1. (*12pts*)

   (a) (*6pts*)

   - The experiment includes one treatment factors, i.e., *glycemic index (GI)*, and one block factor, i.e., *children*. The treatment factor GI has 2 levels and the block factor children has 10 levels.
   - This is a randomized block design.
   - An appropriate statistical model for this data is

   $$Y_{ij} = \overline{\mu} + \alpha_i + \beta_j + \epsilon_{ij}, \quad i = 1, 2, \ j = 1, \ldots, 10,$$

   where $\overline{\mu}$ is the grand mean, $\alpha_i$'s are the treatment effects, $\beta_j$'s are the block effects, and $\epsilon_{ij}$'s are i.i.d. random errors.

   (b) (*6pts*)

   - The experiment had one treatment factor, i.e., *species*, and no block factor. This treatment factor has 3 levels.
   - This is a one-way layout.
   - An appropriate statistical model for this data is

   $$Y_{ij} = \overline{\mu} + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, 3, \ j = 1, \ldots, 20,$$

   where $\overline{\mu}$ is the grand mean, $\alpha_i$'s are the treatment effects, and $\epsilon_{ij}$'s are i.i.d. random errors.

2. (*24pts*) Let $Z_i = D_i - \xi_0$, $i = 1, \ldots, n$. Then, $Z_1, \ldots, Z_n$ are i.i.d. from $f(x + \xi_0)$. Denote the median of $f(x + \xi_0)$ by $\xi^{(Z)}$. Because $f(x)$ has median $\xi_0$ if and only if $f(x + \xi_0)$ has median 0, testing $H_0 : \xi = \xi_0$ using the data $D_1, \ldots, D_n$ is equivalent to testing

   $$H_0 : \xi^{(Z)} = 0 \quad \text{vs.} \quad H_0 : \xi^{(Z)} \neq 0$$

   using the data $Z_1, \ldots, Z_n$. We can apply the sign test or the signed-rank test on the data $Z_i$'s to test $H_0 : \xi^{(Z)} = 0$ (i.e., $\xi = \xi_0$).

   (a) (*12pts*) Sign test using $Z_i$'s

   - test statistic: $N_+ = \#\{Z_i > 0\} = \#\{D_i > \xi_0\}$ (or $N_- = \#\{Z_i < 0\} = \#\{D_i < \xi_0\}$)
   - null distribution: under $H_0$, $N_+ \sim \text{binomial}(n, 1/2)$ (or $N_- \sim \text{binomial}(n, 1/2)$)
   - rejection region: reject $H_0$ if $N_+$ is small (close to 0) or large (close to $n$) (or $N_-$ is small or large), i.e., $|N_+ - n/2| > c$ (or $|N_- - n/2| > c$), where $c$ is the critical value from binomial$(n, 1/2)$.

   (b) (*12pts*) Wilcoxon signed-rank test using $Z_i$'s

   - test statistic
     i. Let $R_1, \ldots, R_n$ be the ranks of $|Z_1|, \ldots, |Z_n|$ (i.e., ranks of $|D_1 - \xi_0|, \ldots, |D_I - \xi_0|$).
     ii. Restore the signs of $Z_i$'s (i.e., signs of $(D_i - \xi_0)$'s) to the ranks $R_i$'s, obtaining signed ranks $R_i' = \text{sign}(Z_i) \times R_i$, $i = 1, \ldots, n$.

iii. The test statistic $W_+$ is the sum of the signed ranks $R_i'$'s that have positive signs, i.e.,

$$W_+ = \sum_{i=1}^{n} I_{[Z_i>0]} R_i' = \sum_{i=1}^{n} I_{[Z_i>0]} R_i = \sum_{i=1}^{n} I_{[D_i>\xi_0]} R_i, \tag{I}$$

where $I$ is the indicator function.

- null distribution
  - Under $H_0$, each possible outcomes of $R_i$'s have equal probability $1/n!$, and $I_{[Z_1>0]}, \ldots, I_{[Z_n>0]} \sim$ i.i.d. Bernoulli(1/2). Furthermore, the two sets of random variables $\{R_i\text{'s}\}$ and $\{I_{[Z_i>0]}\text{'s}\}$ are independent. The exact null distribution of $W_+$ can be exhaustively enumerated from all possible outcomes of $\{R_i\text{'s}\}$ and $\{I_{[Z_i>0]}\text{'s}\}$, each with probability $1/(n! \times 2^n)$, via (I). An alternative way is to randomly assign positive or negative signs to each of the integers $1, \ldots, n$ (the ranks). There are $2^n$ such assignments, and each of them has equal probability $1/2^n$. For each assignment, calculate the value of $W_+$ and assign the value a probability of $1/2^n$ to generate the exact null distribution of $W_+$.
  - When $n > 20$, the asymptotic null distribution of $W_+$ is the normal distribution with mean $\dfrac{n(n+1)}{4}$ and variance $\dfrac{n(n+1)(2n+1)}{24}$.
- rejection region: Let $W_- = \frac{n(n+1)}{2} - W_+$. Reject $H_0$ when $\min(W_+, W_-)$ is small (close to 0), i.e., $\min(W_+, W_-) < w$, where $w$ is the critical value from the (exact or asymptotic) null distribution of $W_+$. An alternative expression of this rejection region is

$$\left| W_+ - \frac{n(n+1)}{4} \right| > w.$$

3. (*22pts*) Notice that for each sample, the sample mean $\overline{Y}_{i\cdot}$ and sample variance $s_i^2$ are independent, and

$$\overline{Y}_{i\cdot} \sim N(\mu_i, \sigma^2/J_i) \quad \text{and} \quad (J_i - 1)s_i^2/\sigma^2 \sim \chi^2_{J_i - 1}\,. \tag{II}$$

Furthermore, we know that

$$\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{I\cdot}, s_1^2, \ldots, s_I^2 \text{ are independent random variables.} \tag{III}$$

(a) (*6pts*) We have

$$\frac{SS_W}{\sigma^2} = \sum_{i=1}^{I} \frac{(J_i - 1)s_i^2}{\sigma^2},$$

where

i. by (II), $\dfrac{(J_i - 1)s_i^2}{\sigma^2} \sim \chi^2_{J_i - 1}$, $i = 1, \ldots, I$, and

ii. by (III),

$$\frac{(J_1 - 1)s_1^2}{\sigma^2}, \ldots, \frac{(J_I - 1)s_I^2}{\sigma^2} \text{ are independent.}$$

Because the sum of $k$ independent chi-square random variables, each with $n_i$ degrees of freedom, follows a chi-square distribution with $n_1 + \cdots + n_k$ degrees of freedom, we know that

$$\frac{SS_W}{\sigma^2} \sim \chi^2_{(J_1 - 1) + \cdots + (J_I - 1) = N - I}\,.$$

(b) (*6pts*) Under the null $H_0 : \mu_1 = \cdots = \mu_I \equiv \mu$, we have that the $N = J_1 + \cdots + J_I$ observations

$$Y_{11}, \ldots, Y_{1J_1}, Y_{21}, \ldots, Y_{2J_2}, \ldots, Y_{I1}, \ldots, Y_{IJ_I} \sim \text{i.i.d. } N(\mu, \sigma^2),$$

which can be regarded as a one-sample model. Because $\overline{Y}_{\cdot\cdot}$ is the sample mean of this one-sample data, we find that $SS_{TOT}$ is the sample variance of this one-sample data. By the property of normal distribution, the sample variance

$$SS_{TOT} \sim \chi^2_{(J_1+\cdots+J_I)-1=N-1}\,.$$

(c) (*4pts*) Note that $SS_W$ is a function of $s_1^2, \ldots, s_I^2$ and $SS_B$ is a function of $\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{I\cdot}$ ($\overline{Y}_{\cdot\cdot}$ is a weighted average of $\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{I\cdot}$). By (III), $SS_W$ and $SS_B$ are independent because they are functions of different random variables that are independent.

(d) (*6pts*) Denote $U = SS_W/\sigma^2$, $V = SS_B/\sigma^2$, and $W = SS_{TOT}/\sigma^2$. Then, we have

$$W = U + V \quad \text{and} \quad U, V \text{ are independent (by (c)).} \qquad \text{(IV)}$$

By (a), the moment generating function (mgf) of $U$ is $M_U(t) = (1 - 2t)^{(N-I)/2}$. By (b), the mgf of $W$ is $M_W(t) = (1 - 2t)^{(N-1)/2}$ under $H_0$. Let $M_V(t)$ be the mgf of $V$. From (IV), we know $M_W(t) = M_U(t) \times M_V(t)$. Hence, under $H_0$,

$$M_V(t) = \frac{M_W(t)}{M_U(t)} = \frac{(1-2t)^{(N-1)/2}}{(1-2t)^{(N-I)/2}} = (1-2t)^{(I-1)/2},$$

which is the mgf of a chi-square distribution with $I - 1$ degrees of freedom. By the uniqueness theorem of mgf, we know that $V = SS_B/\sigma^2 \sim \chi^2_{I-1}$ under $H_0$.

4. (*30pts*)

(a) (*2pts*) Yes, because we have equal replicates (4 observations) in each cell.

(b) (*4pts*) For Treatment B, the estimate of $\beta_j$ is

$$\overline{Y}_{\cdot 2\cdot} - \overline{Y}_{\cdots} = 6.767 - 4.790 = 1.977.$$

For the cell of Poison I and Treatment B, the estimate of $\delta_{ij}$ is

$$\overline{Y}_{14\cdot} - \overline{Y}_{1\cdot\cdot} - \overline{Y}_{\cdot 4\cdot} + \overline{Y}_{\cdots} = 6.1 - 6.175 - 5.325 + 4.790 = -0.610.$$

(c) (*12pts*) The complete ANOVA table is

| Source | $SS$ | $df$ | $MS$ | $F$ | $p$-value |
|---|---|---|---|---|---|
| Treatment | 91.9 | **3** | **30.6333** | **14.013** | 0.000 |
| Poison | **103** | **2** | 51.500 | 23.558 | 0.000 |
| Interaction | 24.7 | **6** | **4.1167** | **1.883** | 0.110 |
| Error | **78.7** | **36** | 2.186 | | |
| Total | 298.3 | **47** | | | |

(d) (*2pts*) The degrees of freedom are 2 and 36.

(e) (*2pts*) $\hat{\sigma}^2 = 2.186$.

(f) (*2pts*) Yes, because the $F$-test of Interaction has a $p$-value of 0.110, which is insignificant.

(g) (*4pts*) Under the main-effect-only model (no interaction), the sum of squares for errors is 24.7+78.7=103.4 with degrees of freedom 6+36=42. Hence,

$$\hat{\sigma}^2 = \frac{24.7 + 78.7}{6 + 36} = 2.4619.$$

(h) (*2pts*) Yes, because the plot shows that the error variance increases with the mean.

5. (*12pts*) Under $\bigcap_{1 \le i_1 < i_2 \le I} H_0^{(i_1, i_2)}$ (i.e., $\alpha_1 = \cdots = \alpha_I = 0$), we have

   (i) $\overline{Y}_{i\cdot} \sim N(\overline{\mu}, \sigma^2/J)$ because $\sum_{j=1}^{J} \beta_j = 0$,

   (ii) $[(I-1)(J-1)]\hat{\sigma}^2/\sigma^2 \sim \chi^2_{(I-1)(J-1)}$,

   (iii) $\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{I\cdot}, \hat{\sigma}^2$ are independent.

   From (i) and (iii), we know that

$$\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{I\cdot} \sim \text{i.i.d. } N(\overline{\mu}, \sigma^2/J)\,,$$

   i.e., $\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{I\cdot}$ is like a one-sample data under $\bigcap_{1 \le i_1 < i_2 \le I} H_0^{(i_1, i_2)}$. Then, by standardizing $\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{I\cdot}$, we get

$$Z_1 \equiv \frac{\overline{Y}_{1\cdot} - \overline{\mu}}{\sigma/\sqrt{J}}, \ldots, Z_I \equiv \frac{\overline{Y}_{I\cdot} - \overline{\mu}}{\sigma/\sqrt{J}} \sim \text{ i.i.d. } N(0, 1) \qquad \text{(V)}$$

   and

$$\hat{\sigma}^2 \text{ is independent of } Z_1, \ldots, Z_I \text{ (by (iii))}. \qquad \text{(VI)}$$

   Let $\overline{Y}_{(1)}, \ldots, \overline{Y}_{(I)}$ be the order statistics of $\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{I\cdot}$. Then,

$$Q \equiv \max_{1 \le i_1 < i_2 \le I} \frac{\left|\overline{Y}_{i_1\cdot} - \overline{Y}_{i_2\cdot}\right|}{\hat{\sigma}/\sqrt{J}} = \frac{\overline{Y}_{(I)} - \overline{Y}_{(1)}}{\hat{\sigma}/\sqrt{J}} = \frac{\frac{\overline{Y}_{(I)} - \overline{\mu}}{\sigma/\sqrt{J}} - \frac{\overline{Y}_{(1)} - \overline{\mu}}{\sigma/\sqrt{J}}}{\sqrt{\frac{[(I-1)(J-1)]\hat{\sigma}^2}{\sigma^2} \times \frac{1}{(I-1)(J-1)}}}\,.$$

   where

   $\frac{\overline{Y}_{(I)} - \overline{\mu}}{\sigma/\sqrt{J}}$ and $\frac{\overline{Y}_{(1)} - \overline{\mu}}{\sigma/\sqrt{J}}$ are the maximum and the minimum of $Z_1, \ldots, Z_I$, respectively. (VII)

   Thus, by (iii) and (V)-(VII), we know that $Q$ follows the studentized range distribution with parameter $I$ and $(I-1)(J-1)$ under $\bigcap_{1 \le i_1 < i_2 \le I} H_0^{(i_1, i_2)}$.