Ole J. Forsberg

STAT 200: Introductory Statistics

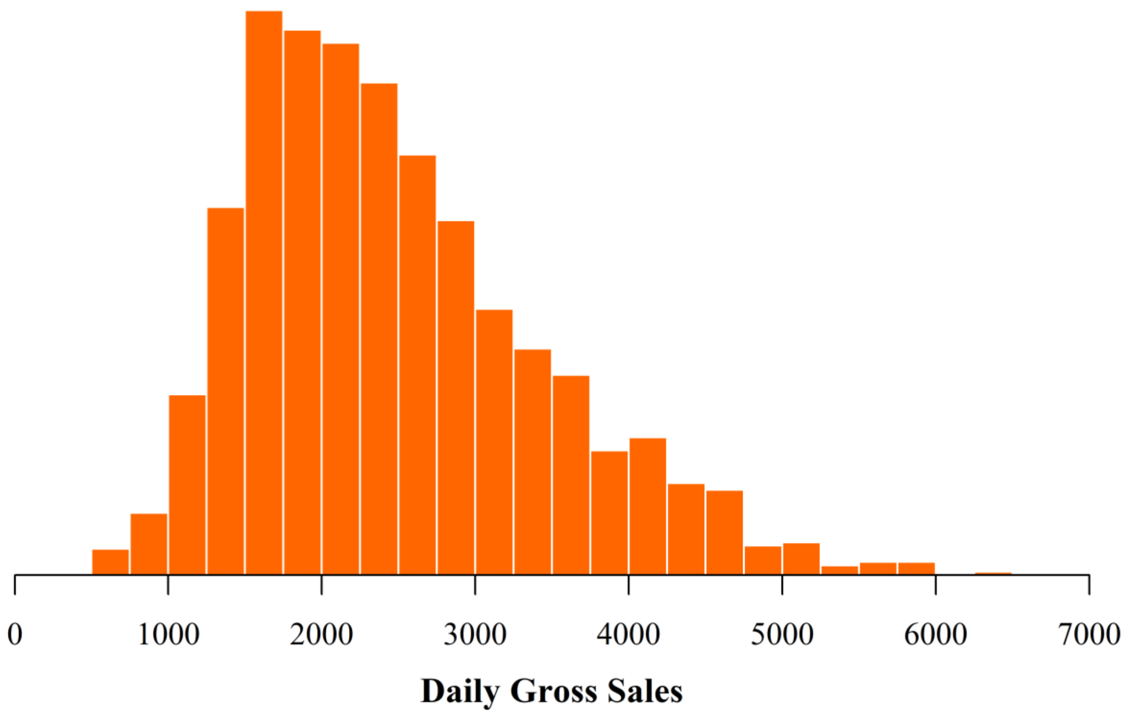
Practicum 2: The Data-Generating Process

February 5, 2018

Gross Sales

Gross sales are defined as the amount of money earned before subtracting costs like taxes, wages, and materials. It is a continuous random variable. As such, there are three distributions that work: Uniform, Exponential, and Normal. Since this random variable is bounded below by 0, it seems as though the Exponential would be the obvious choice.

However, the following is a histogram of the gross sales at the Lamplighter restaurant for the period between January 5, 2014, and December 30, 2017. Note the distinct bell shape. That suggests the Normal distribution may be closest.



Recall from Practicum 1 that the Hildebrand Rule does not indicate sufficient skew (H = 0.19). The Exponential distribution is skewed. The Normal is not. Because of these reasons, it appears as though the distribution is most closely Normal. The Normal distribution has two parameters. From this data, the best estimate of μ is 2482, and the best estimate of σ is 963.

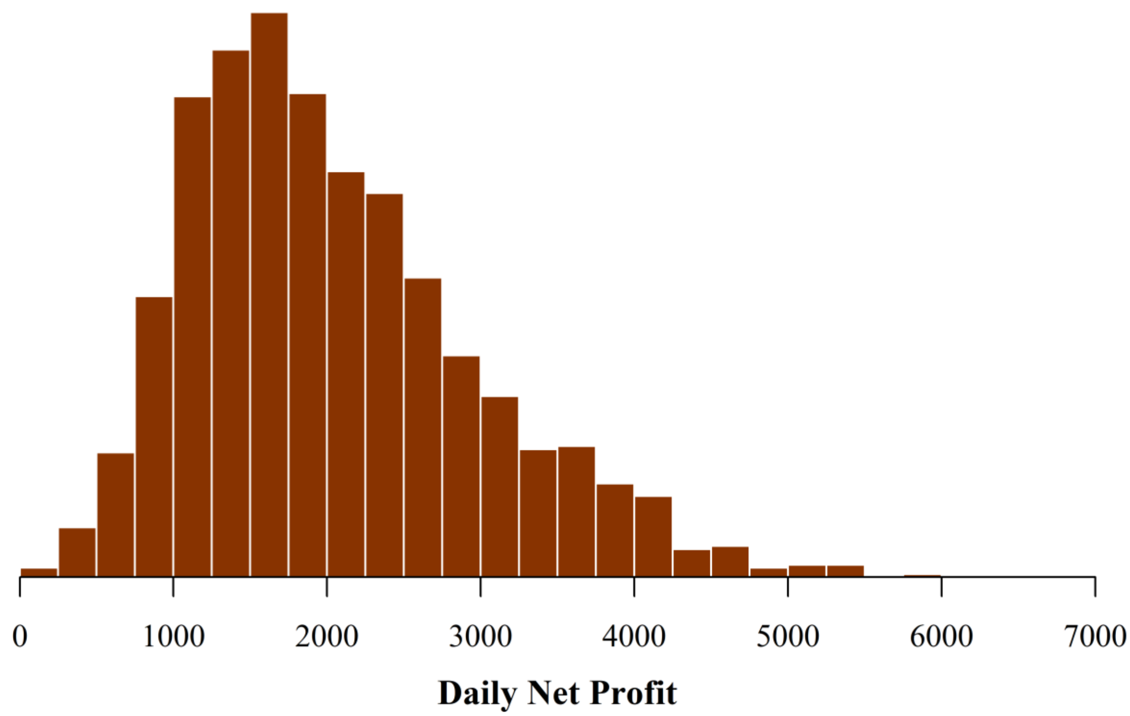
And so, if G is the daily gross sales,

$$\mathbf{G \sim Normal(\mu = 2482; \sigma = 963)}$$

Net Profit

The net profit is the amount of money made after subtracting off expenses from the gross sales. It is a continuous random variable. As such, there are three distributions that work: Uniform, Exponential, and Normal. Since this random variable is bounded below by 0, it seems as though the Exponential would be the obvious choice.

However, the following is a histogram of the net profit made by the Lamplighter restaurant during the period from January 5, 2014, and December 30, 2017. Note the distinct bell shape.



Also, as we discovered in Practicum 1, the Hildebrand Rule indicates it is not sufficiently skewed (H = 0.18). Exponential distributions are skewed. As such, the Normal distribution may be a better estimate. The Normal distribution has two parameters. From this data, the best estimate of μ is 2038, and the best estimate of σ is 940.

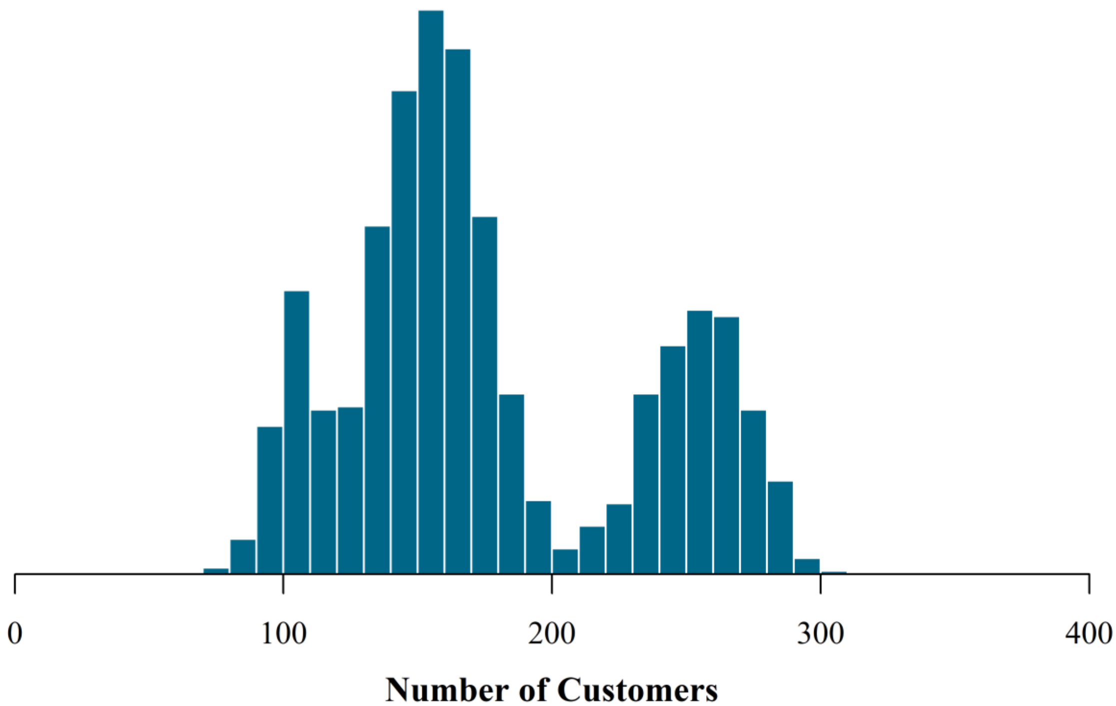
And so, if P is the daily profit,

$$\mathbf{P \sim Normal(\mu = 2038; \sigma = 940)}$$

Number of Customers

The number of customers is a discrete random variable. It is bounded below by 0. There is no upper bound. Furthermore, the variable measures the number of successes (a customer enters) over a time period. As such, logically, this variable most closely follows the Poisson distribution.

The following is a histogram of the net profit made by the Lamplighter restaurant during the period from January 5, 2014, and December 30, 2017.



While it does not look like a Poisson random variable, that is the closest distribution. The Poisson distribution is a count (discrete) of successes (customer entering the Lamp Lighter) over time (a day).

The Poisson distribution has one parameter, λ, the mean. Thus, from this data, the best estimate of λ is 177.86. And so, if N is the number of customers,

$$N \sim \text{Poisson}(\lambda = 177.86)$$

Appendix: R Script

The following is the R script that created the above analysis.

```
##### Practicum Two
#####


### Preamble

source("http://rfs.kvasaheim.com/stat200.R")

dt = read.csv("lamplighterSales.csv")
summary(dt)
attach(dt)




##### Gross Sales

## Nice-looking graphic
par(mar=c(4,1,1,1))
par(family="serif")
par(cex.lab=1.1, font.lab=2)

histogram(grossSales, col="#ff6600", breaks=seq(0,7000,250))

axis(1)
title(xlab="Daily Gross Sales", line=2.5)



## Which distribution?
hildebrand.rule(grossSales)  # Not skewed -> Normal

mean(grossSales)  # 2482
sd(grossSales)    #  963




##### Net Profit

## Nice-looking graphic
par(mar=c(4,1,1,1))
par(family="serif")
par(cex.lab=1.1, font.lab=2)

histogram(netProfit, col="#883300", breaks=seq(0,7000,250))

axis(1)
title(xlab="Daily Net Profit", line=2.5)
```

```
## Which distribution?
hildebrand.rule(netProfit)  # Not skewed -> Normal

mean(netProfit)  # 2038
sd(netProfit)    #  940





##### Number of Customers

## Nice-looking graphic
par(mar=c(4,1,1,1))
par(family="serif")
par(cex.lab=1.1, font.lab=2)

histogram(customers, col="#006688", breaks=seq(0,400,10))

axis(1)
title(xlab="Number of Customers", line=2.5)



## Which distribution? Poisson
mean(customers)  # 178
```