NTHU STAT 3875 Final Examination Jan 8, 2019

Note. There are 5 problems in total. **For problems 2 to 5, to ensure consideration for partial scores, write down necessary intermediate steps.** Correct answers with inadequate or no intermediate steps may result in zero credit.

---

1. (*12pts, 6pts for each*) For each of the following experiments,

   - identify its treatment factor(s) and/or block factor(s), and the number of levels in each factor
   - recognize the design plan(s) (i.e., one-way layout, two-way layout, or randomized block design) of the experiment.
   - suggest an appropriate statistical model for the data.

   (a) It is thought that the glycemic index (GI) of food is an indicator of how sustaining or satisfying it is and may influence appetite. A sample of 10 children were provided with a breakfast of low GI foods on one day and high GI food on another. The two breakfasts contained the same quantities of carbohydrate, fat, and protein. On each day, a buffet lunch was provided. For each child, the difference of the number of calories eaten at the two lunches was recorded. The objective of the experiment is to determine whether the kind of breakfast eaten has an effect on mean calorie intake.

   (b) Three species of mice were tested for "aggressiveness." The species were A/J, C57, and F2. Twenty mice of each species were tested. Each of the mice was placed in a 1-m^2 box, which was marked off into 49 equal squares. The mouse was let go on the center square, and the number of squares traversed in a 5-min period was counted. The objective of this experiment is to determine whether there is a significant difference among species.

2. Suppose that $D_1, D_2, \ldots, D_n$ is a random sample from a continuous distribution with the probability density function (pdf) $f(x)$. It is desired to test a hypothesis concerning the median ξ of $f(y)$, where ξ is a parameter. Assume that $f(x)$ is a pdf *symmetric* about ξ. Construct a test of $H_0 : \xi = \xi_0$ against $H_A : \xi \neq \xi_0$, where ξ_0 is a specified known constant.

   (a) (*12 pts*) Use the sign test. Your final answer must include **(i) a test statistic, (ii) the (exact or asymptotic) null distribution of the test statistic, (iii) a rejection region**.

(b) (*12 pts*) Use the Wilcoxon signed-rank test. Your final answer must include **(i) a test statistic, (ii) the (exact or asymptotic) null distribution of the test statistic, (iii) a rejection region**.

[**Hint**. Under the null, $D_1, D_2, \ldots, D_n$ are i.i.d. from $f(x)$ with median ξ_0. Let $Z_i = D_i - \xi_0$, $i = 1, \ldots, n$. Then, under the null, $Z_1, Z_2, \ldots, Z_n$ are i.i.d. from $f(x + \xi_0)$, where $f(x + \xi_0)$ has median zero. You can use Z_i's to develop the tests.]

3. Suppose that I independent samples of *different* sample sizes $J_1, J_2, \ldots, J_I$ are taken from each of I normally distributed populations with mean $\mu_1, \mu_2, \ldots, \mu_I$ and common variances, all equal to σ^2. Let Y_{ij} denote the jth observation from the population i, for $j = 1, 2, \ldots, J_i$ and $i = 1, 2, \ldots, I$. Let $N = J_1 + J_2 + \cdots + J_I$. Define

$$\overline{Y}_{i\cdot} = \frac{1}{J_i}\sum_{j=1}^{J_i} Y_{ij} \quad \text{and} \quad \overline{Y}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij} = \sum_{i=1}^{I}\frac{J_i}{N}\overline{Y}_{i\cdot}$$

(a) (*6 pts*) Recall that

$$SS_W = \sum_{i=1}^{I}(J_i - 1)s_i^2, \quad \text{where} \quad s_i^2 = \frac{1}{J_i - 1}\sum_{j=1}^{J_i}(Y_{ij} - \overline{Y}_{i\cdot})^2.$$

Show that SS_W/σ^2 has a χ^2 distribution with $(J_1 - 1) + (J_2 - 1) + \cdots + (J_I - 1) = N - I$ degrees of freedom.

(b) (*6 pts*) Notice that under the null hypothesis

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$$

all the Y_{ij}'s are independent, normally distributed random variables with the same mean and variance. Use this property to show that, under the null hypothesis,

$$SS_{TOT} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij} - \overline{Y}_{\cdot\cdot})^2$$

is such that SS_{TOT}/σ^2 has a χ^2 distribution with $N - 1$ degrees of freedom.

(c) (*4 pts*) Show that SS_W and SS_B are independent, where

$$SS_B = \sum_{i=1}^{I}(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2.$$

(d) (*6 pts*) Recall that $SS_{TOT} = SS_B + SS_W$. Use it and (a)-(c) to show that under the hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$, SS_B/σ^2 has a χ^2 distribution with $I - 1$ degrees of freedom.

4. The following tables give the survival times (in hours) for animals in an experiment whose design consisted of three poisons, four treatments, and four observations per cell.

Poison	Treatment							
	A		B		C		D	
I	3.1	4.5	8.2	11.0	4.3	4.5	4.5	7.1
	4.6	4.3	8.8	7.2	6.3	7.6	6.6	6.2
II	3.6	2.9	9.2	6.1	4.4	3.5	5.6	10.0
	4.0	2.3	4.9	12.4	3.1	4.0	7.1	3.8
III	2.2	2.1	3.0	3.7	2.3	2.5	3.0	3.6
	1.8	2.3	3.8	2.9	2.4	2.2	3.1	3.3

The model considered for this data is the two-way ANOVA model:

$$Y_{ijk} = \overline{\mu} + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijk},$$

$i = 1, 2, 3,\ j = 1, 2, 3, 4,\ k = 1, 2, 3, 4,$ where α_i's (with the constraint $\sum_{i=1}^{3} \alpha_i = 0$) are the main effects of Poison, β_j's (with the constraint $\sum_{j=1}^{4} \beta_j = 0$) are the main effects of Treatment, δ_{ij}'s (with the constraints $\sum_{i=1}^{3} \delta_{ij} = 0, \forall j$, and $\sum_{j=1}^{4} \delta_{ij} = 0, \forall i$) are the interactions between Poison and Treatment, and ϵ_{ijk}'s are i.i.d. from $N(0, \sigma^2)$. The cell, row, column, and overall averages are calculated and given below:

Poison	Treatment				
	A	B	C	D	
I	$\overline{Y}_{11\cdot} = 4.125$	$\overline{Y}_{12\cdot} = 8.8$	$\overline{Y}_{13\cdot} = 5.675$	$\overline{Y}_{14\cdot} = 6.1$	$\overline{Y}_{1\cdot\cdot} = 6.175$
II	$\overline{Y}_{21\cdot} = 3.2$	$\overline{Y}_{22\cdot} = 8.15$	$\overline{Y}_{23\cdot} = 3.75$	$\overline{Y}_{24\cdot} = 6.625$	$\overline{Y}_{2\cdot\cdot} = 5.43125$
III	$\overline{Y}_{31\cdot} = 2.1$	$\overline{Y}_{32\cdot} = 3.35$	$\overline{Y}_{33\cdot} = 2.35$	$\overline{Y}_{34\cdot} = 3.25$	$\overline{Y}_{3\cdot\cdot} = 2.7625$
	$\overline{Y}_{\cdot 1\cdot} = 3.142$	$\overline{Y}_{\cdot 2\cdot} = 6.767$	$\overline{Y}_{\cdot 3\cdot} = 3.925$	$\overline{Y}_{\cdot 4\cdot} = 5.325$	$\overline{Y}_{\cdot\cdot\cdot} = 4.790$

(a) (*2 pts*) Is this a balanced data? Explain.

(b) (*4 pts*) What is the estimate of β_j for Treatment B? What is the estimate of δ_{ij} for the cell of Poison I and Treatment D?

(c) (*12 pts*) A two-way ANOVA was performed. Fill in all missing values (marked by ??) in this ANOVA table.

Source	SS	df	MS	F	p-value
Treatment	91.9	??	??	??	0.000
Poison	??	??	51.500	23.558	0.000
Interaction	24.7	??	??	??	0.110
Error	??	??	2.186		
Total	298.3	??			

(d) (*2 pts*) What are the degrees of freedom for the F statistic 23.558 in this table?

(e) (*2 pts*) What is the estimated variance $\hat{\sigma}^2$ of the errors?

(f) (*2 pts*) An interaction plot of this data is given in the left panel of Figure 1. Can we claim that an additive model (i.e., a main-effect-only model) is appropriate for this data? Explain.

(g) (*4 pts*) Under the main-effect-only model, what is the estimated variance $\hat{\sigma}^2$ of the errors?

(h) (*2 pts*) A side-by-side box plots of this data is given in the right panel of Figure 1. Does this plot indicate that the response Y_{ijk}'s should be transformed? Explain.

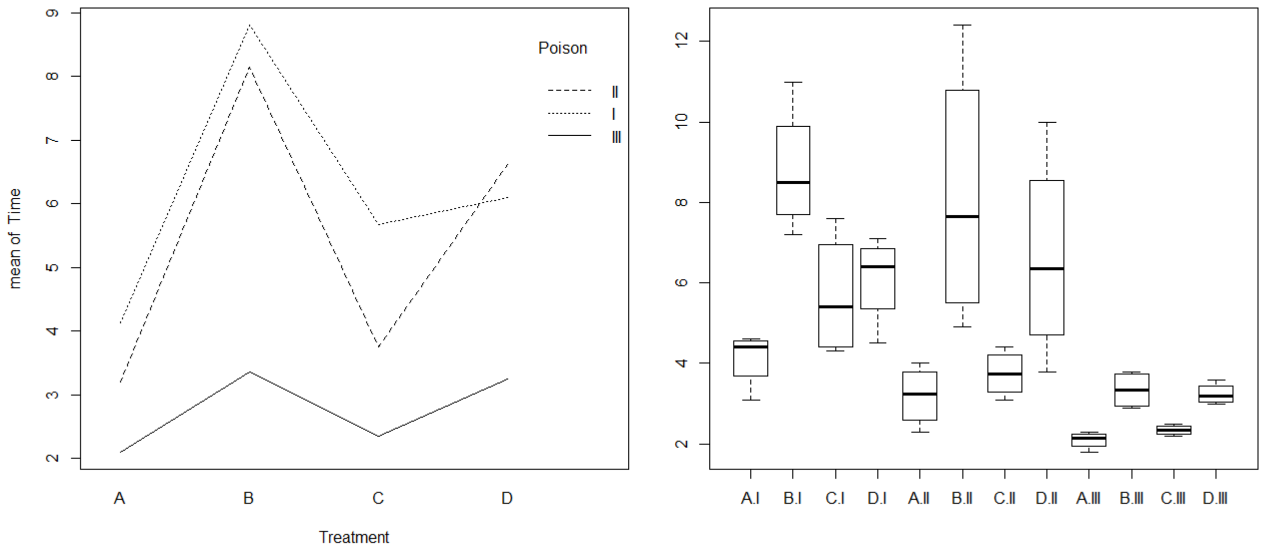

Figure 1: Interaction plot and box plots

5. (*12 pts*) (Tukey's pairwise comparison procedure for randomized block designs) Consider the data Y_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, obtained from a randomized block design. We assume a main-effect-only model for Y_{ij}'s:

$$Y_{ij} = \overline{\mu} + \alpha_i + \beta_j + \epsilon_{ij},$$

where α_i's (with the constraint $\sum_{i=1}^{I} \alpha_i = 0$) are the main effects of the treatment factor, β_j's (with the constraint $\sum_{j=1}^{J} \beta_j = 0$) are the main effects of the block factor, and ϵ_{ij}'s are i.i.d. from $N(0, \sigma^2)$. Recall that under this model, we have

$$SS_{TOT} = SS_A + SS_B + SS_{AB},$$

and the estimator of the error variance σ^2 is

$$\hat{\sigma}^2 = MS_{AB} = \frac{SS_{AB}}{(I-1)(J-1)}.$$

Suppose that we are interested in simultaneously testing the pairwise comparisons of treatments

$$H_0^{(i_1,i_2)} : \alpha_{i_1} = \alpha_{i_2} \quad \text{vs.} \quad H_A^{(i_1,i_2)} : \alpha_{i_1} \neq \alpha_{i_2}$$

for all pairs (i_1, i_2)'s of treatment means. Show that under $\bigcap_{1 \le i_1 < i_2 \le I} H_0^{(i_1,i_2)}$, the statistic

$$\max_{1 \le i_1 < i_2 \le I} \frac{\left|\overline{Y}_{i_1\cdot} - \overline{Y}_{i_2\cdot}\right|}{\hat{\sigma}/\sqrt{J}}$$

follows a studentized range distribution with parameters I and $(I-1)(J-1)$.

[**Hint**. (i) We can write

$$\max_{1 \le i_1 < i_2 \le I} \frac{\left|\overline{Y}_{i_1\cdot} - \overline{Y}_{i_2\cdot}\right|}{\hat{\sigma}/\sqrt{J}} = \frac{\left|\overline{Y}_{(I)} - \overline{Y}_{(1)}\right|}{\hat{\sigma}/\sqrt{J}}$$

where $\overline{Y}_{(1)}$ and $\overline{Y}_{(I)}$ are the minimum and maximum of $\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{I\cdot}$, respectively. (ii) $\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{I\cdot}$ and $\hat{\sigma}^2$ are independent.]