

# Skewness Analysis of Petal Width (Versicolor)

Behrooz Filzadeh

## Introduction

Ich möchte in diesem Bericht die (Skewness) der Variable Petal.Width aus dem Datensatz iris mit der Bootstrap-Methode analysieren.

## a) Understanding Skewness

Skewness ist ein Maß dafür, ob die Verteilung von Daten symmetrisch ist oder nicht. Wenn Skewness positiv ist, dann hat die Verteilung einen langen Schwanz nach rechts (zu große Werte). Wenn Skewness negativ ist, dann geht der Schwanz nach links (zu kleine Werte). Wenn Skewness null ist, dann ist die Verteilung symmetrisch. Ich benutze die Funktion skew() aus dem Paket psych, weil sie einfach die Schiefe von den Daten berechnet. Das hilft mir zu verstehen, ob die Daten normal verteilt sind oder nicht. Wenn die Verteilung nicht symmetrisch ist, müssen wir vielleicht andere Methoden benutzen oder die Daten verändern.

## b) Bootstrap of Skewness

```
# Ich lade die Pakete, die ich brauche:  
# - "psych" für die Berechnung von Skewness (Schiefe)  
# - "ggplot2" für schöne Diagramme  
library(psych)  
library(ggplot2)
```

```
##  
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':  
##  
## %+%, alpha
```

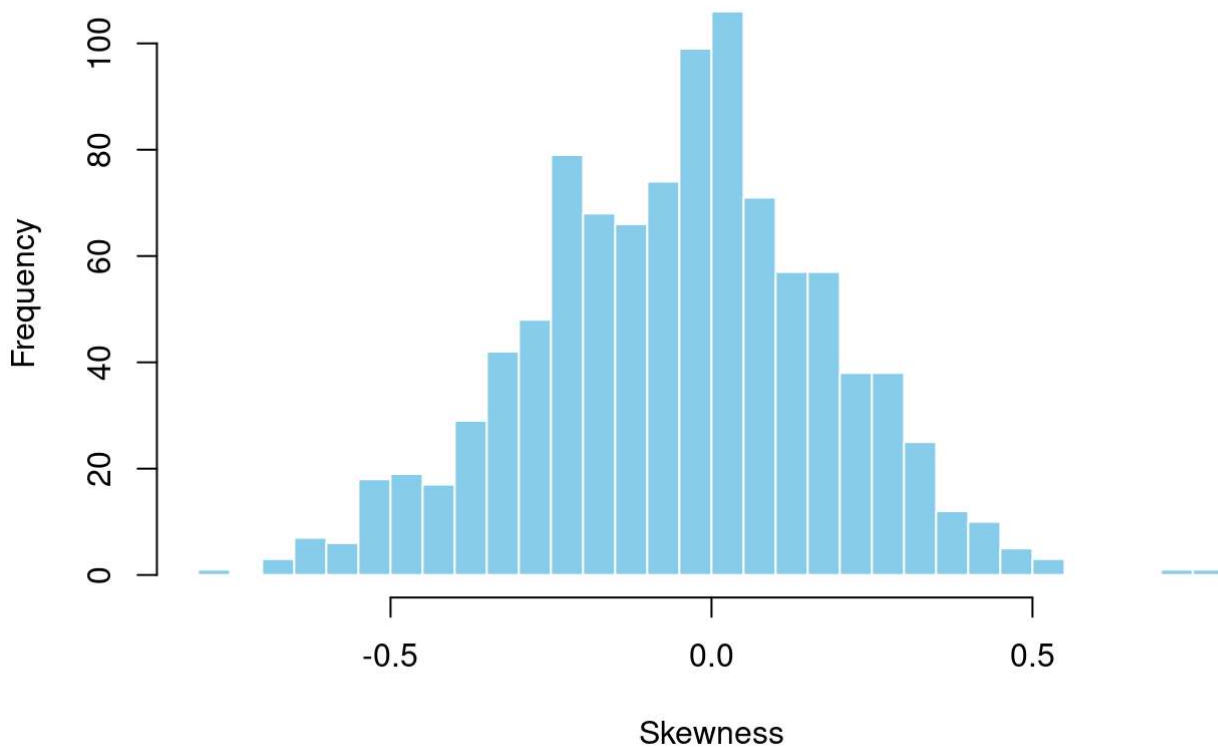
```
# Ich nehme nur die Daten von der Art "versicolor" aus dem iris-Datensatz.  
versicolor_data <- subset(iris, Species == "versicolor")  
  
# Ich wähle nur die Spalte "Petal.Width" (Blütenblatt-Breite).  
petal_width <- versicolor_data$Petal.Width  
  
# Ich berechne die ursprüngliche Skewness der Daten.  
# Skewness zeigt, ob die Werte mehr nach rechts oder links verteilt sind.  
original_skew <- skew(petal_width)  
original_skew # Das ist die ursprüngliche Schiefe
```

```
## [1] -0.02933377
```

```
# Jetzt mache ich eine Bootstrap-Analyse mit 1000 Wiederholungen.
# Das bedeutet: Ich ziehe 1000-mal Stichproben mit Zurücklegen.
# Für jede Stichprobe berechne ich die Skewness.
set.seed(123) # Damit bekomme ich jedes Mal die gleichen Ergebnisse (Zufallskontrolle)
B <- 1000 # Anzahl der Wiederholungen
bootstrap_skews <- replicate(B, {
  sample_data <- sample(petal_width, replace = TRUE) # Ich ziehe eine Stichprobe mit Zurücklegen
  skew(sample_data) # Ich berechne die Skewness dieser Stichprobe
})

# Ich zeichne ein Histogramm (Balkendiagramm) mit den 1000 Skewness-Werten.
# Das zeigt, wie die geschätzten Skewness-Werte verteilt sind.
hist(bootstrap_skews, breaks = 30,
     main = "Bootstrap-Verteilung der Schiefe",
     xlab = "Skewness", col = "skyblue", border = "white")
```

### Bootstrap-Verteilung der Schiefe



```
# Ich berechne die Standardabweichung der 1000 Bootstrap-Skewness-Werte.
# Das zeigt, wie stark die Werte schwanken.
sd_boot <- sd(bootstrap_skews)
sd_boot # Das ist die Streuung der Skewness
```

```
## [1] 0.2295927
```

```
# Ich berechne das 95%-Konfidenzintervall.  
# Das zeigt den Bereich, in dem die „wahre“ Skewness mit 95% Sicherheit liegt.  
ci <- quantile(bootstrap_skews, c(0.025, 0.975))  
ci
```

```
##          2.5%          97.5%  
## -0.5248518  0.3776921
```

## c) Interpretation

The estimated skewness is:

```
original_skew
```

```
## [1] -0.02933377
```

The bootstrap standard deviation is:

```
sd_boot
```

```
## [1] 0.2295927
```

The 95% confidence interval for the skewness is:

```
ci
```

```
##          2.5%          97.5%  
## -0.5248518  0.3776921
```

Wenn die Skewness positiv ist, bedeutet das, dass die Verteilung leicht nach rechts verschoben ist.

Wenn das 95%-Konfidenzintervall 0 einschließt, könnte die Verteilung symmetrisch sein.

Das Histogramm hilft uns, visuell zu sehen, ob die Verteilung schief ist oder nicht.

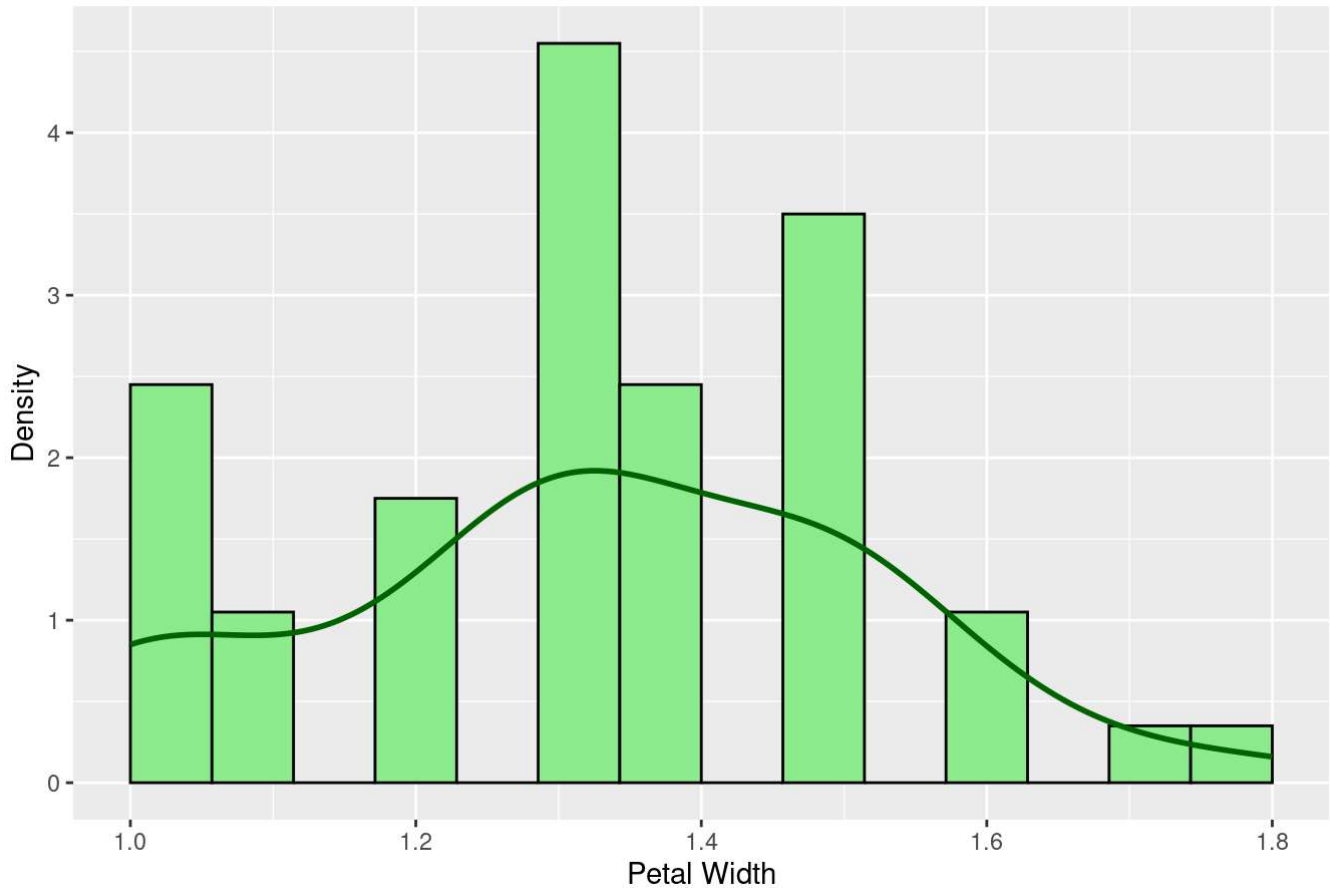
Lass uns die Variable noch einmal visualisieren:

```
# Density plot  
ggplot(versicolor_data, aes(x = Petal.Width)) +  
  geom_histogram(aes(y = ..density..), bins = 15, fill = "lightgreen", color = "black") +  
  geom_density(color = "darkgreen", size = 1) +  
  labs(title = "Distribution of Petal Width for Versicolor",  
       x = "Petal Width", y = "Density")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## i Please use `linewidth` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was  
## generated.
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.  
## i Please use `after_stat(density)` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was  
## generated.
```

Distribution of Petal Width for Versicolor



## Bonus: Interpretation des Konfidenzintervalls

Das Konfidenzintervall zeigt den Bereich von Skewness-Werten, die ich aufgrund der Stichprobenvariation erwarten kann.

Wenn das Intervall eng ist und null ausschließt, kann ich mit hoher Sicherheit sagen, dass die Verteilung schief ist.

Wenn das Intervall null enthält, gibt es nicht genug Beweise, um eine Schiefe zu bestätigen.