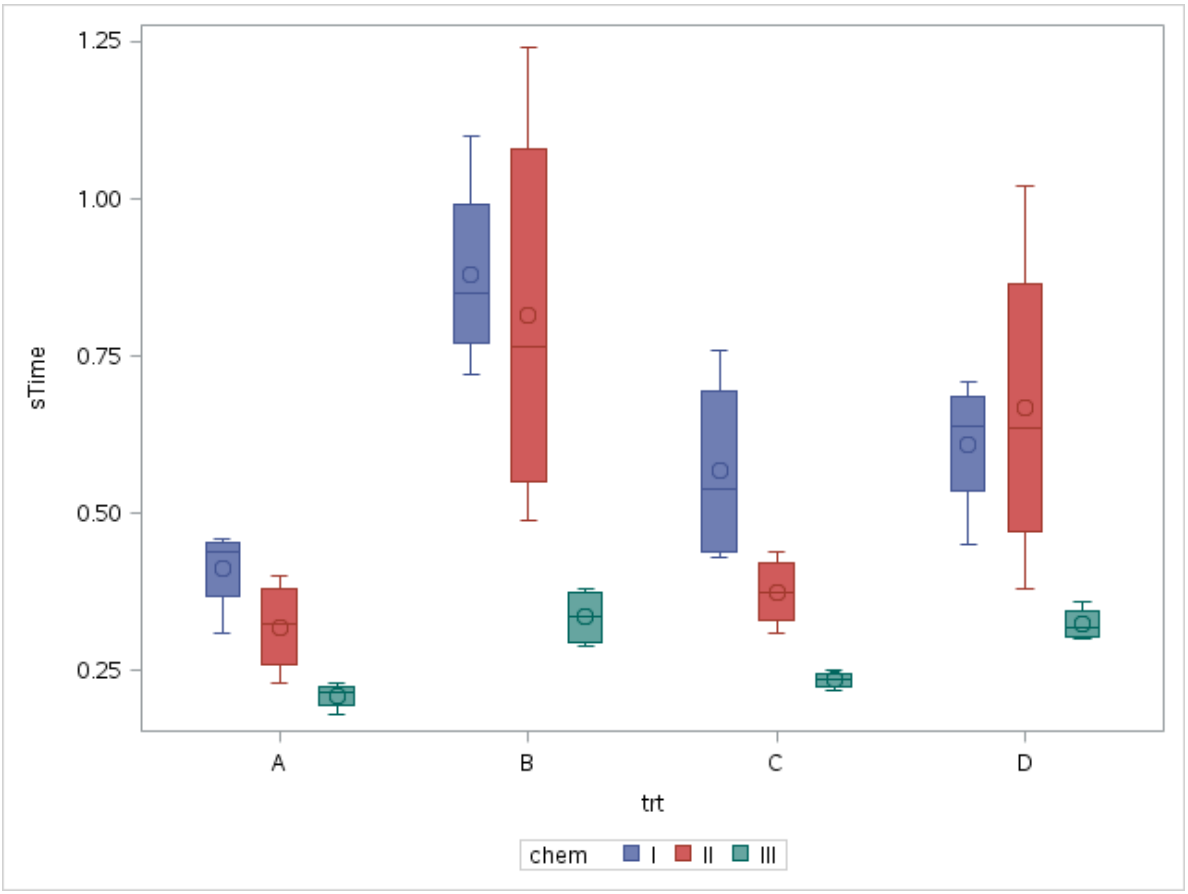
Homework 1 Stat 506

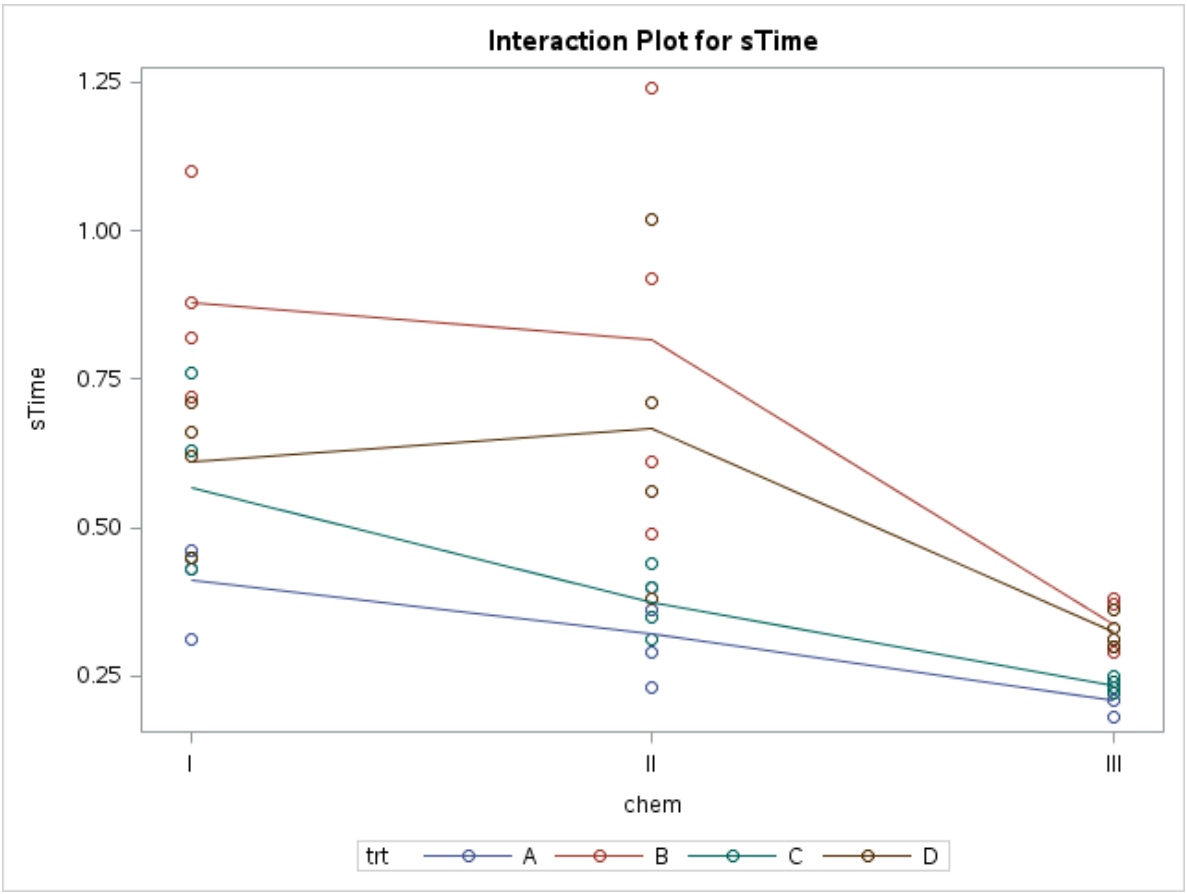
Leslie Gains-Germain

01/23/2015

Refer back to HW3 of Stat 505. Reanalyze the same data in SAS.

1. Plot the data in a manner which allows us to easily compare survival at each combination of treatment and chemical.





2. Write out the full interaction model.

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk},\ i \in \{A, B, C, D\}, j = 1, \ldots, 3,\ k = 1, \ldots, 4$$

3. Fit the interaction model in SAS.

    i. Show the estimated coefficients.

Parameter	Estimate		Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	0.3250000000	B	0.07456937	4.36	0.0001	0.1737663171	0.4762336829
chem I	0.2850000000	B	0.10545701	2.70	0.0104	0.0711232745	0.4988767255
chem II	0.3425000000	B	0.10545701	3.25	0.0025	0.1286232745	0.5563767255
chem III	0.0000000000	B	.	.	.	.	.
trt A	-.1150000000	B	0.10545701	-1.09	0.2827	-.3288767255	0.0988767255
trt B	0.0100000000	B	0.10545701	0.09	0.9250	-.2038767255	0.2238767255
trt C	-.0900000000	B	0.10545701	-0.85	0.3991	-.3038767255	0.1238767255
trt D	0.0000000000	B	.	.	.	.	.
chem*trt I A	-.0825000000	B	0.14913873	-0.55	0.5836	-.3849673658	0.2199673658
chem*trt I B	0.2600000000	B	0.14913873	1.74	0.0898	-.0424673658	0.5624673658
chem*trt I C	0.0475000000	B	0.14913873	0.32	0.7519	-.2549673658	0.3499673658
chem*trt I D	0.0000000000	B	.	.	.	.	.
chem*trt II A	-.2325000000	B	0.14913873	-1.56	0.1278	-.5349673658	0.0699673658
chem*trt II B	0.1375000000	B	0.14913873	0.92	0.3627	-.1649673658	0.4399673658
chem*trt II C	-.2025000000	B	0.14913873	-1.36	0.1830	-.5049673658	0.0999673658
chem*trt II D	0.0000000000	B	.	.	.	.	.
chem*trt III A	0.0000000000	B	.	.	.	.	.
chem*trt III B	0.0000000000	B	.	.	.	.	.
chem*trt III C	0.0000000000	B	.	.	.	.	.
chem*trt III D	0.0000000000	B	.	.	.	.	.

ii. Provide either the Type I or Type III output table, and explain why you think this table is preferred. Is the interaction needed?

Source	DF	Type I SS	Mean Square	F Value	Pr > F
chem	2	1.03301250	0.51650625	23.22	<.0001
trt	3	0.92120625	0.30706875	13.81	<.0001
chem*trt	6	0.25013750	0.04168958	1.87	0.1123

*There is only weak evidence that the difference in mean survival times among levels of the chemical changes across treatment (p-value=*0.1123 *from F-stat=* 1.87 *on* 6 *and* 36 *df). The interaction is not needed.*

I prefer the type I sequential sums of squares in this case. The chemical row in the above table compares a single mean model to a model with chemical as the only predictor. The treatment row in the above table compares a model with chemical and treatment as predictors to a model with chemical as the only predictor.

The last row in the table is the same regardless of whether we use type I or type III. If we use type III sums of squares, the chemical and treatment rows don't make much sense because we usually do not remove main effects when they are included in an interaction term in the model. For this reason, I like type I sums of squares better.

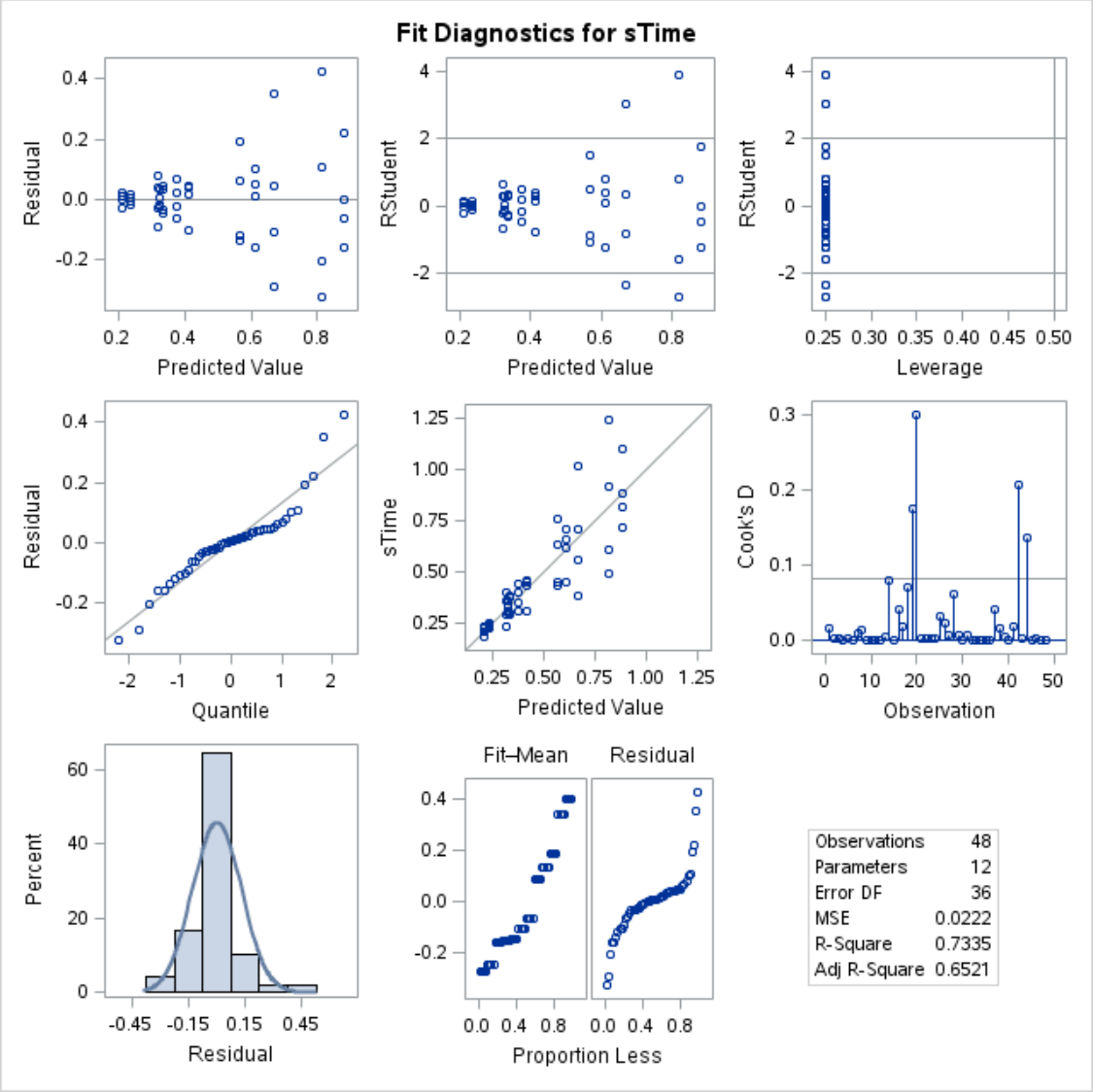
ii. What combination of Greek letters is estimated by each coefficient shown?

Note here that SAS sets the last level of each variable as the reference level. Therefore, treatment D and chemical III are the reference levels. The following table shows what combination of Greek letters are estimated by the coefficients that are not set to 0.

SAS row	Greek Letters
Intercept	$\mu + \alpha_D + \beta_3 + (\alpha\beta)_{D3}$
Chem 1	$\beta_1 - \beta_3 + (\alpha\beta)_{D1} - (\alpha\beta)_{D3}$
Chem 2	$\beta_2 - \beta_3 + (\alpha\beta)_{D2} - (\alpha\beta)_{D3}$
Trt A	$\alpha_A - \alpha_D + (\alpha\beta)_{A3} - (\alpha\beta)_{D3}$
Trt B	$\alpha_B - \alpha_D + (\alpha\beta)_{B3} - (\alpha\beta)_{D3}$
Trt C	$\alpha_C - \alpha_D + (\alpha\beta)_{C3} - (\alpha\beta)_{D3}$
trt*chemA1	$(\alpha\beta)_{A1} - (\alpha\beta)_{D1} - (\alpha\beta)_{A3} + (\alpha\beta\ )_{D3}$
trt*chemB1	$(\alpha\beta)_{B1} - (\alpha\beta)_{D1} - (\alpha\beta)_{B3} + (\alpha\beta)_{D3}$
trt*chemC1	$(\alpha\beta)_{C1} - (\alpha\beta)_{D1} - (\alpha\beta)_{C3} + (\alpha\beta)_{D3}$
trt*chemA2	$(\alpha\beta)_{A2} - (\alpha\beta)_{D2} - (\alpha\beta)_{A3} + (\alpha\beta)_{D3}$
trt*chemB2	$(\alpha\beta)_{B2} - (\alpha\beta)_{D2} - (\alpha\beta)_{A3} + (\alpha\beta)_{D3}$
trt*chemC2	$(\alpha\beta)_{C2} - (\alpha\beta)_{D2} - (\alpha\beta)_{C3} + (\alpha\beta)_{D3}$

iv. Provide the default diagnostic plots and comment on how well the assumptions are met.

The constant variance assumption is clearly not met. We see a funnel pattern in the residuals vs. predicted values plot, with increasing spread across predicted values. There are also several influential observations, and the distribution of observations is long-tailed.

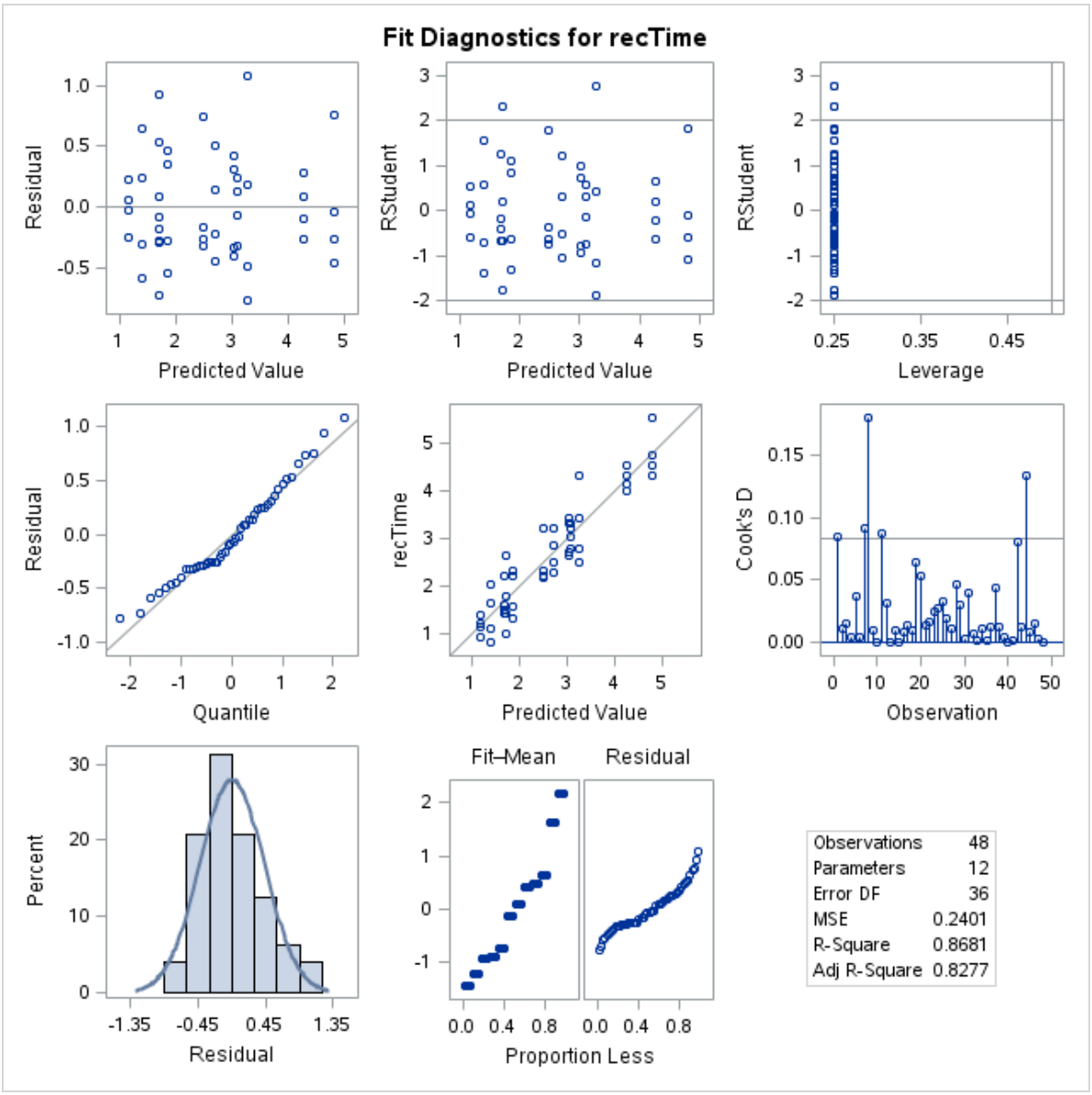


4. We saw that Box and Cox used these data to illustrate transformations. Transform as you did for Stat 505 and redo the table of coefficients, Type II or III table and diagnostic plots.

   There is still no evidence for an interaction effect, and the p-value for the interaction term has changed to 0.3867. *The estimates have also changed. The diagnostic plots look much better and there do not seem to be any problems with the assumptions of normality or constant variance.*

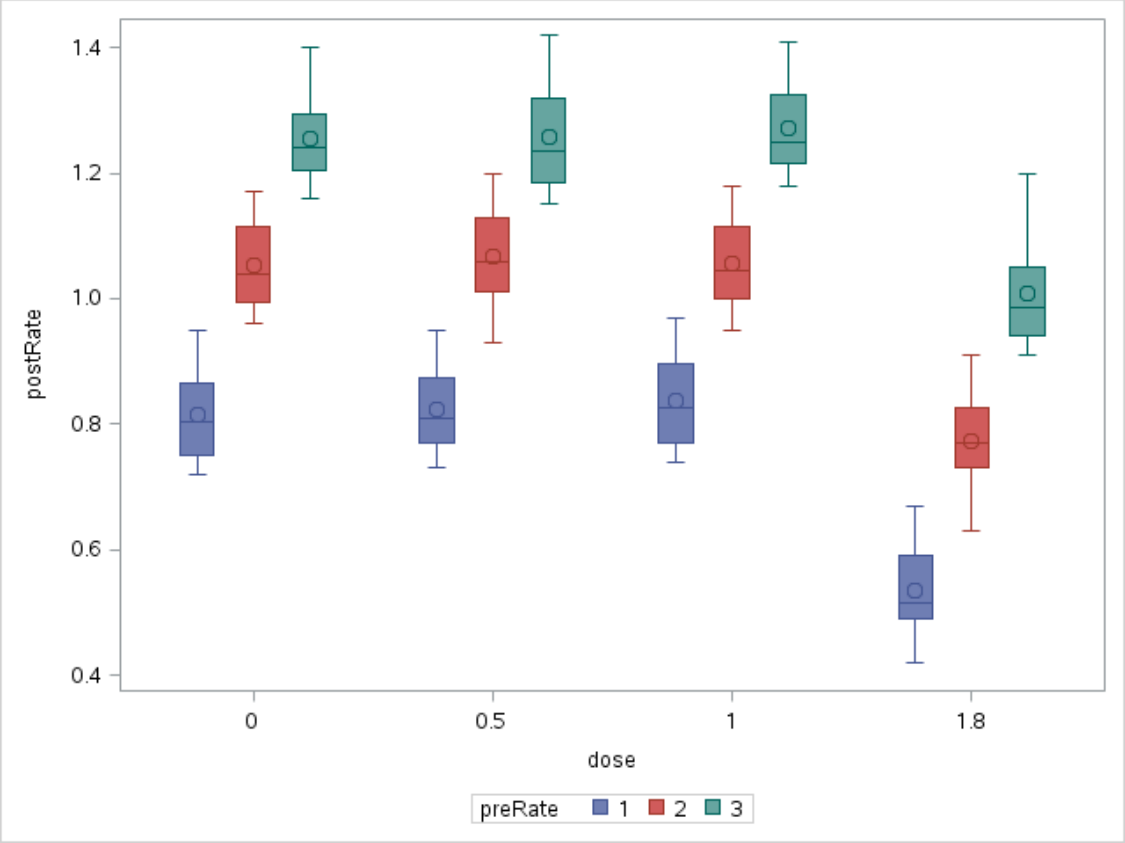
Source	DF	Type III SS	Mean Square	F Value	Pr > F
chem	2	34.87711982	17.43855991	72.63	<.0001
trt	3	20.41428935	6.80476312	28.34	<.0001
chem*trt	6	1.57077226	0.26179538	1.09	0.3867

Parameter	Estimate		Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	3.091805148	B	0.24499267	12.62	<.0001	2.594936976	3.588673320
chem I	-1.402123231	B	0.34647196	-4.05	0.0003	-2.104800939	-0.699445524
chem II	-1.390271125	B	0.34647196	-4.01	0.0003	-2.092948832	-0.687593417
chem III	0.000000000	B	.	.	.	.	.
trt A	1.710880089	B	0.34647196	4.94	<.0001	1.008202382	2.413557797
trt B	-0.062832437	B	0.34647196	-0.18	0.8571	-0.765510144	0.639845271
trt C	1.173181677	B	0.34647196	3.39	0.0017	0.470503969	1.875859384
trt D	0.000000000	B	.	.	.	.	.
chem*trt I A	-0.913681228	B	0.48998535	-1.86	0.0704	-1.907417572	0.080055116
chem*trt I B	-0.463385573	B	0.48998535	-0.95	0.3506	-1.457121916	0.530350771
chem*trt I C	-1.000139924	B	0.48998535	-2.04	0.0486	-1.993876268	-0.006403580
chem*trt I D	0.000000000	B	.	.	.	.	.
chem*trt II A	-0.143944181	B	0.48998535	-0.29	0.7706	-1.137680525	0.849792163
chem*trt II B	-0.245309406	B	0.48998535	-0.50	0.6197	-1.239045750	0.748426938
chem*trt II C	-0.160796555	B	0.48998535	-0.33	0.7447	-1.154532899	0.832939789
chem*trt II D	0.000000000	B	.	.	.	.	.
chem*trt III A	0.000000000	B	.	.	.	.	.
chem*trt III B	0.000000000	B	.	.	.	.	.
chem*trt III C	0.000000000	B	.	.	.	.	.



5. In HW 5 last fall we looked at thirsty albino rats' rate of level pressing.

   i. Provide appropriate plots.



ii. Write out the model.

The random effects model is as follows:

$$\mathbf{y_i} = \mathbf{X_i}\beta + b_i\mathbf{1} + \epsilon_\mathbf{i}$$
$$b_i \sim N(0, \sigma_b^2)$$
$$\epsilon_i \sim N(0, \sigma^2)$$

where the b_i's are the rats with $i \in \{1,2,3,...,12\}$ and

$$\beta = \begin{bmatrix} \mu \\ \tau_L \\ \tau_M \\ \tau_H \\ \alpha_0 \\ \alpha_{0.5} \\ \alpha_1 \\ \alpha_{1.8} \end{bmatrix}$$

where the τ_k's are the preRate effects and the α_j's are the dose effects.

iii. Fit using PROC MIXED and interpret the relevant output.

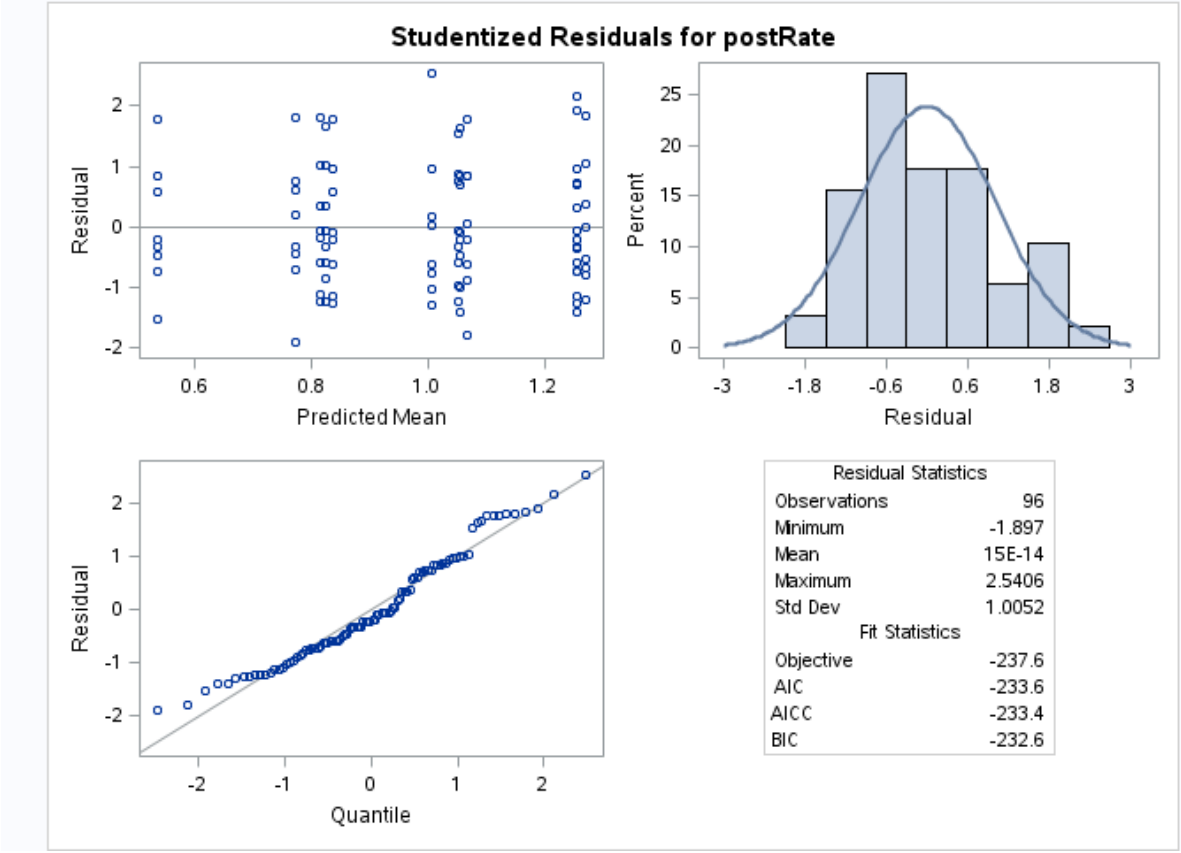
Covariance Parameter Estimates	
Cov Parm	Estimate
ratID	0.005533
Residual	0.001819

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
preRate	2	75	34.46	<.0001
dose	3	75	251.68	<.0001
preRate*dose	6	75	0.59	0.7378

Solution for Fixed Effects										
Effect	dose	preRate	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept			1.0075	0.04013	9	25.10	<.0001	0.05	0.9167	1.0983
preRate		1	-0.4725	0.05675	75	-8.33	<.0001	0.05	-0.5856	-0.3594
preRate		2	-0.2337	0.05675	75	-4.12	<.0001	0.05	-0.3468	-0.1207
preRate		3	0	.	.	.	.	.	.	.
dose	0		0.2475	0.02132	75	11.61	<.0001	0.05	0.2050	0.2900
dose	0.5		0.2488	0.02132	75	11.67	<.0001	0.05	0.2063	0.2912
dose	1		0.2638	0.02132	75	12.37	<.0001	0.05	0.2213	0.3062
dose	1.8		0	.	.	.	.	.	.	.
preRate*dose	0	1	0.03125	0.03016	75	1.04	0.3034	0.05	-0.02882	0.09132
preRate*dose	0.5	1	0.04000	0.03016	75	1.33	0.1887	0.05	-0.02007	0.1001
preRate*dose	1	1	0.03750	0.03016	75	1.24	0.2175	0.05	-0.02257	0.09757
preRate*dose	1.8	1	0	.	.	.	.	.	.	.
preRate*dose	0	2	0.03250	0.03016	75	1.08	0.2846	0.05	-0.02757	0.09257
preRate*dose	0.5	2	0.04375	0.03016	75	1.45	0.1510	0.05	-0.01632	0.1038
preRate*dose	1	2	0.01875	0.03016	75	0.62	0.5360	0.05	-0.04132	0.07882
preRate*dose	1.8	2	0	.	.	.	.	.	.	.
preRate*dose	0	3	0	.	.	.	.	.	.	.
preRate*dose	0.5	3	0	.	.	.	.	.	.	.
preRate*dose	1	3	0	.	.	.	.	.	.	.
preRate*dose	1.8	3	0	.	.	.	.	.	.	.

There is strong evidence that the mean post experiment pressing rate depends on the pre-experiment pressing rate after accounting for rat and dose (p-value$<$ 0.0001 *from F-stat*$=$ 34.46 *on* 2 *and* 9 *df). There is also strong evidence that the mean post experiment pressing rate depends on drug dose after accounting for rat and pre-experiment rate (p=value*$<$ 0.0001 *from F-stat*$=$ 251.68 *on* 3 *and* 75 *df). There is no evidence that the difference in the mean post experiment pressing rates among doses changes across pre-experiment pressing rates, after accounting for rat (p-value*$=$ 0.7378 *from F-stat*$=$ 0.59 *on* 6 *and* 75 *df). We also see that the covariance between observations on the same rat is estimated to be* 0.005533.

iv. Discuss the diagnostic plots.



The studentized residuals are mostly between -2 and 2 and the spread is mostly constant across predicted values. There is a long right tail shown in the normal Q-Q plot, but it is not too extreme and I don't expect it to severely interfere with our tests.

Code Appendix

```
DATA rats;
  INFILE "/folders/myfolders/ratSurvival.csv" firstobs=2 delimiter =',';
  INPUT chem $ trt $ sTime;
  ;
RUN;
```

```
PROC SGPLOT data=rats;
  vbox sTime / category=trt group=chem;
RUN;
```

```
PROC GLM data = rats plots=diagnostics;
  CLASS chem trt;
  MODEL sTime = chem|trt / ss3 SOLUTION clparm;
RUN;
```

```
DATA recrats;
  SET rats;
  recTime = 1/sTime;
```

```
  ;
RUN;

PROC GLM data = recrats plots=diagnostics;
  CLASS chem trt;
  MODEL recTime = chem|trt / ss3 SOLUTION clparm;
RUN;
```

```
DATA drugs;
  INFILE "/folders/myfolders/drugResponse.csv" firstobs=2 delimiter =',';
  INPUT arm ratID $ preRate dose $ rep postRate;
  ;
RUN;
```

```
PROC SGPLOT data=drugs;
  vbox postRate / category=dose group=preRate;
RUN;
```

```
PROC MIXED DATA=drugs plots=all;
  CLASS ratID preRate dose rep;
  MODEL postRate = preRate|dose / SOLUTION CL;
  RANDOM ratID;
RUN;
```