

[1 Fried Chicken Pricing](#)[2 Conclusion](#)[3 Reference](#)[Code ▼](#)

# Fried Chicken - Intro to Data Science overview

GWU

2022-09-13

## 1 Fried Chicken Pricing

### 1.1 The beginning

At a yelp-highly-rated (<https://www.yelp.com>) chicken place one summer, I was waiting for my order, which took **forever** (exaggerated). Without any other better things to do, I re-studied the menu there, trying to understand the pricing structure. See photo:



I was wondering to myself tons of questions: how can I get the best deal? What is the cost per wing, and cost per drum? Does the restaurant use a formula to determine the pricing?

Bugger! "I am *cursed*. Shouldn't have come here. I might not get out of here alive..." I thought at the time.

You can see the set prices in the photo. We will need to re-format our data, from this *wide* format (often in pivot and summary tables), to a flat, *long* format here. Let us practice our Markdown skill at the same time to tabulate them:

Wings	Drums	Price
5	0	6.69
10	0	12.99
15	0	17.99
20	0	23.99
40	0	45.99
0	3	6.69
0	5	10.99
0	10	19.99
0	15	29.99

Wings	Drums	Price
3	2	7.99
7	4	16.99
12	6	27.99
20	9	43.99

Let us now enter these as a dataframe in R.

[Code](#)

So there are 13 data points in our dataset. (Notice the use of inline R codes here.)

## 1.2 Exploratory Data Analysis (EDA)

There are a bit of things we typically look at for EDA.

1. Basic statistics
  - mean, s.d., median, range (four spaces at the start of line for proper sub-list indentation)
2. Simple correlations and tests
  - correlation matrix if applicable
  - z-test, t-test, anova test if applicable
  - chi-squared test if applicable
3. Normality
  - QQ-plot
  - boxplot
  - histogram
  - Shapiro-Wilk test
  - ...

### 1.2.1 Basic statistics

[Code](#)

Table: Statistics summary.

	wing	drum	price
Min	Min. :-5.0	Min. : 0.00	Min. : 6.7
Q1	1st Qu.: 0.0	1st Qu.: 0.00	1st Qu.:11.0
Median	Median : 7.0	Median : 3.00	Median :18.0
Mean	Mean : 9.4	Mean : 4.15	Mean :20.9
Q3	3rd Qu.:15.0	3rd Qu.: 6.00	3rd Qu.:28.0
Max	Max. :40.0	Max. :15.00	Max. :46.0

When I ran this the first time, the minimum number of wings is **negative 5**. That's obviously a mistake. Have an action plan to fix such non-sensible or missing values. For me, it's just getting some more coffee. After I fixed it, re-run the EDA, and all looks good now.

[Code](#)

Table: Statistics summary.

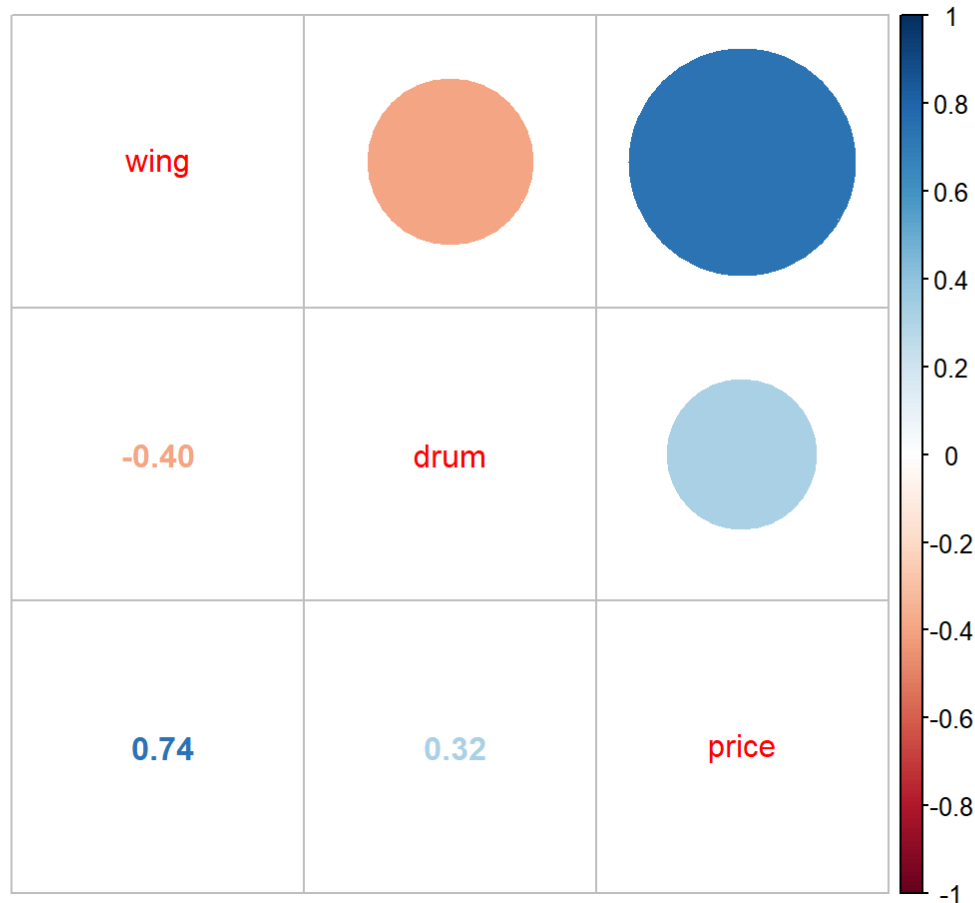
	wing	drum	price
Min	Min. : 0.0	Min. : 0.00	Min. : 6.7
Q1	1st Qu.: 0.0	1st Qu.: 0.00	1st Qu.:11.0
Median	Median : 7.0	Median : 3.00	Median :18.0
Mean	Mean :10.2	Mean : 4.15	Mean :20.9
Q3	3rd Qu.:15.0	3rd Qu.: 6.00	3rd Qu.:28.0
Max	Max. :40.0	Max. :15.00	Max. :46.0

## 1.2.2 Tests (Correlation, ANOVA, ...)

Since the features/variables are all numerical (quantitative), it makes most sense to check their linear correlations.

[Code](#)

## Correlation Matrix for Chicken Meal Price



It is good to see that price has a decent correlation with both wing and drum, with wing seems to have a higher correlation. It is also great that drum and wing do not have too strong a correlation between them. Imagine if each drum I buy, I always will be thrown with 2 dog bones, I will have a hard time knowing if I am really paying for the drum or the dog bones. This is called a problem of collinearity. (We'll check with VIF.)

### 1.2.3 Normality check/test

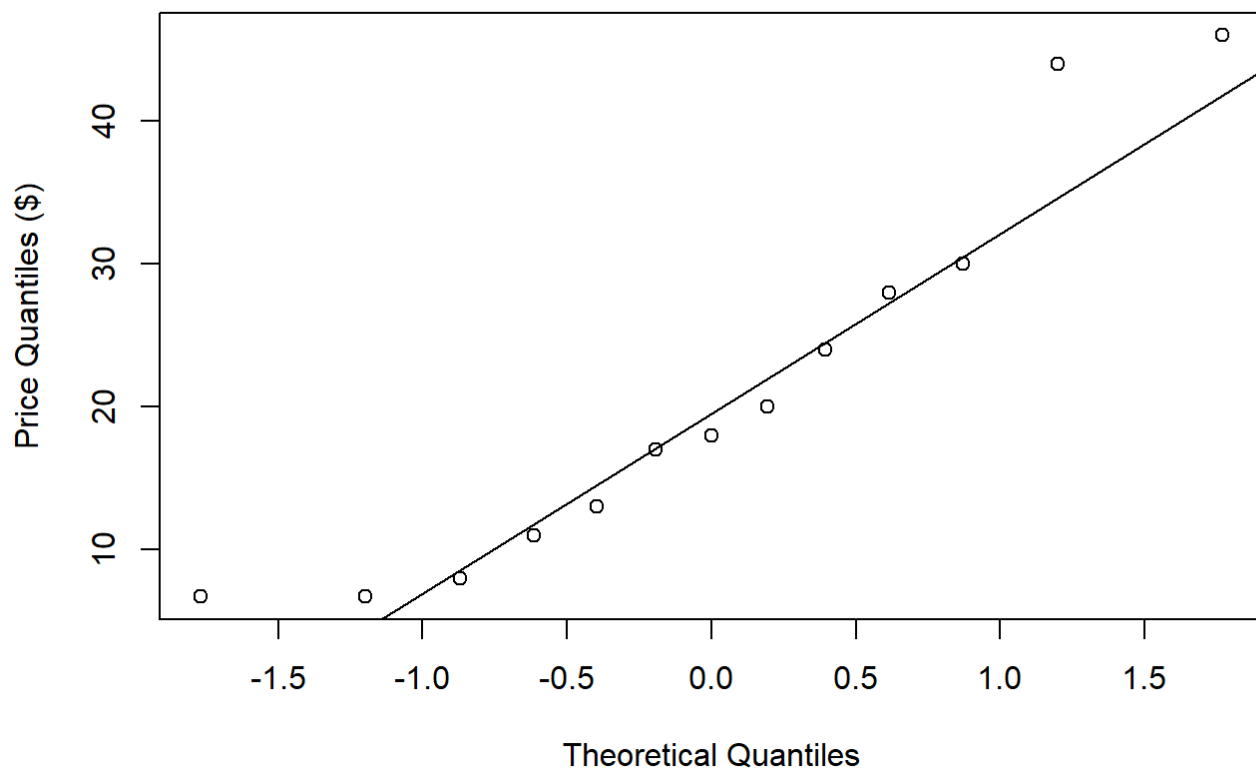
It is usually best if all the variables (numerical ones) are normally distributed. This usually does not happen, but we would still like to know overall the distribution is bell-shaped, and not too awkward.

#### 1.2.3.1 QQ-plot

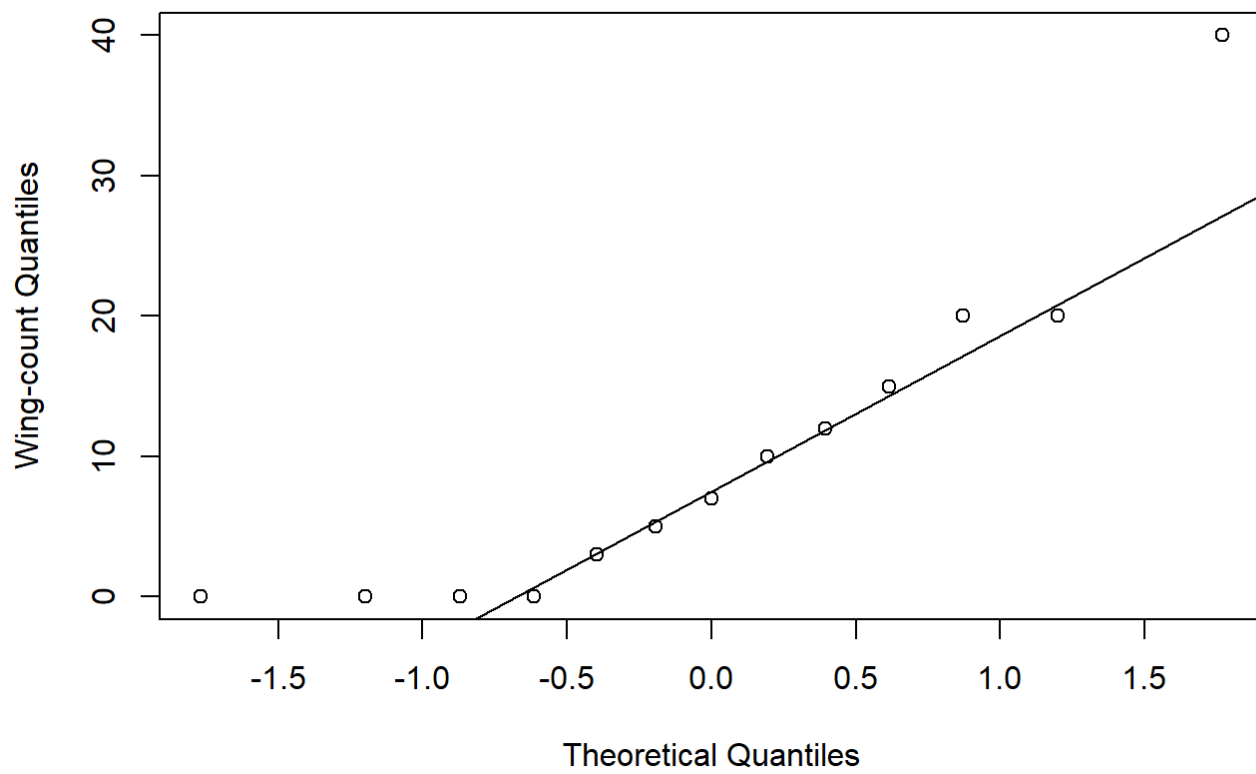
We can first check the QQ-plots for each variable by itself. A straight line means normal distribution.

[Code](#)

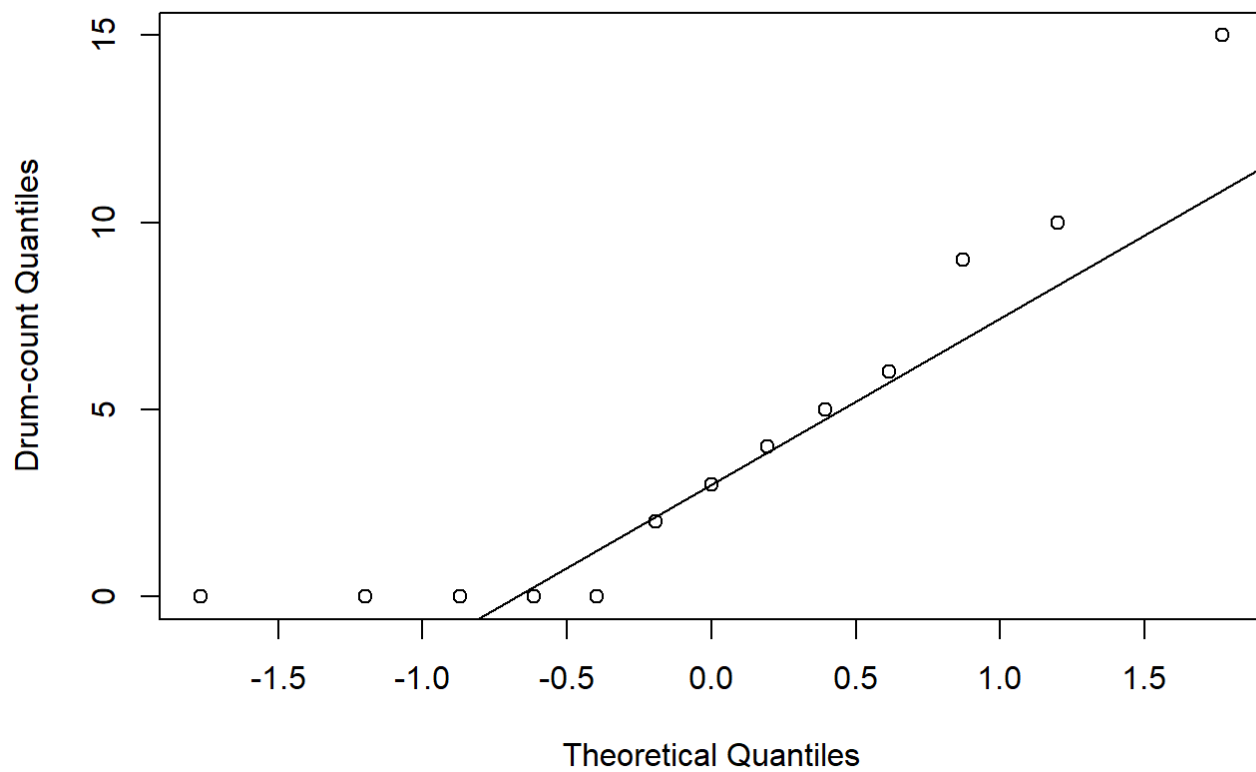
## Price Q-Q Plot

[Code](#)

## Wing Q-Q Plot

[Code](#)

## Drum Q-Q Plot



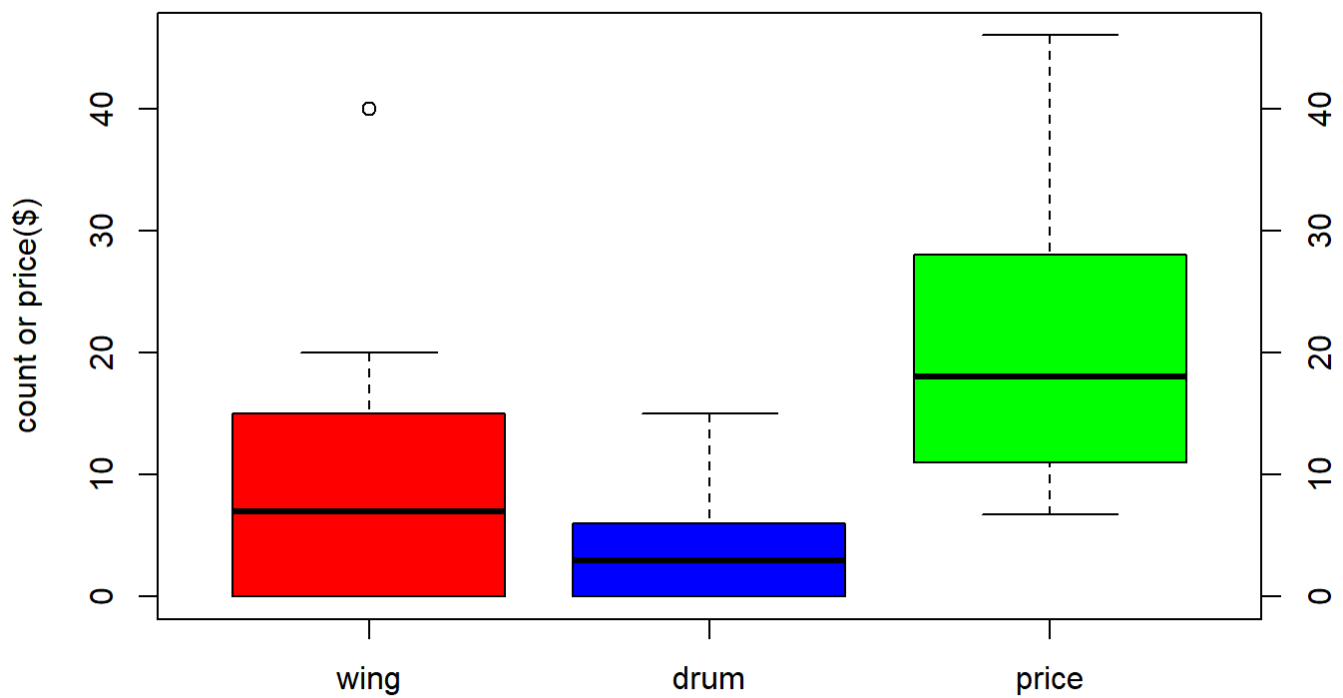
The data values are not close to normal distribution, but with only 13 data points, that is expected, and it's probably okay.

### 1.2.3.2 Boxplot

Next we can check the boxplots for a rough visual.

[Code](#)

## Boxplots for the three variables



Same conclusion: they do not look like normal, but we'll take it.

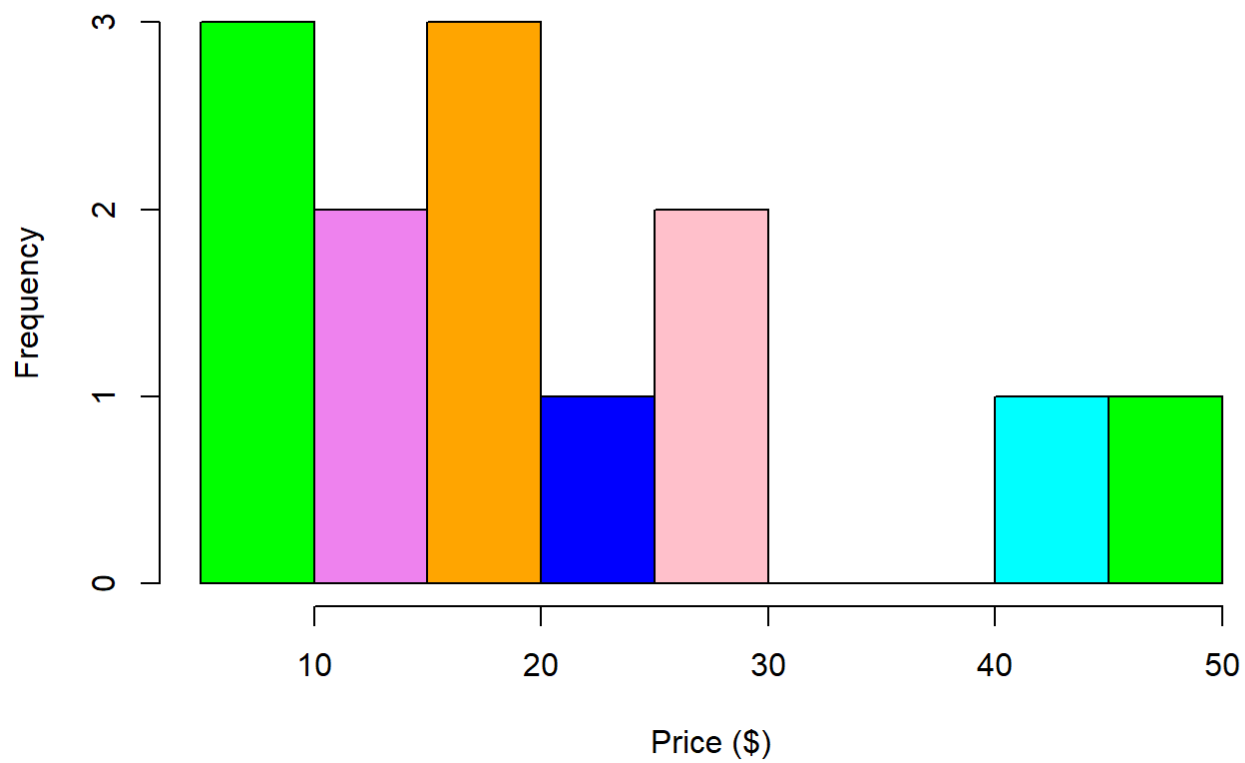
### 1.2.3.3 Histogram

Now histograms:

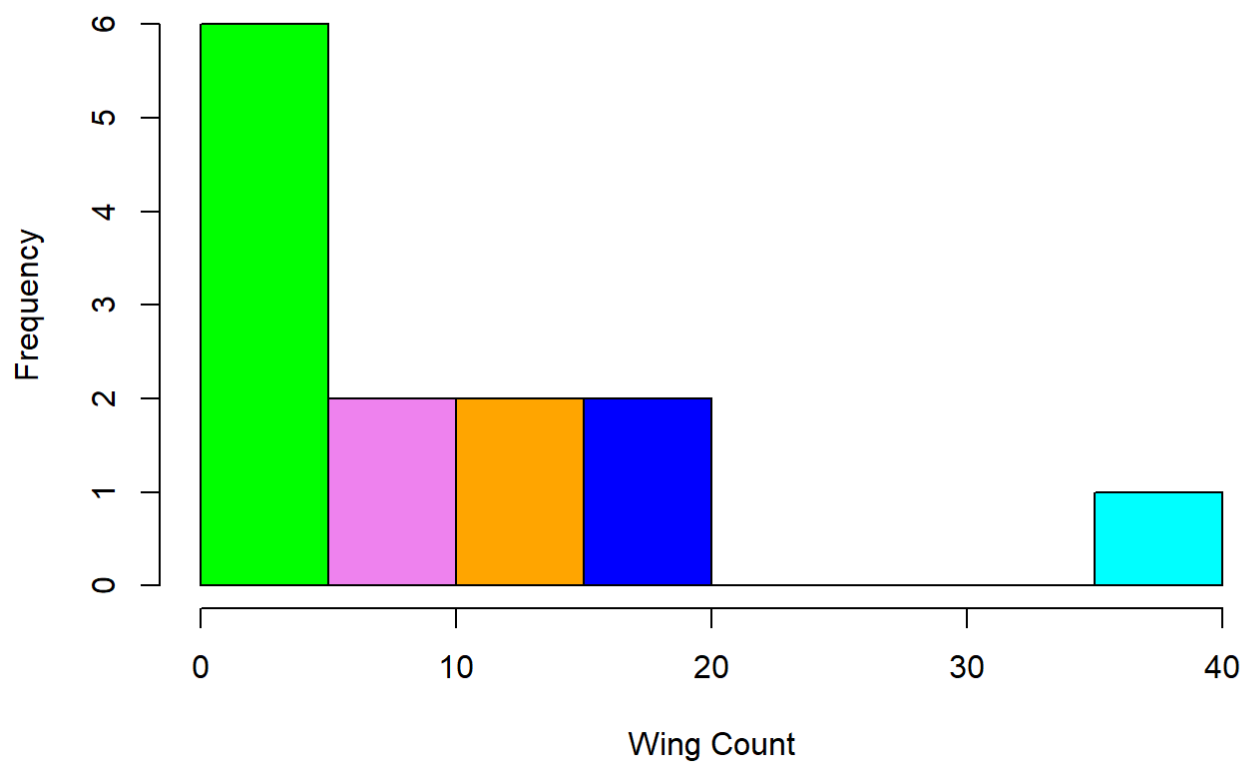
[Code](#)



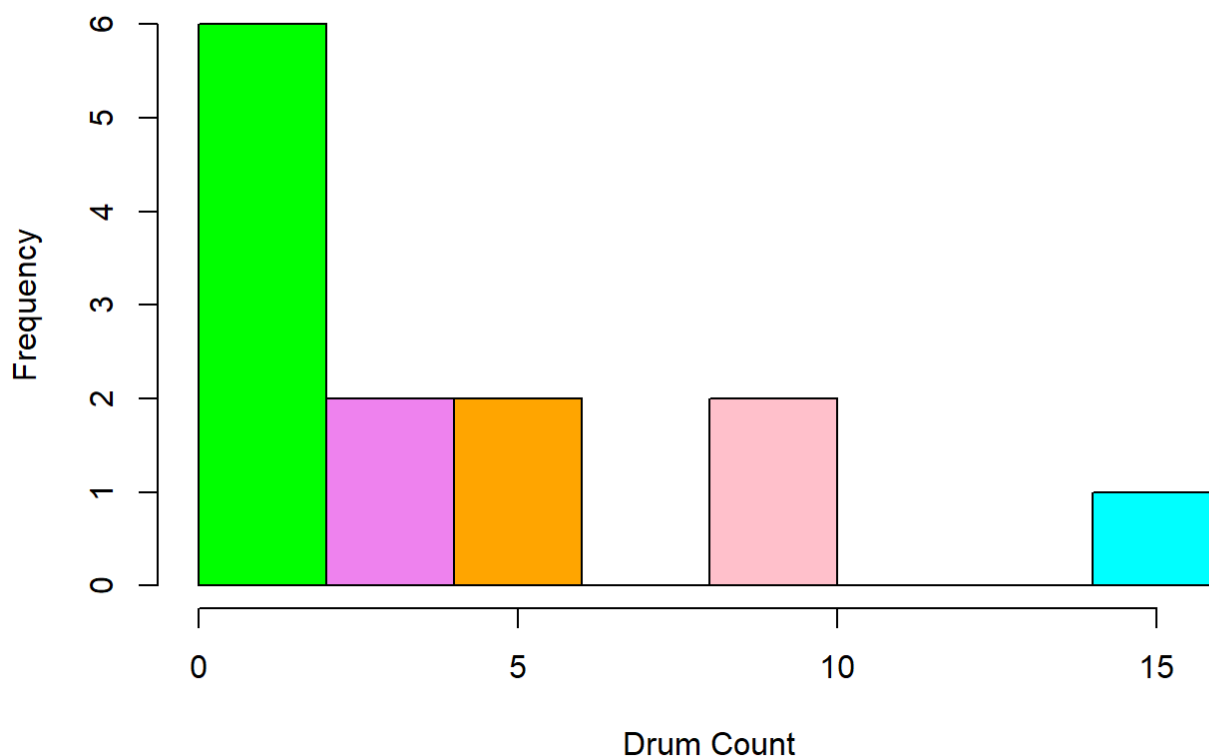
## Histogram for Price distribution

[Code](#)

## Histogram for Wing-count

[Code](#)

## Histogram for Drum-count



### 1.2.3.4 Shapiro-Wilk test

And finally, using shapiro-wilk test:

[Code](#)

The Shapiro-Wilk test p-value on `price` is 0.129.

The Shapiro-Wilk test p-value on `wing` is 0.019.

The Shapiro-Wilk test p-value on `drum` is 0.03.

We'll learn to interpret these soon.

## 1.3 Linear Model

Now if those passed the smell test, we can try build some models. Linear model is first to come to mind. We will model the price with wing and drum as the two independent variables (also called features).

[Code](#)

Summary of linear model for Chicken Meals

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.777	0.5038	1.54	0.154
wing	1.163	0.0255	45.67	0.000
drum	2.012	0.0620	32.44	0.000

As we can see, the  $R^2$  (using *L<sup>A</sup>T<sub>E</sub>X* formatting here) value of the linear model is 0.9957, which shows the prices are set with rather strict per-piece-structure.

If we use the (default 95%) confidence intervals of the coefficients as shown here:

[Code](#)

Coefficient Confidence  
Intervals (CI) at 95%

	2.5 %	97.5 %
(Intercept)	-0.345	1.90
wing	1.106	1.22
drum	1.874	2.15

## 1.3.1 Findings

We find that:

1. Each piece of wing cost about \$1.11 to \$1.22
2. Each piece of drum stick cost about \$1.87 to \$2.15
3. The intercept of about \$-0.345 to \$1.9, or average of \$0.78 probably represent the base per-order or box/bag charge.

Check:

We should always check for multi-collinearity in linear models. Here are the **Variance Inflation factors VIF**.

[Code](#)

Chicken Model  
VIFs

drum	wing
1.19	1.19

[Code](#)

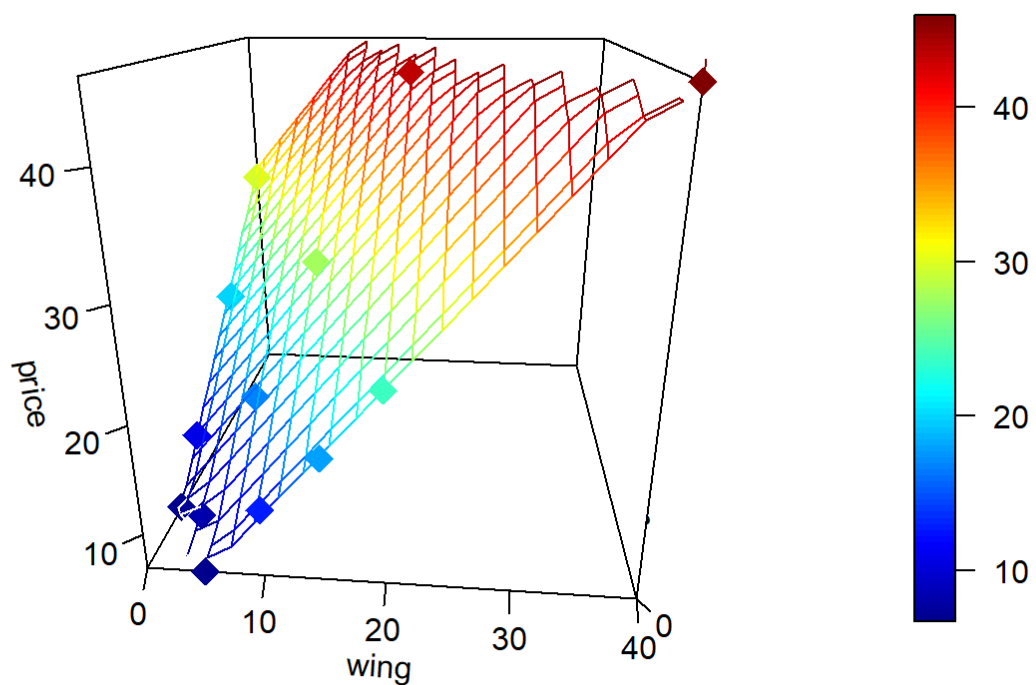
With VIF values less than 5, we can safely conclude there is not much collinearity concerns in the dataset.

## 1.3.2 Plot 3D

Visual is always king in data science. Let's get a bit fancy.

[Code](#)

## Chicken Price Analysis



## 2 Conclusion

So what is the best deal???

### 2.1 First bite

Let me compare the predicted values from the model with the actual values. In other words, look at the residual values.

[Code](#)

Data values and residuals

wing	drum	price	fitval	residuals
5	0	6.69	6.59	0.0976
10	0	12.99	12.41	0.5827
15	0	17.99	18.22	-0.2322
20	0	23.99	24.04	-0.0471
40	0	45.99	47.30	-1.3066

wing	drum	price	fitval	residuals
0	3	6.69	6.81	-0.1236
0	5	10.99	10.84	0.1524
0	10	19.99	20.90	-0.9077
0	15	29.99	30.96	-0.9678
3	2	7.99	8.29	-0.3005
7	4	16.99	16.97	0.0235
12	6	27.99	26.81	1.1846
20	9	43.99	42.15	1.8448

Ah, the residual is most negative with the 40-wing combo at -\$1.31. That saves me a can of soda. I should go for that and kill myself.

But if I am wiser, and insist on having a one-person portion:

1. 5-piece wing is at +\$0.10 (paying a dime too much!) - Not good.
2. 3-piece drum is at -\$0.12, saving a dozen pennies, sounds good.
3. The 3-wing-2-drum deal (at -\$0.30) is the best value for me!

## 2.2 Double dip

Hold on! If we plan to do this regularly, shouldn't we inspect the percentage savings instead?

[Code](#)

Revised Data values and residuals

wing	drum	price	fitval	residuals	res_2_p
5	0	6.69	6.59	0.0976	1.458
10	0	12.99	12.41	0.5827	4.486
15	0	17.99	18.22	-0.2322	-1.291
20	0	23.99	24.04	-0.0471	-0.196
40	0	45.99	47.30	-1.3066	-2.841
0	3	6.69	6.81	-0.1236	-1.848
0	5	10.99	10.84	0.1524	1.386

wing	drum	price	fitval	residuals	res_2_p
0	10	19.99	20.90	-0.9077	-4.541
0	15	29.99	30.96	-0.9678	-3.227
3	2	7.99	8.29	-0.3005	-3.761
7	4	16.99	16.97	0.0235	0.139
12	6	27.99	26.81	1.1846	4.232
20	9	43.99	42.15	1.8448	4.194

The new metric here tells me:

1. The 10-drum deal at -4.5% is best percentage-wise.
2. The 3-wing-2-drum deal at -3.76% is next.

To me, putting all these different metrics and criteria together, the 3-wing-2-drum is my winner.

**Case closed. Mic dropped.**

## 2.3 Confession

Oh wait... I forgot that I am a pescatarian (vegetarian and fish) for over a decade! I just have these cravings for wings once in a while. What if I add a sin-ful parameter as a penalty term, how much should I order next time to minimize my sin (Root-Mean-Squared-Sinfulness).

To be continued...

But yes, the wings there were really really good. Totally worth the wait.

## 3 Reference

APA Style preferred

1. Yelp
2. Google
3. My wallet