STA303: M3 class activity

Return to Statdew Valley

Instructions

To participate in this activity you will need have to two windows readily available to you:

1) Your Zoom window

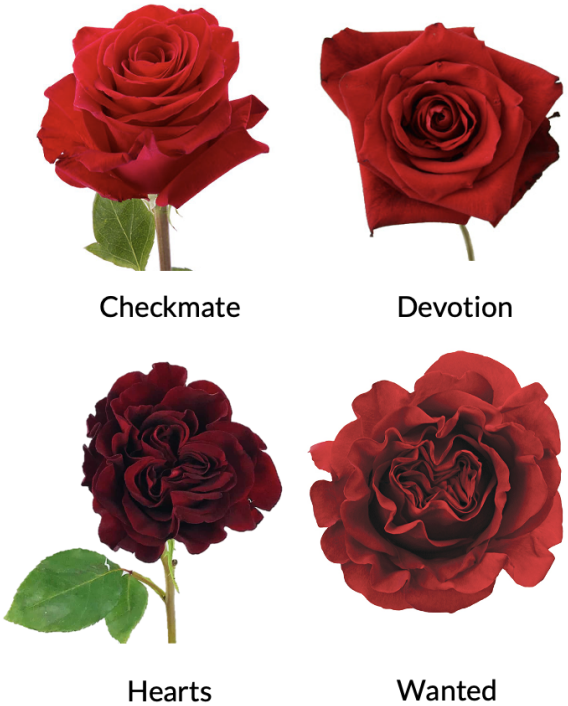
2) The Team Up! activity linked from Quercus in a browser window for voting.

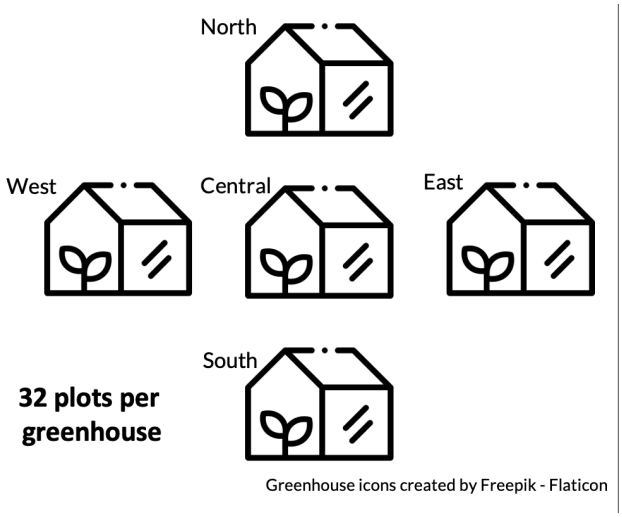
I would recommend that one member of the team shares their screen with the rest of the team and shows this activity where you can see the question and options.

Note 1: In the Team Up! activity you will just see the letters for the questions, not the options themselves.

Note 2: There are hints for some questions, but if you get really stuck, please use the 'Ask for Help' option in Zoom, our join the Zoom main room.

Data





This Team Up! activity uses the date and ideas from the rose example we started to explore in synchronous class on Wednesday (Feb 9). I have improved the description a little so it hopefully makes sense alone, without my additional commentary.

Suppose, you run a series of greenhouses in which you grow hot house roses. In preparation for Valentine's Day LAST year, you planted an equal number of plots of each of the four special varieties of red rose you have in each greenhouse. Then the flowers were cut and sold and it was recorded how much net profit was made per plot (in 000s of dollars).

At the time there were 5 greenhouses, and they were all at different levels of modernization as it can be very expensive to replace and update irrigation systems and other fixtures. You hope to continue making general improvements and have since acquired 3 more greenhouses.

- The 5 original green houses are referred to as North, East, South, West and Central.
- There are 4 varieties of red rose that your buisness specializes in: Devotion, Checkmate, Wanted, and Hearts.
- Within each greenhouse, there are 32 plots.

Now, you want to understand your data from last year and make predictions about your 2022 profits, as well as consider how you might prioritize which varieties to grow more of in your new greenhouses.

```
## Rows: 160
## Columns: 3
## $ greenhouse <chr> "north", "north", "north", "north", "north", "north", "nort~
## $ variety    <chr> "Devotion", "Devotion", "Devotion", "Devotion", "Devotion",~
## $ profit     <dbl> 23.44, 20.42, 23.86, 28.74, 29.10, 24.73, 23.86, 24.11, 26.~
```

Question 1

To answer the following question, wrangle and plot the rose data as follows:

1. Change the order of the levels of the variety factor so that instead of being alphabetical, they go: "north", "south", "east", "west", "central".

- This WON'T change the way the data appears when you view it from the environment pane but WILL change how ggplot plots it.

1. Pipe this changed `roses` data directly into a ggplot.
2. Put `greenhouse` on your x-axis, `profit` on your y, and colour by `variety`.
3. Use points as your geometry. (There is probably a bit of overprinting, but we won't worry about that here.)
4. You can add `theme_minimal()` if you like, but use the default colour palette.

Which greenhouse and variety combination had the lowest profit blue dot on your graph?

A. Devotion variety in Central greenhouse. B. Wanted variety in the West greenhouse. C. Hearts variety in the South greenhouse. D. Impossible to say.

Question 2

Considering the `roses` data described in the Data section and the plot you made, which ONE of the following statements best describes an appropriate analysis for this data?

A. We can approach this analysis with linear regression (all fixed effects) as all our assumptions appear to be valid.
B. We should make variety a fixed effect and greenhouse as a random effect.
C. We should make greenhouse a random effect and variety a fixed effect and consider including a greenhouse-variety interaction (also a random effect).
D. We should make everything a random effect.

Question 3

The below code aggreagtes the data to get the average profit by greenhouse and variety.

```
## # A tibble: 20 x 3
##    variety   greenhouse profit_avg
##    <chr>     <chr>           <dbl>
##  1 Checkmate central          23.8
##  2 Checkmate east             22.1
##  3 Checkmate north            29.5
##  4 Checkmate south            22.3
##  5 Checkmate west             18.1
##  6 Devotion  central          21.4
##  7 Devotion  east             22.2
##  8 Devotion  north            24.8
##  9 Devotion  south            22.4
## 10 Devotion  west             19.1
## 11 Hearts    central          26.8
## 12 Hearts    east             27.7
## 13 Hearts    north            35.3
## 14 Hearts    south            22.3
## 15 Hearts    west             26.3
## 16 Wanted    central          16.0
## 17 Wanted    east             15.7
## 18 Wanted    north            24.3
## 19 Wanted    south            15.9
## 20 Wanted    west             11.0
```

How many values are averaged to get each observation of `profit_avg` in `agg_int`?

A. 4
B. 5
C. 8
D. 32

Question 4

Suppose an appropriate model for this data is:

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + \epsilon_{ijk}$$

where:

- μ is the grand mean profit of strawberries, in thousands of dollars, across all the data.
- α_i is the fixed effect of variety.
- b_j is the random effect of greenhouse, $N(0, \sigma_b^2)$.
- $(\alpha b)_{ij}$ is in interaction between variety and greenhouse, $N(0, \sigma_{\alpha b}^2)$.
- ϵ_{ijk} is the error term, $N(0, \sigma^2)$.

Which ONE of the following statements is FALSE? (Notice, FALSE *not* TRUE!)

A. α_i and $(\alpha b)_{ij}$ are the only fixed effects in this model.
B. μ is the same for all values of i, j and k.
C. There are three random quantities being estimated.
D. $b_j \sim N(0, \sigma_b^2)$ in this model.

Question 5

This question takes you theough the 'teaching and learning' world version of thinking about sources of variance with random effects.

Which of the following correctly describes what each of the values being calculated under A, B and C below is matches with in the model specification? (Scroll!)

A. A is $\hat{\sigma}^2$, B is $\hat{\sigma}_b^2$, C is $\hat{\sigma}_{\alpha b}^2$.
B. A is $\hat{\sigma}_b^2$, B is $\hat{\sigma}^2$, C is $\hat{\sigma}_{\alpha b}^2$.
C. A is $\hat{\sigma}_{\alpha b}^2$, B is $\hat{\sigma}_b^2$, C is $\hat{\sigma}^2$.
D. A is $\hat{\sigma}_{\alpha b}^2$, B is not actually one of the variances for our random effects, C is $\hat{\sigma}^2$.

Question 6

Suppose our data was presented as follows. We are still interested in making the same plot and doing the same models as before.

observation	Checkmate	Devotion	Hearts	Wanted
north, plot group 1	29.49	23.44	32.77	24.64
north, plot group 2	27.60	20.42	36.41	27.38

observation	Checkmate	Devotion	Hearts	Wanted
north, plot group 3	26.67	23.86	34.05	25.21
north, plot group 4	30.53	28.74	35.06	27.29
north, plot group 5	28.60	29.10	38.84	21.98
north, plot group 6	30.01	24.73	35.12	23.56
north, plot group 7	33.30	23.86	33.98	19.85
north, plot group 8	29.52	24.11	35.94	24.88
east, plot group 1	20.67	26.56	32.15	14.46
east, plot group 2	25.45	22.04	26.92	20.31
east, plot group 3	20.08	20.43	28.88	15.07
east, plot group 4	23.47	20.78	27.10	14.06
east, plot group 5	18.79	20.16	28.30	16.04
east, plot group 6	23.34	23.60	25.80	13.80
east, plot group 7	21.17	23.59	25.60	17.74
east, plot group 8	23.82	20.48	26.93	13.91
south, plot group 1	25.01	25.27	22.00	16.59
south, plot group 2	20.34	21.86	24.00	15.70
south, plot group 3	20.33	20.25	19.61	19.11
south, plot group 4	21.08	22.72	25.24	15.12
south, plot group 5	20.58	18.89	24.08	13.81
south, plot group 6	21.11	21.63	20.57	15.47
south, plot group 7	24.35	25.45	23.21	16.38
south, plot group 8	25.81	23.30	19.32	14.93
west, plot group 1	15.34	18.76	28.10	11.77
west, plot group 2	17.62	20.19	29.68	9.60
west, plot group 3	20.22	19.09	24.92	11.26
west, plot group 4	16.87	20.10	22.88	8.88
west, plot group 5	15.04	19.14	27.65	10.18
west, plot group 6	21.45	19.56	24.28	11.33
west, plot group 7	22.27	16.08	25.32	15.54
west, plot group 8	15.75	19.88	27.74	9.70
central, plot group 1	23.03	22.76	25.75	15.02
central, plot group 2	27.94	22.54	28.18	16.20
central, plot group 3	22.52	23.61	26.80	15.21
central, plot group 4	23.64	17.95	28.30	17.06
central, plot group 5	26.36	22.99	29.59	16.72
central, plot group 6	21.95	18.82	23.02	15.56
central, plot group 7	22.46	22.55	26.93	13.51
central, plot group 8	22.68	19.67	26.09	18.35

Which ONE of the following statements best describes our situation with this new version of the dataset?

A. This dataset is also tidy for our purposes.
B. This dataset is NOT tidy for our purposes because each observational unit has too many observations.
C. This dataset is NOT tidy for our purposes because it is not the case that each value is in one and only one cell.
D. This dataset is NOT tidy for our purposes because it is not the case that each variable is in one and only one column.