NTHU STAT 5510 Midterm Examination May 1, 2024

Instructions: Attempt all questions. Short and specific answers are preferred. Given explanation when required, but keep it as short and simple as possible. Give only one answer to each question – if you give alternative answers, the worst answer will be graded.

Question A.

For each of the following experiments,

(a) identify its treatment factor(s) and block factor(s) (or covariate),
(b) determine the types (i.e., qualitative or quantitative) and numbers of levels of these factors,
(c) identify the experimental units,
(d) suggest an appropriate conceptual model,
(e) recognize the design plan(s) (e.g., one-way layout, two-way layout, multi-way layout, completely randomized design, paired comparison design, randomized block design, Latin square design, BIBD, analysis of covariance, ...) that is suitable for the experiment.

(1) (4.5 pts) An experiment was conducted to investigate the impact of adding ethanol to fuels from different brands on CO emissions. Four different brands of fuel were used in this experiment. For each brand of fuel, a fixed portion of a standard fuel was used each time, with a certain amount of ethanol added. After burning this mixture, the CO emissions released were measured. In this experiment, for each brand of fuel, three different ethanol amounts: 0.1, 0.2, 0.3, were added twice each.

(2) (4.5 pts) An experiment was conducted to test the bioequivalence of three formulations A, B, C of a drug, as measured by AUC, i.e, the Area Under the Curve which relates the concentration of the drug in the blood as a function of the time since dosing. Three volunteer subjects took each formulation in succession with a sufficient washout period between. After dosing, blood samples were obtained every half-hour for four hours and analyzed for drug concentration. AUC was calculated with the resulting data. It was expected that there may be a large variation in metabolism of the drug from subject to subject, and the absorption and metabolism of a drug will vary from time (i.e., the number of times taking a drug) to time for a particular subject.

Question B.

The dataset comes from a study of blood coagulation times. Twenty-four animals were randomly assigned to 4 different diets, labelled "A" through "D," and blood samples were taken in a random order. The blood coagulation time was measure. A side-by-side boxplot of the data is shown in Figure 1.

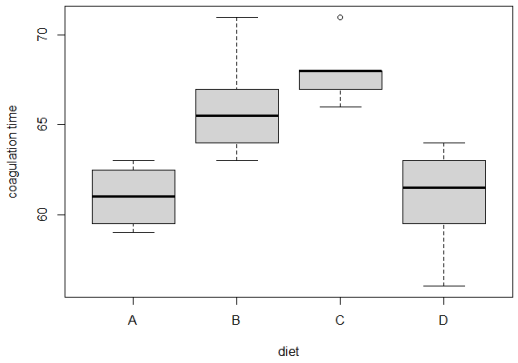

Figure 1. Side-by-side Boxplot

(3) (1 pt) Why are the median and upper quartile the same in the boxplot of diet C?

(4) (1.5 pts) List at least three assumptions of the ANOVA model that can be diagnosed using these boxplots.

A regression model using treatment coding (i.e., (0, 1)-coding) was fit and the following results was obtained:

```
Coefficients:
             Estimate    Std. Error    t value    Pr(>|t|)
(Intercept)  61.00       1.1832        51.55      < 0.00000
dietB        5.00        1.5275        3.27       0.00380
dietC        7.00        1.5275        4.58       0.00018
dietD        0.00        1.4491        0.00       1.00000
---
Residual standard error: 2.3664 on 20 degrees of freedom
F-statistic: 13.6 on 3 and 20 DF,  p-value: 0.00005
```

(5) (1 pt) Why is the standard error of `dietD` different from `dietB` and `dietC`?

(6) (2 pts) Which diet has the smallest number of replicates? Explain.

(7) (1 pt) When performing an ANOVA test using the command "`> anova(lm(coag~diet))`", what is the value of the test statistic and what is the value of the sum of squares for residuals in the ANOVA table?

(8) (1 pt) If sum coding (i.e., (−1, 1)-coding) is used instead of treatment coding to fit a regression model, what is the coefficient estimate for `dietC`?

The Tukey method, which is based on the studentized range distribution, is used for making multiple comparisons, and here are the results:

```
  Tukey multiple comparisons of means
    95% family-wise confidence level
$diet
       Diff      lwr         upr        p adj
B-A    5         0.72455     9.2754     0.01833
C-A    7         2.72455     11.2754    ??
D-A    0         -4.05604    4.0560     ??
C-B    2         -1.82407    5.8241     ??
D-B    -5        -8.57709    -1.4229    ??
D-C    -7        -10.57709   -3.4229    ??
```

(9) (1 pt) What are the multiple comparison conclusions for these diets?

(10) (1 pt) Utilize the provided R outputs above to illustrate that the Bonferroni method in multiple comparisons is more conservative (i.e., less powerful) than the Tukey method.

Question C.

A nutritionist studied the effects of 6 diets, labelled "a" through "f," on weight gain of domestic rabbits. From past experience with sized of litters, it was felt that only 3 homogeneous rabbits could be selected from each available litter. There were 10 litters available forming blocks of size 3. A balanced incomplete block design (BIBD) with t=6, r=5, b=10, k=3 was employed in this experiment, and its experimental layout is as follows:

block 1	block 2	block 3	block 4	block 5
b, c, f	a, b, c	c, d, f	a, c, e	c, d, e
block 6	**block 7**	**block 8**	**block 9**	**block 10**
b, e, f	a, b, d	a, e, f	b, d, e	a, d, f

(11) (1 pt) One characteristic of BIBD is that, during multiple comparisons, any two treatments have the same number, denoted by λ, of within-block comparisons. In this experiment, what is the value of λ?

(12) (1 pt) For the characteristic mentioned in the previous question, provide a benefit it brings, especially in estimating treatment contrasts during multiple comparisons.

The results of two sequential ANOVA analyses are given below:

```
Analysis of Variance Table
Response: gain
          Df  Sum Sq   Mean Sq   F value   Pr(>F)
block     9   730.39   81.154    8.0738    0.0002454
treat     5   158.73   31.745    3.1583    0.0381655
Residuals 15  150.77   10.052
```

and

```
Analysis of Variance Table
Response: gain
          Df  Sum Sq   Mean Sq   F value   Pr(>F)
treat     5   293.38   58.676    5.8375    0.0034544
block     9   595.74   66.193    6.5854    0.0007602
Residuals 15  150.77   10.052
```

(13) (1 pt) In this experiment, what issue would the nutritionist encounter when further seeking to explore interactions between `treat` and `block`? Explain.

(14) (1 pt) Explain why the *F*-statistic for testing whether `treat` is significant differs in the two ANOVA tables.

(15) (1 pt) What are the null and alternative models corresponding to the *p*-value of `0.0034544`?

(16) (3 pts) What is the value of the *F*-statistic in the one-way ANOVA table generated by executing the following command: `aov(gain ~ treat)`? Explain why in this experiment, the *F*-statistic in this one-way ANOVA is much smaller than the *F*-statistics for `treat` in the above two ANOVA tables.

(17) (1.5 pts) Evaluate the relative efficiency, measured as the ratio of variances, in estimating treatment contrasts to determine how much better this BIBD is compared to the completely randomized design.

(18) (2 pts) If the nutritionist wants to add 2 more blocks (i.e., a total of 12 blocks) and still use BIBD, how would you respond? If you can accommodate the nutritionist's request, describe how the 6 treatments would be assigned to the 12 blocks. If not, please explain why.