

Math 484 HW 10 Fall 2022, due Friday, Nov. 18.

Exam 3 is Wednesday, Nov. 30.

The Final is Friday, Dec. 16, 10:15-12:15.

Wed. Nov. 16: Q9 covers HW9 and HW10. 3 pages, 6 problems A)-F)

A) This is like HW8 C) with the BigMac data. Assume that a multiple linear regression model holds for $Y = t(Z)$ and for some terms (usually powers or logs of the predictors). Using the techniques learned in class, find such a model. Do parts a), b) and c) below.

student	file	Z
1 Aziz, Kuan	paddle.lsp	Weight
2 Dhanushka	highway.lsp	rate (from the menu Highway, select "Add a variate" and type sigsp1 = sigs + 1. Then you can transform sigsp1.)
3 Bethany	allomet.lsp	BRAIN
4 Lakni	evaporat.lsp	Evap (log rule is bad for Z = Evap)
5 Sanjuka	sniffer.lsp	Y
6 Yunjia	casuarin.lsp	W
7 Maryam	ozone.lsp	ozone
8 Itria	ais.lsp	%Bfat
9 Dianqui	water.lsp	Y
10 Mina	fuel90.lsp	inc
11 Kasun	bodfat.lsp	y (Do not use x1 as a predictor. The file is at (www.parker.ad.edu/olive/regbk.htm).)
12 Felicia, Wyatt	haystack.lsp	Vol
13 Zander, Yangge	hald.lsp	Y
14 Howard	landrent.lsp	Y
15 Jie Shi	boston2.lsp	y (The file is at (www.parker.ad.edu/olive/regbk.htm).)

a) Write down the full model that you use and include a response plot for the full

model. (This plot should be linear). Also give R^2 for the full model. (Output showing coefficients, pvalues and R^2 is fine.)

b) Write down the final submodel that you use.

c) Include the least squares output for your model and 5 more plots that justify that your multiple linear regression model is reasonable. Use the residual plot for the full model, the RR plot, the FF plot, the residual plot for your model, and the response plot for your model. Below or beside each plot, give a brief explanation for how the plot gives support for your model. I do not want to see intermediate steps or competing models.

B) 5.13 This problem uses data from Kuehl (1994, p. 128).

a) Get *lregdata* and *lregpack* into *R*. Type the following commands. Then simultaneously press the *Ctrl* and *c* keys. In *Word* use the menu command “Paste.” Print out the figure.

```
y <- ycrab+1/6
aovtplt(crabhab,y)
```

b) From the figure, what response transformation should be used: $Y = 1/Z$, $Y = 1/\sqrt{Z}$, $Y = \log(Z)$, $Y = \sqrt{Z}$, or $Y = Z$?

C) 5.14 The following data set considers the number of warp breaks per loom, where the factor is tension (low, medium, or high).

a) Copy and paste the commands for this problem into *R*.

Highlight the ANOVA table by pressing the left mouse key and dragging the cursor over the ANOVA table. Then use the menu commands “Edit> Copy.” Enter *Word* and use the menu command “Paste.”

b) To place the residual plot in *Word*, get into *R* and click on the plot, hit the *Ctrl* and *c* keys at the same time. Enter *Word* and use the menu command “Paste” or hit the *Ctrl* and *v* keys at the same time.

c) Copy and paste the commands for this part into *R*.

Click on the response plot, hit the *Ctrl* and *c* keys at the same time. Enter *Word* and use the menu command “Paste.”

D) 5.15ab Obtain the Box et al. (2005, p. 134) blood coagulation data from *lregdata* and the *R* program *ganova* from *lregpack*. The program does graphical Anova for the one way Anova model.

a) Enter the following command and include the plot in *Word* by simultaneously pressing the *Ctrl* and *c* keys, then using the menu command “Paste” in *Word*, or hit the *Ctrl* and *v* keys at the same time.

```
ganova(bloodx,bloody)
```

The scaled treatment deviations are on the top of the plot. As a rule of thumb, if all of the scaled treatment deviations are within the spread of the residuals, then population treatment means are not significantly different (they all give response near the grand mean). If some deviations are outside of the spread of the residuals, then not all of the

population treatment means are equal. Box et al. (2005, p. 137) state ‘The graphical analysis discourages overreaction to high significance levels and avoids underreaction to “very nearly” significant differences.’

b) From the output, which two treatments means were approximately the same?

E) 5.6

output for Problem 5.6	level			
1	2	3	4	5
15 percent	20 percent	25 percent	30 percent	35 percent

\bar{y}_1	\bar{y}_5	\bar{y}_2	\bar{y}_3	\bar{y}_4
9.8	10.8	15.4	17.6	21.6
—	—	—	—	

5.6. The tensile strength of a cotton nylon fiber used to make women’s shirts is believed to be affected by the percentage of cotton in the fiber. The 5 levels of cotton percentage that are of interest are tabled above. Also shown is a (Tukey pairwise) comparison of means. Which groups of means are not significantly different? Data is from Montgomery (1984, pp. 51, 66).

F) 5.10

Source	df	SS	MS	F	P
treatment	3	89.19	29.73	15.68	0.0002
error	12	22.75	1.90		

A textile factory weaves fabric on a large number of looms. They would like to obtain a fabric of uniform strength. Four looms are selected at random and four samples of fabric are obtained from each loom. The strength of each fabric sample is measured.

- a) State whether this is a random or fixed effects one way Anova. Explain briefly.
- b) Using the output above, perform the appropriate 4 step Anova F test.