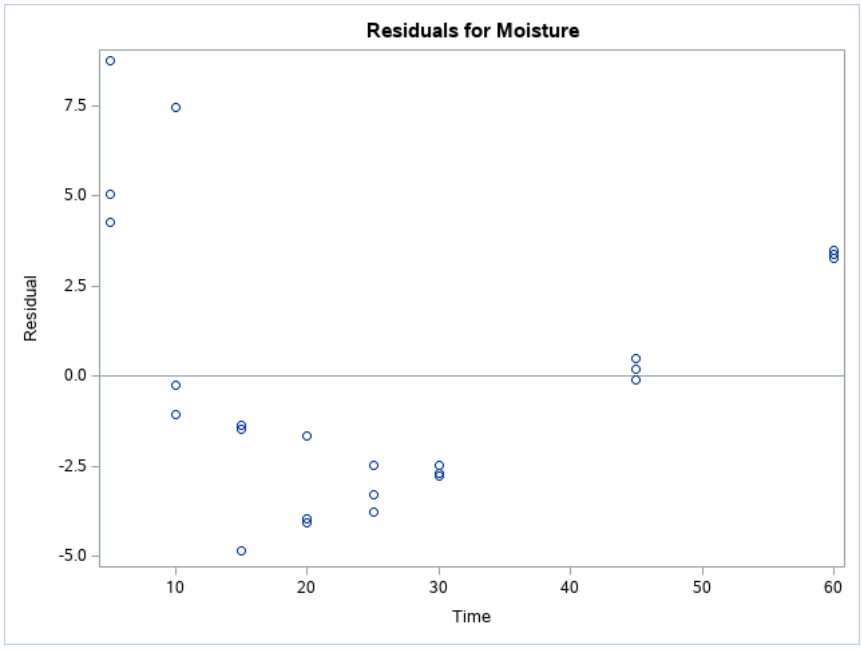
Parameter Estimates

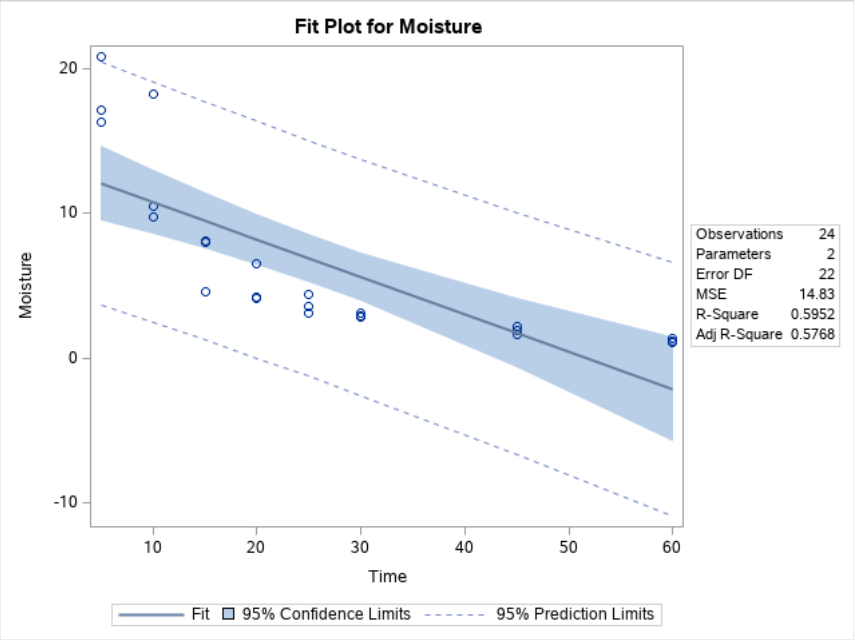
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	13.34756	1.42988	9.33	<.0001
Time	1	-0.25880	0.04550	-5.69	<.0001

2d. Linear Regression of Soggy Chips Data

The REG Procedure
Model: MODEL1
Dependent Variable: Moisture


Fit Diagnostics for Moisture
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Moisture
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 24
Parameters 2
Error DF 22
MSE 14.83
R-Square 0.5952
Adj R-Square 0.5768


Residuals for Moisture
Residual
Time



2e. Power Regression of Soggy Chips Data

The REG Procedure
Model: MODEL1
Dependent Variable: MoisturePower

Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	17.17739	17.17739	364.96	<.0001
Error	22	1.03546	0.04707		
Corrected Total	23	18.21285			

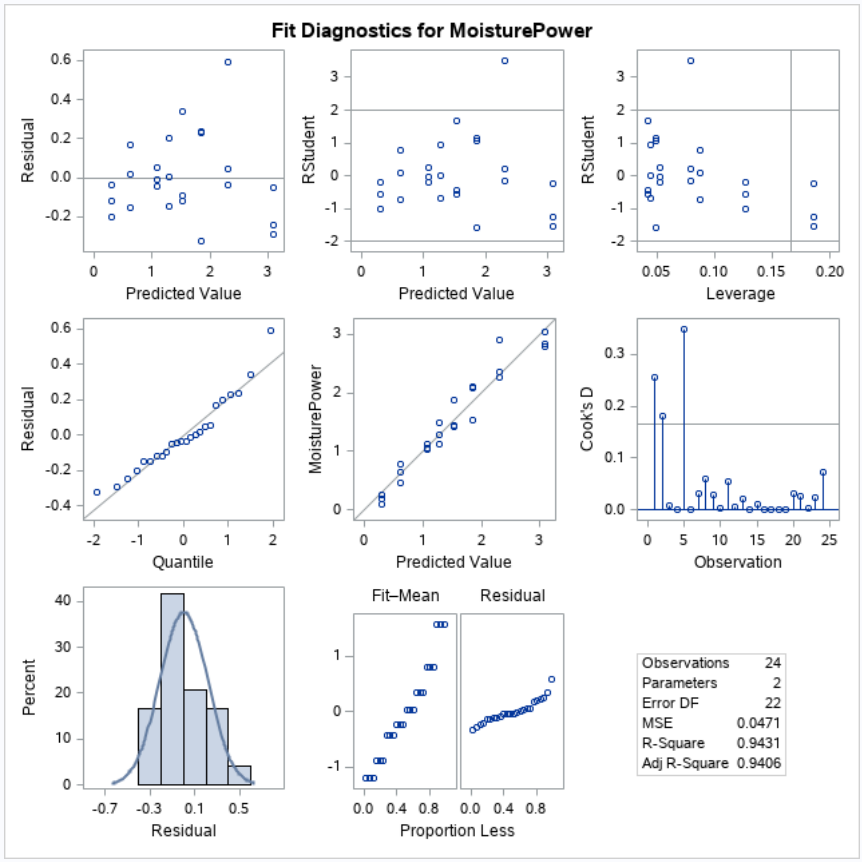
Root MSE	0.21695	R-Square	0.9431
Dependent Mean	1.50689	Adj R-Sq	0.9406
Coeff Var	14.39708		

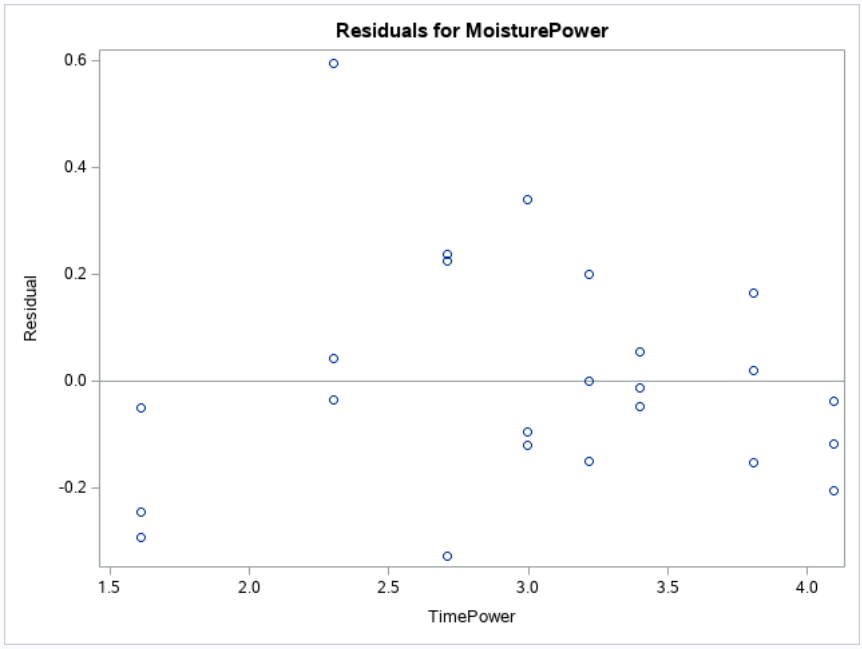
Parameter Estimates

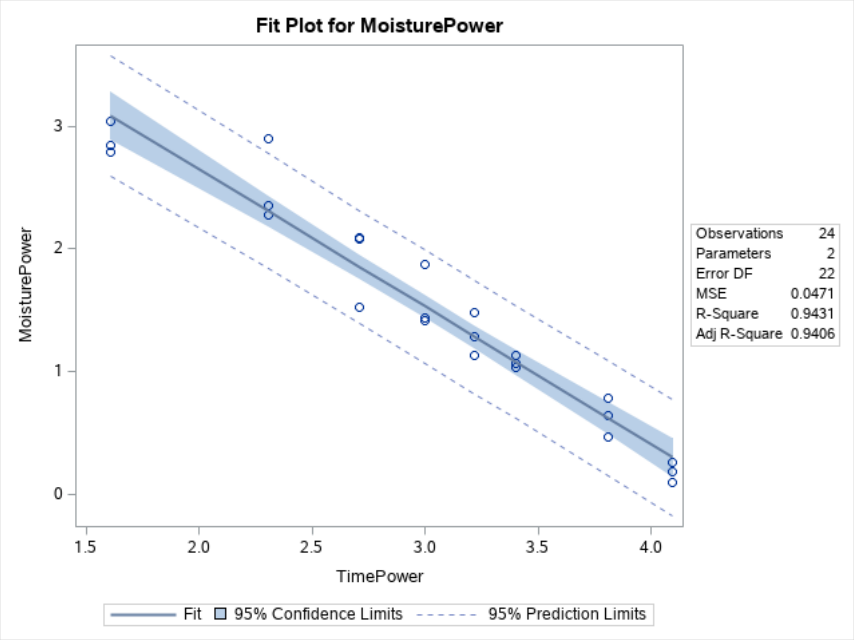
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	4.88770	0.18243	26.79	<.0001
TimePower	1	-1.12055	0.05866	-19.10	<.0001

2e. Power Regression of Soggy Chips Data

The REG Procedure
Model: MODEL1
Dependent Variable: MoisturePower

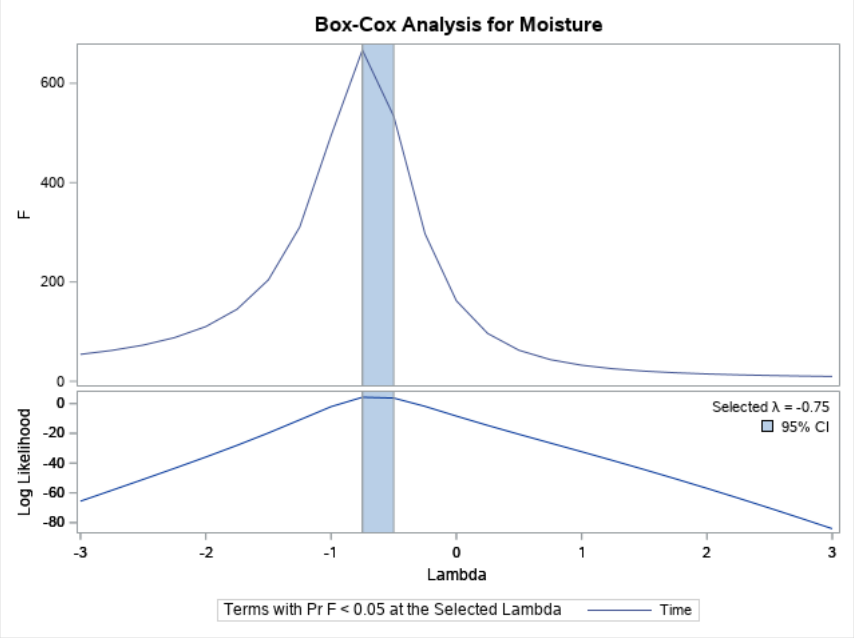

Fit Diagnostics for MoisturePower
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
MoisturePower
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 24
Parameters 2
Error DF 22
MSE 0.0471
R-Square 0.9431
Adj R-Square 0.9406


Residuals for MoisturePower
Residual
TimePower



2f. Box-Cox Transformation of Soggy Chips Data

The TRANSREG Procedure



2f(ii). Power Regression of Soggy Chips Data with Box-Cox Transformation

The REG Procedure
Model: MODEL1
Dependent Variable: MoistureBC

Number of Observations Read	24
Number of Observations Used	24

Analysis of Variance

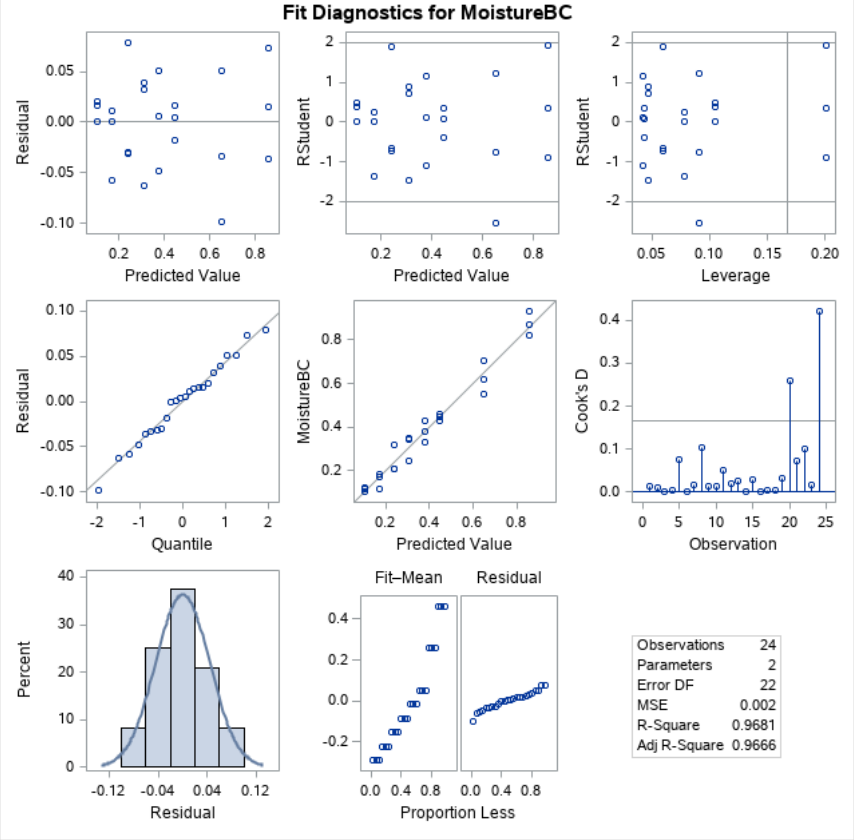
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.34872	1.34872	666.96	<.0001
Error	22	0.04449	0.00202		
Corrected Total	23	1.39321			

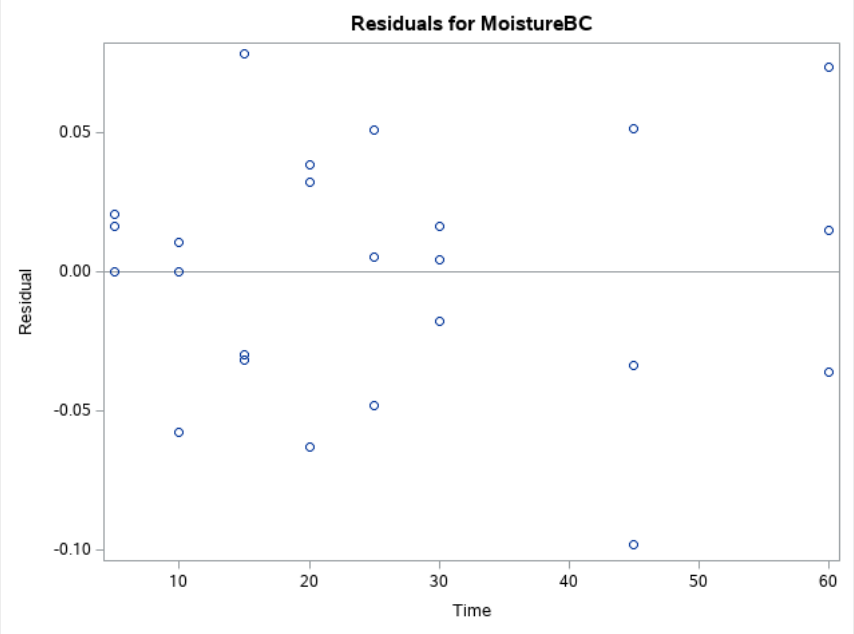
Root MSE	0.04497	R-Square	0.9681
Dependent Mean	0.39421	Adj R-Sq	0.9666
Coeff Var	11.40736		

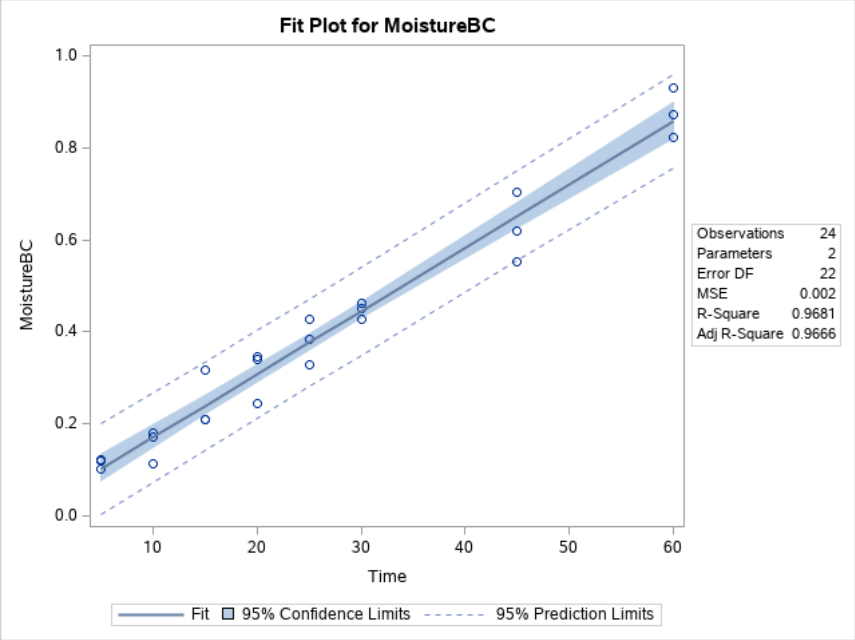
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.03400	0.01670	2.04	0.0540
Time	1	0.01372	0.00053135	25.83	<.0001

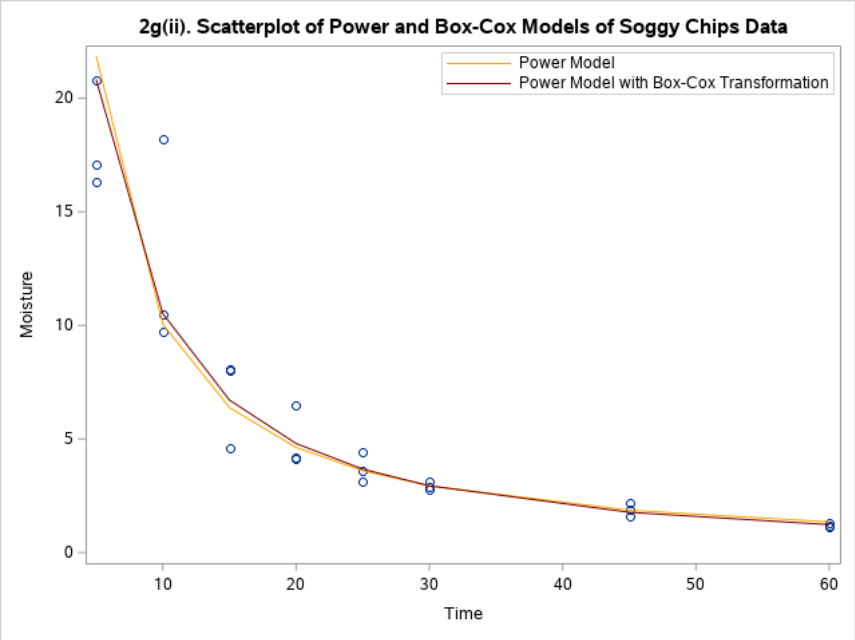
2f(ii). Power Regression of Soggy Chips Data with Box-Cox Transformation

The REG Procedure
Model: MODEL1
Dependent Variable: MoistureBC

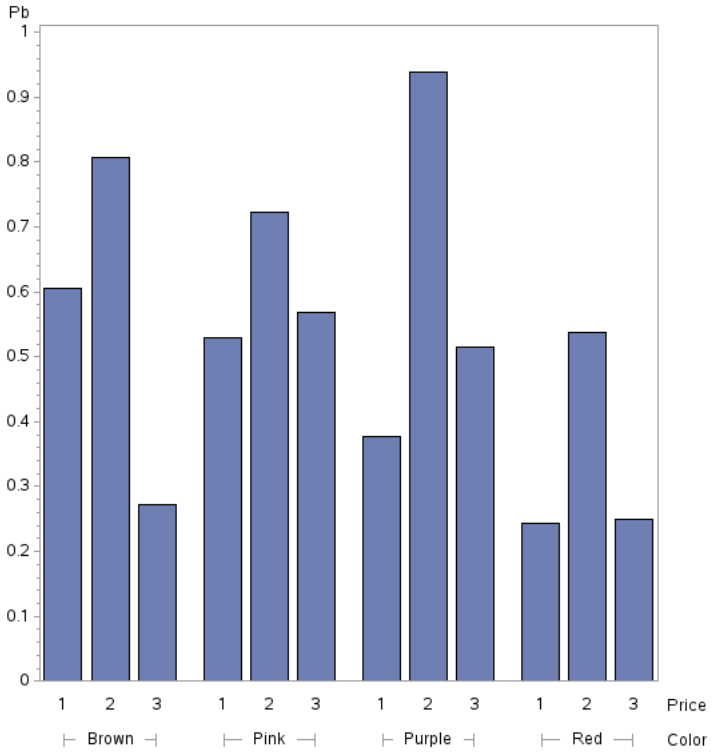





Fit Plot for MoistureBC
MoistureBC
Time
1.0
0.8
0.6
0.4
0.2
0.0
10
20
30
40
50
60
Observations 24
Parameters 2
Error DF 22
MSE 0.002
R-Square 0.9681
Adj R-Square 0.9666
Fit
95% Confidence Limits
95% Prediction Limits


2g(ii). Scatterplot of Power and Box-Cox Models of Soggy Chips Data
Moisture
Time
Power Model
Power Model with Box-Cox Transformation
20
15
10
5
0
10
20
30
40
50
60

3d. Lead Content of Lipstick by Color and Price Category



3e. ANOVA of Lipstick Data With Interaction Term

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
Color	4	Brown Pink Purple Red
Price	3	1 2 3

Number of Observations Read	223
Number of Observations Used	223

3e. ANOVA of Lipstick Data With Interaction Term

The GLM Procedure

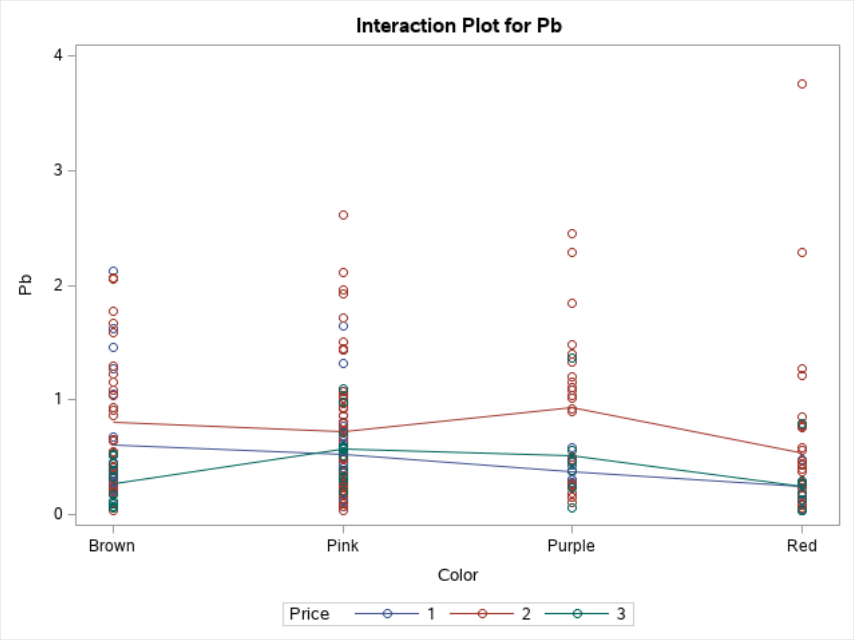
Dependent Variable: Pb Pb

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	7.71746766	0.70158797	2.18	0.0165
Error	211	67.81932875	0.32141862		
Corrected Total	222	75.53679641			

R-Square	Coeff Var	Root MSE	Pb Mean
0.102168	90.18272	0.566938	0.628655

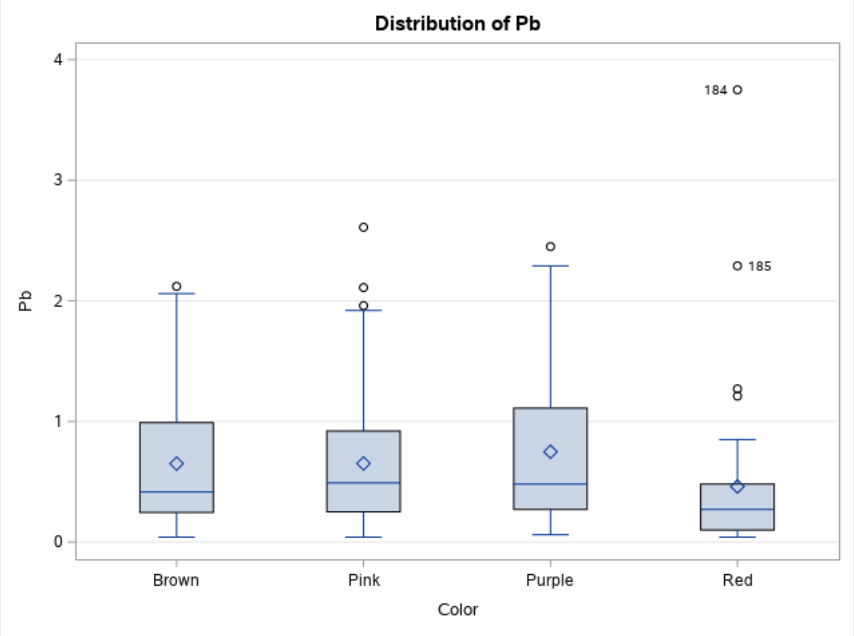
Source	DF	Type I SS	Mean Square	F Value	Pr > F
Color	3	1.87932225	0.62644075	1.95	0.1228
Price	2	4.49279407	2.24639704	6.99	0.0012
Color*Price	6	1.34535134	0.22422522	0.70	0.6518

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Color	3	1.31059937	0.43686646	1.36	0.2563
Price	2	4.92181842	2.46090921	7.66	0.0006
Color*Price	6	1.34535134	0.22422522	0.70	0.6518



3e. ANOVA of Lipstick Data With Interaction Term

The GLM Procedure



3e. ANOVA of Lipstick Data With Interaction Term

The GLM Procedure

Tukey's Studentized Range (HSD) Test for Pb

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	211
Error Mean Square	0.321419
Critical Value of Studentized Range	3.66228

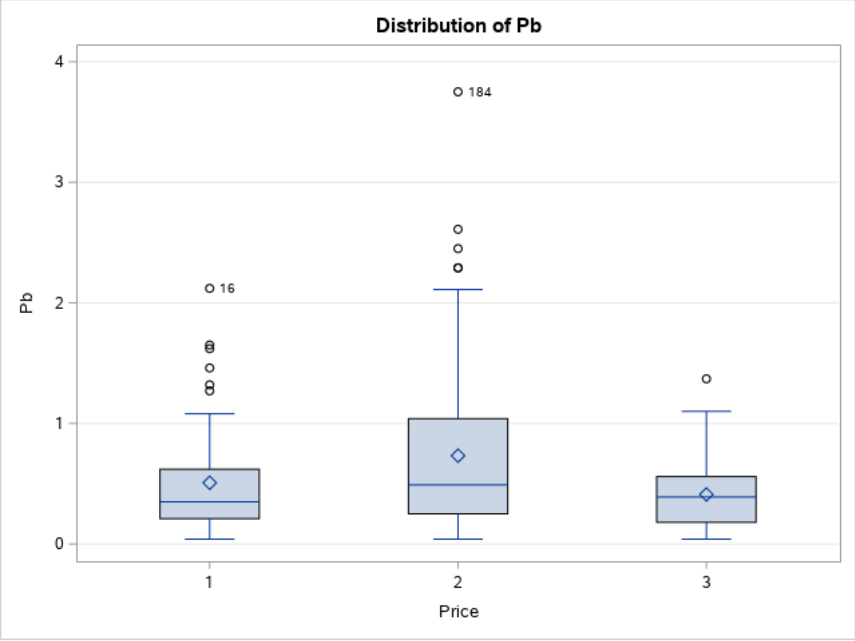
Comparisons significant at the 0.05 level are indicated by *.**

Color Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Purple - Pink	0.09729	-0.19403	0.38861	
Purple - Brown	0.09832	-0.20857	0.40520	
Purple - Red	0.28843	-0.03739	0.61424	
Pink - Purple	-0.09729	-0.38861	0.19403	
Pink - Brown	0.00102	-0.24905	0.25110	
Pink - Red	0.19114	-0.08183	0.46410	

Comparisons significant at the 0.05 level are indicated by ***.				
Color Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Brown - Purple	-0.09832	-0.40520	0.20857	
Brown - Pink	-0.00102	-0.25110	0.24905	
Brown - Red	0.19011	-0.09941	0.47964	
Red - Purple	-0.28843	-0.61424	0.03739	
Red - Pink	-0.19114	-0.46410	0.08183	
Red - Brown	-0.19011	-0.47964	0.09941	

3e. ANOVA of Lipstick Data With Interaction Term

The GLM Procedure



3e. ANOVA of Lipstick Data With Interaction Term

The GLM Procedure

Tukey's Studentized Range (HSD) Test for Pb

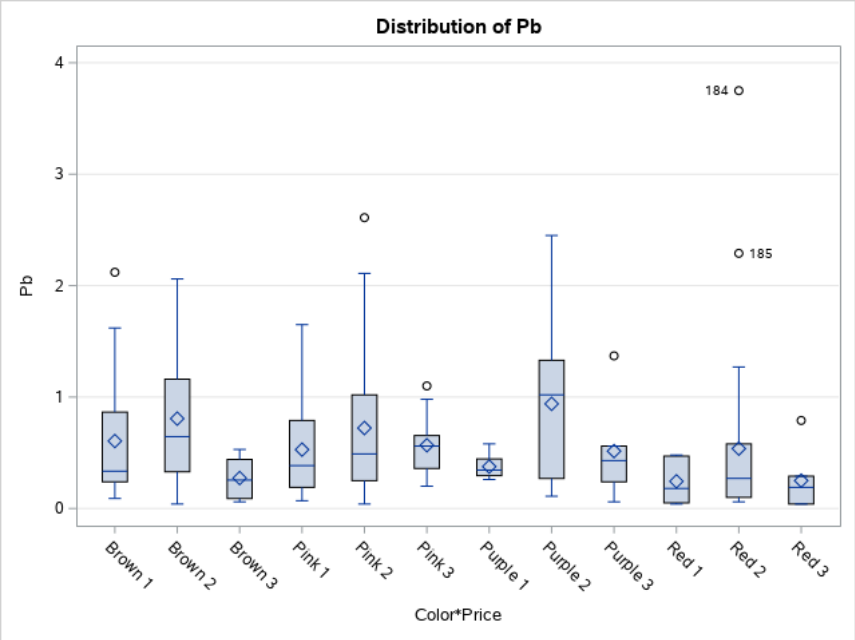
Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	211
Error Mean Square	0.321419
Critical Value of Studentized Range	3.33807

Comparisons significant at the 0.05 level are indicated by ***.				
Price Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
2 - 1	0.22512	0.00820	0.44203	***
2 - 3	0.32181	0.06798	0.57564	***
1 - 2	-0.22512	-0.44203	-0.00820	***
1 - 3	0.09669	-0.19477	0.38815	
3 - 2	-0.32181	-0.57564	-0.06798	***
3 - 1	-0.09669	-0.38815	0.19477	

3e. ANOVA of Lipstick Data With Interaction Term

The GLM Procedure



Level of Color	Level of Price	N	Pb	
			Mean	Std Dev
Brown	1	20	0.60500000	0.57626292
Brown	2	30	0.80633333	0.57777506
Brown	3	10	0.27300000	0.18973959
Pink	1	20	0.52850000	0.44187728
Pink	2	49	0.72204082	0.60942931
Pink	3	12	0.56750000	0.26608696
Purple	1	8	0.37625000	0.10888231
Purple	2	23	0.93913043	0.67529060
Purple	3	6	0.51500000	0.45460972
Red	1	5	0.24400000	0.21801376
Red	2	33	0.53757576	0.73845713
Red	3	7	0.25000000	0.25813433

3e(iii). ANOVA of Lipstick Data Without Interaction Term

The GLM Procedure

Class Level Information		
Class	Levels	Values
Color	4	Brown Pink Purple Red
Price	3	1 2 3

Number of Observations Read	223
Number of Observations Used	223

3e(iii). ANOVA of Lipstick Data Without Interaction Term

The GLM Procedure

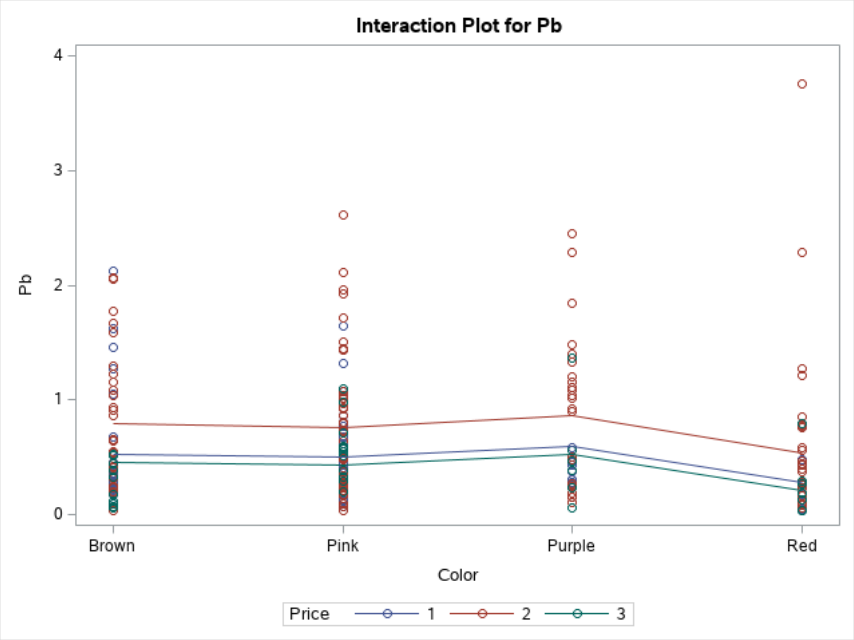
Dependent Variable: Pb Pb

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	6.37211632	1.27442326	4.00	0.0017
Error	217	69.16468009	0.31873124		
Corrected Total	222	75.53679641			

R-Square	Coeff Var	Root MSE	Pb Mean
0.084358	89.80492	0.564563	0.628655

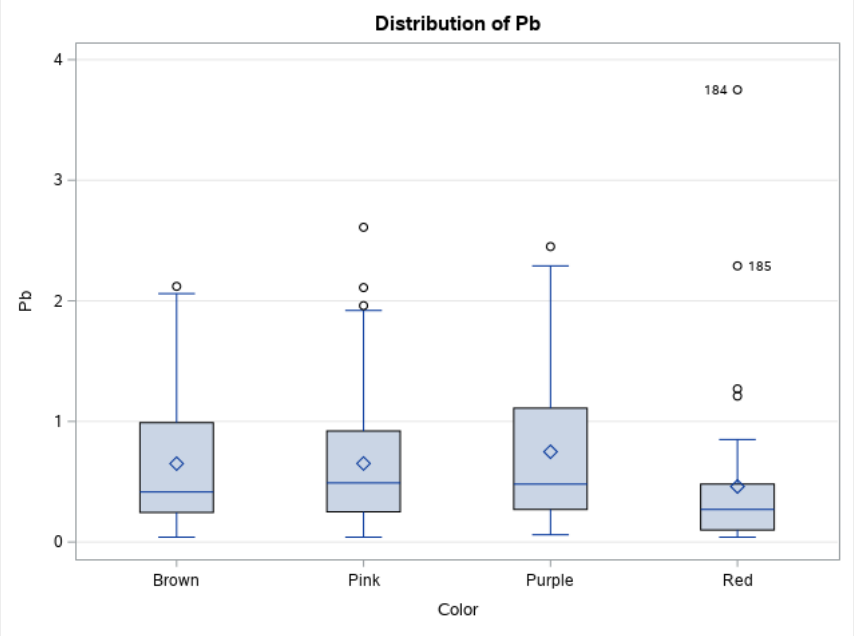
Source	DF	Type I SS	Mean Square	F Value	Pr > F
Color	3	1.87932225	0.62644075	1.97	0.1201
Price	2	4.49279407	2.24639704	7.05	0.0011

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Color	3	2.47401533	0.82467178	2.59	0.0540
Price	2	4.49279407	2.24639704	7.05	0.0011



3e(iii). ANOVA of Lipstick Data Without Interaction Term

The GLM Procedure



3e(iii). ANOVA of Lipstick Data Without Interaction Term

The GLM Procedure

Tukey's Studentized Range (HSD) Test for Pb

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	217
Error Mean Square	0.318731
Critical Value of Studentized Range	3.66147

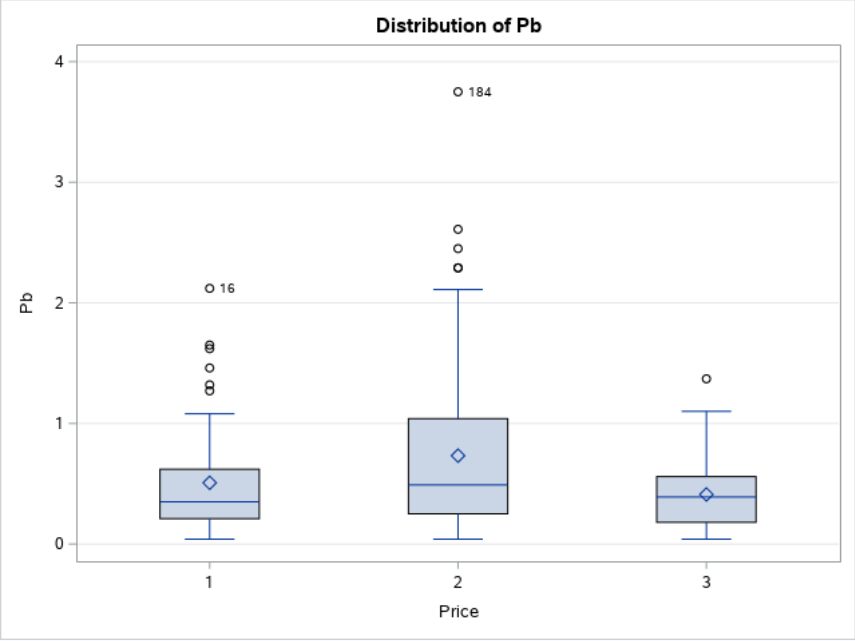
Comparisons significant at the 0.05 level are indicated by *.**

Color Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
Purple - Pink	0.09729	-0.19274	0.38733
Purple - Brown	0.09832	-0.20722	0.40385
Purple - Red	0.28843	-0.03595	0.61281
Pink - Purple	-0.09729	-0.38733	0.19274
Pink - Brown	0.00102	-0.24794	0.24999
Pink - Red	0.19114	-0.08063	0.46290

Comparisons significant at the 0.05 level are indicated by ***.				
Color Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Brown - Purple	-0.09832	-0.40385	0.20722	
Brown - Pink	-0.00102	-0.24999	0.24794	
Brown - Red	0.19011	-0.09814	0.47836	
Red - Purple	-0.28843	-0.61281	0.03595	
Red - Pink	-0.19114	-0.46290	0.08063	
Red - Brown	-0.19011	-0.47836	0.09814	

3e(iii). ANOVA of Lipstick Data Without Interaction Term

The GLM Procedure



3e(iii). ANOVA of Lipstick Data Without Interaction Term

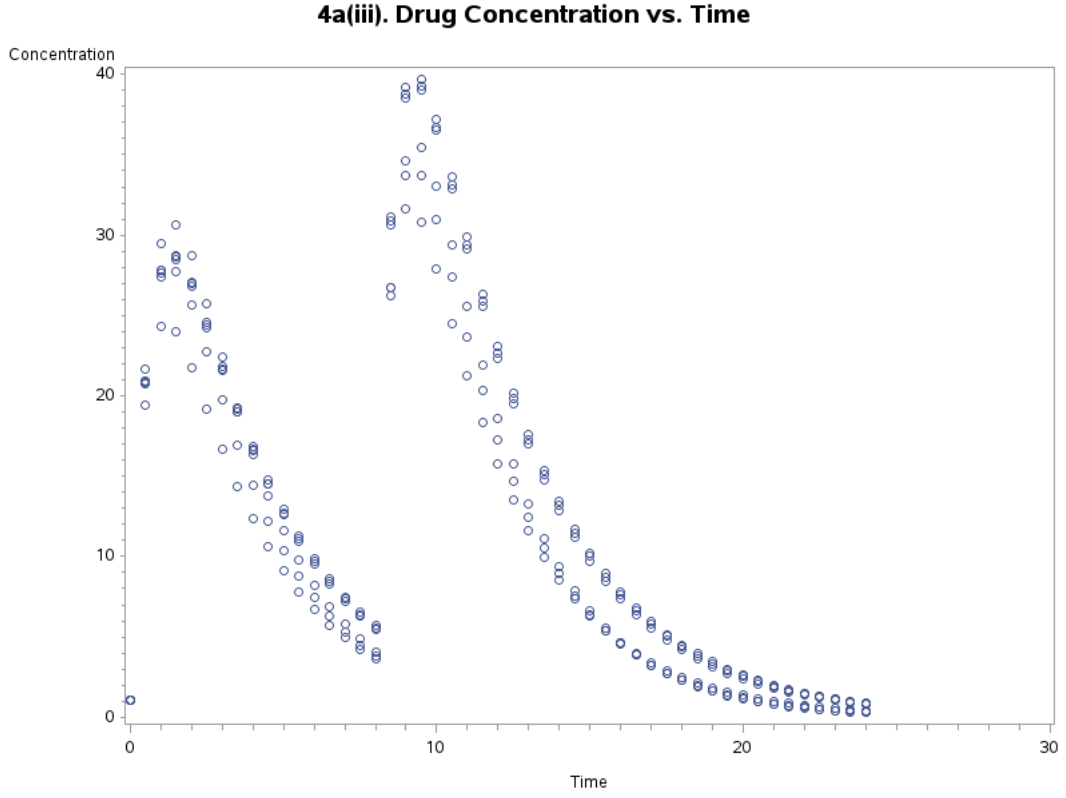
The GLM Procedure

Tukey's Studentized Range (HSD) Test for Pb

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	217
Error Mean Square	0.318731
Critical Value of Studentized Range	3.33742

Comparisons significant at the 0.05 level are indicated by ***.				
Price Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
2 - 1	0.22512	0.00916	0.44108	***
2 - 3	0.32181	0.06909	0.57452	***
1 - 2	-0.22512	-0.44108	-0.00916	***
1 - 3	0.09669	-0.19350	0.38688	
3 - 2	-0.32181	-0.57452	-0.06909	***
3 - 1	-0.09669	-0.38688	0.19350	

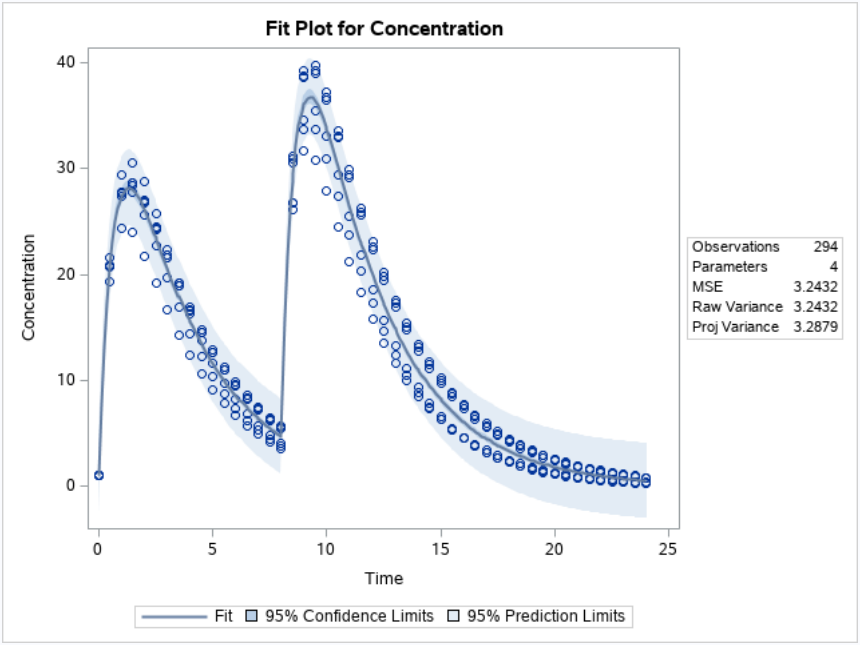


4b. Nonlinear Model of Drug Concentration Data

The NLIN Procedure
Dependent Variable Concentration
Method: Gauss-Newton

Iterative Phase					
Iter	α	β	θ1	θ2	Sum of Squares
0	1.0000	64.8000	0.3000	1.5000	1696.8
1	1.1849	64.8331	0.2987	1.5516	941.1
2	1.1903	64.6932	0.2988	1.5491	940.5
3	1.1903	64.6935	0.2988	1.5491	940.5

NOTE: Convergence criterion met.



Estimation Summary	
Method	Gauss-Newton
Iterations	3
R	6.201E-8
PPC(θ1)	1.777E-8

Estimation Summary	
RPC(θ2)	8.97E-6
Object	1.506E-9
Objective	940.5222
Observations Read	294
Observations Used	294
Observations Missing	0

Note: An intercept was not specified for this model.

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	4	78611.6	19652.9	6059.76	<.0001
Error	290	940.5	3.2432		
Uncorrected Total	294	79552.1			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
α	1.1903	0.0181	1.1548	1.2259
β	64.6935	1.9200	60.9145	68.4725
θ1	0.2988	0.00650	0.2860	0.3116
θ2	1.5491	0.0607	1.4297	1.6685

Approximate Correlation Matrix				
	α	β	θ1	θ2
α	1.0000000	-0.3525342	0.2738070	-0.1585911
β	-0.3525342	1.0000000	-0.5062918	0.8969524
θ1	0.2738070	-0.5062918	1.0000000	-0.7554343
θ2	-0.1585911	0.8969524	-0.7554343	1.0000000