My Final Paper

Captain Jack[1]

[1] University of Basel

Author Note

R is awesome

Correspondence concerning this article should be addressed to Captain Jack, Missionsstrasse 62A, Basel, Switzerland. E-mail: jack@rpirates.com

Abstract

Does the growth rate of chickens depend on their diet?

Keywords: keywords

Word count: X

My Final Paper

Dataset Description

I obtained the data from the international chicken data repository. I don't know how the data were originally collected. The main columns in the data were as follows: *weight*, the weight of chickens in grams, *Time* the age in weeks of the chick at the time of measurement, *Chick* a unique number for each chicken, and *Diet* the diet given to the chick.

Questions

I will answer the following questions in my paper.

1. How did the chicken weights generally change over time?
2. Was there a difference in the the average chicken weights as a result of the different diets?
3. Were the chicken weights at time 1 normally distributed?
4. Was there a difference in weights between time 2 and time 4?
5. Did more chickens die in one diet than another?

Analyses

```
# Task 1: Load data
# INSERT CODE HERE
```

Summary statistics from the data are presented in the following table.

```
# Task 2: Summary statistics from a dataframe
# INSERT CODE HERE
```

```
##      weight            Time           Chick            Diet
##  Min.   : 35.0   Min.   : 0.00   Min.   : 1.00   Min.   :1.000
```

```
##  1st Qu.: 63.0   1st Qu.: 4.00   1st Qu.:13.00   1st Qu.:1.000
##  Median :103.0   Median :10.00   Median :26.00   Median :2.000
##  Mean   :121.8   Mean   :10.72   Mean   :25.75   Mean   :2.235
##  3rd Qu.:163.8   3rd Qu.:16.00   3rd Qu.:38.00   3rd Qu.:3.000
##  Max.   :373.0   Max.   :21.00   Max.   :50.00   Max.   :4.000
```

The data had 4 columns: weight, Time, Chick, and Diet:

```
# Task 3: Printing variable names
# INSERT CODE HERE
```

```
## [1] "weight" "Time"   "Chick"  "Diet"
```

The data for Diet were originally coded as numbers, I recoded the Diet data as string variables

```
# Task 4: Recoding a variable
# INSERT CODE HERE
```

```
# Task 5: Calculate simple summary statistics
# INSERT CODE HERE
```

The mean weight of chickens across all data was 121.82, the median weight was 103 and the standard deviation was 71.07.

A table of frequencies showing how many observations there were for each diet is displayed in the following table:

```
# Task 6: Print a table
# INSERT CODE HERE
```

```
# Task 7: Count outliers
# INSERT CODE HERE
```

To see if there were any outliers in the weight data, I counted how many chicks had weights greater than 3 standard deviations above the mean, or less than 3 standard deviations below the mean. Using this procedure, I counted 3 outliers.

A scatterplot showing the relationship between time and weight is shown in the following figure

```
# Task 8: Scatterplot with regression line.
# INSERT CODE HERE
```

A histogram of the weight data are presented in the next figure

```
# Task 9: Histogram
# INSERT CODE HERE



# Task 20: Custom Function: my.hist()
# Insert code here
```

Histograms separately for each diet are displayed in the next figure

```
# Task 20: Loop
# INSERT CODE HERE
```

A pirateplot showing the relationship between diet and weight is shown here:

```
# Task 10: pirateplot
# INSERT CODE HERE
```

The mean weight of chicks on each diet is shown in the following table:

```
# Task 11: Descriptive statistics across groups
# INSERT CODE HERE
```

```
# Task 12: 1 sample t-test
# INSERT CODE HERE
```

```
##
##  One Sample t-test
##
## data:  ChickWeight$weight
## t = 7.3805, df = 577, p-value = 5.529e-13
## alternative hypothesis: true mean is not equal to 100
## 95 percent confidence interval:
##  116.0121 127.6246
## sample estimates:
## mean of x
##  121.8183
```

A one sample t-test comparing the weights of chickens to a null hypothesis of 100 was significant $M = 121.82$, 95% CI [116.01, 127.62], $t(577) = 7.38$, $p < .001$. The mean weight of chickens was significantly larger than 100 grams.

```
# Task 13: t-test with subset
# INSERT CODE HERE
```

```
##
##  Welch Two Sample t-test
##
## data:  weight by Diet
```

```
## t = -2.6378, df = 201.38, p-value = 0.008995
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -34.899942  -5.042482
## sample estimates:
## mean in group 1 mean in group 2
##        102.6455        122.6167
```

A two sample t-test comparing the weights of chickens between diets 1 and 2 was significant $\Delta M = 19.97$, 95% CI $[-34.90, -5.04]$, $t(201.38) = -2.64$, $p = .009$, the weights of chickens was significantly higher in diet 2 compared to diet 1.

```
# Task 14: correlation test
# INSERT CODE HERE
```

```
##
##  Pearson's product-moment correlation
##
## data:  Time and weight
## t = 36.725, df = 576, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8109073 0.8599481
## sample estimates:
##       cor
## 0.8371017
```

A correlation test detecting a relationship between time and weight was significant $r = .84$, 95% CI $[.81, .86]$, $t(576) = 36.73$, $p < .001$, as time increased, the weight of chickens increased.

```
# Task 15: correlation test with subset
# INSERT CODE HERE
```

```
##
##  Pearson's product-moment correlation
##
## data:  Time and weight
## t = 15.449, df = 118, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7485471 0.8697470
## sample estimates:
##       cor
## 0.8180325
```

A correlation test detecting a relationship between time and weight only for chickens on diet 2 was significant $r = .82$, 95% CI [.75, .87], $t(118) = 15.45$, $p < .001$, as time increased, the weight of chickens on diet 2 increased.

```
# Task 16: Chi-Square test
# INSERT CODE HERE
```

```
##
##  Chi-squared test for given probabilities
##
## data:  table(ChickWeight$Diet)
## X-squared = 52.616, df = 3, p-value = 2.214e-11
```

To see if there was a significant difference in the number of chickes on each diet. I performed a chi-square test. The test was significant $\chi^2(3, n = 578) = 52.62$, $p < .001$, indicating that chickens were not equally distributed amongst the diets.

```
# Task 17: ONE-WAY ANOVA
# INSERT CODE HERE
```

```
## Call:
##    aov(formula = weight ~ factor(Diet), data = ChickWeight)
##
## Terms:
##                 factor(Diet) Residuals
## Sum of Squares      155862.7 2758693.3
## Deg. of Freedom            3       574
##
## Residual standard error: 69.32594
## Estimated effects may be unbalanced
```

```
##               Df  Sum Sq Mean Sq F value   Pr(>F)
## factor(Diet)   3  155863   51954   10.81 6.43e-07 ***
## Residuals    574 2758693    4806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To see if there was a significant difference in weights between diets, I performed a one-way ANOVA. The test was significant , indicating that there was a significant difference between diets. Post-hoc tests showed significant differences between Diets 1 and 3 (diff = 40.30, p < .01), and Diets 1 and 4 (diff = 32.62, p < .01).

```
# Task 18: TWO-WAY ANOVA
# INSERT CODE HERE
```

```
## Call:
```

```
##    aov(formula = weight ~ factor(Diet) + factor(Time), data = ChickWeight)
##
## Terms:
##                 factor(Diet) factor(Time) Residuals
## Sum of Squares      155862.7    2040908.0  717785.2
## Deg. of Freedom            3           11       563
##
## Residual standard error: 35.70615
## Estimated effects may be unbalanced
```

```
##               Df  Sum Sq Mean Sq F value Pr(>F)
## factor(Diet)   3  155863   51954   40.75 <2e-16 ***
## factor(Time)  11 2040908  185537  145.53 <2e-16 ***
## Residuals    563  717785    1275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To see if there was a significant difference in weights between diets and time points, I performed a two-way ANOVA. The effect of both diet $F(3, 563) = 40.75$, $\mathrm{MSE} = 1,274.93$, $p < .001$, $\eta^2_G = .178$ and time $F(11, 563) = 145.53$, $\mathrm{MSE} = 1,274.93$, $p < .001$, $\eta^2_G = .740$ were significant. Post-hoc tests showed significant differences between all Diets except for 4 and 3 (diff = -7.69, p = 0.34). There were significant differences between almost all pairs of time periods.

```r
# Task 19: REGRESSION
# INSERT CODE HERE
```

```
##
## Call:
```

```
## lm(formula = weight ~ Time, data = ChickWeight)
##
## Coefficients:
## (Intercept)         Time
##      27.467        8.803
```

```
##
## Call:
## lm(formula = weight ~ Time, data = ChickWeight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -138.331  -14.536    0.926   13.533  160.669
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27.4674     3.0365   9.046   <2e-16 ***
## Time          8.8030     0.2397  36.725   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 38.91 on 576 degrees of freedom
## Multiple R-squared:  0.7007, Adjusted R-squared:  0.7002
## F-statistic:  1349 on 1 and 576 DF,  p-value: < 2.2e-16
```

To see if time was related to weight, I regressed weight on time. Results showed a significant positive effect of time $b = 8.80$, 95% CI [8.33, 9.27], $t(576) = 36.73$, $p < .001$, $R^2 = .70$, $F(1, 576) = 1,348.74$, $p < .001$

Conclusion

The two most important results were that chickens gain weight over time, and Diet 3 lead to the highest weights while Diet 1 lead to the lowest weights.

References

Table 1

Number of chicks on each diet

Diet	Frequency
1	220.00
2	120.00
3	120.00
4	118.00

Table 2

Mean weights of chickens on each diet

Diet	Mean Weight
1.00	102.65
2.00	122.62
3.00	142.95
4.00	135.26

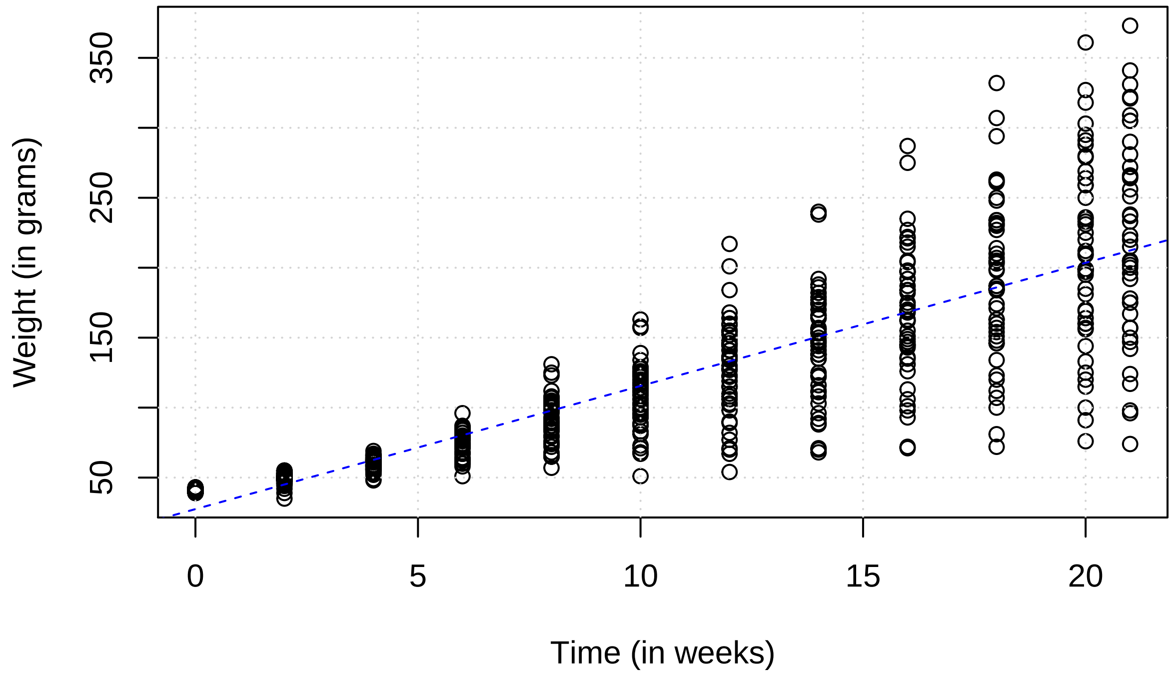

Figure 1. Scatterplot of chicken weights over time.

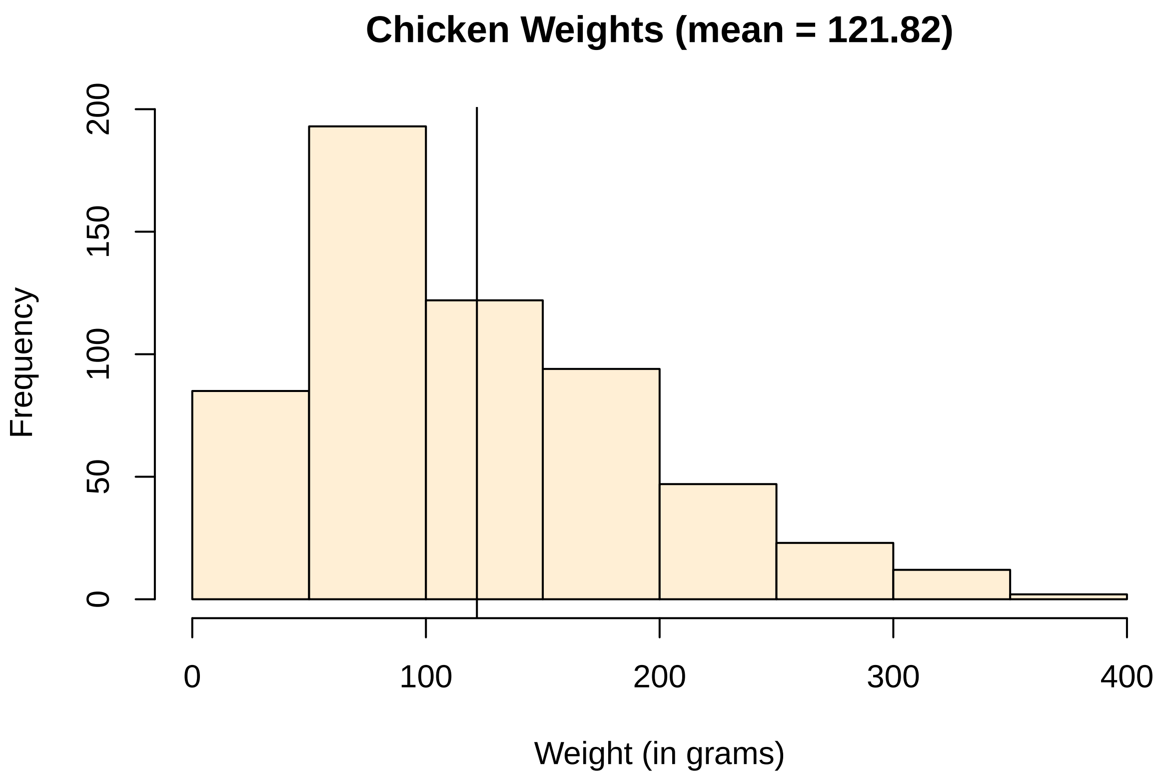

Figure 2. Distribution of weights across all data points.

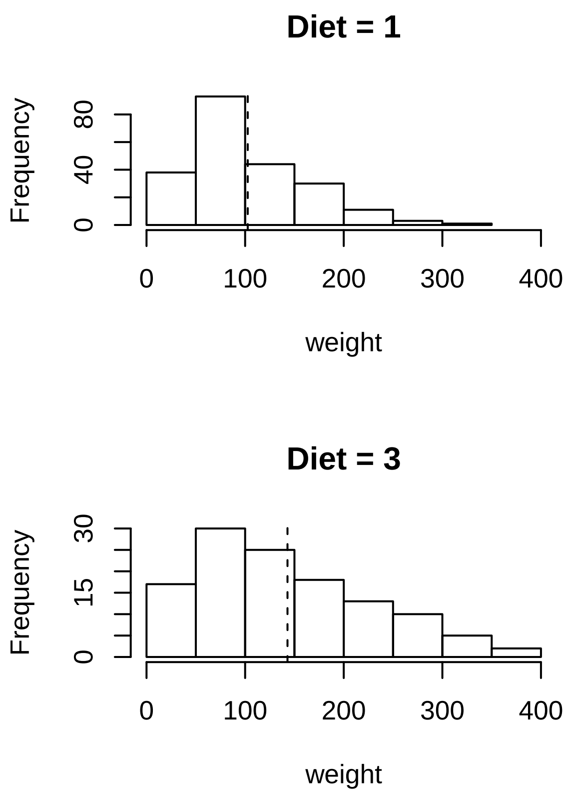

Figure 3. Histograms of the distribution of weights across time for each diet. Vertical lines are means.

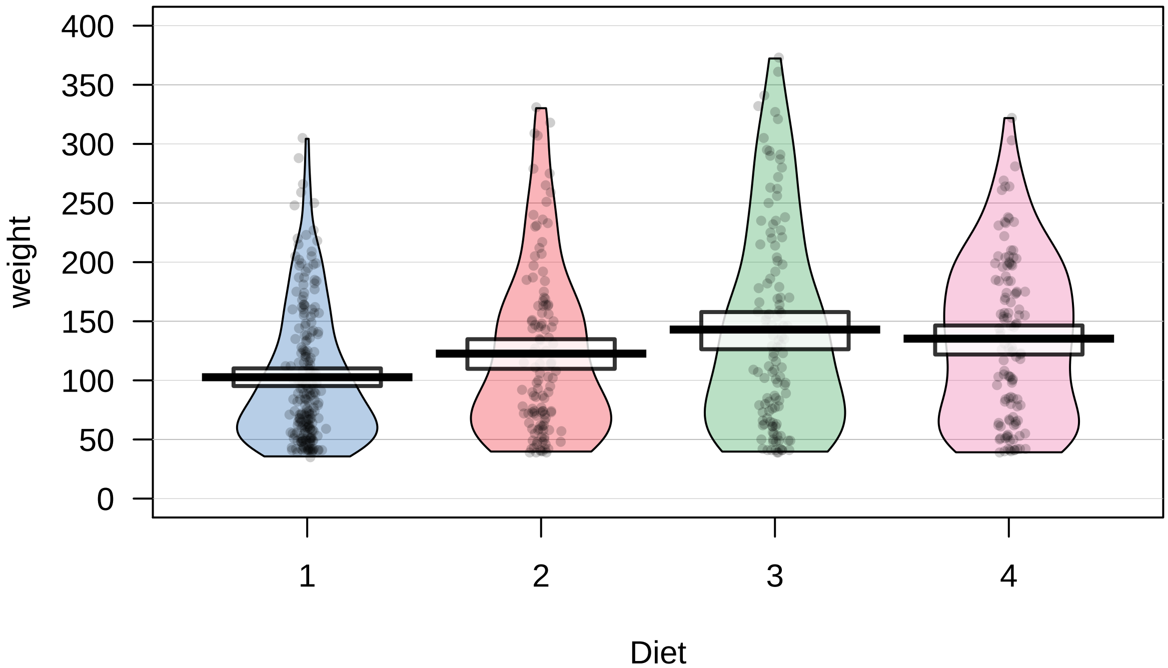

Figure 4. Pirateplot showing the distribution of chicken weights by diet. Horizontal lines show means while white boxes show Bayesian 95% highest density intervals.