(1, 3pts) **(a)** The experiment includes one treatment factor (i.e., *amount* of sulfamerazine added to fish food). Notice that if *trough* is treated as a block factor, it is fully confounded with *amount*, and we cannot study the effects of *amount* under the circumstance.

(b) *Amount* is a quantitative factor with 4 equally spaced levels

(c) a fish (i.e., a trout)

(d) An acceptable conceptual model is

$$response \sim amount_l + amount_q + \cdots,$$

where $amount_l$, $amount_q, \ldots$ are the linear, quadratic, ..., effects of *amount*, each with 1 degree of freedom.

(e) This is a completely randomized design with one-way layout. Each treatments have 10 replicates.

(2, 3pts) **(a)** The experiment includes two treatment factors (i.e., *brand* and percentage *capacity*) and one block factor (i.e., the *wattage*). Notice that *wattage* is confounded with *day*. The use of the 2 blocks can simultaneously "remove" the wattage-to-wattage and day-to-day variation.

(b) *Brand* is qualitative and has 3 levels. Both *capacity* and *wattage* are quantitative and have 2 levels. The block size of *wattage* is 24.

(c) a bulb

(d) An acceptable conceptual model is

$$response \sim \underbrace{wattage}_{1\text{ d.f.}} + \underbrace{brand}_{2\text{ d.f.}} + \underbrace{capacity_l}_{1\text{ d.f.}} + \underbrace{brand \times capacity_l}_{2\text{ d.f.}} + \underbrace{\epsilon}_{41\text{ d.f.}}$$

(e) Because each block has enough experimental units (24) to let all level combinations of the treatment factors appear 4 times, this is a randomized (complete) block design with a 2-way layout for the treatment factors.

(3, 1pt) No, 194 is not divisible by the number of level combinations $2\times2\times2=8$ so it cannot be balanced.

(4, 2pts) Look for skewness, non-constant variance and outliers - none are evident here.

(5, 1pt) No, these boxplots are not the ones within each level combinations. Instead, data in several distinct level combinations are merged to draw one boxplot here so the factorial effects would have some influence on these boxplots. They are not suitable for drawing conclusion about the distribution of experimental errors.

(6, 1pt) The null distributions in various tests might not be t- or F-distribution anymore. However, for this data the impact should be little since the sample size (194) is large, i.e., the distribution of the effect estimates should still be close to normal.

(7, 1pt) $\sqrt{3.95088} = 1.987682$.

(8, 2pts) The null model is the main-effect-only model while the alternative model is the full model. From the sequential ANOVA table, we can get the test statistic

$$F = \frac{(6.3244 + 5.2185 + 6.6583 + 0.0136)/4}{3.95088} = 1.152579 < 2.42,$$

so that we would not reject the null and the main-effect-only model is preferred. Notice that if this ANOVA is of drop-one type, rather than of sequential type, the test cannot be done in this way.

(9, 2pts) The ANOVA table (together with the conclusion in problem (8)) shows that only temperature is a significant factor. Furthermore, the boxplot of temperature shows that warm level would increase the bill length.

(10, 2pts) Nothing. This factor is confounded with other changes that take place over time as the description makes clear. It might be important or not.

(11, 1pt) The two spaces are not geometrically orthogonal. It can be easily discovered by examining the inner product between block columns and dough columns. Their inner products are not zero.

(12, 1pt) Their correlation can be obtained from $(X^TX)^{-1}$, where X is the model matrix. But, the calculation of matrix inverse is rather complicated. An alternative method is given as follows. Let the response vector be $y = (y_1, y_2, \ldots, y_9)$. Because treatment coding is used for an orthogonal design, $\hat{\beta}_{\text{blockII}} = \frac{y_4+y_5+y_6}{3} - \frac{y_1+y_2+y_3}{3}$ and $\hat{\beta}_{\text{dough2}} = \frac{y_2+y_5+y_8}{3} - \frac{y_1+y_4+y_7}{3}$. So,

$$\begin{aligned} cov(\hat{\beta}_{\text{blockII}}, \hat{\beta}_{\text{dough2}}) &= \frac{1}{9}(-cov(y_4, y_4) + cov(y_5, y_5) + cov(y_1, y_1) - cov(y_2, y_2)) \\ &= \frac{1}{9}(-\sigma^2 + \sigma^2 + \sigma^2 - \sigma^2) = 0. \end{aligned}$$

The correlation is zero. Actually, the correlation between any pair of block effect estimator and treatment effect estimator is zero.

(13, 2pts) Even though the block and dough spaces (under treatment coding) are not geometrically orthogonal as shown in the answer to problem (11), the ANOVA table is exactly the same as the one given in question sheet with the rows for dough and block being exchanged. The reason is that the design is a balanced 2-way layout, which guarantees that the model matrix under *sum coding* has the property of geometrical orthogonality. So, the Df and Sum Sq values in the ANOVA table is:

Source	Df	Sum Sq
block	2	47.18 (=2×23.59)
dough	2	2620.23
Residuals	4 (=9-1-2-2)	2286.31

(14, 1pt) Remove the outliers. (FYI. The regression output shows that two observations were removed. Can you see why two? Which two would you expect?)

(15, 1pt) Suppliers 2 and 3 because they have the same Std. Error value. Supplier 1 might be the same as suppliers 2 and 3 or other suppliers, but there is no way to tell from the regression output. (FYI. Supplier 1 actually has a different number of observations from any of the other suppliers. Can you see why? It can be found from the Tukey HSD output by calculating and comparing the lengths of the confidence intervals.)

(16, 1pt) Only suppliers 2 and 5.

(17, 1pt) Block factor because the effect of week is not of interest. Furthermore, weeks are properties of the experimental units - we cannot randomly assign them to experimental units.

(18, 1pt) null: y $\sim$ week, and alternative: y $\sim$ week + supplier, where week is treated as a qualitative factor with 21 degrees of freedom.

(19, 1pt) No, the results would be different because week and supplier are not orthogonal due to the unbalanced structure in design. When no orthogonality, the estimates of supplier effects are different under models with or without week.

(20, 2pts) A linear term for week would use only one df. This analysis is ANCOVA (analysis of covariance).