

BIOSCI738 Assignment 2

2022-03-31

Question 1 [15 Marks]

Experimental design:

Experiment	Potato Type	Treatment Type	Cover Type
1	Store-bought potatoes that had gone green	Wrist-deep symmetric planting method	Mulched
2	Nursery Seed potatoes (x2 varieties)	Buried in furrows	Backfilling
3	Nursery Seed potatoes (x2 varieties)	Wrist-deep symmetric planting method	Mulched

Results:

- 1: Visibly describes many surface potatoes; states the potato plants did not look healthy in their growth stage; numerous small potatoes and a couple of 'decent sized' potatoes
- 2: Sprouted earlier; 'eyeball-method' looks like a "decent haul"
- 3: Second-healthiest looking potato plants while growing; 'eyeball-method' says the potatoes are better quality and higher yield

He claims using the 'eyeball method' that the piles don't look quite as different as he thought. Treatment group 1 certainly wasn't of the same quality or yield but it was not significantly different. He claimed that treatment group 2 was likely the best method based on output. Not all that much visible difference between the planting methods of the two seed potatoes

Discussion of the experimental design

His methods seem reasonable. As the plots are in the same locations the weather and rainfall variables are considered. In saying that, it would have been more experimentally-thorough if he had included another treatment group where the old potatoes were planted using the furrow method to ensure this is not where the variation lies, rather than in the type of potato.

Critiques:

As well as the above criticism regarding the variable not accounted for, he also could have been more thorough in quantifying his results. I think weighing the potatoes would have been a really simple but effective way to quantify a result, ie mass of potatoes produced per treatment method. He also could have used average diameter of potato or number of potatoes/m².

Question 2 [35 Marks]

a) What is the experimental and the observational unit in this experiment?

The experimental unit is the plot. The observational unit is each quadrat (corner) of the plot, with 4 total observational units/experimental unit.

b) Treatment factor and levels

The treatment factor in this case is the method of control for heather growth. The levels of this treatment factor are: control, biocontrol, herbicide and herbicide + biocontrol. There are 6 replications of each treatment level across the experiment, with a total number of 4 observations per replication therefore 24 observations per level. The treatment levels are randomized across the 6 blocks to avoid any biased effect of placement. They have also controlled for the sprays applied by applying a water spray to the treatment level without any herbicide or insecticide.



Interpreting the violin plots:

Violin plots are useful for the visualisation of the spread of data over a given time period or experiment. The length of the violin indicates the range of the data within the data set, while the width shows the frequency at any given data point. In the above plot, you can see that the number of natives most likely differs between the treatment levels. The two treatment levels containing herbicide show lower total natives across a number of the years. From this data you could tentatively say that the biocontrol-only treatment is the least likely to reduce total natives species richness.

c) Fitting a one-way analysis of variance model

```
lm(formula = natives ~ treatment, data = data)
```

Coefficients:

```
Estimates: (Intercept) 10.79
treatmentControl -0.88
treatmentHerbicide -0.58
treatmentHerbicide and Biocontrol -2.25
```

d) Pair-wise comparison of means

##	differences	LSD	tstatistic	SED
## Biocontrol-Control	0.87500	1.42923	1.2159	0.7196192
## Biocontrol-Herbicide	0.58333	1.42923	0.8106	0.7196192
## Control-Herbicide	-0.29167	1.42923	-0.4053	0.7196192
## Biocontrol-Herbicide and Biocontrol	2.25000	1.42923	3.1267	0.7196192
## Control-Herbicide and Biocontrol	1.37500	1.42923	1.9107	0.7196192
## Herbicide-Herbicide and Biocontrol	1.66667	1.42923	2.3160	0.7196192

##	Lower95CI	Upper95CI	pvalue
## Biocontrol-Control	-0.55423	2.30423	0.2271
## Biocontrol-Herbicide	-0.84590	2.01256	0.4197
## Control-Herbicide	-1.72090	1.13756	0.0024
## Biocontrol-Herbicide and Biocontrol	0.82077	3.67923	0.6862
## Control-Herbicide and Biocontrol	-0.05423	2.80423	0.0592
## Herbicide-Herbicide and Biocontrol	0.23744	3.09590	0.0228

The (intercept) of our regression line is essentially an estimation of the mean of our first treatment level (Biocontrol). The estimated coefficients for treatmentControl, treatmentHerbicide, treatmentHerbicide and Biocontrol, are all have negative coefficients. This suggests that compared to the Biocontrol treatment, natives species tend to decrease with all other treatment levels, including no treatment. We then pairwise compare the predicted means of the treatment groups to ascertain which treatment groups show significant evidence of difference between each other. The only groups that show enough evidence of a difference using this model at confidence $\alpha = 0.05$ is Control v Herbicide and Herbicide v Herbicide & Biocontrol

The predicted treatment means (using my predicted means function) are as follows,:

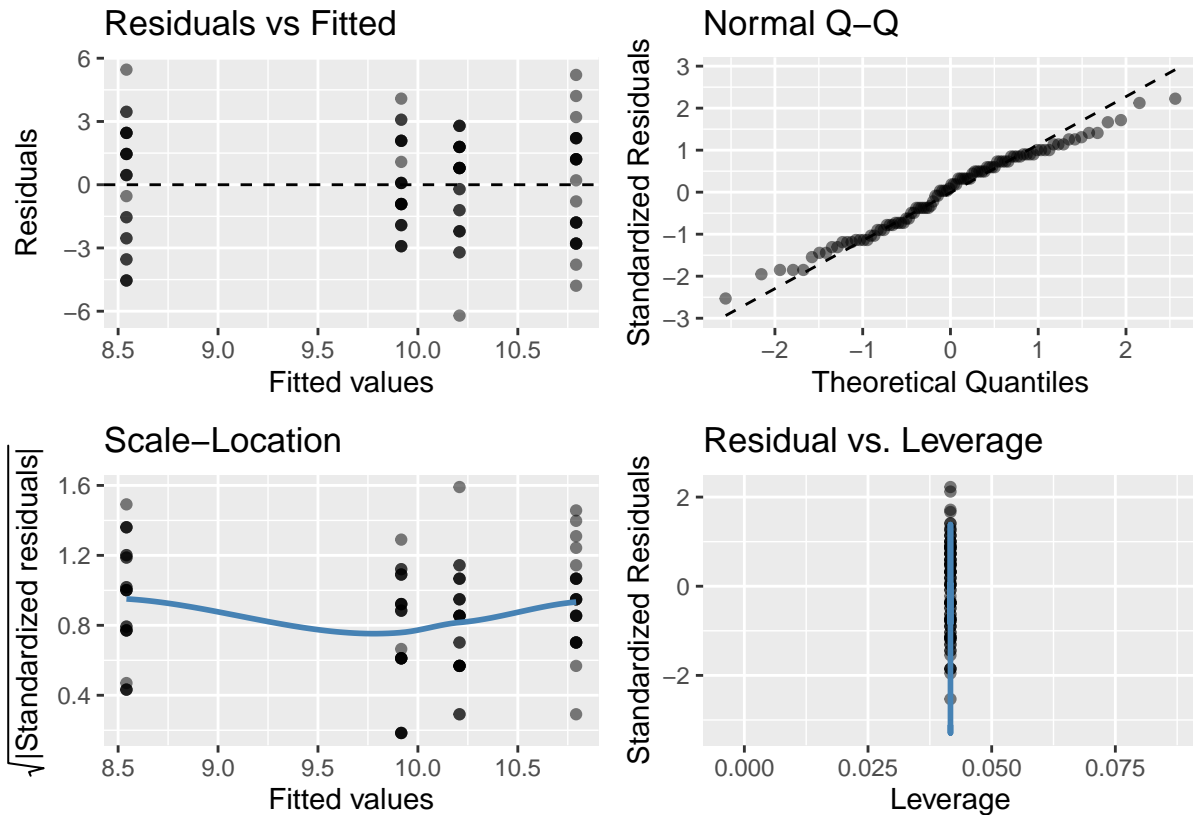
```
Treatment Predicted means Control 9.916667 Biocontrol 10.791667
Herbicide 10.208333
Herbicide & Biocontrol 8.541667
```

The above means show an average of the total native species richness across the 4 treatment groups, with the 4 years and all replications of the experiment combined into a single data point. From this very surface information, we may start to think that biocontrol treatments and herbicide treatments could increase the number of native species while herbicide & biocontrol look as though it reduces species richness.

e) Is this model appropriate for the data?

The residuals in the linear model show relative symmetry in distribution ie the residuals show a normal distribution, which can indicate the model is a good fit.

We can also test this using plots as follows, to check whether the underlying assumptions of the model are met:



Residuals v Fitted plot This plot is used to measure whether there is equal variance in the experiment. Grouping around the zero line is a sign that your model isn't fit for purpose. In the above residuals v fitted values plot, the data forms straight lines due to the discrete treatment variable. The data does not seem to form a distinguishable pattern and they are mostly equally distributed away from the zero line, except for a couple of points - indicating that the variance can mostly be explained by this model but there is some variance this model cannot explain/fit.

Q-Q plot Is used to diagnose the underlying assumption that our data fits a reasonably normal distribution. Our data mostly fits the line except for some deviation at the tail ends, but I think it mostly fits the normal distribution.

Scale-location plot Here we are looking at the average magnitude of the standardised residuals and whether they change because of the fitted values. We are looking for the line to be approximately horizontal. For this plot, the assumption of homoscedasticity is probably satisfied. We can also see there is no visible pattern, i.e. roughly equal variability.

Residuals v leverage plot Tracks how the spread of standardised residuals changes as the leverage changes. It can help detect outliers. It would seem that the leverage for all data is the same but I am unsure how to interpret due to the discrete nature of the treatment used.

Based on the above, I believe that the linear model is acceptable but that we likely can fit a better model. This is supported by the R-squared value I got when I ran the model, which was:

Multiple R-squared: 0.1, Adjusted R-squared: 0.074

You generally use R-squared as a measure of how well the model is fitting the data, and approximates how much of the variance in the response variable can be explained by the explanatory variable. An R-squared closer to 1 (than 0) is considered to explain more of the variance seen in the results and therefore the model is a good fit. My R-squared value indicates that maybe there is another variable that is acting on the results that should be combined.

f) Fit two full interaction models where year is treated as a factor

```
## Analysis of Variance Table
##
## Response: natives
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treatment    3  65.53  21.8438   3.4727 0.01942 *
## year          1   6.09   6.0860   0.9676 0.32799
## treatment:year 3  12.09   4.0312   0.6409 0.59072
## Residuals    88 553.53   6.2901
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table
##
## Response: natives
##           Df Sum Sq Mean Sq F value    Pr(>F)
## year          1   6.09   6.0860   0.9676 0.32799
## treatment      3  65.53  21.8437   3.4727 0.01942 *
## year:treatment 3  12.09   4.0313   0.6409 0.59072
## Residuals    88 553.53   6.2901
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For both models, the residual sum of squares is as follows:

553.53

Evidently, the residual sums of squares for both interaction models are the same. The residual sum of squares is a measure of how much of the response variable's variation is not explained by our interaction model. The above value is quite high which leads me to believe that these models do not fit the data well and do not explain variance adequately.

Below is the results of a type II ANOVA on the two-way model:

```
## Anova Table (Type II tests)
##
## Response: natives
##           Sum Sq Df F value    Pr(>F)
## treatment    65.53  3   3.4727 0.01942 *
## year          6.09  1   0.9676 0.32799
## treatment:year 12.09  3   0.6409 0.59072
## Residuals    553.53 88
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see that the above table is almost identical to the original anova's we ran. For Type II anovas, the order of the factors does not matter which is the opposite of the type I anova previously. Therefore, it makes sense that the values are the same as we have already established this model is not a good fit and that the interaction model doesn't work for the data we have.

g) Fit linear mixed-effect models to the data

In this experiment, the experimental units are split into blocks to control variation. Here we are aiming to partition potential sources of variation that are not useful to the hypothesis.

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```
##
## Correlation matrix not shown by default, as p = 16 > 12.
## Use print(x, correlation=TRUE) or
##     vcov(x)           if you need it
##
##
## Correlation matrix not shown by default, as p = 16 > 12.
## Use print(x, correlation=TRUE) or
##     vcov(x)           if you need it
```

The below is a pair-wise comparison of four linear mixed-effect models using the block as a random effect with different orders the variables **treatment** and **year**.

```
## refitting model(s) with ML (instead of REML)

## Data: data
## Models:
## lmer_one: natives ~ treatment + (1 | block)
## lmer_two: natives ~ treatment + year + (1 | block)
## lmer_three: natives ~ treatment:year + (1 | block)
## lmer_four: natives ~ treatment * year + (1 | block)
##           npar    AIC    BIC logLik deviance   Chisq Df Pr(>Chisq)
## lmer_one      6 416.13 431.52 -202.06   404.13
## lmer_two      9 416.79 439.87 -199.40   398.79  5.3359  3    0.1488
## lmer_three    18 423.86 470.02 -193.93   387.86 10.9296  9    0.2806
## lmer_four    18 423.86 470.02 -193.93   387.86  0.0000  0
```

You can see the error message ‘refitting model(s) with ML (instead of REML)’. This means that maximum likelihood fit better for the data than residual maximum likelihood. The above results show that models 1 and 2 are a better fit than models 3 & 4 through the AIC and BIC values for the latter being lower overall. Of models 1 & 2, model 1 has the lowest BIC, though an identical AIC. This leads me to believe that model 1 is the best model for the data. The deviance of each model follows the same principle: ie lower value = better fit. This is interesting as the deviance does not have the same result as the other two parameters, with model 3 & 4 having the lowest. Based on the above information and the high p-values of the other models, I believe model 1 is the best model.

- h) Use the predictmeans function to perform pairwise comparisons of mean counts between the treatment groups for model i.

```
##                               differences    LSD tstatistic    SED
## Biocontrol-Control           0.87500 1.06841    1.6278 0.537534
## Biocontrol-Herbicide         0.58333 1.06841    1.0852 0.537534
## Control-Herbicide            -0.29167 1.06841   -0.5426 0.537534
## Biocontrol-Herbicide and Biocontrol  2.25000 1.06841    4.1858 0.537534
## Control-Herbicide and Biocontrol    1.37500 1.06841    2.5580 0.537534
## Herbicide-Herbicide and Biocontrol   1.66667 1.06841    3.1006 0.537534
##                               Lower95CI Upper95CI pvalue
## Biocontrol-Control           -0.19341  1.94341 0.1072
## Biocontrol-Herbicide         -0.48508  1.65174 0.2808
## Control-Herbicide            -1.36008  0.77674 0.0001
```

```
## Biocontrol-Herbicide and Biocontrol    1.18159    3.31841 0.5888
## Control-Herbicide and Biocontrol        0.30659    2.44341 0.0123
## Herbicide-Herbicide and Biocontrol      0.59826    2.73508 0.0026
```

From the above output, I can draw conclusions regarding the data. The treatment levels in which evidence of a difference is strong enough to draw conclusions is the following: Comparison p-value
Control v Herbicide 0.0001 Control v Herbicide & Biocontrol 0.0123 Herbicide v Herbicide & Biocontrol 0.0026 The above p values show we have reasonable evidence to conclude that the treatments affect native species richness in different ways, ie that we can reject the null hypothesis that the treatments all have the same effect. Particularly, compared to the control, herbicide increases native species richness. Herbicide & Biocontrol treatment, however, reduces native species richness compared to both the control and the herbicide treatment.

- i) Perform pairwise comparison for HvH&B using Bonferroni's and Tukeys HSD and present them. Consider how the three CIs are calculated. In addition, briefly explain how Fisher's LSD, the Bonferroni correction and Tukey's HSD control the family-wise error rate.

Bonferroni's Adjustment:

```
## Comparison Calculated difference SED Bon LSDs
## 1 Herbicide - Herbicide and Biocontrol 1.66667 0.5894084 1.186418
## Lower 95% CI Upper 95% CI t-statistic P-value
## 1 0.4802522 2.853088 2.8277 0.0072
```

Tukey's HSD Adjustment:

HSD value as below:

```
## [1] 3.026046
```

Table of Tukey's adjustment with CI's:

```
## treatment
## Herbicide Herbicide and Biocontrol
## 10.2083 8.5417

## Comparisons differences upper lower HSD
## 1 Herbicide and Biocontrol - Herbicide -1.6666 1.359446 -4.692646 3.026046
```

Analysis of the Confidence Intervals:

The confidence interval is determined by the estimate value \pm the scale factor multiplied by the standard error of the estimate. Therefore, choosing our scale factor affects the confidence intervals of our estimate. Fisher's LSD (the default), Bonferroni's adjustment and Tukey's Honest Significant Difference are all scale factors we can use. Fisher's LSD uses the t-distribution (calculated as a function of the $\alpha/2$ and degrees of freedom) multiplied by the standard error to get the least significant difference, which is then used to calculate the intervals. Bonferroni's is similar to Fisher's except that it includes divides the alpha value by an additional argument, k, that divides the number of treatment groups by 2. Therefore, it adjusts for treatment levels and that number will affect the outcome. Tukey's HSD does not use the t-distribution, and instead uses a studentised range which uses the alpha value, number of treatment groups and degrees of freedom. This number is then divided by the square root of 2 and multiplied by the square root of the residual mean square error and the number of replicates in the experiment.

The family-wise error rate is the risk of making at least one Type I error among the family of comparisons and this number increases with the number of hypothesis tests carried out. Bonferroni's adjustment is a single-step procedure that controls the FWER by adjusting the alpha level based on the number of treatment levels ie alpha/number of tests. Tukey's method controls the family-wise error rate across all pair-wise comparisons at a specified level across all tests.

j) model all of the plant groups against treatments Herbicide and Biocontrol

P values of Herbicide v Biocontrol on the species richness of different plant groupings:

##	Plant	P.values	Adjusted.p.values
## 1	Exotic Dicots	0.0000	0.0000000
## 2	Total Exotics	0.0001	0.0004000
## 3	Native Dicots	0.0042	0.0112000
## 4	Total Native Monocots	0.0430	0.0860000
## 5	Total Bryophytes	0.1281	0.2049600
## 6	Total Natives	0.1781	0.2374667
## 7	Calluna vulgaris	0.3232	0.3693714
## 8	Total Exotic Monocots	0.4333	0.4333000

The number of "metabolites" (which I assume means the different groupings of plants above) treated with Herbicide or Biocontrol declared 'statistically significant' with unadjusted p-values is 4 groups: Exotic Dicots Total Exotics Native Dicots Total Native Monocots When we adjust for false discovery rate, Total Native Monocots no longer shows enough evidence of a significant difference between the treatments.