

STAT416 Assignment 2

Ashish Jain

Question 1:

- a). The treatment factors considered in the experiment are temperature and breed of the mouse.
- b). The treatment levels considered in the experiment for temperature are normal, cold and hot and for genotype, it is breed 1 and breed 2.
- c). Experimental units in this experiment are the houses (6 houses) for temperature as treatment factor and chickens (24 chickens) for the breed as the treatment factor.
- d). The observational units in this experiment are the chickens.
- e). There is no blocking involved in this experiment.
- f). This experiment follows a split-plot design. In this experiment, the houses are the whole plot units because the levels of the temperature (whole-plot factor) are randomly assigned to it. The split-plots units are the chickens because the levels of the genotype (split-plot factor) is randomly assigned to the chickens.

Question 2:

The design for this experiment is shown in figure 1. In the figure, $C_{i,j}$ represents the chicken with i as the temperature such that $i \in (\text{Hot, Normal, Cold})$ and j as the breed of chicken such as $j \in (\text{Breed 1, Breed 2})$. In order to answer the questions effectively, in our design, we sequence houses 1, 2, and 3 in one flow cell and the houses 4, 5, and 6 in the other flow cell.

Our question is to find genes differentially expressed between the two breeds under each temperature condition. In our design, we are sequencing the samples from the same house in the same flow cell in such a way that we can eliminate the adapter effect. In addition to that, we are also eliminating the flow cell effect by sequencing the biological replicates (houses) in terms of temperature on the different flow cell. By using this design, we can eliminate almost all the effects except the lane effect. For the secondary goal, we have to compare the temperature conditions for each breed. In this case, the only variance that will come in the estimation of this goal will be due to the house and the lane effect. In this case, the temperature difference and the house effect will be confounding as per our design.

Question 3:

In this experiment, we will design the phase 2 in such a way that we will confound the house effect and the session effect. For that, we will prepare each sample in a house in one session (6 sessions in total for 6 houses). In a session, we can take processing order to be random such that it will eliminate any bias related to processing. After that, for phase 3 we will match the order of RNA extraction from samples within phase 3 sessions to that of the tissue collection order used in phase 2 and again take the random processing order. These confounding will be applied to the design which will help us to answer the first research question in which we need to find genes differentially expressed between the two breeds under each temperature condition. In this case, again we can confound the session effect and the house effect. For the secondary goal, we have to compare the temperature conditions for each breed. In this case, the variance/effect that will come in the estimation of this goal will be due to the house and the lane effect. In this case, the temperature difference and the house effect will also be confounding as per our design.

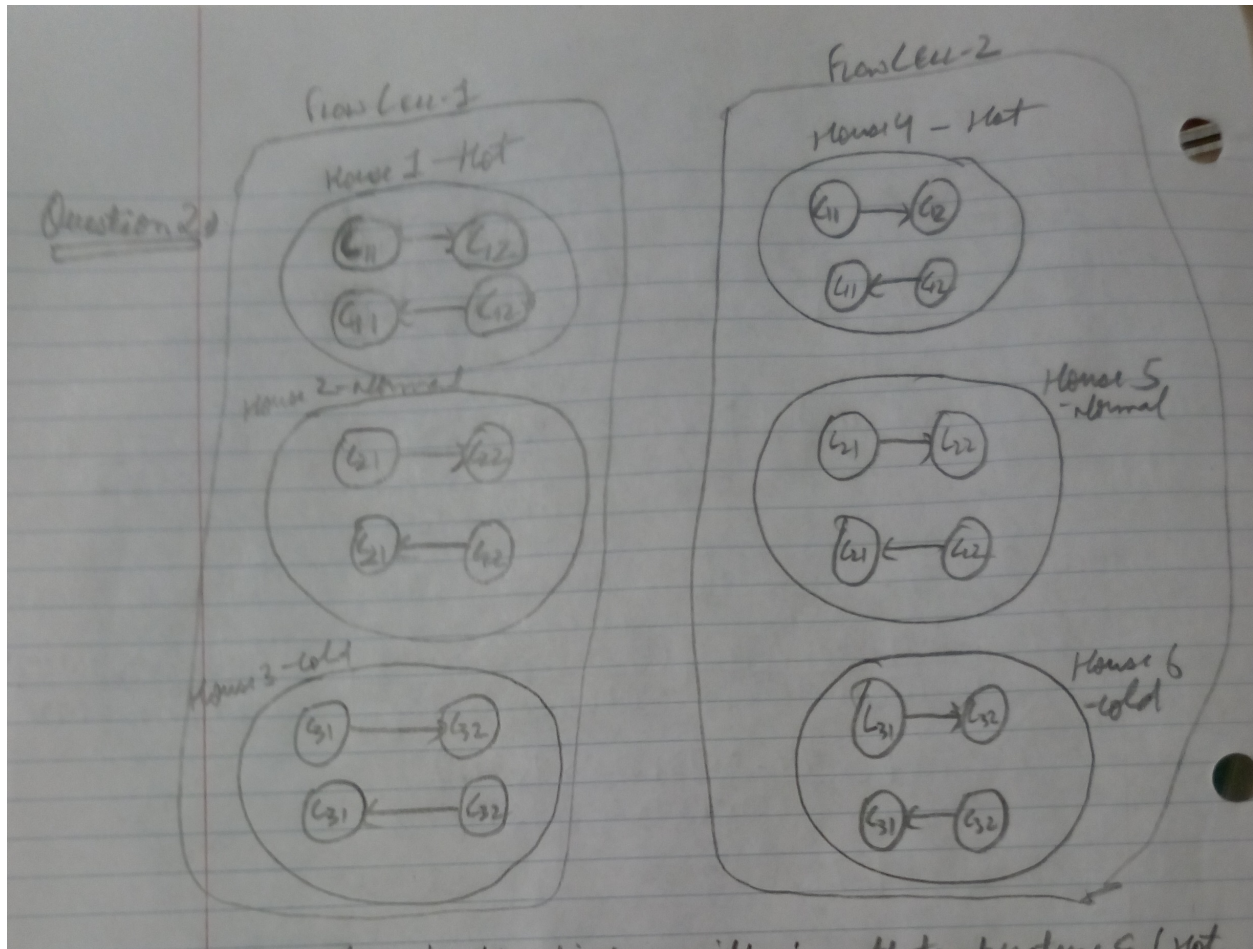


Figure 1: Experimental Design - Question 2

Question 4:

a)

i).

$$\begin{aligned} E(2Y_1 + 3) &\stackrel{ind.}{=} E(2Y_1) + E(3) \\ &= 2E(Y_1) + 3 \\ &= 2\mu_1 + 3 \end{aligned}$$

ii).

$$\begin{aligned} E(Y_1 + Y_2 + Y_3) &\stackrel{ind.}{=} E(Y_1) + E(Y_2) + E(Y_3) \\ &= \mu_1 + \mu_2 + \mu_3 \end{aligned}$$

b).

i).

$$E(c) = c$$

ii).

$$\begin{aligned} Var(c) &= E[\{c - E(c)\}^2] \\ &= E[\{c - c\}^2] \\ &= 0 \end{aligned}$$

c).

i).

$$\begin{aligned} Var(Y_1 - Y_2) &\stackrel{ind.}{=} Var(Y_1) + Var(Y_2) \\ &= \sigma^2 + \sigma^2 \\ &= 2\sigma^2 \end{aligned}$$

ii).

$$\begin{aligned} Var\left(\frac{Y_1 + Y_2}{2}\right) &= \frac{Var(Y_1 + Y_2)}{4} \\ &\stackrel{ind.}{=} \frac{Var(Y_1)}{4} + \frac{Var(Y_2)}{4} \\ &= \frac{\sigma^2}{4} + \frac{\sigma^2}{4} \\ &= 2 * \frac{\sigma^2}{4} \\ &= \frac{\sigma^2}{2} \end{aligned}$$

d).

i).

$$\begin{aligned} Var\left(\sum_{i=1}^n Y_i\right) &\stackrel{ind.}{=} Var(Y_1) + Var(Y_2) + \dots + Var(Y_n) \\ &= \sigma^2 + \sigma^2 + \dots + \sigma^2 \\ &= n\sigma^2 \end{aligned}$$

ii).

$$\begin{aligned}
Var\left(\frac{\sum_{i=1}^n Y_i}{n}\right) &\stackrel{ind.}{=} Var\left(\frac{Y_1}{n}\right) + Var\left(\frac{Y_2}{n}\right) + \dots + Var\left(\frac{Y_n}{n}\right) \\
&= \frac{Var(Y_1)}{n^2} + \frac{Var(Y_2)}{n^2} + \dots + \frac{Var(Y_n)}{n^2} \\
&= \frac{\sigma^2}{n^2} + \frac{\sigma^2}{n^2} + \dots + \frac{\sigma^2}{n^2} \\
&= n * \frac{\sigma^2}{n^2} \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

e).

$$\begin{aligned}
Var(Y_{1..}^- - Y_{2..}^-) &\stackrel{FromSlides}{=} Var(T_1 + T_2 + b_{1.}^- + b_{2.}^- + e_{1..}^- + e_{2..}^-) \\
&\stackrel{ind.}{=} Var(T_1) + Var(T_2) + Var(b_{1.}^-) + Var(b_{2.}^-) + Var(e_{1..}^-) + Var(e_{2..}^-) \\
&= 0 + 0 + \frac{\sigma_b^2}{n} + \frac{\sigma_b^2}{n} + \frac{\sigma_e^2}{m} + \frac{\sigma_e^2}{m} \\
&= 2 * \frac{\sigma_b^2}{n} + 2 * \frac{\sigma_e^2}{m}
\end{aligned}$$

i). In Design 1, n is equal to 8 and m is equal to 8,

$$\begin{aligned}
Var(Y_{1..}^- - Y_{2..}^-) &= 2 * \frac{\sigma_b^2}{n} + 2 * \frac{\sigma_e^2}{m} = 2 * \frac{\sigma_b^2}{8} + 2 * \frac{\sigma_e^2}{8} \\
&= \frac{\sigma_b^2}{4} + \frac{\sigma_e^2}{4} \\
&= \frac{\sigma_b^2 + \sigma_e^2}{4}
\end{aligned}$$

ii).

In Design 2, n is equal to 4 and m is equal to 8,

$$\begin{aligned}
Var(Y_{1..}^- - Y_{2..}^-) &= 2 * \frac{\sigma_b^2}{n} + 2 * \frac{\sigma_e^2}{m} = 2 * \frac{\sigma_b^2}{4} + 2 * \frac{\sigma_e^2}{8} \\
&= 2 * \frac{\sigma_b^2}{4} + \frac{\sigma_e^2}{4} \\
&= \frac{2\sigma_b^2 + \sigma_e^2}{4}
\end{aligned}$$