Stat 505 Assignment 3 Solutions

1. Lots of latitude here. One easy option is to use a 4 by 4 diagonal matrix with two columns of zero attached. $\boldsymbol{A} = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 \end{bmatrix}$ and $\boldsymbol{G} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. Then

   $\boldsymbol{AGA} = \boldsymbol{I}_4\boldsymbol{A} = \boldsymbol{A}$.

2. Matrices with 4 rows or columns must have rank ≤ 4, whereas $\boldsymbol{I}_6$ has rank 6. It is impossible to multiply together two rank 4 (or less) matrices to get a matrix of rank 6, because you cannot increase rank through multiplication.

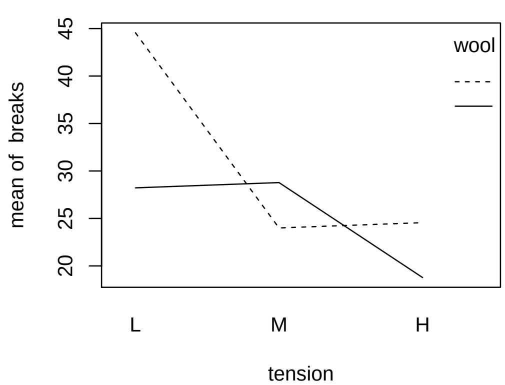
3. The `warpbreaks` data included in R provides the results of an experiment wherein two types of wool (A or B) were used on several looms under three tensions settings (low, medium, high). The outcome of interest is the number of breaks in the warps.

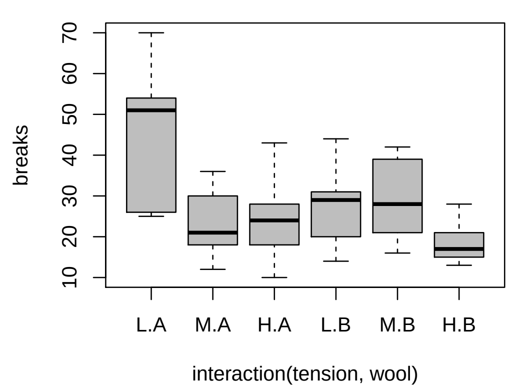
   (a) We use a full interaction model:

   $$y_{ijk} = \mu + \alpha_i + \tau_j + \gamma_{ij} + \epsilon_{ijk}$$

   and will start with a table of mean breakage and some exploratory plots.

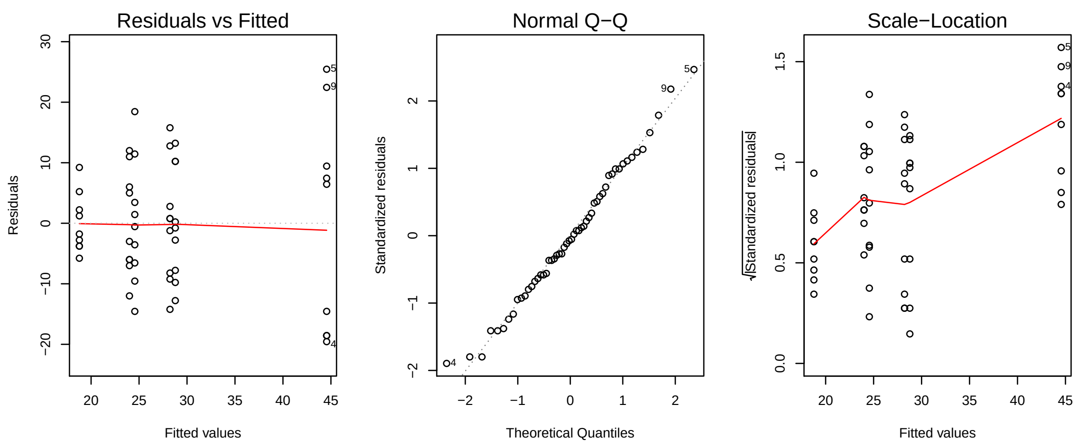
	L	M	H
A	44.56	24.00	24.56
B	28.22	28.78	18.78



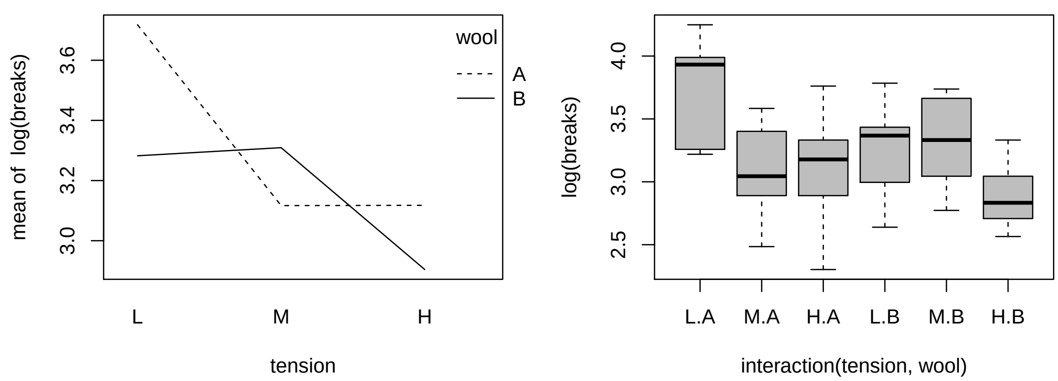


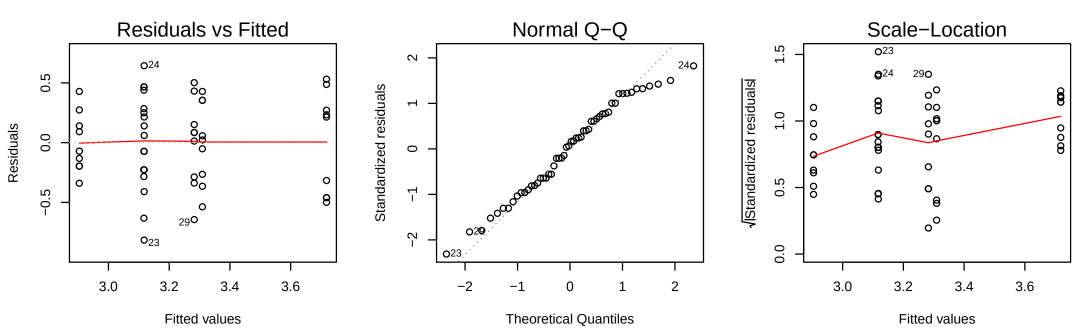
   From the plots and table of means, we see that lowest mean number of breaks occurs at high tension with wool B. There may be an interaction between wool type and tension, because wool B is better and High and Low tension, but A is better at medium tension. High tension is always better than low, but for type A wool, medium tension is slightly better than high tension. (I doubt that this is a substantial difference.)

These data do exhibit non-constant variance as shown in the spread location plot. I took care of the problem by using log(breaks) instead of the original response. his makes some sense since it is a count which might have a Poisson distribution. Taking log on a Poisson helps to stabilize the variance.



Now I'll plot and summarize the data again in the log scale.





Diagnostic plots now show only a slight increase in variance as means increase, so the transformation has fixed the problem.

	L	M	H
A	3.72	3.12	3.12
B	3.28	3.31	2.90

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
wool	1	0.31	0.31	2.23	0.1415
tension	2	2.18	1.09	7.78	0.0012
wool:tension	2	0.91	0.46	3.26	0.0469
Residuals	48	6.71	0.14		

The anova table provides tests of

- The main effect of tension. With $F_{2,48} = 7.78$ and a p-value of 0.001, we conclude that there is a significant difference between mean log breakage of the three tensions.
- The main effect of wool type is not significant ($F_{1,48} = 2.23$, p-value = 0.14); we fail to reject the hypothesis of equal mean log breakage for type A and type B wools. However, for the full story on wool type, we need to include interactions. It does not make sense to remove wool type because it is part of a "significant" interaction – read on.
- There is a moderately strong interaction between wool types and tensions. The test for no interaction effect gives $F_{2,48} = 3.26$, p-value = 0.047, so we could reject the null hypothesis of no treatment effect (at the always popular 5% significance level). We can make different conclusions about which tension is best for wool A and for wool type B.

(b) The full model matrix $\boldsymbol{X}$ is 54 by 12 having 6 unique rows:

	int	A	B	L	M	H	A.L	A.M	A.H	B.L	B.M	B.H
1	1	1	0	1	0	0	1	0	0	0	0	0
2	1	1	0	0	1	0	0	1	0	0	0	0
3	1	1	0	0	0	1	0	0	1	0	0	0
4	1	0	1	1	0	0	0	0	0	1	0	0
5	1	0	1	0	1	0	0	0	0	0	1	0
6	1	0	1	0	0	1	0	0	0	0	0	1

	int	A	B	L	M	H	A.L	A.M	A.H	B.L	B.M	B.H
int	54	27	27	18	18	18	9	9	9	9	9	9
A	27	27	0	9	9	9	9	9	9	0	0	0
B	27	0	27	9	9	9	0	0	0	9	9	9
L	18	9	9	18	0	0	9	0	0	9	0	0
M	18	9	9	0	18	0	0	9	0	0	9	0
H	18	9	9	0	0	18	0	0	9	0	0	9
A.L	9	9	0	9	0	0	9	0	0	0	0	0
A.M	9	9	0	0	9	0	0	9	0	0	0	0
A.H	9	9	0	0	0	9	0	0	9	0	0	0
B.L	9	0	9	9	0	0	0	0	0	9	0	0
B.M	9	0	9	0	9	0	0	0	0	0	9	0
B.H	9	0	9	0	0	9	0	0	0	0	0	9

(c) $\boldsymbol{X}^\mathsf{T}\boldsymbol{X}$ is 12 by 12 with rank 6.

(d) The model matrix R uses is 54 by 6 with rank 6 and can be obtained by omitting columns 2, 4, 7, 8, 9, 10. of the full $\boldsymbol{X}$ matrix. The crossproduct matrix is only 6 by 6 and is missing those same rows and columns (also of rank 6, so it is of fulll column rank). It has a real inverse, which is the only generalized inverse.

	(Intercept)	woolB	tensionM	tensionH	woolB:tensionM	woolB:tensionH
(Intercept)	0.11	-0.11	-0.11	-0.11	0.11	0.11
woolB	-0.11	0.22	0.11	0.11	-0.22	-0.22
tensionM	-0.11	0.11	0.22	0.11	-0.22	-0.11
tensionH	-0.11	0.11	0.11	0.22	-0.11	-0.22
woolB:tensionM	0.11	-0.22	-0.22	-0.11	0.44	0.22
woolB:tensionH	0.11	-0.22	-0.11	-0.22	0.22	0.44

(e) By setting the above inverse matrix into the non-omitted rows and columns of a 12 by 12 matrix of zeroes, we obtain this matrix

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.11	0.00	-0.11	0.00	-0.11	-0.11	0.00	0.00	0.00	0.00	0.11	0.11
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	-0.11	0.00	0.22	0.00	0.11	0.11	0.00	0.00	0.00	0.00	-0.22	-0.22
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	-0.11	0.00	0.11	0.00	0.22	0.11	0.00	0.00	0.00	0.00	-0.22	-0.11
6	-0.11	0.00	0.11	0.00	0.11	0.22	0.00	0.00	0.00	0.00	-0.11	-0.22
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	0.11	0.00	-0.22	0.00	-0.22	-0.11	0.00	0.00	0.00	0.00	0.44	0.22
12	0.11	0.00	-0.22	0.00	-0.11	-0.22	0.00	0.00	0.00	0.00	0.22	0.44

This is a generalized inverse of full $\boldsymbol{X}^\top\boldsymbol{X}$ as shown when I summarize the differences $(\boldsymbol{X}^\top\boldsymbol{X})\boldsymbol{G}(\boldsymbol{X}^\top\boldsymbol{X}) - (\boldsymbol{X}^\top\boldsymbol{X})$

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0       0       0       0       0       0 
```

All differences are 0, so we have an identity.

(f) Computing the solutions to the normal equationas using this generalized inverse gives:

```
    (Intercept)           woolB        tensionM        tensionH woolB:tensionM 
          3.718          -0.436          -0.601          -0.600          0.628 
woolB:tensionH 
          0.222 
```

	1	2	3	4	5	6	7	8	9	10	11	12
1	3.72	0.00	-0.44	0.00	-0.60	-0.60	0.00	0.00	0.00	0.00	0.63	0.22

with zeroes in positons 2, 4, 7, 8, 9, 10, which are the omitted columns of $\boldsymbol{X}^\top\boldsymbol{X}$.

(g) The estimates are for these parameters:

Line Label:	Estimate	Parameter
(Intercept)	3.72	$\mu + \alpha_A + \tau_L + \gamma_{AL}$
woolB	-0.44	$\alpha_B - \alpha_A + \gamma_{BL} - \gamma_{AL}$
tensionM	-0.60	$\tau_M - \tau_L + \gamma_{AM} - \gamma_{AL}$
tensionH	-0.60	$\tau_H - \tau_L + \gamma_{AH} - \gamma_{AL}$
woolB:tensionM	0.63	$\gamma_{BM} - \gamma_{BL} - \gamma_{AM} + \gamma_{AL}$
woolB:tensionH	0.22	$\gamma_{BH} - \gamma_{BL} - \gamma_{AH} + \gamma_{AL}$

(h) Refitting with SAS contrast options, we now omit rows 3, 6, 9, 10, 11, 12 to get this generalized inverse matrix:

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.11	-0.11	0.00	-0.11	-0.11	0.00	0.11	0.11	0.00	0.00	0.00	0.00
2	-0.11	0.22	0.00	0.11	0.11	0.00	-0.22	-0.22	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	-0.11	0.11	0.00	0.22	0.11	0.00	-0.22	-0.11	0.00	0.00	0.00	0.00
5	-0.11	0.11	0.00	0.11	0.22	0.00	-0.11	-0.22	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.11	-0.22	0.00	-0.22	-0.11	0.00	0.44	0.22	0.00	0.00	0.00	0.00
8	0.11	-0.22	0.00	-0.11	-0.22	0.00	0.22	0.44	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Again I confirmed that we have a generalized inverse by subtracting $(\boldsymbol{X}^\top\boldsymbol{X})\boldsymbol{G}_2(\boldsymbol{X}^\top\boldsymbol{X}) - \boldsymbol{X}^\top\boldsymbol{X}$ and getting a summary of all zeroes. Then we get these coefficient estimates from `coef(warpfit3)`:

```
  (Intercept)          woolA       tensionL        tensionM woolA:tensionL
        2.904          0.213          0.378           0.405          0.222
woolA:tensionM
       -0.406
```

and the same from matrix multiplication, with the `omit.rows` values set to 0.

	1	2	3	4	5	6	7	8	9	10	11	12
1	2.90	0.21	0.00	0.38	0.41	0.00	0.22	-0.41	0.00	0.00	0.00	0.00

(i) The estimates above are for the following quantities:

Line Label:	Estimate	Parameter
Intercept	2.904	$\mu + \alpha_B + \tau_H + \gamma_{BH}$
wool A	0.213	$\alpha_A - \alpha_B + \gamma_{AH} - \gamma_{BH}$
tension L	0.378	$\tau_L - \tau_H + \gamma_{BL} - \gamma_{BH}$
tension M	0.405	$\tau_M - \tau_H + \gamma_{BM} - \gamma_{BH}$
woolA*tensionL	0.222	$\gamma_{AL} - \gamma_{AH} - \gamma_{BL} + \gamma_{BH}$
woolA*tensionM	-0.406	$\gamma_{AM} - \gamma_{AH} - \gamma_{BM} + \gamma_{BH}$

(j) Finally, we need to check that fitted values are the same for both sets of coefficient estimates. I used just the 6 uniques rows of $\boldsymbol{X}$, and matrix multiplied by each coefficient estimate vector. Values agree precisely.

	1	2
1	3.72	3.72
2	3.12	3.12
3	3.12	3.12
4	3.28	3.28
5	3.31	3.31
6	2.90	2.90