EMAT30007 Applied Statistics

Lab 9: Experimental Design and ANOVA

Nikolai Bode

Parts (1) and (2) of this lab cover the two types of ANOVA we looked at in lectures. In addition, we will make an explicit comparison between ANOVA and Linear models. Part (3) is a group exercise where you will have to address a statistical experimental design challenge.

(1) One-way ANOVA

We will analyse the data we have already seen in lecture 9. The data are from a study of the strength of structural beams in Hogg (1987). Reference:

[Hogg, R. V., and J. Ledolter. *Engineering Statistics*. New York: MacMillan, 1987]

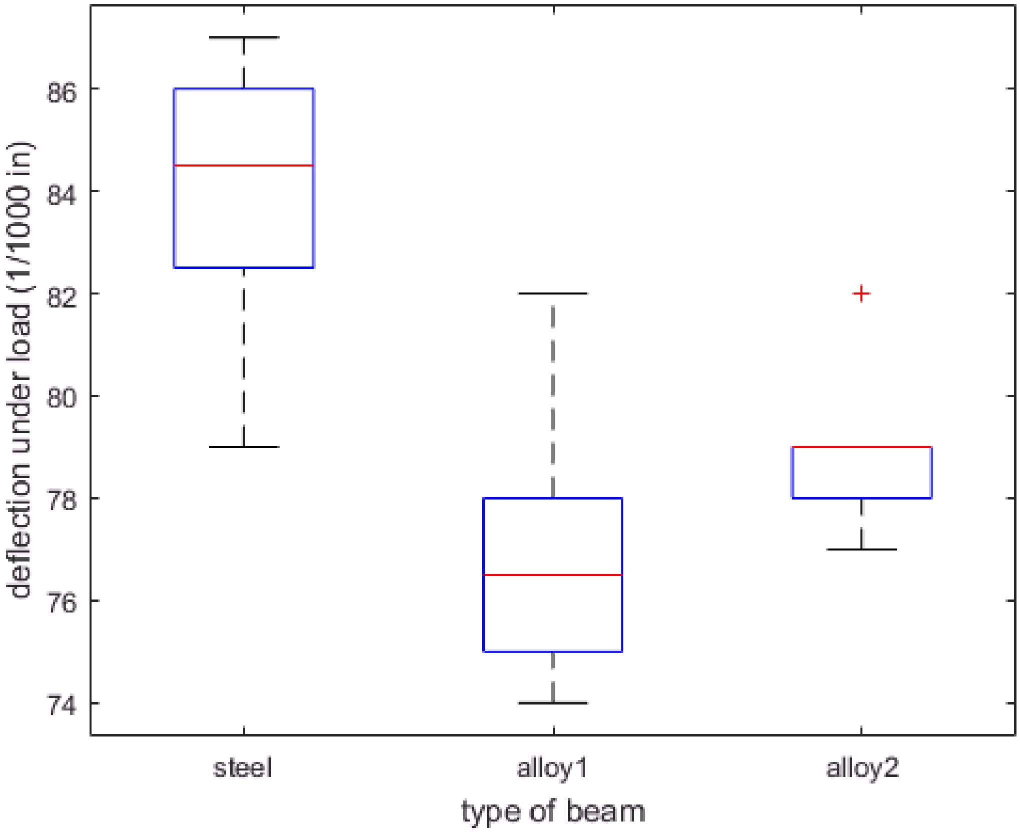
You can copy the data out of the document below, or find it online under the link shown below.

```
% one-way ANOVA example

% https://uk.mathworks.com/help/stats/anova1.html

strength = [82 86 79 83 84 85 86 87 74 82 ...
            78 75 76 77 79 79 77 78 82 79];
alloy = {'steel','steel','steel','steel','steel',...
         'steel','steel','steel','alloy1','alloy1',...
         'alloy1','alloy1','alloy1','alloy1',...
         'alloy2','alloy2','alloy2','alloy2','alloy2','alloy2'};
```

The vector `strength` measures deflections of beams in thousandths of an inch under 3000 pounds of force. The vector `alloy` identifies each beam as steel ('steel'), alloy 1 ('alloy1'), or alloy 2 ('allloy2'). Although alloy is sorted in this example, grouping variables do not need to be sorted.

Plot the mean strength for the three different alloys tested using the Matlab command 'boxplot'.

```
clf
boxplot(strength,alloy)
xlabel('type of beam') % x-axis label
ylabel('deflection under load (1/1000 in)') % y-axis label
```

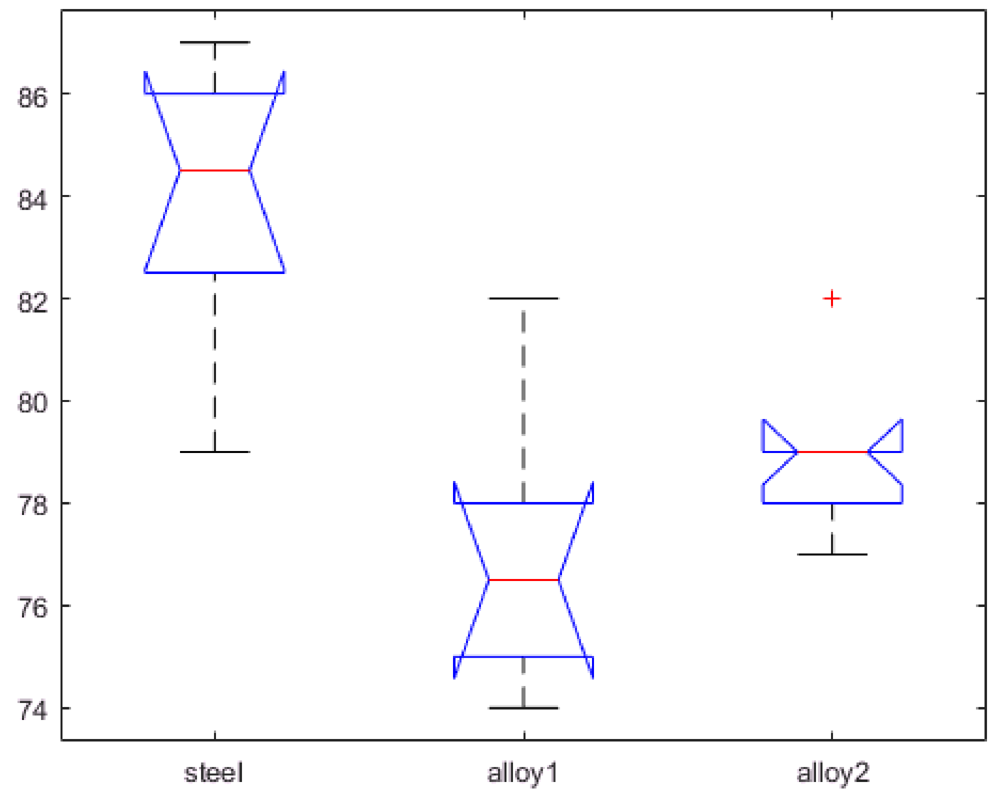


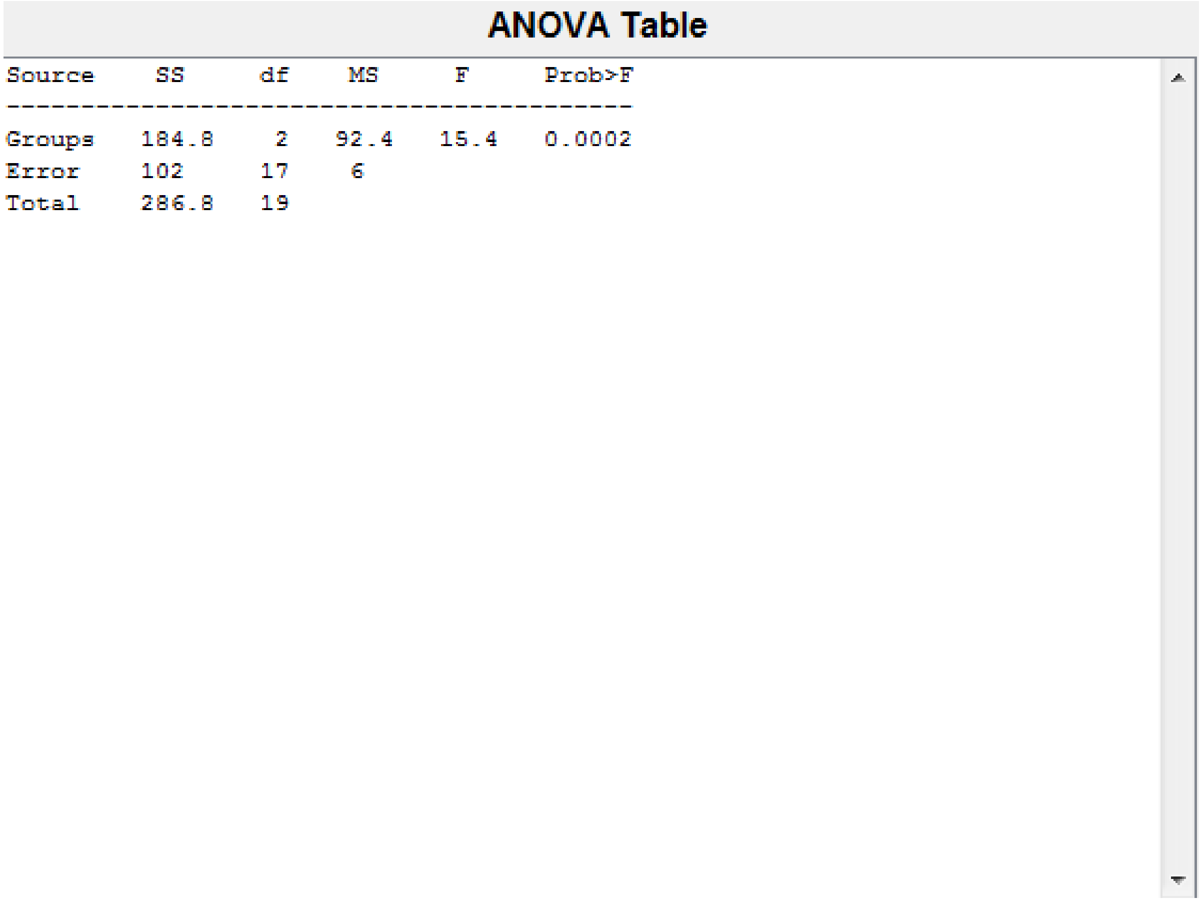
This plot should show that the deflection of alloys 1 and 2 seems lower than the deflection of steel beams.

1.1. anova1()

In this experiment, we only consider one factor ('alloy'). This means the experiment is equivalent to a completely randomised design and we can use a one-way ANOVA to analyse the data. Use the command `anova1` in Matlab to perform a one-way ANOVA on this data, making sure that the command shows an ANOVA table. Use the Matlab help page or the link above if you get stuck.

```
clf
anova1(strength,alloy);
```





ANOVA Table

```
Source     SS      df     MS       F      Prob>F
------------------------------------------------
Groups   184.8     2    92.4    15.4    0.0002
Error    102      17     6
Total    286.8    19
```

You should get the same ANOVA table as the one shown in lecture 9. Remember, this only shows that the mean strength of beams for at least two alloys differs.

1.2. Analysis using a Linear Model

In lecture 9 we saw that performing ANOVA is equivalent to analysing data using Linear Models. Fit the linear model corresponding to the analysis in 1.1. to the data. Remember that you will have to tell Matlab that the variable 'alloy' is qualitative. You may find it easier to create a vector with numbers indicating the alloy type.

```
% create a vector indicating alloy type:
alloytype = [0 0 0 0 0 0 0 0 1 1 1 1 1 1 2 2 2 2 2 2];
% create a table for the data:
data = table(alloytype',strength','VariableNames',{'alloy','strength'});
data.alloy = nominal(data.alloy);
% show first few rows of table
data(1:5,:)
```

```
ans =
    alloy    strength
    _____    ________

    0        82
    0        86
    0        79
    0        83
    0        84
```

```
% fit linear model with one qualitative predictor:
m1 = fitlm(data, 'strength~alloy')
```

```
m1 = 

Linear regression model:
    strength ~ 1 + alloy

Estimated Coefficients:
                   Estimate      SE        tStat       pValue  
                   ________    _______    _______    __________

    (Intercept)    84          0.86603     96.995    9.0805e-25
    alloy_1        -7           1.3229    -5.2915    5.9823e-05
    alloy_2        -5           1.3229    -3.7796     0.0014955


Number of observations: 20, Error degrees of freedom: 17
Root Mean Squared Error: 2.45
R-squared: 0.644,  Adjusted R-Squared 0.603
F-statistic vs. constant model: 15.4, p-value = 0.000153
```

Compare the F-test of this model to the test result shown in the ANOVA table above. They should be the same. Why?

The parameter-specific t-tests in the Linear Model summary table also give an indication of how alloy1 and alloy2 differ in the strength measure compared to the baseline (which is steel here).

(2) Two-way ANOVA

In this section we will use the popcorn data that is included in Matlab to perform a two-way ANOVA.

In Matlab, read in the data using the code below:

```
load popcorn
popcorn
```

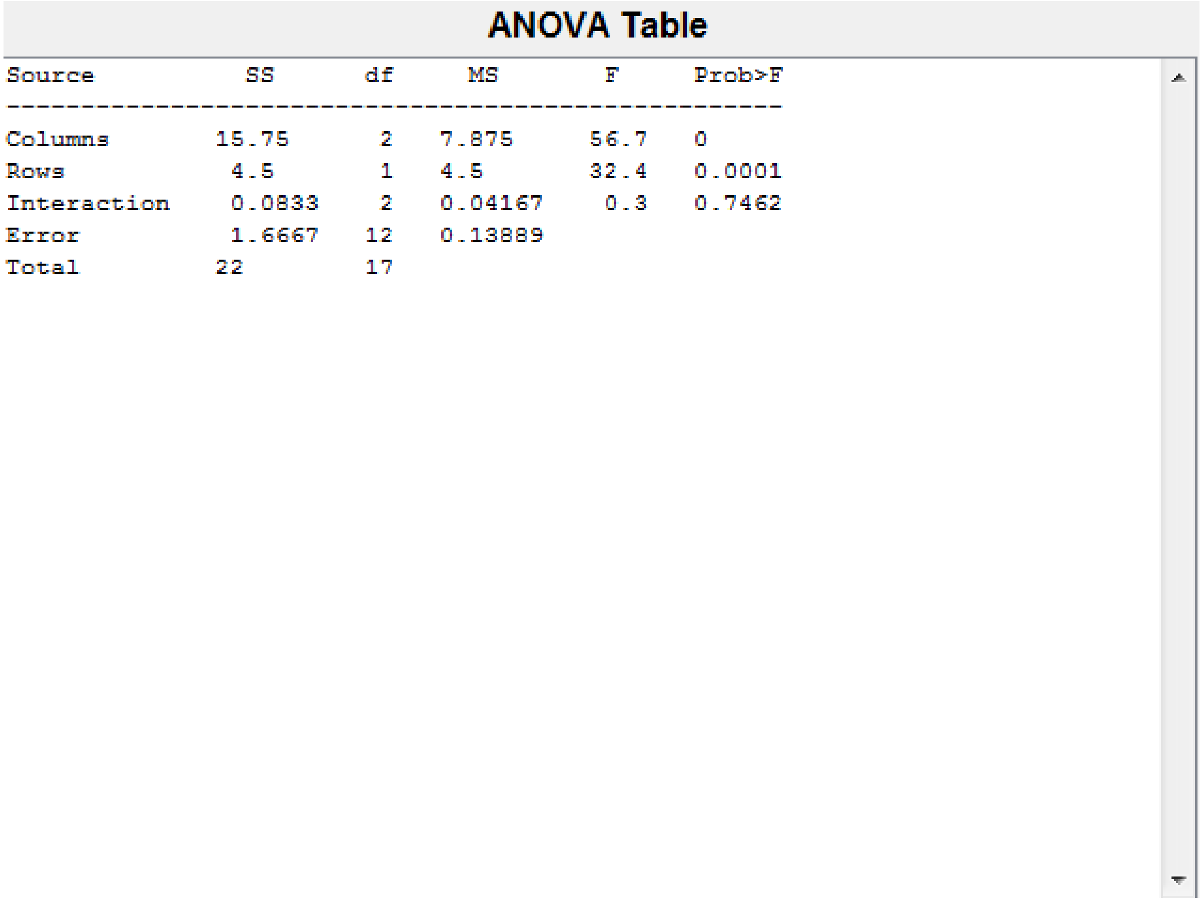
```
popcorn = 6x3 double

    5.5000    4.5000    3.5000
    5.5000    4.5000    4.0000
    6.0000    4.0000    3.0000
    6.5000    5.0000    4.0000
    7.0000    5.5000    5.0000
    7.0000    5.0000    4.5000
```

The Matlab help pages tell us that the data is from a study of popcorn brands and popper types (Hogg 1987). The three columns of the matrix `popcorn` are the brands, Gourmet, National, and Generic, respectively. The rows are popper types: oil and air. To get the data researchers popped a batch of each brand three times with each popper, that is, the number of replicates is 3. The first three rows correspond to the oil popper, and the last three rows correspond to the air popper. The response values are the yield in cups of popped popcorn.

2.1. anova2()

Use the Matlab command `anova2` to analyse the data. Have a look at the input arguments of this function - you need to let it know how many replicates are included in the rows of the data matrix `popcorn`.

```
clf
anova2(popcorn,3);
```



```
Source          SS      df     MS        F      Prob>F
-----------------------------------------------------
Columns        15.75     2   7.875     56.7    0
Rows            4.5      1   4.5       32.4    0.0001
Interaction     0.0833   2   0.04167    0.3    0.7462
Error           1.6667  12   0.13889
Total          22       17
```

You should obtain the same ANOVA table as shown in the lecture notes. Notice how anova2 tests for interactions between factors by default.

2.2. Analysis using a Linear Model

The lecture notes show what the linear model equivalent to a two-way ANOVA is. Fit this model to the popcorn data. You will have to re-format the data into separate predictor and response vectors.

```
% re-format data:
response = popcorn(:); % turns matrix into a vector (by column)
popper = [0 0 0 1 1 1 0 0 0 1 1 1 0 0 0 1 1 1];
% variable capturing popper type; 0-oil, 1-air.
brand = [repelem(0,6) repelem(1,6) repelem(2,6)];
% popcorn brand; 0-gourmet, 1-national, 2-generic.

% create table for data:
data = table(popper',brand',response,'VariableNames',...
    {'popper_type','brand','popcorn_yield'});
data.popper_type = nominal(data.popper_type);
data.brand = nominal(data.brand);
data(1:5,:)
```

```
ans =
    popper_type    brand    popcorn_yield
    ___________    _____    _____________

    0              0        5.5
    0              0        5.5
    0              0          6
    1              0        6.5
    1              0          7
```

```
% fit linear model with interaction between two
% qualitative predictors:
m1 = fitlm(data, 'popcorn_yield~popper_type*brand')
```

```
m1 = 

Linear regression model:
    popcorn_yield ~ 1 + popper_type*brand

Estimated Coefficients:
                             Estimate      SE        tStat       pValue   
                             ________    _______    _______    __________

    (Intercept)                5.6667    0.21517     26.336    5.4984e-12
    popper_type_1              1.1667    0.30429     3.8341      0.002378
    brand_1                   -1.3333    0.30429    -4.3818    0.00089351
    brand_2                   -2.1667    0.30429    -7.1204    1.2129e-05
    popper_type_1:brand_1    -0.33333    0.43033    -0.7746       0.45357
    popper_type_1:brand_2    -0.16667    0.43033    -0.3873       0.70532


Number of observations: 18, Error degrees of freedom: 12
Root Mean Squared Error: 0.373
R-squared: 0.924,  Adjusted R-Squared 0.893
F-statistic vs. constant model: 29.3, p-value = 2.51e-06
```

Compare the F-test provided in this model fit to the tests in the ANOVA table produced by `anova2`. None of the tests are the same. Why?

(3) Group exercise

Ideally, statisticians are already involved when experiments are designed, but often they only get approached by other scientists to help with analysing experimental data that has already been collected. Sometimes it is possible to develop an elegant statistical analysis *post hoc*, but sometimes the data does not lend itself to statistical analysis and sometimes a statistician does not have the experience or expertise to deal with an already existing data set. It is an important skill to identify and admit when this is the case.

In groups, have a look at one of the following eamples of proposed experimental designs. Come up with answers to the following questions, if possible:

a) Can you use a statistical model we have covered already to assess the statistical design described? (write down a formula for the structure of the model)

b) If you could change the experimental design, what would you do and why? (assume you cannot change the number of data points)

c) What statistical model would you use for your new experimental design? (write down a formula for the structure of the model)

Example 1 - computer chips

A computer chip manufacturer wants to assess the efficacy of a new photolithographic printing process to form transistors and circuits on silicon wafers. The company owns 10 printing machines and each machine can be programmed to use the old or new process. Past experience shows that not all machines perform identically when new software is introduced. The company decides to test two machines. For each machine, the silicon wafer printing is first performed twice using the old process and then four times using the new process. The measure the company assesses is the number of viable chips a printed wafer produces (in hundreds of chips).

Answer:

(a) This is a complete factorial design. It can be assessed using a linear model of the form: $Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 m_{1,i} + \beta_3 x_{1,i} m_{1,i} + \varepsilon_i$*, where* $x_{1,i}$ *is a dummy variable capturing the printing process and* $m_{1,i}$ *is a dummy variable capturing which machine is being used.*

(b) The experiment could be improved by testing more machines and by balancing the replicate tests across the two treatments.

(c) A similar model to the one shown in (a) could be used.

Example 2 - indoor drones

A magazine dedicated to electronic toys wants to investigate the responsiveness of small indoor drones of a particular make. The team have a total of 10 drones each from two different manufacturers available for testing - the products by both manufacturers are known to vary a lot in quality, even across items. In the responsiveness experiment, a test pilot has to wait for a remotely activated light signal and then steer the drone as quickly as possible from a starting hovering position across a finishing line. The experiment is filmed at a high frame rate and the time it takes for the drone to cross the finishing line after the light signal is activated is measured (in milliseconds). As all 20 staff of the magazine are keen to act as test pilots, they are each randomly assigned a drone to test.

Answer:

(a) This is a completely randomised design. It can be investigated using a one-way ANOVA of the form $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$*, where* x_i *is the dummy variable capturing which manufacturer produced the drone.*

(b) As the performance of test pilots may vary, it may be an idea to change the experiment to a randomised block design, where 10 test pilots perform the experiment with one drone from each manufacturer.

(c) The structure for the linear model for the experimental design suggested in (b) is $Y_i = \beta_0 + \beta_1 x_i + \beta_2 w_{1,i} + \beta_3 w_{2,i} + \beta_4 w_{3,i} + \beta_5 w_{4,i} + \beta_6 w_{5,i} + \beta_7 w_{6,i} + \beta_8 w_{7,i} + \beta_9 w_{8,i} + \beta_1 0 w_{9,i} + \varepsilon_i$*, where the dummy variables* $w_{j,i}$ *capture which test pilot is flying the drone (*x_i *is defined as before).*

Example 3 - health apps

A health agency wants to assess the accuracy of two apps that use image analysis to determine people's blood pressure. One hundred patients are included into the experiment. Half of them have high blood pressure and the other half low blood pressure. In each of these patient groups, half of the patients are randomly assigned to either app. The health agency workers measure the blood pressure of all patients using established and reliable approached and subsequently record a binary measure indicating whether high/low blood pressure was identified correctly.

Answer:

(a) The experiment has a complete factorial design. However, as the outcome is a binary measure, it is not appropriate to use a Linear Model to investigate this experiment. We could use a Generalised Linear Model, but we will only discuss those in lecture 10.

(b) Measure a continuous measure, e.g. difference in systolic blood pressure, instead of a binary outcome.

(c) If a continuous measure is available as a response, we can use a standard two-way ANOVA to analyse the data.

Example 4

A gas company wants to reduce the peak in demand on winter mornings (most peoples' boilers are programmed to start heating at 6am). The company has obtained permission to perform a trial with twelve comparable sets of ten households each. In the trial, the follwing factors are considered in a complete factorial design: price inducement (3 levels: low, medium, high), educational campaign (2 levels: present, not present). Households are randomly allocated to the treatments, ensuring that the same number of households per treatment. The company measures peak demand in cubic feet for each set of households.

Answer:

(a) As this is a complete factorial design, it can be analysed using a standard two-way ANOVA fitting a Linear Model of the form $Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 w_{1,i} + \beta_4 x_{1,i} w_{1,i} + \beta_5 x_{2,i} w_{1,i} + \varepsilon_i$.

(b) This is a good experimental design, although the number of replicates is quite low per treatment.

(c) Not necessary.