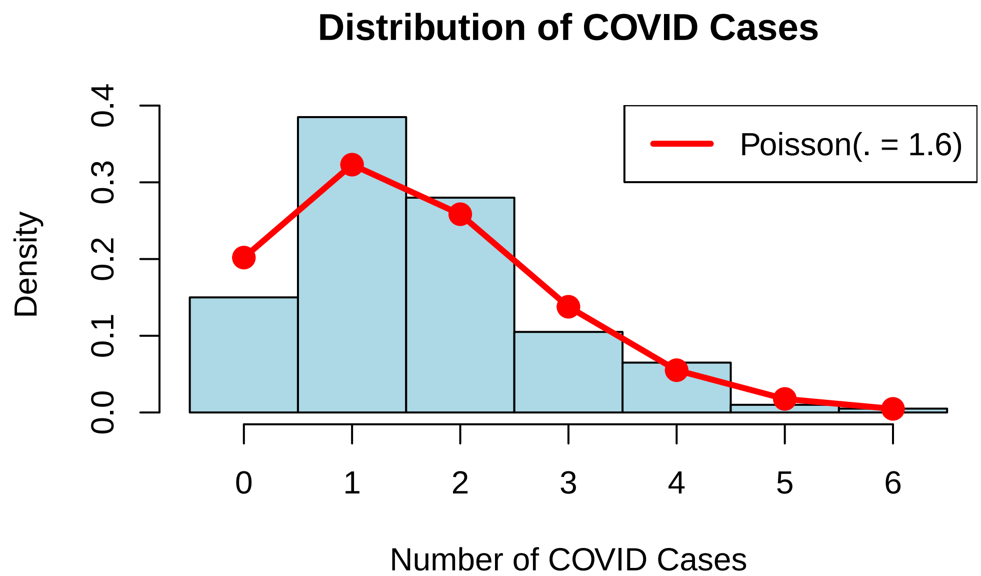
Lab 3

Christopher Munoz

```
load("StudentData.dat")
hist(covid, prob=TRUE, breaks=seq(-0.5, max(covid)+0.5, 1), main="Distribution of COVID Cases
lambda_covid <- mean(covid)
x_vals <- 0:max(covid)
y_vals <- dpois(x_vals, lambda=lambda_covid)
points(x_vals, y_vals, col='red', pch=19, cex=1.5)
lines(x_vals, y_vals, col='red', lwd=3)
legend("topright", legend=paste0("Poisson(  = ", round(lambda_covid, 2), ")"), col="red", lwd
```

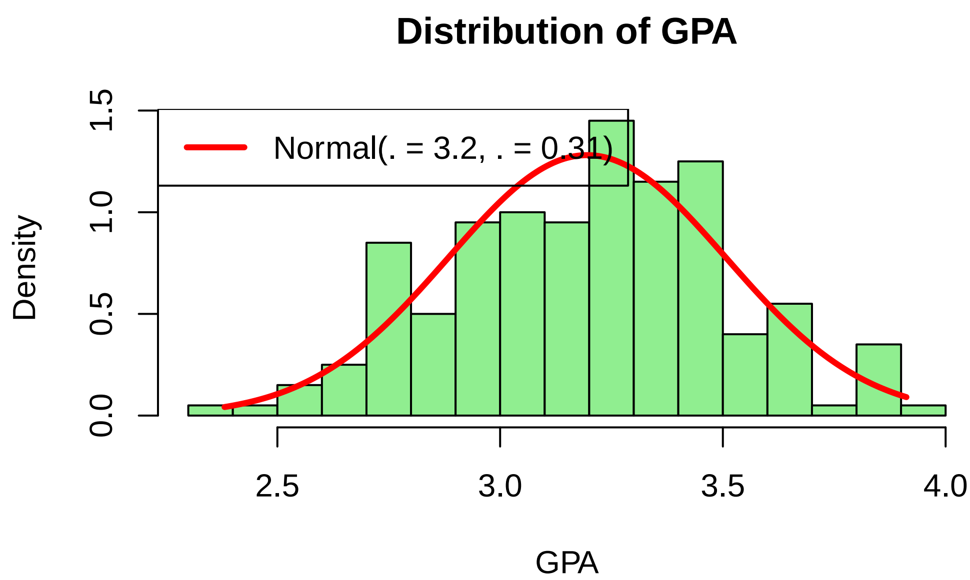



```
cat("COVID: Poisson(  =", round(lambda_covid, 2), ")\n")
```

```
COVID: Poisson(  = 1.6 )
```

The variable covid is best modeled by a is best modeled by a random variable **Poisson because our data is discrete, non-negative and represents the count of independent rare events. Poisson is also a natural choice for biological phenomenon, as my colleague says, its the "fish" distribution.**

```
load("StudentData.dat")
hist(GPA, prob=TRUE, breaks=20, main="Distribution of GPA", xlab="GPA", col="lightgreen", bo
mean_gpa <- mean(GPA)
sd_gpa <- sd(GPA)
x_vals <- seq(min(GPA), max(GPA), length=200)
y_vals <- dnorm(x_vals, mean=mean_gpa, sd=sd_gpa)
lines(x_vals, y_vals, col='red', lwd=3)
legend("topleft", legend=paste0("Normal(  = ", round(mean_gpa, 2), ",   = ", round(sd_gpa, 2)
```

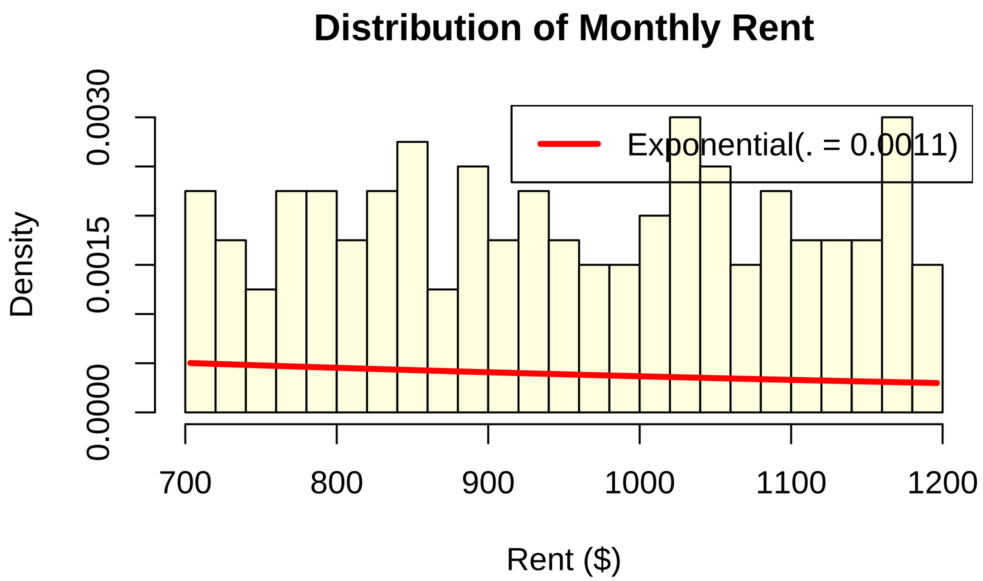



```
cat("GPA: N(  =", round(mean_gpa, 2), ",   =", round(sd_gpa, 2), ")\n")
```

```
GPA: N(  = 3.2 ,   = 0.31 )
```

The variable GPA is best modeled by a **Normal** random variable because **GPA is a continuous metric that results from averaging many independent events(assignments, tests, homework). From the Central Limit Theorem we know when many independent factors combine we get a Normal distribution.**

```
load("StudentData.dat")
hist(rent, prob=TRUE, breaks=20, main="Distribution of Monthly Rent", xlab="Rent ($)", col="l
rate_rent <- 1 / mean(rent)
x_vals <- seq(min(rent), max(rent), length=200)
y_vals <- dexp(x_vals, rate=rate_rent)
lines(x_vals, y_vals, col='red', lwd=3)
legend("topright", legend=paste0("Exponential(  = ", round(rate_rent, 4), ")"), col="red", lw
```



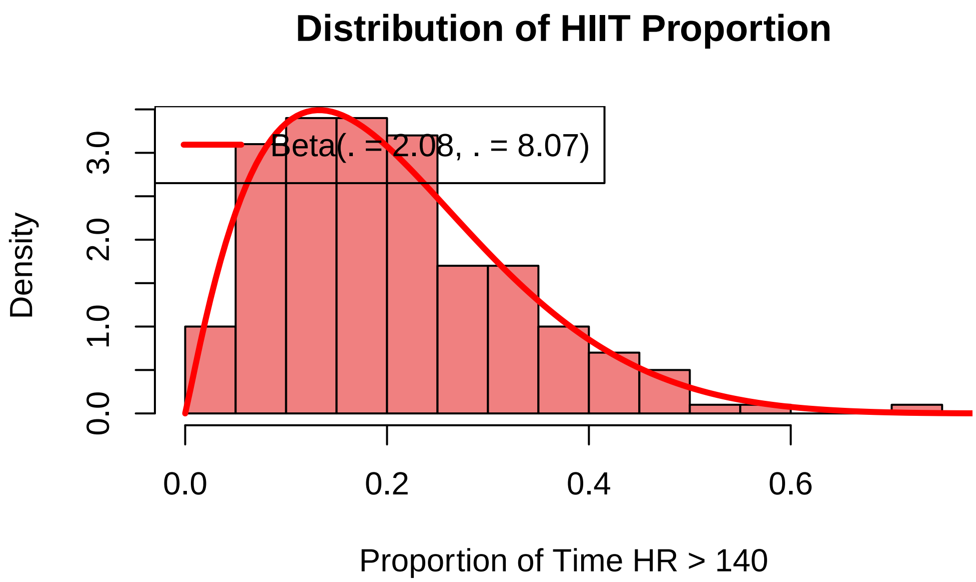

```
cat("Rent: Exp(  =", round(rate_rent, 4), ")\n")
```

```
Rent: Exp(  = 0.0011 )
```

The variable rent is best modeled by a **Exponential** random variable
because **Monthly rent is continuous and strictly positive with a right skew, its commonly for monetary values and costs associated data.**

```
load("StudentData.dat")
hist(HIIT, prob=TRUE, breaks=20, main="Distribution of HIIT Proportion", xlab="Proportion of
m <- mean(HIIT)
v <- var(HIIT)
alpha_hiit <- m * (m * (1 - m) / v - 1)
```

```
beta_hiit <- (1 - m) * (m * (1 - m) / v - 1)
x_vals <- seq(0, 1, length=200)
y_vals <- dbeta(x_vals, shape1=alpha_hiit, shape2=beta_hiit)
lines(x_vals, y_vals, col='red', lwd=3)
legend("topleft", legend=paste0("Beta(  = ", round(alpha_hiit, 2), ",   = ", round(beta_hiit,
```




```
cat("HIIT: Beta(  =", round(alpha_hiit, 2), ",   =", round(beta_hiit, 2), ")\n")
```

```
HIIT: Beta(  = 2.08 ,   = 8.07 )
```

The variable HIIT is best modeled by a **Beta** random variable because **High Intensity interval training represents a proportion, in this case the fraction of workouts with heart rates above 140 bmp. Since proportions are bounded between 0 and 1, the Beta distribution is the the best fit.**

Summary Statistics for each distribution and dataset:

```
=== SUMMARY STATISTICS ===

COVID:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0     1.0     1.0     1.6     2.0     6.0
Mean: 1.6
Variance: 1.366834
SD: 1.169117

GPA:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.381   2.967   3.208   3.195   3.413   3.912
Mean: 3.195423
Variance: 0.09702458
SD: 0.3114877

RENT:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  703.3   828.3   948.5   951.1  1069.2  1196.3
Mean: 951.103
Variance: 20668.04
SD: 143.7638

HIIT:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.01157 0.10490 0.18417 0.20504 0.27874 0.70344
Mean: 0.205038
Variance: 0.01461413
SD: 0.1208889
```