m3-autograded

February 13, 2022

0.1 Module 3 - Autograded

0.1.1 Outline:

In this homework assignment, here are learning objectives:

1. Motivate the use of two-way ANOVA through real data analysis examples.
2. Define a two-way ANOVA model with and without interaction terms.
3. Formulate three important research questions that can be answered using two-way ANOVA.
4. Interpret the two-way ANOVA model, with and without interaction terms.
5. Construct and interpret interaction plots to visually assess the importance of an interaction term.
6. Conduct hypothesis tests to decide whether a two-way ANOVA interaction term is statistically significant.
7. Use the two-way ANOVA and ANCOVA models to answer research questions using real data.

General tips:

1. Read the questions carefully to understand what is being asked.
2. When you feel that your work is completed, feel free to hit the `Validate` button to see your results on the visible unit tests.
3. Before submitting, we recommend restarting the kernel and running all the cells in order that they appear to make sure that there are no additional bugs in your code.

---

[1]:
```
# Run this cell to import the required packages for this assignment.
library(testthat)
library(tidyverse)
```

```
Error in get(genname, envir = envir) : object 'testthat_print' not found

  Attaching packages                                    tidyverse
1.3.0

 ggplot2 3.3.0      purrr   0.3.4
 tibble  3.0.1      dplyr   0.8.5
 tidyr   1.0.2      stringr 1.4.0
 readr   1.3.1      forcats 0.5.0

  Conflicts
```

```
tidyverse_conflicts()
  dplyr::filter()  masks stats::filter()
  purrr::is_null() masks
testthat::is_null()
  dplyr::lag()     masks stats::lag()
  dplyr::matches() masks
tidyr::matches(), testthat::matches()
```

0.1.2 Problem #1: Does stretching protocol impact vertical jump height? (26 Points)

The experiment used in this problem is a modified version of the one conducted in the 2013 paper "Acute Effects of Warm-Up Stretch Protocols on Balance, Vertical Jump Height and Range of Motion in Dancers"

(Citation: N. Morrin and E. Redding (2013). "Acute Effects of Warm-Up Stretch Protocols on Balnce, Vertical Jump Height and Range of Motion in Dancers," Journal of Dance Medicine & Science, Vol. 17, #1, pp. 34-40)

The research question being asked in this experiment is: Does stretching protocol impact vertical jump height? The variables measured include:

1. dancer = dancer (1-10)
2. stretch = stretch protocol (1=Static Stretching, 2=Dynamic Stretching, 3=Combination Stretching, 4=No Stretch)
3. condition = the physical condition of the dancer (1 = less conditioned, 2 = more conditioned)
4. height = vertical jump height (cm)

Dancers were first sorted into a condition group, based on how much dance training that they had recently done. Then, stretching routines were randomly assigned to dancers within each condition group. In this case, condition served as the blocking variable, and stretch as the treatment variable. Vertical jump height is the response.

Here's the data...

[2]:
```
dance = read.table("dancer_warmup.dat", sep = "")
names(dance) = c("stretch", "dancer", "height")
dance = dance %>%
    mutate(condition = ifelse(dancer == 1 | dancer == 2 | dancer == 3 | dancer␣
 ↪== 4 |
                              dancer == 5, 1,2))%>%
    select(dancer, stretch, condition, height) %>%
    mutate(stretch = as.factor(stretch), dancer = as.factor(dancer), condition␣
 ↪= as.factor(condition))

head(dance)
```

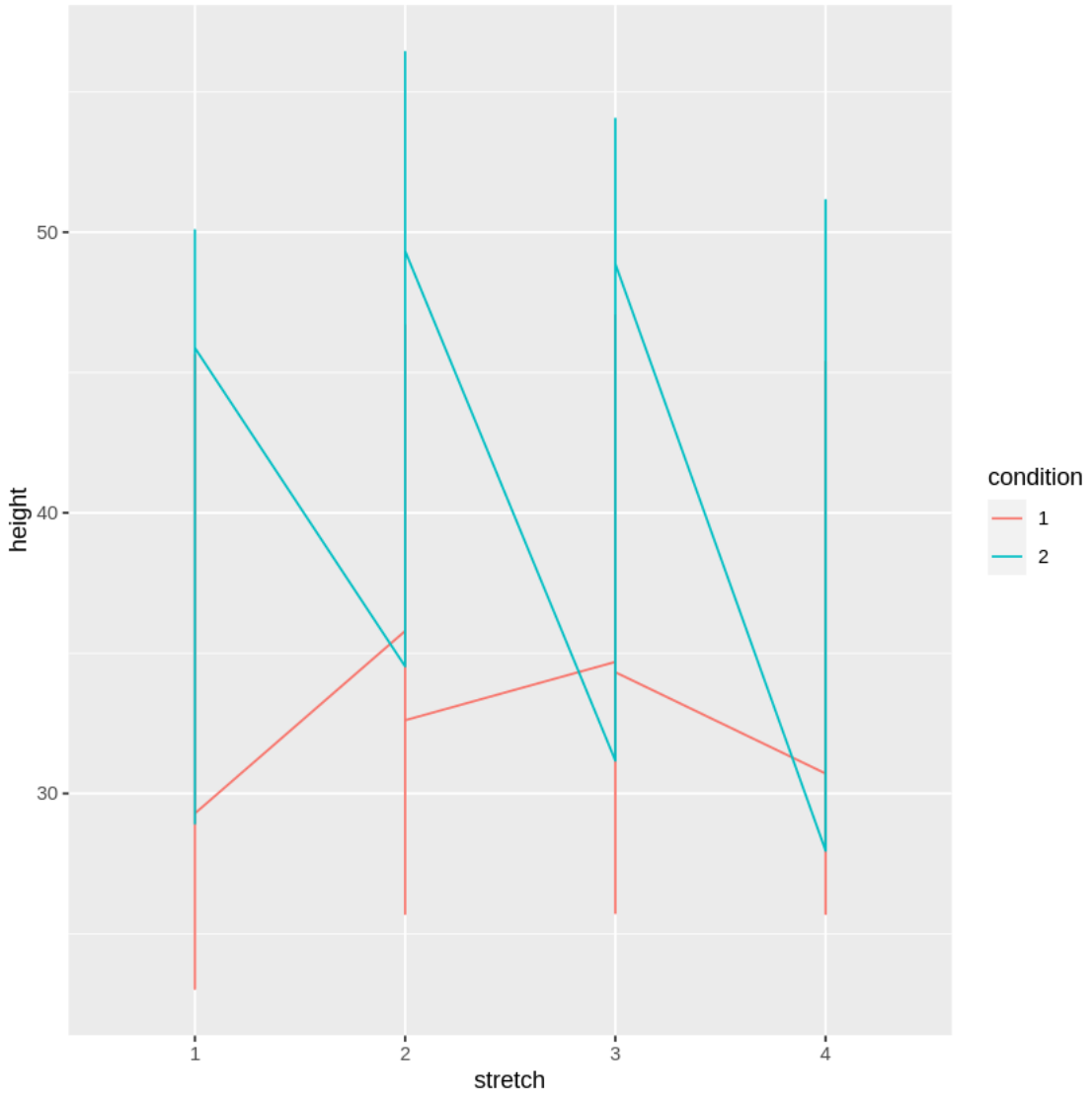
A data.frame: 6 × 4

	dancer <fct>	stretch <fct>	condition <fct>	height <dbl>
1	1	1	1	32.76
2	2	1	1	32.67
3	3	1	1	23.04
4	4	1	1	45.63
5	5	1	1	29.29
6	6	1	2	28.90

1. (a) Produce an interaction plot to detect whether there are interactions between `stretch` and `condition`. Specifically, store a `ggplot` object called `p1` looking at the relationship between `stretch` and `height` grouping on `condition`.

[7]:
```
p1 = NULL

# your code here
p1 = ggplot(data=dance, aes(x=stretch, y=height, group=condition))
p1= p1 + geom_line(aes(color=condition))
p1
```



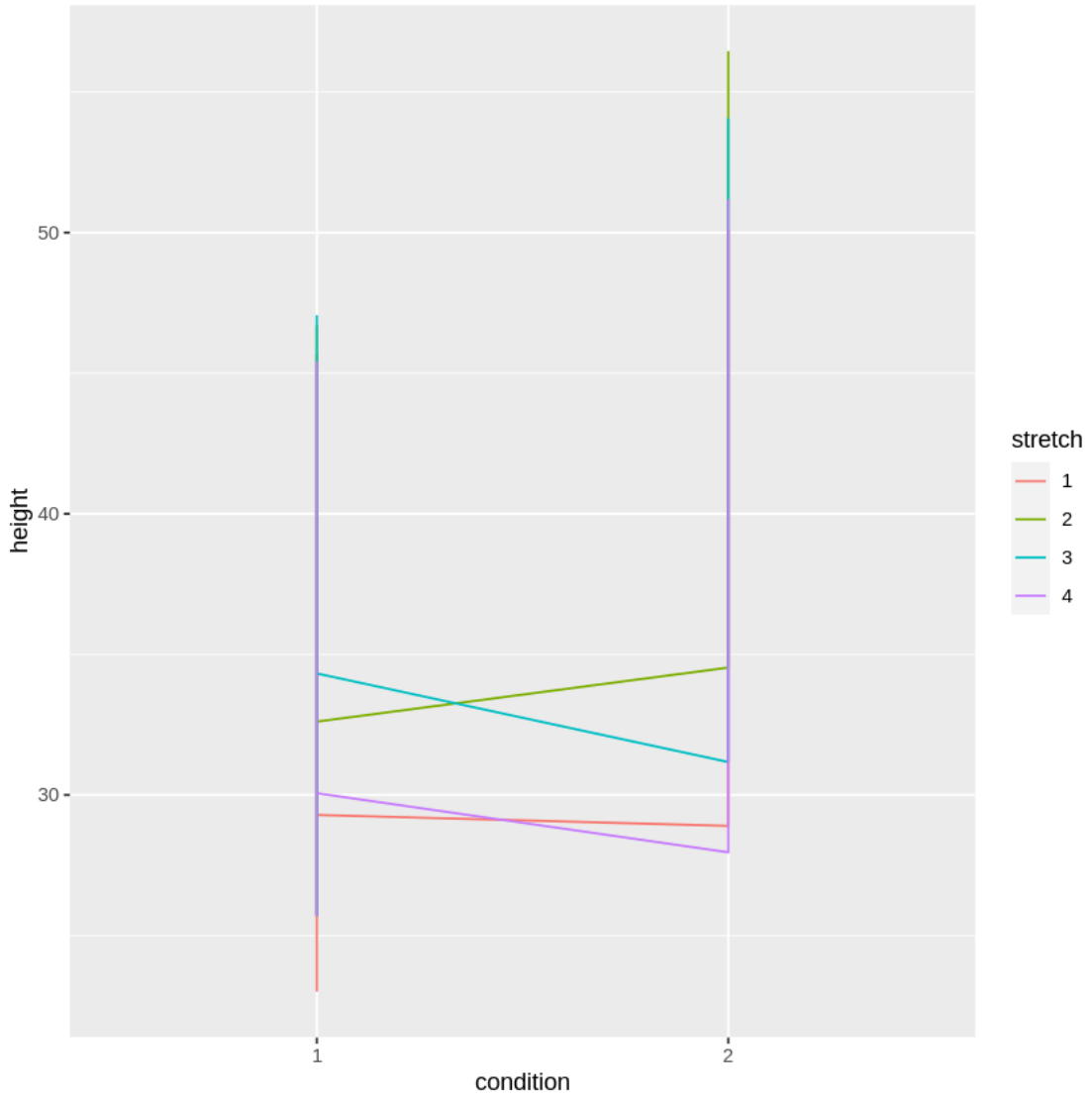
[ ]:
```
# Test Cell
# This cell has hidden test cases that will run after submission.
```

1. (b) Produce an interaction plot to detect whether there are interactions between `stretch` and `condition`. Specifically, store a `ggplot` object called `p2` looking at the relationship between `condition` and `height` grouping on `stretch`.

[8]:
```
p2 = NULL

# your code here
p2 = ggplot(data=dance, aes(x=condition, y=height, group=stretch))
p2= p2 + geom_line(aes(color=stretch))
```

```
p2
```



[9]:
```
# Test Cell
# This cell has hidden test cases that will run after submission.
```

1. (c) Create two linear models:

- A model, stored as `rcbd_int`, that uses `stretch`, `condition`, and an interaction term to predict jump height.
- A model, stored as `rcbd`, that uses `stretch`, `condition`, and ***NO*** interaction term to predict jump height.

[11]:
```
rcbd_int = NULL
rcbd = NULL

# your code here
rcbd_int = lm(height ~ stretch + condition + stretch:condition, data=dance)
rcbd = lm(height ~ stretch + condition, data=dance)
summary(rcbd_int)
print("No Interaction Model Below------------------")
summary(rcbd)
```

```
Call:
lm(formula = height ~ stretch + condition + stretch:condition,
    data = dance)

Residuals:
    Min      1Q  Median      3Q     Max
-15.124  -3.092   0.037   6.742  12.952

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)           32.678      3.790   8.622 7.48e-10 ***
stretch2               2.926      5.360   0.546   0.5889
stretch3               3.126      5.360   0.583   0.5638
stretch4               0.106      5.360   0.020   0.9843
condition2            10.664      5.360   1.990   0.0552 .
stretch2:condition2    0.988      7.580   0.130   0.8971
stretch3:condition2   -0.174      7.580  -0.023   0.9818
stretch4:condition2   -0.714      7.580  -0.094   0.9255
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.475 on 32 degrees of freedom
Multiple R-squared:  0.3541,Adjusted R-squared:  0.2128
F-statistic: 2.506 on 7 and 32 DF,  p-value: 0.03579



[1] "No Interaction Model Below------------------"

Call:
lm(formula = height ~ stretch + condition, data = dance)

Residuals:
     Min       1Q   Median       3Q      Max
-15.2235  -3.4005  -0.1455   6.4913  12.9655

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)   32.666      2.867  11.392 2.52e-13 ***
stretch2       3.420      3.627   0.943 0.352178
stretch3       3.039      3.627   0.838 0.407779
stretch4      -0.251      3.627  -0.069 0.945221
condition2    10.689      2.565   4.168 0.000192 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.11 on 35 degrees of freedom
Multiple R-squared:  0.353,Adjusted R-squared:  0.2791
F-statistic: 4.774 on 4 and 35 DF,  p-value: 0.00352
```

[ ]:
```r
# Test Cell
if(test_that('checking 2 models: ',{
expect_is(rcbd_int,"lm")
expect_is(rcbd,"lm")})){
    print('Two linear models: PASS')
    print('Reminder: make sure that you have the correct predictors/response!')
}
# This cell has hidden test cases that will run after submission.
```

1. (d) Run an F-test between `rcbd_int` and `rcbd` to determine if the interaction term is statistically significant. Store the p-value from the F-test in the object `rcbd_ftest_pvalue`.

[17]:
```r
rcbd_ftest_pvalue = NULL

# your code here
rcbd_ftest_pvalue = anova(rcbd, rcbd_int)[2, "Pr(>F)"]
rcbd_ftest_pvalue
```

0.996764942169362

[ ]:
```r
# Test Cell
# This cell has hidden test cases that will run after submission.
```

1. (e) Run a full F-test on the better model (determined by 1.(d)) in order to decide if there is a statistically significant difference in mean vertical jump height across stretching routines. Store the p-value of the `stretch` factor from the F-test in the object `stretch_ftest_pvalue`.

[28]:
```r
stretch_ftest_pvalue = NULL

# your code here
anova(rcbd)
stretch_ftest_pvalue = 0.6348088114
```

```
#p-value is 0.6348088114. Therfore, we cannot confidently say there is a␣
↪difference in means.
```

A anova: 3 × 5

	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
stretch	3	113.6011	37.86702	0.5757083	0.6348088114
condition	1	1142.5472	1142.54721	17.3706269	0.0001916087
Residuals	35	2302.1134	65.77467	NA	NA

[22]:
```
# Test Cell
# This cell has hidden test cases that will run after submission.
```

[ ]:
```
```