Data Analysis 3

Maksuda Aktar Toma, Jo Charbonneau, Ryan Lalicker

November 18, 2024

Introduction

In this paper we will be looking at data related to calves. The data comes from an experiment designed to study the impact dietary treatments given to pregnant heifers had on the development of the calves. The study was conducted over a three year period and involved three different dietary treatments given to select groups of heifers in the final trimester. In total the data has 22 variables for 120 entries, though some data points are missing.

For more information on the experiment, the data, or any other files used in this paper see our Github page which can be found at https://github.com/RyanLalicker/Data-Analysis-2-STAT-325-825. The coding languages used in the paper are R and SAS. The corresponding code can be found in *Appendix A - R Code* and *Appendix B - SAS Code* respectively.

Exploring the Data

Variables

As mentioned above the experiment used three different dietary treatments. These were DDG, CON, and MET. For the first two trimesters the heifers were given one of seven developmental treatments, found in `Development.Treatment`, and then in the final trimester the each was given one of the three treatments mentioned above. This is recorded in the `Calan.Treatment` column of the data set.

The heifers were placed into one of four pens by weight, which can be seen in the column `Pen #`. They were then artificially inseminated from an assigned sire, which we will assume was done randomly since the client says weight was not a factor. The sire is represented by the column of the same name and has six unique entries.

Upon the birth of the calves, several measurements were taken. These include the sex of the calf, weights taken at both birth and slaughter, and scores of both the calf's vigor and the ease of birth. The variable names line up with these descriptions.

Other variables, such as the id of the calf, length of gestation for the heifer, and postmortem scoring such as hot carcass weight (HCW) are included as well. (Saner (2024)). Note two birthdays are included in the data, `Birth.date` and `Birth.date.1`. These variables will not be used in the models below so no further investigation was done on our part to determine the differences.

The client's main focus is the effect the third trimester treatment and the sex of a calf have on the calf's vigor score, ease of birth score, and final body weight. Therefore, these are the variables we will place more of an emphasis on, while exploring the effect some of the other variables may have.

Missing Values

UPDATE THIS AFTER SEEING WHAT VARIABLES ARE NEEDED FOR THE MODEL

The data contains some missing values. In regards to the five variables the client is most interested in, 19 entries are missing one or more values. Since that still leaves 101 entries with all five variables

Initially the dataset has missing values in the following columns: Pen #: 4 missing values DMI: 36 missing values SEX: 1 missing value

The heatmap above visualizes the distribution of missing data in the dataset. Columns with missing values (e.g., "Pen #", "DMI", and "SEX") are marked with red line gaps.

```
# Might want this later.
# May need to add more later.
neededvars <- c("Calan.Treatment", "SEX", "Final.Calf.BW", "Calving.Ease", "Calf.Vigor")
data <- data[complete.cases(data[, neededvars]), ]
```

Figure 1: Chart counting the number of missing values for each variable within the data.

Cleaning the dataset

This code cleans a data set by replacing all occurrences of . with NA to standardize missing values. It ensures columns are assigned the correct data types, converting numeric-like columns to numeric and others to factor. Missing values are handled by imputing the median for numeric columns and the mode for factor columns. After cleaning, the code verifies that no missing values remain in the dataset.

```
[1] "Remaining missing values:  0"
```

Summary Statistics

```
[1] "Summary Statistics for Numerical Variables"
```

```
   ID_mean   ID_sd ID_min ID_max ID_median Birth.date_mean Birth.date_sd
1 744.2583 80.9808    600    898       744          22.925      14.97795
```

```
  Birth.date_min Birth.date_max Birth.date_median Development.Treatment_mean
1              1             50              22.5                    3.908333
  Development.Treatment_sd Development.Treatment_min Development.Treatment_max
1                 1.842522                         1                         7
  Development.Treatment_median Calan.Treatment_mean Calan.Treatment_sd
1                            3                 2.05          0.8285859
  Calan.Treatment_min Calan.Treatment_max Calan.Treatment_median Pen.._mean
1                   1                   3                      2   2.508333
  Pen.._sd Pen.._min Pen.._max Pen.._median Initial.BW_mean Initial.BW_sd
1 1.100006         1         4            3        50.79167      28.14843
  Initial.BW_min Initial.BW_max Initial.BW_median Final.BW_mean Final.BW_sd
1              1             99              49.5      48.16667    25.83475
  Final.BW_min Final.BW_max Final.BW_median DMI_mean   DMI_sd DMI_min DMI_max
1            1           94            49.5 40.66667 18.57681       1      75
  DMI_median ADG_mean   ADG_sd ADG_min ADG_max ADG_median Gest.Length_mean
1         46   51.675 27.73378       1      95         55         9.016667
  Gest.Length_sd Gest.Length_min Gest.Length_max Gest.Length_median
1       3.775158               1              21                  8
  Calved.First.21.Days_mean Calved.First.21.Days_sd Calved.First.21.Days_min
1                  1.966667               0.1802581                         1
  Calved.First.21.Days_max Calved.First.21.Days_median X1st.Calf.BW_mean
1                        2                           2          16.64167
  X1st.Calf.BW_sd X1st.Calf.BW_min X1st.Calf.BW_max X1st.Calf.BW_median
1          7.3035                1               35                  15
  Calving.Ease_mean Calving.Ease_sd Calving.Ease_min Calving.Ease_max
1             1.125       0.4213324                1                3
  Calving.Ease_median Calf.Vigor_mean Calf.Vigor_sd Calf.Vigor_min
1                   1        1.341667     0.8043561              1
  Calf.Vigor_max Calf.Vigor_median Birth.date.1_mean Birth.date.1_sd
1              5                 1            21.775        9.579795
  Birth.date.1_min Birth.date.1_max Birth.date.1_median Sire_mean  Sire_sd
1                1               43                  23  2.883333 1.768935
  Sire_min Sire_max Sire_median SEX_mean    SEX_sd SEX_min SEX_max SEX_median
1        1        6           3     1.55 0.4995797       1       2          2
  Final.Calf.BW_mean Final.Calf.BW_sd Final.Calf.BW_min Final.Calf.BW_max
1           42.68333         21.14217                  1                83
  Final.Calf.BW_median HCW_mean   HCW_sd HCW_min HCW_max HCW_median REA_mean
1                   44   44.225 21.38069       1      83         49 50.11667
   REA_sd REA_min REA_max REA_median  YG_mean    YG_sd YG_min YG_max YG_median
1 26.7777       1     100         45 32.93333 18.45036      1     70      29.5
  BF_mean    BF_sd BF_min BF_max BF_median
1    44.9 22.85556      1     88        44
```

```
[1] "Summary Statistics for Categorical Variables"
```

```
data frame with 0 columns and 1 row
```

Exploring the Data

Relationships among variables

Potential models

Model 1

```
                     Df Sum Sq Mean Sq F value   Pr(>F)
calan_treatment       2   1840     920   2.347 0.100267
sex                   1   4802    4802  12.254 0.000666 ***
initial_bw            1    800     800   2.041 0.155884
calan_treatment:sex   2   1472     736   1.879 0.157560
Residuals           113  44279     392
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = final_calf_bw ~ calan_treatment * sex + initial_bw, data = data_cleaned)

$calan_treatment
         diff        lwr        upr     p adj
2-1  5.210526  -5.575086 15.9961383 0.4871443
3-1 -4.287081 -14.698511  6.1243487 0.5922780
3-2 -9.497608 -19.909038  0.9138223 0.0814415

$sex
        diff      lwr      upr     p adj
2-1 12.63931 5.443081 19.83554 0.0007141
```

```
                     Df Sum Sq Mean Sq F value   Pr(>F)
calan_treatment       2   1840     920   2.347 0.100267
sex                   1   4802    4802  12.254 0.000666 ***
initial_bw            1    800     800   2.041 0.155884
```

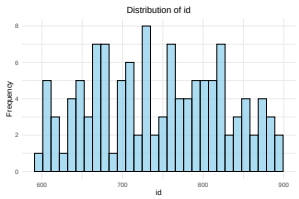

Figure 2

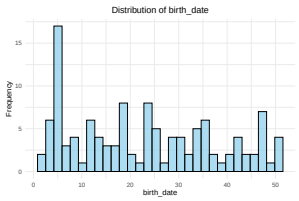

Figure 3

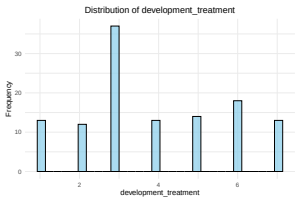

Figure 4

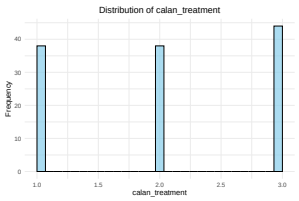

Figure 5

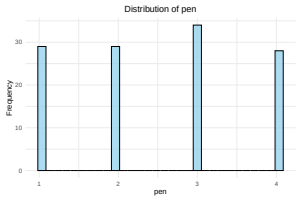

Figure 6

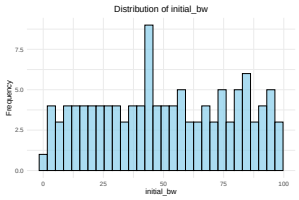

Figure 7

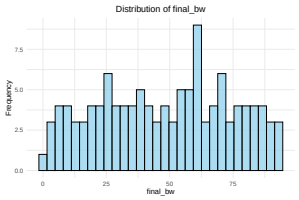

Figure 8

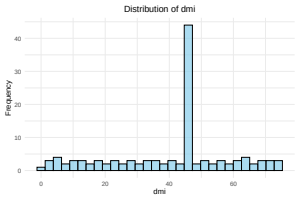

Figure 9

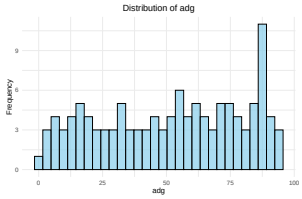

Figure 10

Figure 11

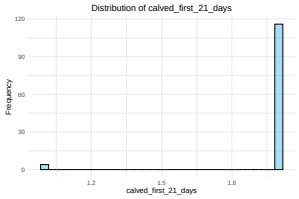

Figure 12

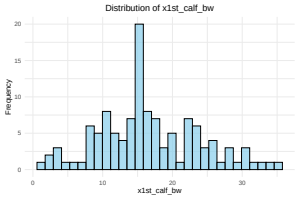

Figure 13

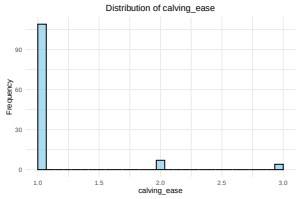

Figure 14

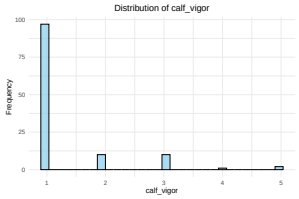

Figure 15

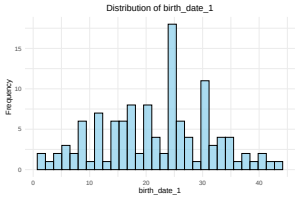

Figure 16

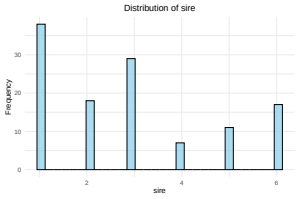

Figure 17

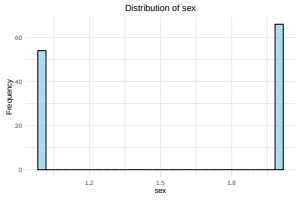

Figure 18

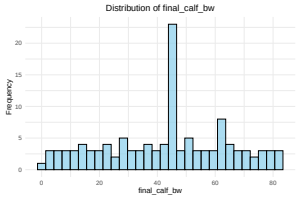

Figure 19

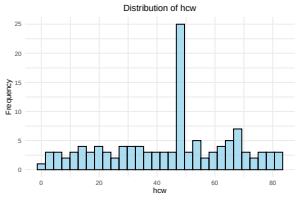

Figure 20

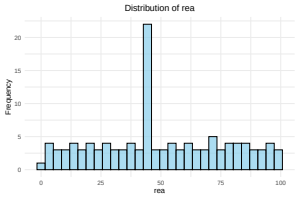

Figure 21

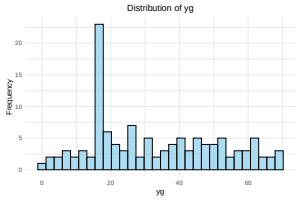

Figure 22

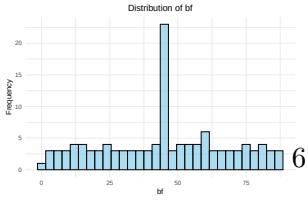

Figure 23

```
calan_treatment:sex   2   1472     736   1.879 0.157560
Residuals           113  44279     392
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$y_{ijklmn} = ENTER - MODEL - HERE$$

where y_{ijklm} represents the *dependent variable*, ...

```
![Picture of SAS Output](filename.png){width="3in"}
```

Conclusion

Recomendation

References

Saner, Brianna, Randy & Buseman. 2024. "How Many Pounds of Meat Can We Expect from a Beef Animal?" 2024. https://beef.unl.edu/beefwatch/2020/how-many-pounds-meat-can-we-expect-beef-animal.

Appendix A - R Code

Appendix B - SAS Code

Appendix C - Additional SAS Output

```
![](filename.png)
```