# Laboration 2: Statistisk hypotesprövning

Sebastian Babic och Svedlund

2024-12-05

## Contents


**1 Sammanfattning** **1**

**2 Uppgift 1** **1**

2.1 Uppgift 1.1: Teoretisk uppgift . . . . . . . . . . . . . . . . . . . . . . . 2

2.2 Uppgift 1.2: Praktisk uppgift . . . . . . . . . . . . . . . . . . . . . . . 3

**3 Uppgift 2** **6**


## 1 Sammanfattning

I detta inledande stycke ska du förklara vad labben går ut på, utan att gå in på detaljer.

## 2 Uppgift 1

Varje sommar delades in i ett antal mindre tvådagarsperioder -> under varje sådan valdes en av dagarna slumpmässigt -> molnsådd på den dagen, inget på andra dagen

Om dag 1 molnsåddes så varade arbetet kl 12-14. Mätte sedan mha 29 mätstationer.

## 2.1 Uppgift 1.1: Teoretisk uppgift

Anledningen till att dela in sommaren i tvådagarsperioder är att säkerställa en rättvis jämförelse mellan dagar med molnsådd och utan molnsådd under liknande miljöförhållanden. Genom att jämföra dagar som ligger nära varandra i tid minskar vi effekten av andra faktorer, såsom långsiktiga väderförändringar, som skulle kunna påverka nederbörden.

Om vi i stället singlade slant varje dag skulle fördelningen mellan dagar med och utan molnsådd kunna bli obalanserad på grund av slumpens variation, särskilt med ett begränsat antal dagar (till exempel de 92 dagarna i en meteorologisk sommar). Detta skulle kunna resultera i en sned fördelning som inte är representativ för en jämn 50/50-fördelning, vilket gör det svårare att dra slutsatser från resultaten.

Slumpmässigheten i att singla slant på dem två dagar säkerställer att valet av dag för molnsådd inte påverkas av några systematiska faktorer, exempelvis om den första dagen i en period tenderar att ha annorlunda väderförhållanden än den andra. Detta eliminerar en potentiell källa till bias och garanterar att resultaten endast reflekterar effekten av molnsådd, inte en skillnad i grundläggande förhållanden mellan de två dagarna.

Ett lämpligt test för att undersöka om molnsåddning verkligen har ett effekt på regnmängd är Wilcoxons teckenrangtest ty dem två dagarna i tvådagarsperioderna är genom rimlig antagande beroende ty miljöförhållanden förändras om en av dagar såddes. Ännu en anledning till att välja Wilcoxons teckenrangtest är att datan kan antas inte vara normalfördelade

## 2.2 Uppgift 1.2: Praktisk uppgift

```r
arizona <- read.csv("arizona.csv", header = FALSE) # läs csv filen

# skapa variabler
year <- arizona$V1
seed <- arizona$V2
nonseed <- arizona$V3
```

```r
seed_mean <- mean(seed)
seed_sd <- sd(seed)
nonseed_mean <- mean(nonseed)
nonseed_sd <- sd(nonseed)
```

```r
old_par <- par(mfrow = c(2,3))

# sådda
hist(seed, prob = TRUE, main = "Histogram - Sådda dagar",
     xlab = "Regnmängd (inches)", ylab = "Densitet")
curve(dnorm(x, mean = seed_mean, sd = seed_sd),
      col = "red", lwd = 2, add = TRUE) # Normalfördelning
curve(dexp(x, rate = 1 / seed_mean),
      col = "blue", lwd = 2, add = TRUE,) # Exponentiell fördelning
legend("topright", legend = c("Normal", "Exponentiell"),
       col = c("red", "blue"), lty = c(1, 2), cex = 0.8)
```

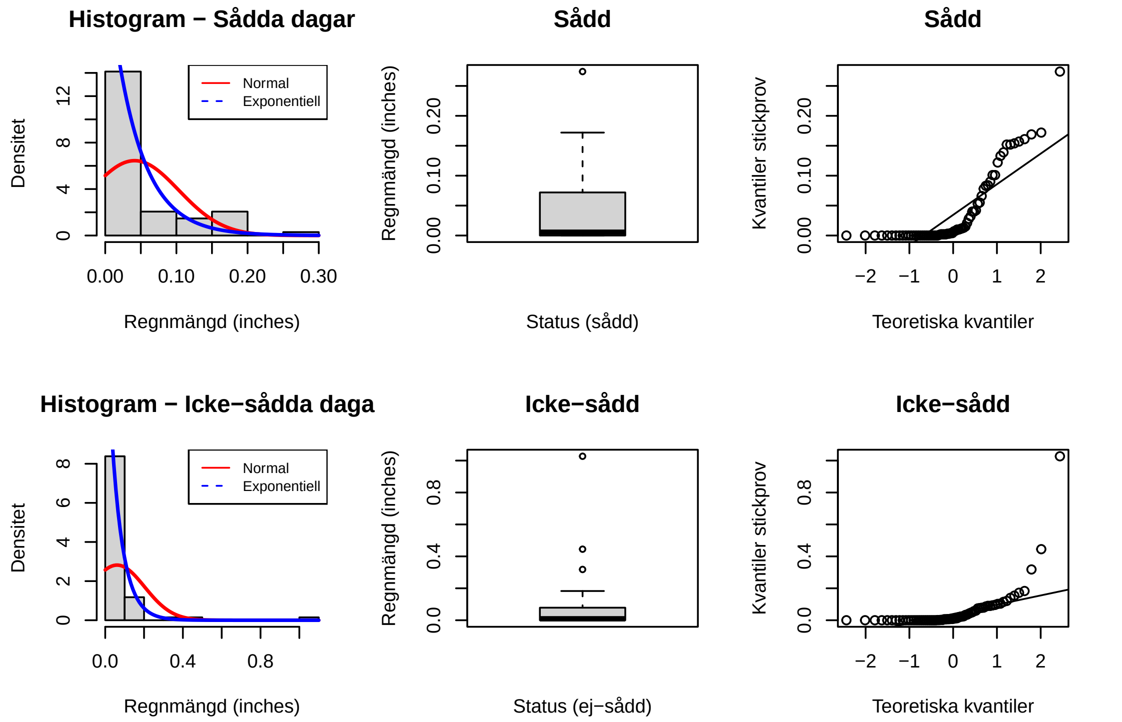
```r
boxplot(seed, main = "Sådd", xlab = "Status (sådd)", ylab ="Regnmängd (inches)")


qqnorm(seed, main = "Sådd", xlab = "Teoretiska kvantiler", ylab ="Kvantiler stickp
qqline(seed, col = "black")


# icke sådda
hist(nonseed, prob = TRUE, main = "Histogram - Icke-sådda dagar",
     xlab = "Regnmängd (inches)", ylab = "Densitet")
curve(dnorm(x, mean = nonseed_mean, sd = nonseed_sd),
      col = "red", lwd = 2, add = TRUE) # Normalfördelning
curve(dexp(x, rate = 1 / nonseed_mean),
      col = "blue", lwd = 2, add = TRUE) # Exponentiell fördelning
legend("topright", legend = c("Normal", "Exponentiell"),
       col = c("red", "blue"), lty = c(1, 2), cex = 0.8)


boxplot(nonseed, main = "Icke-sådd", xlab = "Status (ej-sådd)", ylab ="Regnmängd (


qqnorm(nonseed, main = "Icke-sådd", xlab = "Teoretiska kvantiler", ylab ="Kvantile
qqline(nonseed, col = "black")
```




```r
par(old_par)
```

Kan se en exponentialfördelning hos både sådda och ej sådda dagar och regnmängden på respektive dagar.

Där nollhypotesen $H_0$ är här att det inte finns en skillnad i genomsnittlig regnmängd mellan dagar med och utan molnsådd. Det vill säga att $\Delta = \mu_1 - \mu_0$ där $\mu_1$ är en sådd dag, $\mu_0$ ej sådd dag. Alltså är alternativa hypotesen $H_1$ att det finns en skillnad. Det vill säga

$$\begin{cases} H_0 : \mu_\Delta = 0 \\ H_1 : \mu_\Delta \neq 0 \end{cases} \tag{1}$$

```r
# Wilcoxon teckenrangtest
wilcoxon_test <- wilcox.test(seed, nonseed,
                             alternative = "two.sided",
                             paired = TRUE,
                             conf.level = 0.95)

# Utskrift av resultat
print(wilcoxon_test)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  seed and nonseed
## V = 691, p-value = 0.5107
## alternative hypothesis: true location shift is not equal to 0
```

Vi använder alltså t-testet för två parvist beroende stickprov med felrisk $\alpha = 0.05$. Vi förkastar nollhypotesen om p-värdet är större än felrisken. Vi ser här att enligt beräkningen ovan så är $0.3024 =$`p-value`$> \alpha = 0.05$. Vi kan därmed säga att det inte finns någon statistikt signifikant skillnad i genomsnittlig regnmängd.

# 3 Uppgift 2

Mer text. Använd underrubriker för delfrågor om du tycker det är motiverat.

Et cetera om det är fler än 2 uppgifter.