Laboration 2: Statistisk hypotesprövning

Sebastian Babic

2024-12-30

Contents

Sammanfattning

För att göra linjär regression använder vi oss av funktionen `lm()` i R. Vi kan använda oss av `summary()` för att få en sammanfattning av modellen. För att göra en hypotesprövning använder vi oss av `anova()` och `summary()`. Vi kan också använda oss av `confint()` för att få konfidensintervall för parametrarna.

I `lm()` så skriver vi in vår modell som `lm(y ~ 1 + x, data = data)`. Där `y` är den beroende variabeln och `x` är den oberoende variabeln.

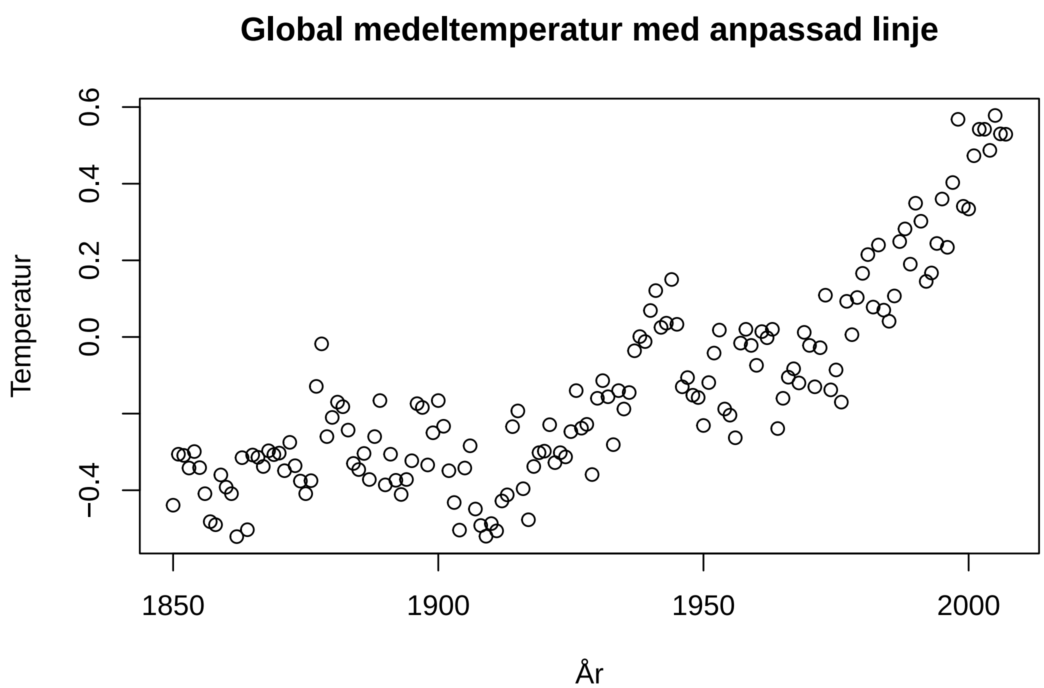
Använd `summary()` för att få fram värde.

För att få residueerna så kan vi använda oss av `residuals <- modell$residuals` och sedan plotta residuelaplotten med `plot(df$age, residuals)` ofta tsm med `abline(a = 0, b= 0, lty = "dotted"`.

Uppgift 1

```
df <- read.csv("temperatur.csv", header = TRUE)
```

```
temperature <- df$temperatur
age <- df$år

plot(age, temperature, xlab = "År", ylab = "Temperatur", main = "Global medeltemperatur med anpas
```



Vi ser att det absolut inte följer en linjär trend på grund av den exponentiala utseendet i datapunkterna. Vi ser även att datapunkterna har en stor spridning som tyder på att det inte är en linjär trend.

För att vi ska kunna göra en linjär regression så har vi kraven: 1. **Linjäritet**: Sambandet mellan variablerna bör vara linjärt.

2. Residualerna bör ha **konstant varians**.
3. **Oberoende observationer**: Inga observationer bör påverka varandra.
4. **Normalfördelade residualer**: Residualerna bör följa en normalfördelning.

Vi genomför linjär regression för att se om det finns något samband mellan temperatur och år och sedan ser om det är rimligt med ett linjär regression genom att utföra residualanalys och analys via normalfördelningsplot.

```r
old_par <- par(mfrow = c(1, 3))
modell <- lm(temperature ~ 1 + age, data = df) # vill ha intercept så + 1

# scatterplot med linje
plot(age, temperature, xlab = "År", ylab = "Temperatur", main = "Global medeltemperatur")
abline(modell, col = "red")

# residual plot
residual <- modell$residuals
plot(age, residual, xlab = "År", ylab = "Residualer", main = "Residualplot")
abline(a = 0, b = 0, lty = "dotted")

# normalfördelningsplot
qqnorm(residual, main = "Normalfördelningsplot residualer")
qqline(residual)
```

```r
par(old_par)
```

Vi ser redan i första plotten till vänster att vi verkar inte följa en linjär trend. Vi ser även i residualplotten att det inte finns någon konstant värde eftersom vi verkar inte följa något mönster men ser att spridningen verkar öka mot slutet av plotten. Residualerna verkar följa en normalfördelning förutom vid dem extrema värden i svansarna. Vi har därmed inte uppfyllt kraven på linjär regression och vi kan inte dra några slutsatser från denna modell.

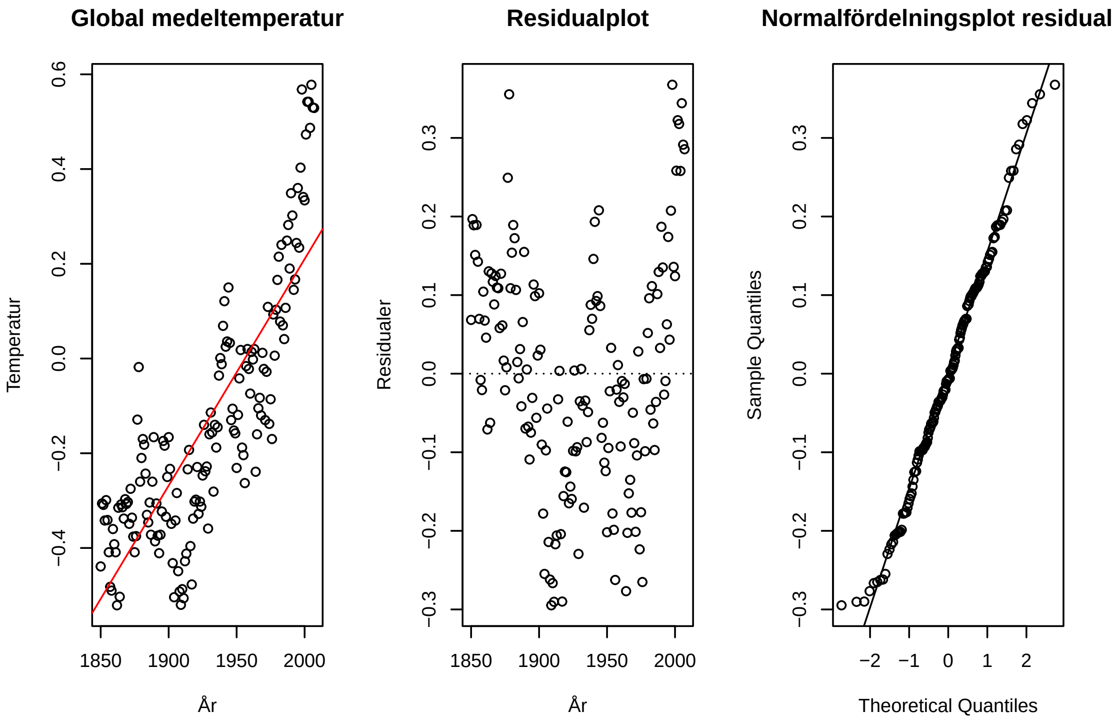
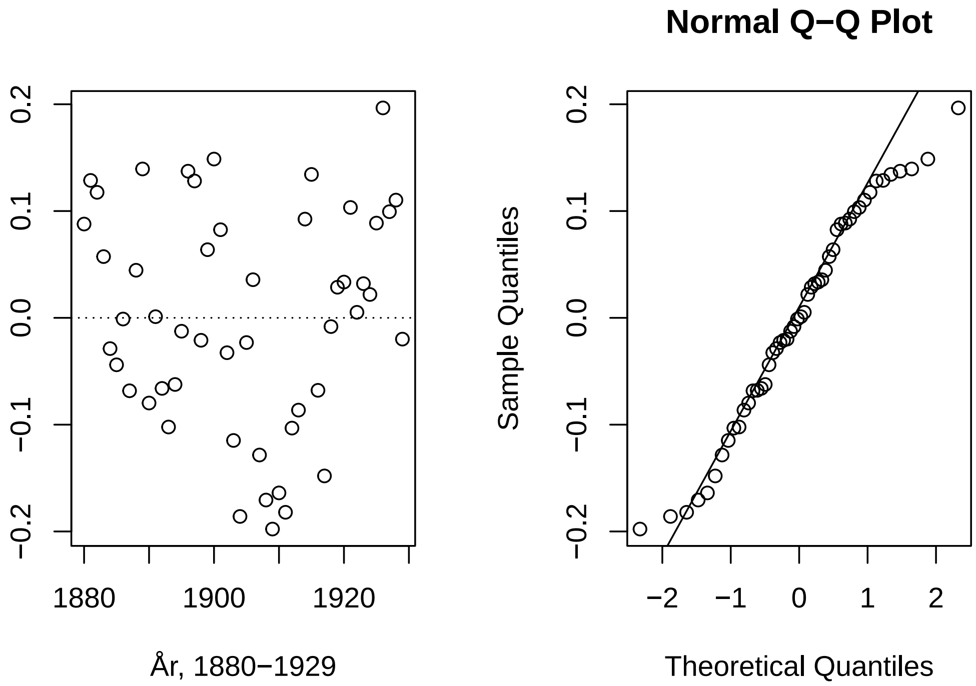


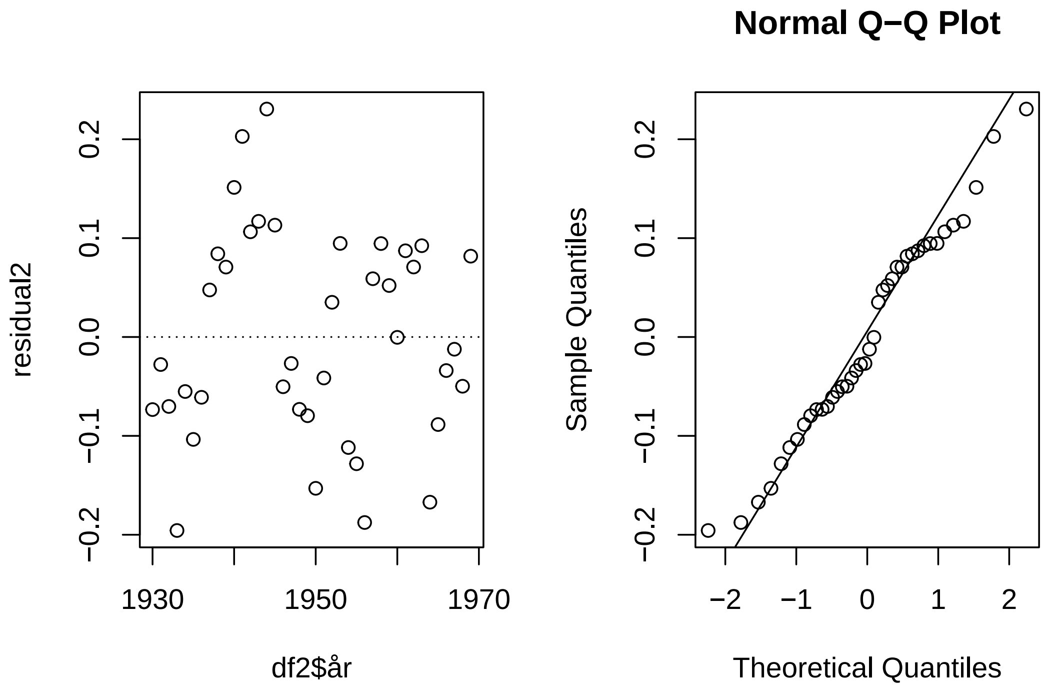
Figure 1: Linjär regression, residualanalys och QQ-plot analys för global temperaturdata

Uppgift 2

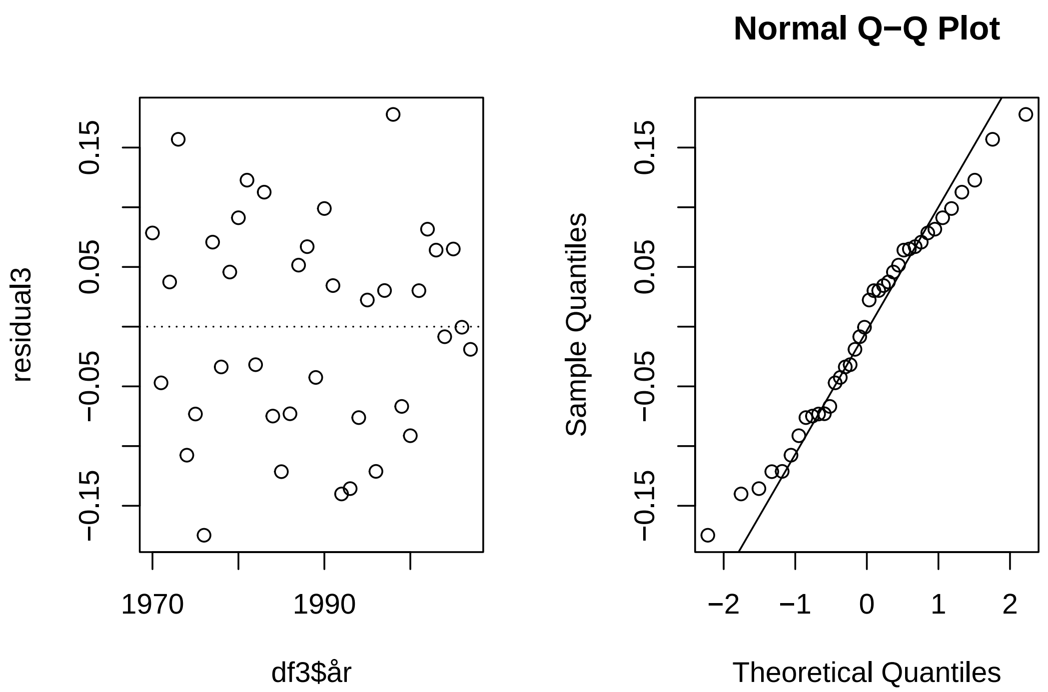
```
# delar in i tre årsperioder
df1 <- subset(df, år >= 1880 & år <= 1929)
df2 <- subset(df, år >= 1930 & år <= 1969)
df3 <- subset(df, år >= 1970 & år <= 2007)

# utför linjär regression för varje årsperiod för att se om kraven uppfylls bättre
modell1 <- lm(temperatur ~ år, data = df1)
modell2 <- lm(temperatur ~ år, data = df2)
modell3 <- lm(temperatur ~ år, data = df3)

old_par <- par(mfrow = c(1, 2))
# summering av modellerna
residual1 <- modell1$residuals
plot(df1$år, residual1, xlab = "År, 1880-1929", ylab ="")
abline(a = 0, b = 0, lty = "dotted")
qqnorm(residual1)
qqline(residual1)
```



```
old_par <- par(mfrow = c(1, 2))
residual2 <- modell2$residuals
plot(df2$år, residual2)
abline(a = 0, b = 0, lty = "dotted")
qqnorm(residual2)
qqline(residual2)
```

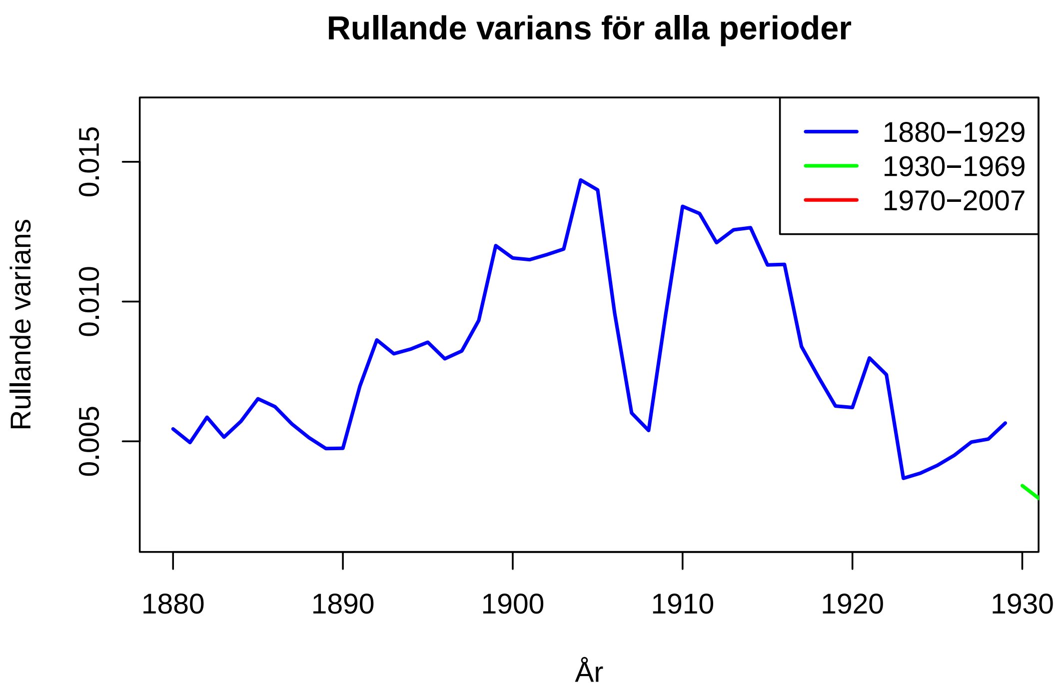



```
old_par <- par(mfrow = c(1, 2))
residual3 <- modell3$residuals
plot(df3$år, residual3)
abline(a = 0, b = 0, lty = "dotted")
qqnorm(residual3)
qqline(residual3)
```




```
# Funktion för att beräkna rullande varians
calculate_rolling_variance <- function(residuals, år, window_size = 10) {
  rolling_var <- numeric(length(år))
  half_window <- floor(window_size / 2)

  for (i in seq_along(år)) {
    # Definiera fönstergränser
    lower <- max(1, i - half_window)
    upper <- min(length(år), i + half_window)

    # Beräkna varians inom fönstret
    rolling_var[i] <- var(residuals[lower:upper], na.rm = TRUE)
  }
  return(rolling_var)
}

# Residualer och år för alla tre modeller
residual1 <- modell1$residuals
residual2 <- modell2$residuals
residual3 <- modell3$residuals
```

```r
år1 <- df1$år
år2 <- df2$år
år3 <- df3$år

# Beräkna rullande varians
rolling_var1 <- calculate_rolling_variance(residual1, år1)
rolling_var2 <- calculate_rolling_variance(residual2, år2)
rolling_var3 <- calculate_rolling_variance(residual3, år3)

# Plotta alla perioder i en enda graf
plot(år1, rolling_var1, type = "l", col = "blue", lwd = 2,
     ylim = range(c(rolling_var1, rolling_var2, rolling_var3)),
     main = "Rullande varians för alla perioder", xlab = "År", ylab = "Rullande varians")
lines(år2, rolling_var2, col = "green", lwd = 2)
lines(år3, rolling_var3, col = "red", lwd = 2)

# Lägg till en legend
legend("topright", legend = c("1880-1929", "1930-1969", "1970-2007"),
       col = c("blue", "green", "red"), lwd = 2)
```



Där vi använder `summary` för att få fram interceptet och lutningskoefficienten för varje modell.

```r
summary1 <- summary(modell1)
summary2 <- summary(modell2)
summary3 <- summary(modell3)

# extrahera alpha
coef1 <- coef(modell1)
coef2 <- coef(modell2)
coef3 <- coef(modell3)

# extrahera beta
intercept1 <- coef1[1]
slope1 <- coef1[2]

intercept2 <- coef2[1]
slope2 <- coef2[2]

intercept3 <- coef3[1]
slope3 <- coef3[2]


cat("Model 1: Intercept =", intercept1, ", Lutningskoefficient =", slope1, "\n")
```

```
## Model 1: Intercept = 1.284978 , Lutningskoefficient = -0.0008419208
```

```r
cat("Model 2: Intercept =", intercept2, ", Lutningskoefficient =", slope2, "\n")
```

```
## Model 2: Intercept = -0.9199713 , Lutningskoefficient = 0.0004318011
```

```r
cat("Model 3: Intercept =", intercept3, ", Lutningskoefficient =", slope3, "\n")
```

```
## Model 3: Intercept = -34.62775 , Lutningskoefficient = 0.01752653
```

Vi vill nu göra en hypotesprövning för att se om det finns en trend mot varmare klimat ($\beta > 0$) med en ensidig hypotesprövning för perioden 1970-2007, dvs modell 3. Vi använder oss av ett t-test för β för att se om det finns en signifikant trend.

Vi har alltså nollhypotesen $H_0 : \beta \leq 0$ och alternativhypotesen $H_1 : \beta > 0$. Vi använder oss av `summary()` för att få fram p-värdet.

```r
p_value <- summary(modell3)$coefficients[2, 4] / 2
if (p_value < 0.05) {
  print("Vi förkastar nollhypotesen. Direktören har fel.")
} else {
  print("Vi kan inte förkasta nollhypotesen. Direktören kan ha rätt.")
}
```

```
## [1] "Vi förkastar nollhypotesen. Direktören har fel."
```

Uppgift 3

Uppgift 3.1

Uppgift 3.2