

ST 518 Homework 9

Eric Warren

November 11, 2023

Contents

1	Problem 1	1
1.1	Part A	1
1.2	Part B	2
1.3	Part C	2
1.4	Part D	2
1.5	Part E	2
1.6	Part F	2
1.7	Part G	2
1.8	Part H	3
1.9	Part I	3
2	Problem 2	3
2.1	Part A	3
2.2	Part B	3
2.3	Part C	3
2.4	Part D	4

1 Problem 1

Consider a random effects model to analyze data from a completely randomized experimental design, $Y_{ij} = \mu + T_i + E_{ij}$, with $i = 1, \dots, t$ and $j = 1, \dots, n$ where $T_i \sim^{iid} N(0, \sigma_T^2)$ and $E_{ij} \sim^{iid} N(0, \sigma^2)$ with $T_i \perp E_{ij}$.

1.1 Part A

Give the mean and variance of an observation, Y_{ij} .

The mean of Y_{ij} , given that $Y_{ij} = \mu + T_i + E_{ij}$, is $E(Y_{ij}) = E(\mu + T_i + E_{ij}) = E(\mu) + E(T_i) + E(E_{ij}) = \mu + 0 + 0 = \mu$. The variance of Y_{ij} is $Var(Y_{ij}) = Var(\mu + T_i + E_{ij}) = Var(\mu) + Var(T_i) + Var(E_{ij}) = 0 + \sigma_T^2 + \sigma^2 = \sigma_T^2 + \sigma^2$.

1.2 Part B

Write a treatment mean, $\bar{Y}_{i.}$ as a function of μ and (averages of) random effects.

We know that $\bar{Y}_{i.} = \frac{1}{n} \sum (\mu + T_i + E_{ij}) = \mu + \frac{1}{n} (T_i + T_i + \dots + T_i) + \frac{1}{n} (E_{i1} + E_{i2} + \dots + E_{in}) = \mu + T_i + \bar{E}_{i.}$

1.3 Part C

Give the mean and variance of a sample treatment mean, $\bar{Y}_{i.}$

We can find the mean of $\bar{Y}_{i.}$ by saying that $E(\bar{Y}_{i.}) = E(\mu + T_i + \bar{E}_{i.}) = E(\mu) + E(T_i) + E(\bar{E}_{i.}) = \mu + 0 + 0 = \mu$. The variance of $\bar{Y}_{i.}$ is $Var(\mu + T_i + \bar{E}_{i.}) = Var(\mu) + Var(T_i) + Var(\bar{E}_{i.}) = 0 + \sigma_T^2 + \frac{\sigma^2}{n} = \sigma_T^2 + \frac{\sigma^2}{n} = \frac{n\sigma_T^2 + \sigma^2}{n} = \frac{1}{n} E(MS(Trt))$.

1.4 Part D

Write the grand mean, $\bar{Y}_{..}$ as a function of μ and averages of random effects.

$\bar{Y}_{..} = \frac{1}{nt} (\mu + T_i + E_{ij}) = \mu + \frac{nT_1 + \dots + nT_t}{nt} + \frac{1}{nt} (E_{11} + E_{12} + \dots + E_{1n} + E_{t1} + \dots + E_{tn}) = \mu + \bar{T}_{.} + \bar{E}_{..}$

1.5 Part E

Give the mean and variance of the grand mean, $\bar{Y}_{..}$. Be clear about where the assumption that $T_i \perp E_{ij}$ is used.

We can first find the mean for $\bar{Y}_{..}$ which is $E(\bar{Y}_{..}) = E(\mu + \bar{T}_{.} + \bar{E}_{..}) = E(\mu) + E(\bar{T}_{.}) + E(\bar{E}_{..}) = \mu + 0 + 0 = \mu$. The variance for $\bar{Y}_{..}$ can be found by first using the assumption that comes from $T_i \perp E_{ij}$ being true. In this case, since this property of both random variables T and E being independent than the average of those two random variables $\bar{T}_{.}$ and $\bar{E}_{..}$ are also independent. Because of this we can say the variance of this sum of averages is the same of the sum of the variance of averages. Therefore we can say that $Var(\bar{Y}_{..}) = Var(\mu + \bar{T}_{.} + \bar{E}_{..}) = Var(\mu) + Var(\bar{T}_{.}) + Var(\bar{E}_{..}) = 0 + \frac{\sigma_T^2}{t} + \frac{\sigma^2}{nt} = \frac{1}{nt} (n\sigma_T^2 + \sigma^2) = \frac{1}{nt} E(MS(T))$.

1.6 Part F

Derive $E(\frac{1}{t-1} \sum \sum (\bar{Y}_{i.} - \bar{Y}_{..})^2)$

We know that $\sum \sum (\bar{Y}_{i.} - \bar{Y}_{..})^2 = SS(Trt)$ and that $MS(Trt) = \frac{SS(Trt)}{t-1}$ and we will use these two things to get $E(\frac{1}{t-1} \sum \sum (\bar{Y}_{i.} - \bar{Y}_{..})^2)$. Therefore, $E(\frac{1}{t-1} \sum \sum (\bar{Y}_{i.} - \bar{Y}_{..})^2) = E(\frac{1}{t-1} * SS(Trt)) = E(\frac{SS(Trt)}{t-1}) = E(MS(Trt)) = n\sigma_T^2 + \sigma^2$.

1.7 Part G

What is the sampling distribution of the statistic W where $W = \frac{\sum \sum (\bar{Y}_{i.} - \bar{Y}_{..})^2}{n\sigma_T^2 + \sigma^2}$?

We know in this case that $\sum \sum (\bar{Y}_{i.} - \bar{Y}_{..})^2 = SS(Trt) = (t-1)MS(Trt)$. Thus, $W = \frac{\sum \sum (\bar{Y}_{i.} - \bar{Y}_{..})^2}{n\sigma_T^2 + \sigma^2} = \frac{SS(Trt)}{n\sigma_T^2 + \sigma^2} = \frac{(t-1)MS(Trt)}{n\sigma_T^2 + \sigma^2} = (t-1) \frac{MS(Trt)}{n\sigma_T^2 + \sigma^2}$ in which we know that this statistic W follows a sampling distribution of χ_{t-1}^2 or χ^2 with $t-1$ degrees of freedom.

1.8 Part H

Derive $E(\frac{1}{t(n-1)} \sum \sum (Y_{ij} - \bar{Y}_{i.})^2)$

We know that $\sum \sum (Y_{ij} - \bar{Y}_{i.})^2 = SS(E)$ and that $MS(E) = \frac{SS(E)}{t(n-1)}$ and we will use these two things to get $E(\frac{1}{t(n-1)} \sum \sum (Y_{ij} - \bar{Y}_{i.})^2)$. Therefore, $E(\frac{1}{t(n-1)} \sum \sum (Y_{ij} - \bar{Y}_{i.})^2) = E(\frac{1}{t(n-1)} SS(E)) = E(\frac{SS(E)}{t(n-1)}) = E(MS(E)) = \sigma^2$.

1.9 Part I

What is the sampling distribution of the statistic X where $X = \frac{\sum \sum (Y_{ij} - \bar{Y}_{i.})^2}{\sigma^2}$

Note that $N - t = t(n - 1)$. We know in this case that $\sum \sum (Y_{ij} - \bar{Y}_{i.})^2 = SS(E) = (t(n - 1))MS(E)$. Thus, $W = \frac{\sum \sum (Y_{ij} - \bar{Y}_{i.})^2}{\sigma^2} = \frac{SS(E)}{\sigma^2} = \frac{(t(n-1))MS(E)}{\sigma^2} = (t(n - 1)) \frac{MS(E)}{\sigma^2} = (N - t) \frac{MS(E)}{\sigma^2}$ in which we know that this statistic X follows a sampling distribution of χ^2_{N-t} or χ^2 with $N - t$ degrees of freedom or lastly the same as χ^2 with $t(n - 1)$ degrees of freedom.

2 Problem 2

Consider a mixed model for a crossed two factor design, $Y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + E_{ijk}$ with $i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, n$. $B_j \sim^{iid} N(0, \sigma_B^2)$ and $(\alpha B)_{ij} \sim^{iid} N(0, \sigma_{\alpha B}^2)$ and $E_{ijk} \sim^{iid} N(0, \sigma^2)$. Assume B_j and $(\alpha B)_{ij}$ and E_{ijk} are mutually independent random samples.

2.1 Part A

Consider the difference between two A treatment means, $\bar{Y}_{2..} - \bar{Y}_{1..}$. Derive the variance of this difference.

We know that $\bar{Y}_{2..} = \frac{1}{nb} \sum_{j=1}^b \sum_{k=1}^n (\mu + \alpha_2 + B_j + (\alpha B)_{2j} + E_{2jk}) = \mu + \alpha_2 + \bar{B} + (\alpha \bar{B})_2 + \bar{E}_{2..}$ and $\bar{Y}_{1..} = \frac{1}{nb} \sum_{j=1}^b \sum_{k=1}^n (\mu + \alpha_1 + B_j + (\alpha B)_{1j} + E_{1jk}) = \mu + \alpha_1 + \bar{B} + (\alpha \bar{B})_1 + \bar{E}_{1..}$. Thus, $\bar{Y}_{2..} - \bar{Y}_{1..} = \mu + \alpha_2 + \bar{B} + (\alpha \bar{B})_2 + \bar{E}_{2..} - (\mu + \alpha_1 + \bar{B} + (\alpha \bar{B})_1 + \bar{E}_{1..}) = \alpha_2 - \alpha_1 + (\alpha \bar{B})_2 - (\alpha \bar{B})_1 + \bar{E}_{2..} - \bar{E}_{1..}$. Therefore, the $Var(\bar{Y}_{2..} - \bar{Y}_{1..}) = Var(\alpha_2 - \alpha_1 + (\alpha \bar{B})_2 - (\alpha \bar{B})_1 + \bar{E}_{2..} - \bar{E}_{1..}) = \frac{\sigma_{\alpha B}^2}{b} + \frac{\sigma_{\alpha B}^2}{b} + \frac{\sigma^2}{nb} \frac{\sigma^2}{nb} = \frac{2\sigma_{\alpha B}^2}{b} + \frac{2\sigma^2}{nb} = \frac{2}{nb} (n\sigma_{\alpha B}^2 + \sigma^2) = \frac{2}{nb} E(MS(\alpha B))$.

2.2 Part B

Use the rules given on slides 17 to specify $E[MS(\alpha B)]$.

We know from our table that for a mixed model for a crossed two factor design, $E[MS(\alpha B)] = \sigma^2 + n\sigma_{\alpha B}^2$.

2.3 Part C

Consider this model for the data from the experiment to compare campylobacter counts across locations in chicken plants, with data collected at several randomly chosen days. Report a 95% confidence interval for the mean difference in bacteria count (on a log-scale) before and after the washer. Though a mixed model was not fit, the output from the fitted model below should supply the necessary statistics to compute the interval.

In this case, we know that $\bar{Y}_{1..}$ is the mean of bacteria count (on a log-scale) after the washer and $\bar{Y}_{2..}$ is the mean of bacteria count (on a log-scale) before the washer. We also know from our output that $\bar{Y}_{1..} =$

10.4953141 and $\bar{Y}_{2..} = 10.8869897$. Thus, $\bar{Y}_{2..} - \bar{Y}_{1..} = 10.8869897 - 10.4953141 = 0.3916756$. Now we know the confidence interval for this is $\bar{Y}_{2..} - \bar{Y}_{1..} \pm t_{df, 1 - \frac{\alpha}{2}} SE(\bar{Y}_{2..} - \bar{Y}_{1..}) = \bar{Y}_{2..} - \bar{Y}_{1..} \pm t_{df, 1 - \frac{\alpha}{2}} \sqrt{Var(\bar{Y}_{2..} - \bar{Y}_{1..})} = \bar{Y}_{2..} - \bar{Y}_{1..} \pm t_{df, 1 - \frac{\alpha}{2}} \sqrt{\frac{2}{nb} E(MS(\alpha B))}$. From the chart, we know that $E(MS(\alpha B)) = MS(\hat{\alpha}B) = 0.75559410$. Now note that in a 95% confidence interval, $\alpha = 1 - .95 = .05$ so $\frac{\alpha}{2} = \frac{.05}{2} = .025$. Lastly, our df in this case is the same as the df of $MS(\hat{\alpha}B)$ which is $(a - 1)(b - 1) = (3 - 1)(4 - 1) = 2 * 3 = 6$ (as this matches our output in the directions as well). So we can now plug in our confidence interval to get $\bar{Y}_{2..} - \bar{Y}_{1..} \pm t_{df, 1 - \frac{\alpha}{2}} \sqrt{\frac{2}{nb} E(MS(\alpha B))} = 0.3916756 \pm t(6, 1 - .025) \sqrt{\frac{2}{30} 0.75559410} = 0.3916756 \pm t(6, .975) \sqrt{\frac{2}{30} 0.75559410} = 0.3916756 \pm 2.446912 \sqrt{\frac{2}{30} 0.75559410}$. By plugging in this final confidence interval we get a lower bound of -0.1575073 and an upper bound of 0.9408585. Therefore, we are 95% confident that the true mean difference in bacteria count (on a log-scale) before and after the washer is between -0.1575073 and 0.9408585.

2.4 Part D

Report a 95% confidence interval for the error variance.

We know that the confidence interval for error variance is represented by $(\frac{(ab(n-1))MS(E)}{\chi^2_{1 - \frac{\alpha}{2}, (ab(n-1))}}, \frac{(ab(n-1))MS(E)}{\chi^2_{\frac{\alpha}{2}, (ab(n-1))}})$. In this case we know that $ab(n - 1) = 4(3)(10 - 1) = 4(3)(9) = 108$ and $MS(E) = 0.5486569$ from our output in the directions. We also know from **Part C** that $1 - \frac{\alpha}{2} = 0.975$ and $\frac{\alpha}{2} = 0.025$ for 95% confidence intervals. Therefore, we can say that $(\frac{(ab(n-1))MS(E)}{\chi^2_{1 - \frac{\alpha}{2}, (ab(n-1))}}, \frac{(ab(n-1))MS(E)}{\chi^2_{\frac{\alpha}{2}, (ab(n-1))}}) = (\frac{108 * 0.5486569}{\chi_{.975, 108}}, \frac{108 * 0.5486569}{\chi_{.025, 108}}) = (\frac{59.25495}{138.6506}, \frac{59.25495}{81.13292}) = (0.4273689, 0.7303441)$. Therefore, we are 95% confident that the true error variance is between 0.4273689 and 0.7303441.