Day 1, Problem 4 Solution

a) $Y_{ij} = \mu_{\cdot\cdot} + \rho_i + \tau_j + \varepsilon_{ij}$

$i = 1, 2, 3, \quad j = 1, 2, \ldots, 7$

$\mu_{\cdot\cdot}$ is a constant

ρ_i's are block effects, $\sum_i \rho_i = 0$

τ_j's are treatment effects, $\sum_j \tau_j = 0$

$\varepsilon_{ij} \overset{indep}{\sim} N(0, \sigma^2)$

b) Balanced, since the Type I SS and Type III SS are the same.

c) Letting $\mu_{ij} = \mu_{\cdot\cdot} + \rho_i + \tau_j$, we test

$H_0: \mu_{11} = \mu_{12} = \cdots = \mu_{37}$

We use $F^* = \dfrac{MS(\text{cells})}{MSE}$

where $SS(\text{Cells}) = SSTR + SS(Blk)$

Reject H_0 if $F^* > F_{.95, 8, 12} = 2.85$

Our $SS(\text{cells}) = 39.037 + 103.151 = 142.188$

$\Rightarrow MS(\text{cells}) = \dfrac{142.188}{8} = 17.774$

$F^* = \dfrac{17.774}{1.6452} = 10.8 > 2.85$

(since $\sqrt{MSE} = 1.282668 \Rightarrow MSE = 1.6452$) — and error d.f $= 20 - 2 - 6 = 12$

So reject H_0 and conclude the mean stem length is not equal for all treatment-block combinations.

Day 1, Problem 4 Solution (Continued)

d) Contrast (1): $H_0: \mu_{comp} - \frac{1}{6}\sum_{j \neq comp} \mu_j = 0$

$$t^* = \frac{29.67 - 33.04}{\sqrt{1.6452\left(\frac{1^2 + (-\frac{1}{6})^2 + \cdots + (-\frac{1}{6})^2}{3}\right)}} = \frac{29.67 - 33.04}{\sqrt{1.6452\left(\frac{1^2 + 6(\frac{1}{6})^2}{3}\right)}}$$

$$= \frac{-3.37}{0.80} = -4.21$$

$$|t^*| > t_{(1 - \frac{.05}{2(2)}, 12)} = 2.56, \text{ so reject } H_0$$

↑ Bonferroni correction

Conclude the mean stem length for the compost type differs from the mean for the non-Compost types.

Contrast (2): $H_0: \mu_{clar} - \mu_{web} = 0$

$$t^* = \frac{32.17 - 31.1}{\sqrt{1.6452\left(\frac{1^2 + (-1)^2}{3}\right)}} = \frac{1.07}{1.0473} = 1.022$$

$$|t^*| < t_{(1 - \frac{.05}{2(2)}, 12)} = 2.56, \text{ so fail to reject } H_0.$$

Conclude the mean stem length for Clarion type <u>may</u> <u>not</u> <u>differ</u> from the mean stem length for Webster type.

Note: There are $n_T - 1 = 20$ total d.f., so there are 21 total observations, or 3 <u>per</u> <u>type</u>.

Day 1, Problem 4 Solution (Continued)

e) From the output, note $MSE = (1.282668)^2 = 1.6452$
and there are $20 - 2 - 6 = 12$ error d.f.

Hence $SSE = (12)(1.6452) = 19.7424$.

If Region is ignored, SSE becomes

$$19.7424 + SS(Blk) = 19.7424 + 39.03714$$
$$= 58.7795$$

The new error d.f. become $12 + 2 = 14$.

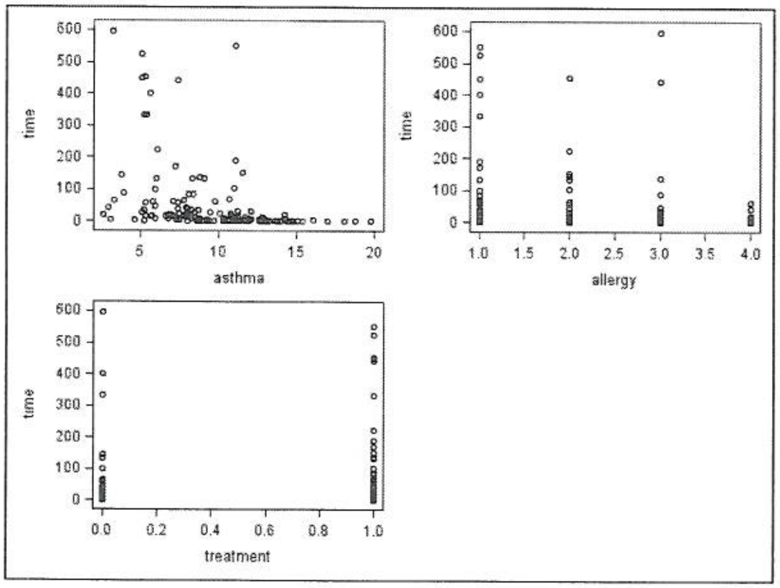
$\Rightarrow$ new MSE is $\dfrac{58.7795}{14} = 4.199$

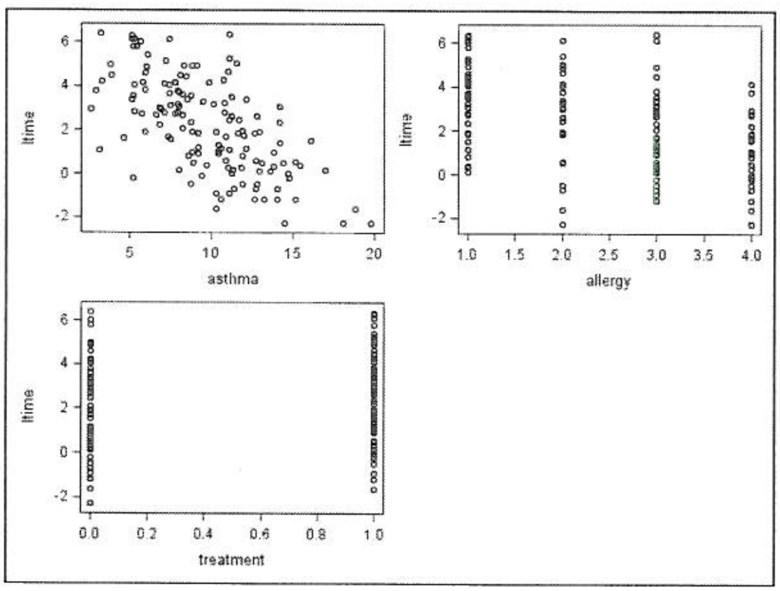
We reject $H_0: \tau_1 = \tau_2 = \cdots = \tau_7 = 0$ if

$$F^* = \frac{MS(Trt)}{MSE} > F_{.95,6,14} = 2.85$$

Our $F^* = \dfrac{17.192}{4.199} = 4.094 > 2.85$, so
we would reject H_0 and conclude the mean lengths are not the same across soil types.

Day 1, Problem 5 Solution (1)





2 Asthma medication

The following is a basic analysis. Students can certainly go far beyond what is below.

Clearly one or more transformations are needed in the response and/or covariates:

A common transformation for time-to-event data is the natural log. Using log-time to wheezing instead of time to wheezing yields (marginally) constant variance and approximately linear effects in asthma and allergy scores.

One can also formally check for the "best" Box-Cox transformation; SAS PROC TRANSREG picks the natural log. Simultaneously checking for transformations in allergy and asthma yields approximate identity transformations (added variable plots should show the same thing for the log response). So we work with the log-time as response, and linear effects in asthma, allergy, and treatment.

Working backwards from a model with main effects, quadratic in asthma, and interactions yields the final model

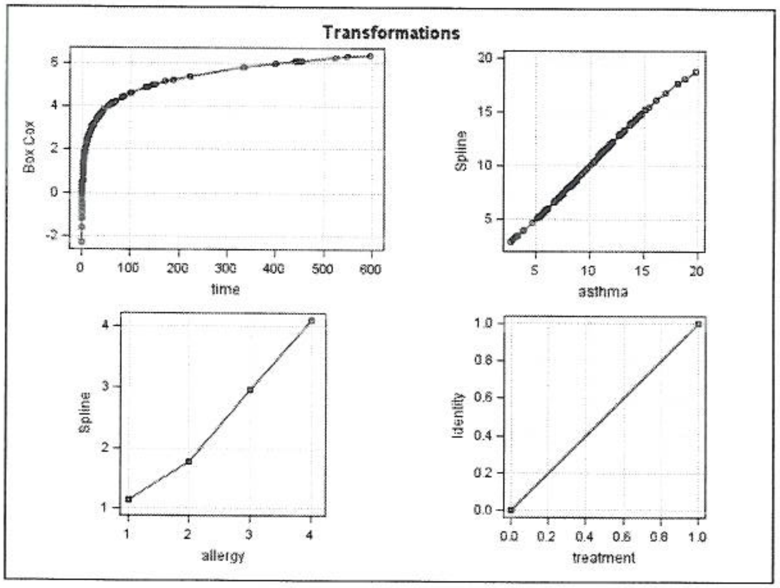
$$Log(T_i) = \beta_0 + \beta_1 \text{asthma}_i + \beta_2 \text{allergy}_i + \beta_3 \text{treatment}_i + e_i,$$

where the e_i are iid $N(0, \sigma^2)$. The fitted model is

$$\widehat{\log(T_i)} = 8.091 - 0.406\text{asthma}_i - 0.970\text{allergy}_i + 0.818\text{treatment}_i.$$

D1, P5

(2)




```
                          The REG Procedure
                           Model: MODEL1
                     Dependent Variable: ltime

            Number of Observations Read            138
            Number of Observations Used            138

                        Analysis of Variance

                                 Sum of           Mean
Source                  DF      Squares         Square    F Value    Pr > F
Model                    3    444.34159      148.11386     114.01    <.0001
Error                  134    174.08391        1.29913
Corrected Total        137    618.42550

          Root MSE              1.13980    R-Square     0.7185
          Dependent Mean        2.16545    Adj R-Sq     0.7122
          Coeff Var            52.63541

                        Parameter Estimates

                     Parameter       Standard
Variable     DF       Estimate          Error    t Value    Pr > |t|
Intercept     1        8.09128        0.38748      20.88      <.0001
asthma        1       -0.40647        0.02820     -14.42      <.0001
allergy       1       -0.96959        0.08813     -11.00      <.0001
treatment     1        0.81772        0.19458       4.20      <.0001
```

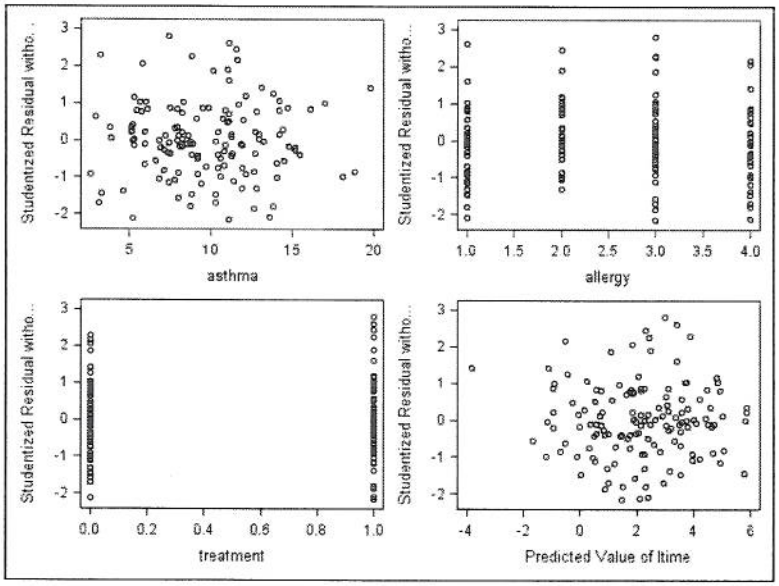
The residual plots show good fit of the model. There is one fitted log wheezing time that is substantially smaller than the rest.

The main goal of the study is to estimate the treatment effect while controlling for asthma & allergy severity. Note that increasing asthma & allergy severity decrease the time to wheezing – that is, the worse a subject's asthma or allergies are, the less time it takes to start wheezing. The treatment effect is given by $e^{0.81772} = 2.27$. Under the treatment, the time to wheezing is at least double, and highly significant.

Obs	ltime	time	asthma	allergy	treatment	c	r
1	-1.60944	0.2	18.8	2	1	0.012737	-0.85024
2	4.23700	69.2	10.7	1	1	0.002244	0.57366
3	3.04452	21.0	8.0	2	0	0.000077	0.12719
4	6.10613	448.6	5.1	1	1	0.000477	0.21385
5	4.11578	61.3	7.1	1	1	0.005512	-0.83465
6	4.92071	137.1	8.8	3	1	0.020875	2.24272
7	3.82647	45.9	5.9	3	1	0.000252	0.19863
8	4.98292	145.9	3.8	2	0	0.001129	0.33476
9	4.41401	82.6	8.4	1	1	0.000068	-0.09845
10	0.18232	1.2	8.0	3	0	0.010689	-1.55583
11	1.16315	3.2	12.1	1	0	0.007142	-0.92724
12	0.99325	2.7	8.8	4	1	0.001293	-0.40889

D1, P5

(3)



13	0.09531	1.1	12.9	3	0	0.000110	0.13810
14	0.83291	2.3	8.6	4	0	0.000075	0.10247
15	1.56862	4.8	7.5	4	1	0.001120	-0.36755
16	3.60278	36.7	8.2	1	0	0.000220	-0.16508
17	1.50408	4.5	16.1	1	0	0.009682	0.83417
18	-0.51083	0.6	11.9	3	0	0.002780	-0.75737
19	-0.10536	0.9	9.4	4	1	0.010450	-1.17300
20	4.14946	63.4	7.8	2	1	0.000427	0.30882
21	0.53063	1.7	15.1	2	1	0.000617	-0.26810
22	0.18232	1.2	11.3	3	0	0.000579	-0.35919
23	0.26236	1.3	11.8	3	0	0.000057	-0.10928
24	0.58779	1.8	12.7	2	0	0.000708	-0.35561
25	4.88658	132.5	9.1	1	1	0.002230	0.57296
26	4.88204	131.9	6.0	2	0	0.007009	1.03894
27	1.02962	2.8	10.4	4	1	0.000315	0.20082
28	1.85630	6.4	12.6	1	1	0.005796	-0.85605
29	0.00000	1.0	14.7	4	1	0.008758	0.84761
30	-0.69315	0.5	12.7	4	0	0.000449	0.22767
31	6.26015	523.3	5.1	1	1	0.001288	0.35137
32	0.58779	1.8	10.9	3	0	0.000090	-0.14480
33	0.47000	1.6	8.9	4	0	0.000087	-0.11112
34	0.47000	1.6	13.1	4	0	0.018066	1.41800
35	1.09861	3.0	3.1	4	0	0.038548	-1.68253
36	3.52342	33.9	8.7	1	0	0.000024	-0.05504
37	-1.60944	0.2	10.3	4	0	0.014957	-1.46142
38	1.13140	3.1	10.6	3	1	0.001067	-0.49441
39	2.26176	9.6	6.8	4	1	0.000000	-0.00436
40	2.33214	10.3	14.2	1	1	0.000211	0.14663
41	3.37074	29.1	5.1	3	0	0.000397	0.23169
42	0.99325	2.7	13.8	3	0	0.010805	1.26542
43	-2.30259	0.1	19.8	4	0	0.050283**	1.41614
44	-0.51083	0.6	12.8	2	0	0.009498	-1.29881
45	4.19268	66.2	3.3	1	0	0.030479	-1.43938
46	-0.69315	0.5	14.0	2	0	0.007446	-1.02751
47	1.87180	6.5	11.6	4	0	0.034269	2.14256
48	5.14575	171.7	7.2	1	1	0.000109	0.11797
49	1.33500	3.8	10.4	3	1	0.000643	-0.38631
50	2.44235	11.5	11.1	2	0	0.002261	0.70862
51	5.80664	332.5	5.4	1	0	0.006763	0.78745
52	3.49347	32.9	11.2	1	1	0.000062	0.09437
53	0.00000	1.0	11.2	4	1	0.001474	-0.42501
54	5.81054	333.8	5.2	1	1	0.000002	-0.01359
55	2.48491	12.0	11.7	2	0	0.004456	0.96382
56	3.37417	29.2	8.5	2	1	0.000064	-0.12398
57	3.01553	20.4	6.8	2	1	0.005812	-1.05562
58	1.41099	4.1	14.2	3	1	0.008417	1.05360
59	3.64545	38.3	7.4	1	1	0.010026	-1.14682
60	3.69635	40.3	8.0	1	1	0.005621	-0.88148
61	4.12390	61.8	9.8	1	0	0.005795	0.87686
62	4.59815	99.3	5.9	1	0	0.000128	-0.11181
63	3.09104	22.0	7.5	3	0	0.003423	0.84671

D1, P5 (4)

64	2.59525	13.4	11.2	2	1	0.000104	0.15678
65	2.68785	14.7	6.6	3	1	0.001746	-0.55718
66	1.09861	3.0	11.1	4	1	0.002152	0.51535
67	-0.51083	0.6	8.7	4	1	0.024594	-1.80108
68	5.98746	398.4	5.6	1	0	0.011072	1.02308
69	0.47000	1.6	14.4	3	1	0.000654	0.28674
70	3.74715	42.4	2.9	4	0	0.005893	0.64169
71	1.90211	6.7	5.9	4	1	0.004131	-0.65169
72	0.91629	2.5	10.4	3	1	0.002460	-0.75682
73	0.33647	1.4	13.8	1	1	0.029214	-1.79582
74	2.75366	15.7	5.7	4	0	0.005488	0.76516
75	2.69463	14.8	8.2	3	1	0.000003	0.02418
76	1.66771	5.3	10.9	3	0	0.002808	0.80917
77	0.18232	1.2	17.0	3	1	0.012733	0.98288
78	0.26236	1.3	11.0	4	0	0.001571	0.46199
79	-0.22314	0.8	5.2	4	0	0.043251**	-2.10443
80	2.76632	15.9	7.9	4	1	0.005703	0.84293
81	-0.91629	0.4	10.3	3	0	0.011721	-1.70332
82	3.19458	24.4	10.8	1	1	0.000679	-0.31460
83	2.66723	14.4	8.2	3	0	0.002286	0.72239
84	6.09108	441.9	7.4	3	1	0.037198	2.81542**
85	1.85630	6.4	9.2	2	0	0.001034	-0.49083
86	3.38439	29.5	12.1	2	1	0.006643	1.18219
87	0.83291	2.3	11.9	1	0	0.013675	-1.29769
88	1.91692	6.8	12.9	1	0	0.000011	0.03438
89	0.09531	1.1	13.6	1	1	0.038308	-2.09412
90	5.23963	188.6	11.1	1	1	0.017976	1.62138
91	3.18221	24.1	10.1	3	0	0.014082	1.87928
92	-1.20397	0.3	15.1	3	0	0.000435	-0.22135
93	2.80940	16.6	7.1	3	1	0.000377	-0.26933
94	-1.20397	0.3	12.6	3	1	0.019192	-1.86471
95	-0.91629	0.4	11.1	3	1	0.020576	-2.15772
96	1.90211	6.7	8.8	1	0	0.016539	-1.46962
97	1.72277	5.6	12.0	3	1	0.001443	0.53044
98	1.66771	5.3	7.3	4	1	0.001043	-0.35180
99	-2.30259	0.1	18.1	2	0	0.015416	-0.99264
100	0.95551	2.6	9.2	3	0	0.000755	-0.42987
101	4.63084	102.6	11.0	2	1	0.014685	1.90442
102	2.97553	19.6	6.9	3	1	0.000203	-0.19443
103	5.02059	151.5	11.5	2	1	0.025422	2.45659
104	3.17805	24.0	7.9	3	1	0.000549	0.34322
105	2.95491	19.2	2.6	4	1	0.013138	-0.92069
106	3.75887	42.9	7.9	2	1	0.000000	0.00012
107	4.12390	61.8	5.8	4	0	0.037869	2.05072
108	1.97408	7.2	11.9	2	1	0.000091	-0.14008
109	5.40807	223.2	6.1	2	1	0.003950	0.81364
110	2.07944	8.0	8.2	3	0	0.000181	0.20277
111	2.86220	17.5	5.3	3	0	0.000154	-0.14718
112	1.62924	5.1	4.6	4	1	0.021726	-1.37898
113	4.44969	85.6	8.1	1	1	0.000221	-0.17497
114	3.04927	21.1	14.2	1	1	0.006066	0.78738
115	4.03954	56.8	7.4	1	0	0.000038	-0.06605
116	4.89110	133.1	8.3	1	0	0.008246	1.01884
117	-1.20397	0.3	14.0	3	0	0.002704	-0.61772
118	0.40547	1.5	15.4	1	0	0.002106	-0.40925
119	4.02177	55.8	5.3	2	0	0.000003	0.02124
120	-1.20397	0.3	13.2	4	0	0.000019	-0.04587
121	2.64617	14.1	12.8	2	1	0.003267	0.77840
122	3.52636	34.0	5.2	3	0	0.001193	0.40583
123	4.47961	88.2	3.9	3	1	0.000031	0.05751
124	-2.30259	0.1	14.5	4	0	0.003429	-0.55596
125	1.90211	6.7	10.3	2	0	0.000013	-0.05588
126	-1.20397	0.3	10.6	4	0	0.006957	-0.98613
127	6.30828	549.1	11.1	1	1	0.045431**	2.61742**
128	2.32239	10.2	8.7	3	1	0.000068	-0.12482
129	-0.22314	0.8	14.8	3	1	0.000294	-0.18462
130	2.74727	15.6	7.8	1	0	0.009510	-1.07422
131	0.40547	1.5	9.7	3	0	0.002195	-0.73661
132	1.22378	3.4	9.2	3	1	0.003535	-0.91685
133	-0.69315	0.5	11.4	3	0	0.005457	-1.10028
134	2.76632	15.9	10.8	1	1	0.003306	-0.69514
135	3.25424	25.9	9.5	2	0	0.003071	0.85163
136	1.85630	6.4	9.2	2	0	0.001034	-0.49083
137	6.11501	452.6	5.3	2	1	0.009335	1.15737
138	6.38806	594.7	3.2	3	0	0.054375**	2.28070

I (arbitrarily) flag $c_i > 0.04$ are particularly influential and $|r_i| > 2.5$ as particularly ill-fit by the model. The largest two Cook's distances correspond to the largest and smallest times in the data set. The effect of these observations on the fitted coefficients could be refined via dfbetas. None of the studentized residuals are over 3

D1, P5

(5)

in magnitude, but two are over 2.5. Observation 127 is both ill-fit and influential. It's worth while to drop this observation from the data set and refit the model to see if the treatment effect changes appreciably.

Note that allergy could be included as a categorical predictor (baseline plus 3 offsets) instead of a linear effect. If this is done the p-value for dropping the allergy by asthma interaction drops from p=0.103 to p=0.053, and the interaction probably should be retained. Under this scenario, the treatment effect is is slightly magnified, going from 0.82 to 0.87, (but standard error is 0.19 in both cases).

```
data wheezing;
input time asthma allergy treatment; treatment=treatment-1;
datalines;
0.20 18.80 2 2
69.20 10.70 1 2
21.00 8.00 2 1
448.60 5.10 1 2
61.30 7.10 1 2
137.10 8.80 3 2
45.90 5.90 3 2
145.90 3.80 2 1
82.60 8.40 1 2
1.20 8.00 3 1
3.20 12.10 1 1
2.70 8.80 4 2
1.10 12.90 3 1
2.30 8.60 4 1
4.80 7.50 4 2
36.70 8.20 1 1
4.50 16.10 1 1
0.60 11.90 3 1
0.90 9.40 4 2
63.40 7.80 2 2
1.70 15.10 2 2
1.20 11.30 3 1
1.30 11.80 3 1
1.80 12.70 2 1
132.50 9.10 1 2
131.90 6.00 2 1
2.80 10.40 4 2
6.40 12.60 1 2
1.00 14.70 4 2
0.50 12.70 4 1
523.30 5.10 1 2
1.80 10.90 3 1
1.60 8.90 4 1
1.60 13.10 4 1
3.00 3.10 4 1
33.90 8.70 1 1
0.20 10.30 4 1
3.10 10.60 3 2
9.60 6.80 4 2
10.30 14.20 1 2
29.10 5.10 3 1
2.70 13.80 3 1
0.10 19.80 4 1
0.60 12.80 2 1
66.20 3.30 1 1
0.50 14.00 2 1
6.50 11.60 4 1
171.70 7.20 1 2
3.80 10.40 3 2
11.50 11.10 2 1
332.50 5.40 1 1
32.90 11.20 1 2
1.00 11.20 4 2
333.80 5.20 1 2
12.00 11.70 2 1
29.20 8.50 2 2
20.40 6.80 2 2
4.10 14.20 3 2
38.30 7.40 1 2
40.30 8.00 1 2
61.80 9.80 1 1
99.30 5.90 1 1
22.00 7.50 3 1
13.40 11.20 2 2
14.70 6.60 3 2
3.00 11.10 4 2
0.60 8.70 4 2
398.40 5.60 1 1
```

D1, P5

(6)

```
1.60 14.40 3 2
42.40 2.90 4 1
6.70 5.90 4 2
2.50 10.40 3 2
1.40 13.80 1 2
15.70 5.70 4 1
14.80 8.20 3 2
5.30 10.90 3 1
1.20 17.00 3 2
1.30 11.00 4 1
0.80 5.20 4 1
15.90 7.90 4 2
0.40 10.30 3 1
24.40 10.80 1 2
14.40 8.20 3 1
441.90 7.40 3 2
6.40 9.20 2 1
29.50 12.10 2 2
2.30 11.90 1 1
6.80 12.90 1 1
1.10 13.60 1 2
188.60 11.10 1 2
24.10 10.10 3 1
0.30 15.10 3 1
16.60 7.10 3 2
0.30 12.60 3 2
0.40 11.10 3 2
6.70 8.80 1 1
5.60 12.00 3 2
5.30 7.30 4 2
0.10 18.10 2 1
2.60 9.20 3 1
102.60 11.00 2 2
19.60 6.90 3 2
151.50 11.50 2 2
24.00 7.90 3 2
19.20 2.60 4 2
42.90 7.90 2 2
61.80 5.80 4 1
7.20 11.90 2 2
223.20 6.10 2 2
8.00 8.20 3 1
17.50 5.30 3 1
5.10 4.60 4 2
85.60 8.10 1 2
21.10 14.20 1 2
56.80 7.40 1 1
133.10 8.30 1 1
0.30 14.00 3 1
1.50 15.40 1 1
55.80 5.30 2 1
0.30 13.20 4 1
14.10 12.80 2 2
34.00 5.20 3 1
88.20 3.90 3 2
0.10 14.50 4 1
6.70 10.30 2 1
0.30 10.60 4 1
549.10 11.10 1 2
10.20 8.70 3 2
0.80 14.80 3 2
15.60 7.80 1 1
1.50 9.70 3 1
3.40 9.20 3 2
0.50 11.40 3 1
15.90 10.80 1 2
25.90 9.50 2 1
6.40 9.20 2 1
452.60 5.30 2 2
594.70 3.20 3 1
;

ods jpg; ods graphics on;
proc sgscatter data=wheezing;
 plot time*(asthma allergy treatment);
run;
ods graphics off; ods jpg close;

data wheezing; set wheezing; ltime=log(time); run;
ods jpg; ods graphics on;
proc sgscatter data=wheezing;
```

D1, P5

7

```
 plot ltime*(asthma allergy treatment);
run;
ods graphics off; ods jpg close;

ods jpg; ods graphics on;
proc transreg solve ss2 plots=(transformation obp residuals);
 model boxcox(time)=spline(asthma) spline(allergy) identity(treatment);
run;
ods graphics off; ods jpg close;

proc glm;
class allergy;
 model ltime=asthma allergy treatment asthma*allergy asthma*treatment allergy*treatment asthma*asthma;
run;

proc glm;
class allergy;
 model ltime=asthma allergy treatment asthma*allergy;
run;

proc glm; * can drop asthma*allergy as p=0.1;
 class allergy;
 model ltime=asthma allergy treatment asthma*allergy;
run;

proc reg;
 model ltime=asthma allergy treatment;
 output out=out p=p student=r;
run;

ods jpg; ods graphics on;
proc sgscatter data=out;
 plot r*(asthma allergy treatment p);
run;
ods graphics off; ods jpg close;

proc reg;
 model ltime=asthma allergy treatment;
 output out=out cookd=c rstudent=r;
run;

proc print data=out; var ltime time asthma allergy treatment c r;
```

Day 1, Problem 6

1. **Poisson regression through the origin**: For regression data $\{(x_i, Y_i)\}_{i=1}^n$ assume the model

$$Y_i \overset{ind.}{\sim} \text{Poisson}(x_i\beta), \quad i = 1, \ldots, n.$$

The $x_1, \ldots, x_n$ are univariate and strictly positive. Let $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$, $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, and $\mathbf{Y} = (Y_1, \ldots, Y_n)$.

(a) Show that the MLE of β is $\hat{\beta} = \bar{Y}/\bar{x}$.

(b) Find the mean and variance of $\hat{\beta}$.

(c) Now assume that β has a gamma prior distribution $\beta \sim \Gamma(wb_0, w)$ where b_0 is our prior best guess and $w > 0$ is a weight attached to this guess. To be precise, β has the prior density

$$f(\beta) = \frac{w^{wb_0}}{\Gamma(wb_0)}\beta^{wb_0-1}\exp(-w\beta)I_{(0,\infty)}(\beta).$$

Find the posterior distribution of $\beta|\mathbf{Y}$.

(d) Show that the posterior mean is a weighted average of the prior mean and the MLE. What does the posterior mean converge to as $w \to 0+$?

(e) $n = 173$ female horseshoe crabs were sampled and the width of their carapice measured (cm) and number of satellites (nestmates besides their husbands) recorded. SAS PROC means was used to get $\bar{x}$ and $\bar{y}$. The model above was fit in SAS PROC GENMOD using `model satell=width / noint link=identity dist=pois;` giving the following output:

```
                          The MEANS Procedure

Variable     N          Mean        Std Dev        Minimum        Maximum
--------------------------------------------------------------------------
satell     173     2.9190751      3.1483357              0     15.0000000
width      173    26.2988439      2.1090610     21.0000000     33.5000000
--------------------------------------------------------------------------

                         The GENMOD Procedure

                          Model Information
                    Distribution              Poisson
                    Link Function            Identity
                    Dependent Variable         satell

             Analysis Of Maximum Likelihood Parameter Estimates
```

Day 1, Problem 6

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	0	0.0000	0.0000	0.0000	0.0000	.	.
width	1	0.1110	0.0049	0.1013	0.1207	505.00	<.0001

Show how SAS obtains the estimate 0.1110, standard error 0.0049, and 95% confidence interval $(0.1013, 0.1207)$.

(a) The log-likelihood is

$$L(\beta) = -\beta \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \log(x_i \beta) - \sum_{i=1}^{n} \log(y_i!).$$

Then

$$L'(\beta) = n\bar{x} + \frac{n\bar{y}}{\beta}.$$

Setting equal to zero gives local extreme $\hat{\beta} = \frac{\bar{Y}}{\bar{x}}$. Note that

$$L''(\beta) = \frac{-n\bar{y}}{\beta^2} < 0,$$

so $\hat{\beta}$ is a (global) max.

(b)

$$\begin{aligned} E(\hat{\beta}) &= E(\bar{Y}/\bar{x}) \\ &= \frac{\sum_{i=1}^{n} E(Y_i)}{n\bar{x}} \\ &= \frac{\sum_{i=1}^{n} x_i \beta}{n\bar{x}} \\ &= \beta. \end{aligned}$$

$$\begin{aligned} var(\hat{\beta}) &= var(\bar{Y}/\bar{x}) \\ &\overset{ind.}{=} \frac{\sum_{i=1}^{n} var(Y_i)}{n^2\bar{x}^2} \\ &= \frac{\sum_{i=1}^{n} x_i \beta}{n^2\bar{x}^2} \\ &= \frac{\beta}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \end{aligned}$$

Day 1, Problem 6

(c)

$$\begin{aligned} p(\beta|\mathbf{y}) &\propto p(\mathbf{y}|\beta)p(\beta) \\ &= e^{-\beta n\bar{x}}\beta^{n\bar{y}}\beta^{wb_0-1}e^{-w\beta} \\ &= \beta^{n\bar{y}+wb_0-1}e^{-\beta(n\bar{x}+w)}. \end{aligned}$$

So then $\beta|\mathbf{y} \sim \Gamma(n\bar{y}+wb_0, n\bar{x}+w)$.

(d)

$$\begin{aligned} E(\beta|\mathbf{y}) &= \frac{n\bar{y}+wb_0}{n\bar{x}+w} \\ &= \frac{n}{n+w/\bar{x}}\hat{\beta} + \frac{w/\bar{x}}{n+w\bar{x}}b_0 \\ &= \left[\frac{\sum_{i=1}^n x_i}{w+\sum_{i=1}^n x_i}\right]\hat{\beta} + \left[\frac{w}{w+\sum_{i=1}^n x_i}\right]b_0 \end{aligned}$$

$E(\beta|\mathbf{y}) \to \hat{\beta}$ as $w \to 0^+$.

(e) Use your best judgment as to how well they explained how the numbers were obtained.