Izaak King

Professor Eleish

Data Science

September 19, 2025

"Lab 1"

For this lab the two variables I chose to examine are ECO (new) which refers to
Ecosystem Vitality and BDH (new) which refers to biodiversity and habitat.

1. Variable summaries
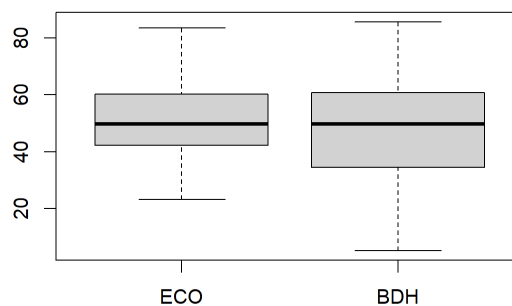
Summary for ECO:

```
> summary(ECO)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  23.10   42.27   49.75   51.10   60.08   83.60
```

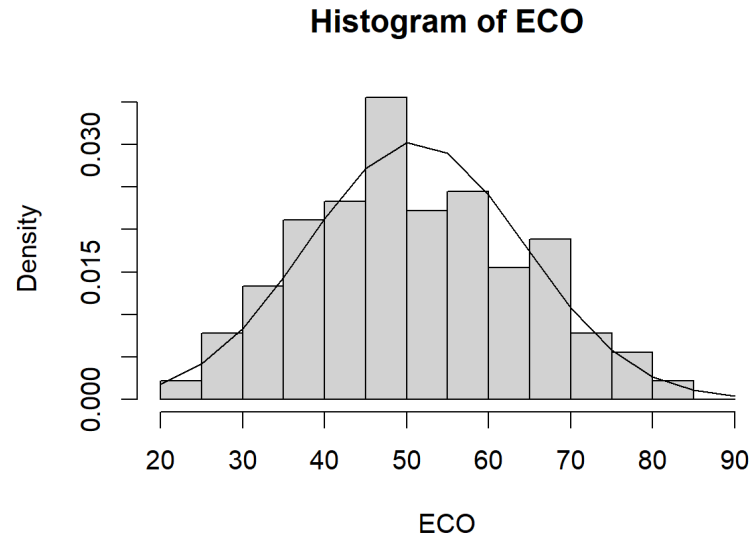Summary for BDH:

```
> summary(BDH)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   5.00   34.67   49.90   47.71   60.55   85.80
```
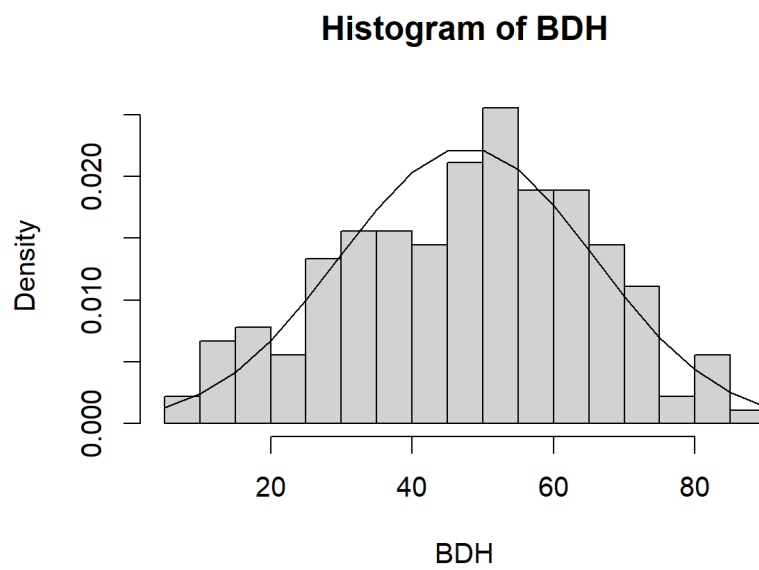
2. Variable boxplots

3. Histograms with overlayed theoretical probability distributions
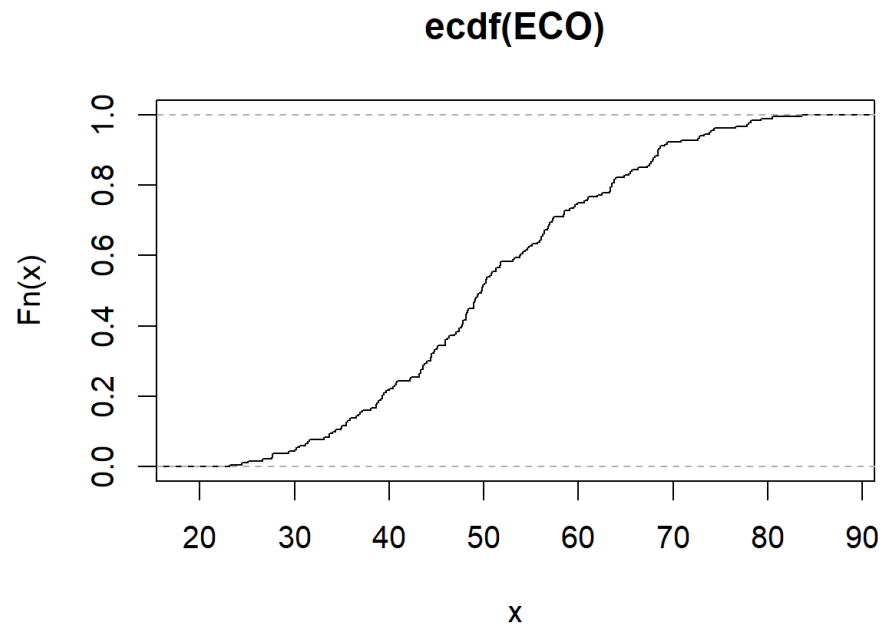
For ECO:
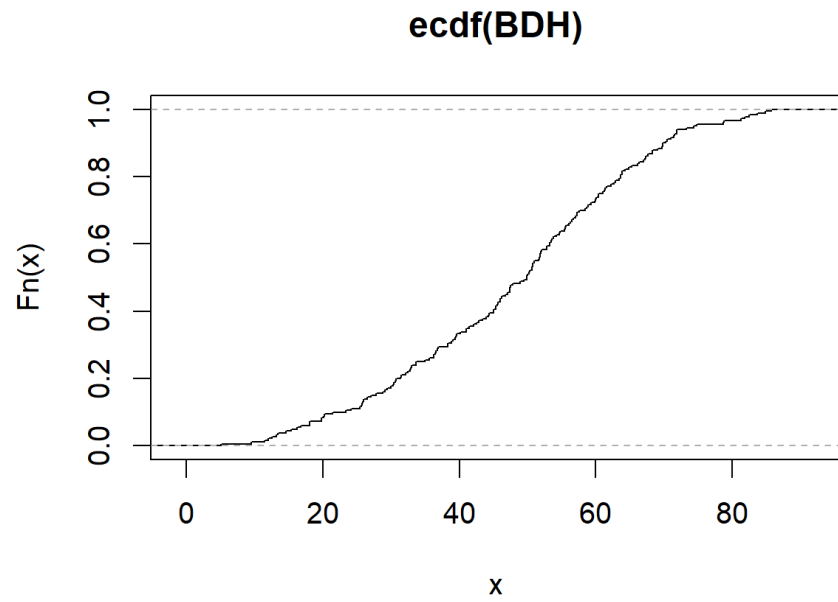
Histogram of ECO

For BDH:

Histogram of BDH

4. ECDF plots

For Eco:

ecdf(ECO)

For BDH:

ecdf(BDH)

5. QQ plots of each variable against the normal distribution

For ECO:

Normal Q-Q Plot
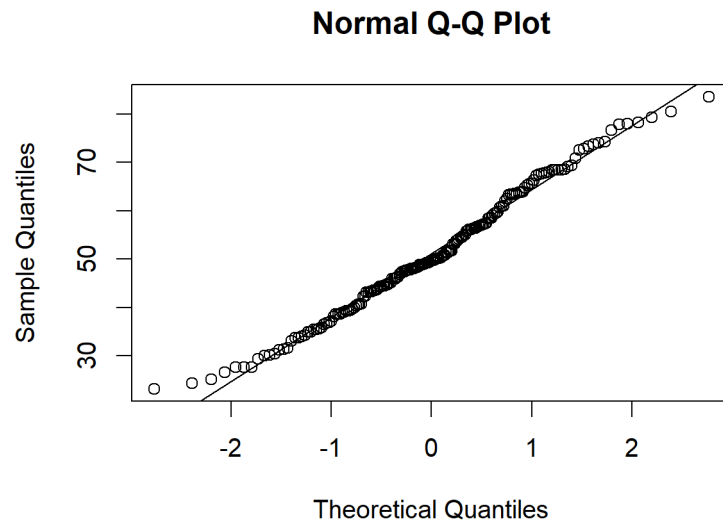
For BDH:
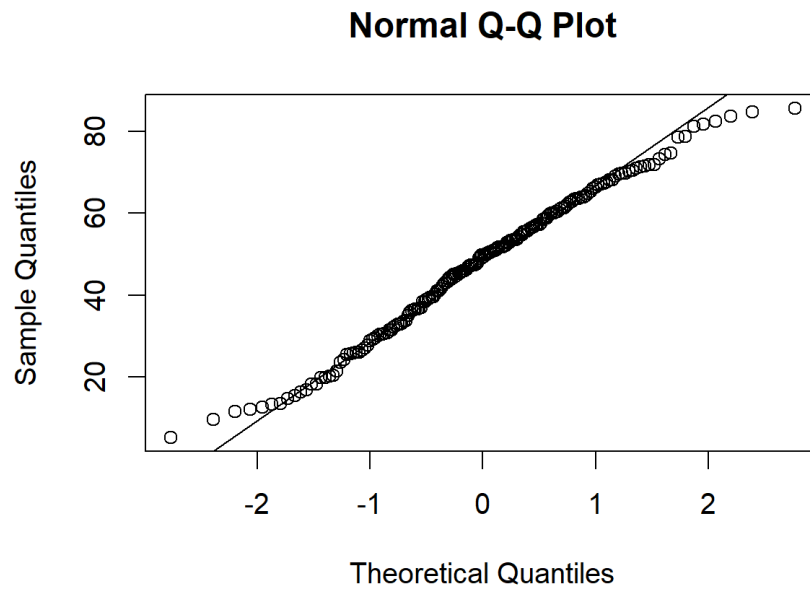
Normal Q-Q Plot

6. QQ plot of the 2 variables against each other

Q-Q plot for ECO vs BDH
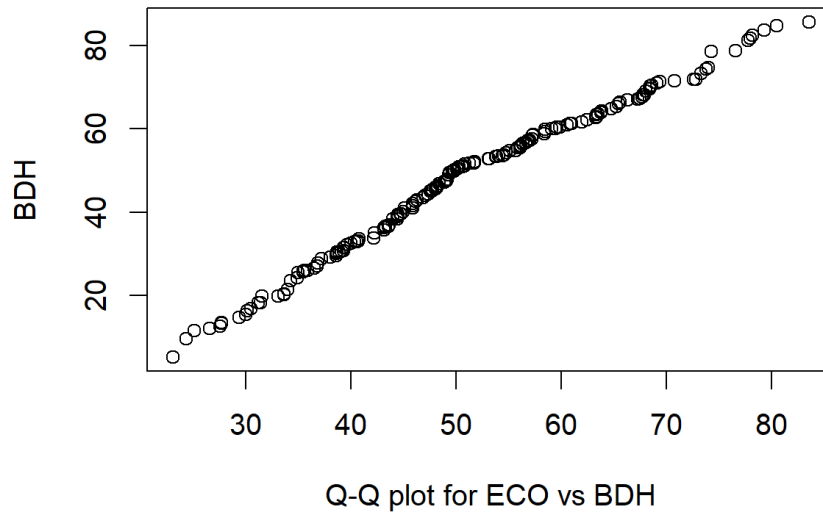
7. Normality statistical tests for each variable (Using both Shapiro-Wilk as well as

   Anderson-Darling tests).

   For ECO:

```
> shapiro.test(ECO)

        Shapiro-Wilk normality test

data:  ECO
W = 0.98825, p-value = 0.1409
```

```
> ad.test(ECO)

        Anderson-Darling normality test

data:  ECO
A = 0.53432, p-value = 0.1694
```

   For BDH:

```
> shapiro.test(BDH)

        Shapiro-Wilk normality test

data:  BDH
W = 0.98759, p-value = 0.1149
```

```
> ad.test(BDH)

        Anderson-Darling normality test

data:  BDH
A = 0.51809, p-value = 0.186
```

Thus, based on our statistical tests, our p-value > 0.05 and we fail to reject the

Null Hypothesis, indicating a normal distribution for both ECO and BDH

8. Statistical test for the variables having identical distributions (Using both Kolmogorov-Smirnov and Wilcox tests)

```
> ks.test(ECO,BDH)

        Asymptotic two-sample Kolmogorov-Smirnov test

data:  ECO and BDH
D = 0.16111, p-value = 0.0187
alternative hypothesis: two-sided
```

```
> wilcox.test(ECO,BDH)

        Wilcoxon rank sum test with continuity correction

data:  ECO and BDH
W = 17639, p-value = 0.1452
alternative hypothesis: true location shift is not equal to 0
```

Thus, based on our statistical tests, our p-value < 0.05 and we reject the Null Hypotheses, indicating ECO and BDH are not from the same distributions