

# Computer Lab 2

*Thomas Zhang*

*2015-11-24*

## Assignment 1

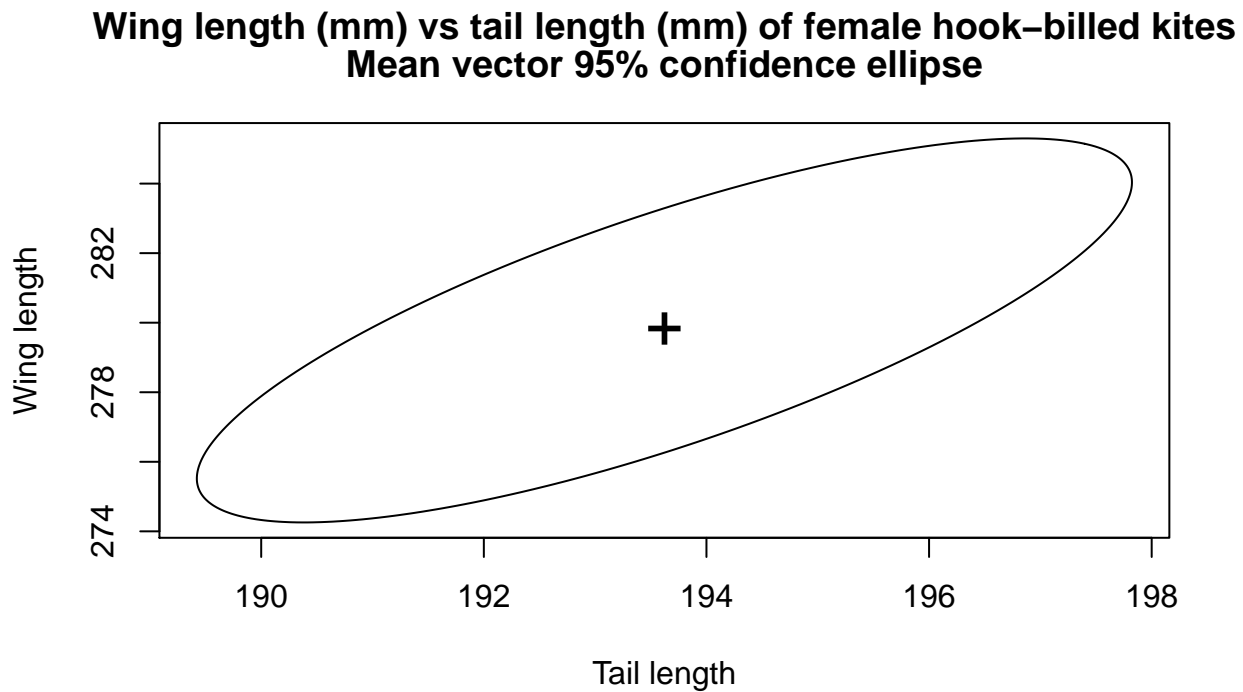
We calculate the chi-square test statistic for seven degrees of freedom at 0.1% significance level and compare them to the squared Mahalanobis distances of countries from the sample mean vector in order to determine which countries can be considered outliers.

```
## [1] "KOR, N is an outlier according to chisq test at 0.1% significance level"
## [1] "PNG is an outlier according to chisq test at 0.1% significance level"
## [1] "SAM is an outlier according to chisq test at 0.1% significance level"
```

North Korea probably does not perform very extremely (i.e. different from the sample mean) in any one distance record, but their records must not exhibit the same covariance with each other as that of other nations. One interpretation is that their training regimen produced vastly different results from that of most countries. Another interpretation could be that the records are fabricated by somebody without knowledge of running record covariances.

## Assignment 2

Based on the data given, we plot the 95% confidence ellipse for the tail length and wing length population means for female hook-billed kites.



According to wikipedia, “Most accipitrids exhibit sexual dimorphism in size, unusually for birds, it is the females that are larger than the males”.

We can easily see that our mean vector 95% confidence ellipse barely touches the population means for male hook-billed kites in both variables, so our data may reflect sexual dimorphism in hook-billed kites.

Let us calculate simultaneous confidence intervals and Bonferroni confidence intervals for the population means and compare them.

```
## [1] "Simultaneous confidence interval for Tail length: ( 189.42 , 197.82 )"
```

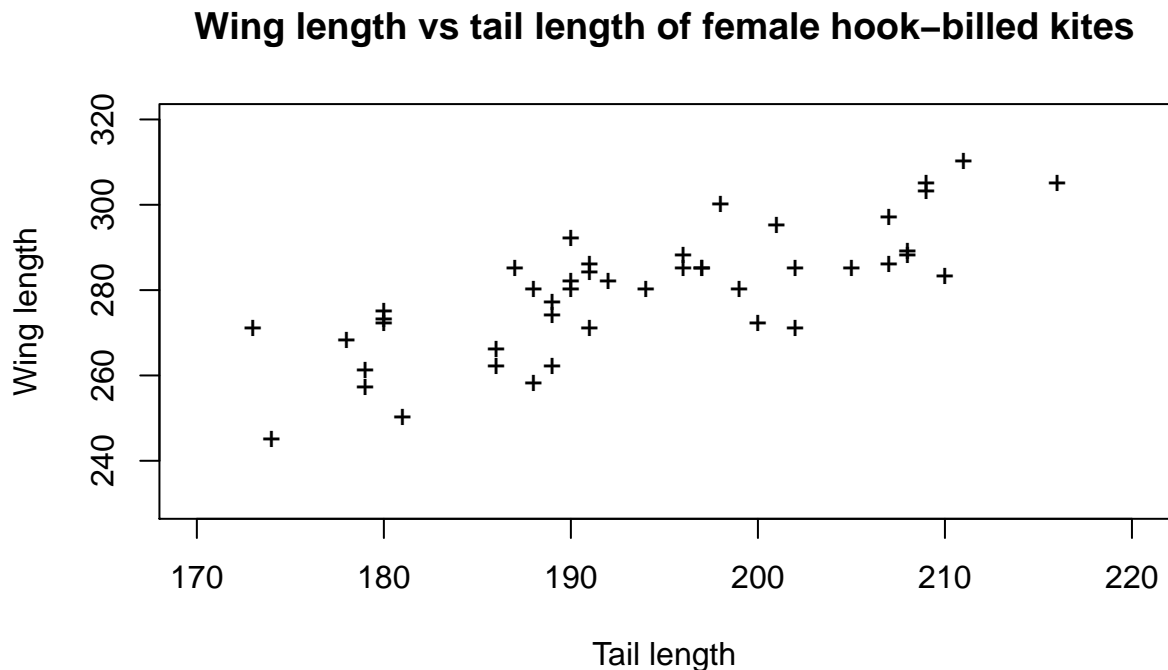
```
## [1] "Simultaneous confidence interval for Wing length: ( 274.26 , 285.3 )"
```

```
## [1] "Bonferroni confidence interval for Tail length: ( 189.82 , 197.42 )"
```

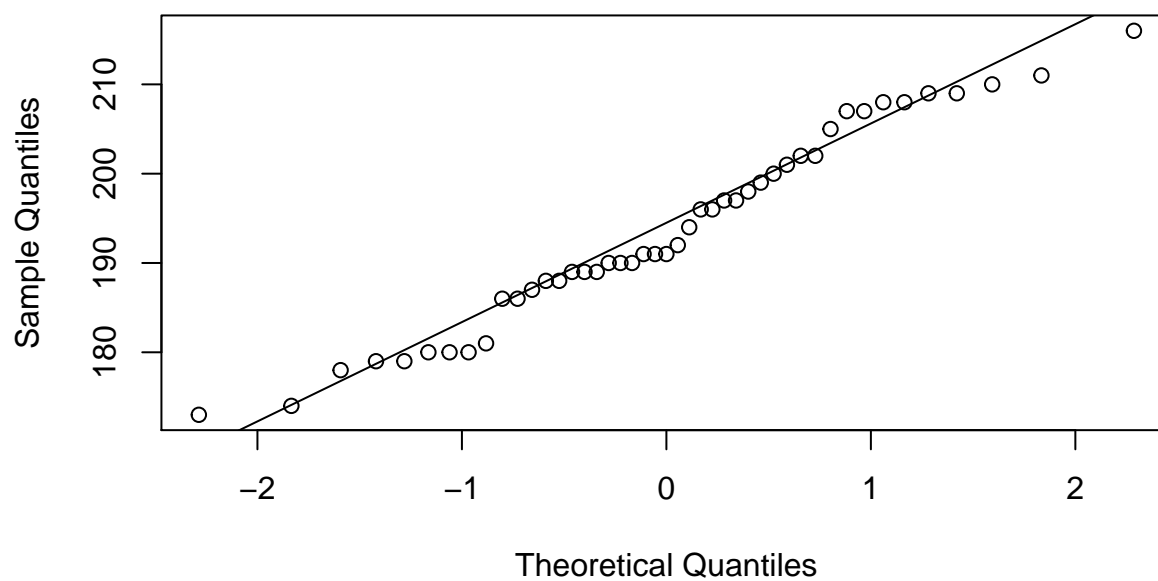
```
## [1] "Bonferroni confidence interval for Wing length: ( 274.78 , 284.77 )"
```

It looks as if the Bonferroni intervals are a little smaller than the simultaneous confidence intervals, and therefore more precise. The downside about using Bonferroni intervals is maybe that we lose the ability to easily make mean confidence intervals around the differences of mean vector components, like we could using simultaneous confidence intervals.

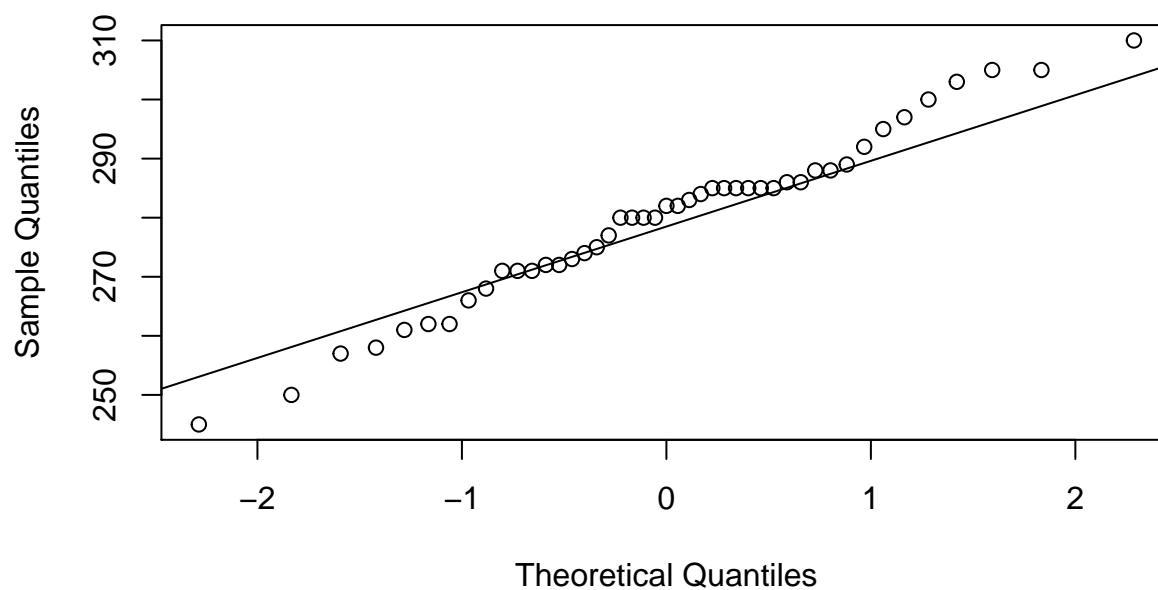
Now we consider whether the data can be described by a bivariate normal distribution by looking at the scatter-plot and Q-Q plots of the data variables. In the Q-Q plots the straight line indicate the theoretical quantile distribution of normally distributed data with the line passing through the first and third quartile sample data points.



**Q-Q Plot of tail length**



**Q-Q Plot of wing length**



The scatter-plot as well as the Q-Q plots suggest that this data is well described by a bivariate normal distribution, and thus it is a viable population model. One can easily imagine an oval shape in the scatter-plot encapsulating all the data points. The Q-Q plot indicate that the data variables are each approximately normally distributed.