

# Statistik för Biologer

## F7: Ickeparametriska metoder

Shaobo Jin

Matematiska institutionen

# Parametriska och ickeparametriska metoder

Test och konfidensintervall som bygger på att vi antar att data följer en viss fördelning (t.ex. normalfördelningen) kallas **parametriska**.

- t-testet `t.test()`
- `cor.test(, method = "pearson")`

Test och konfidensintervall som inte bygger på ett sådant antagande kallas **ickeparametriska**.

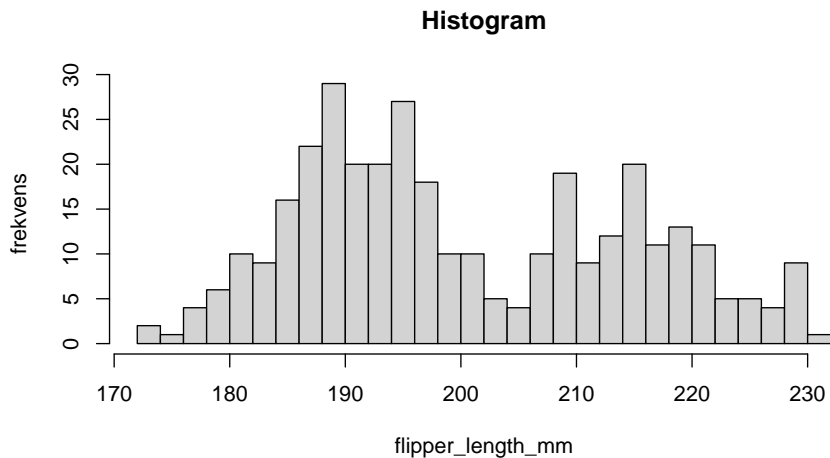
- Teckentestet från Föreläsning 3.

# t-Test Är Parametriska

t-Testet fungerar dåligt om:

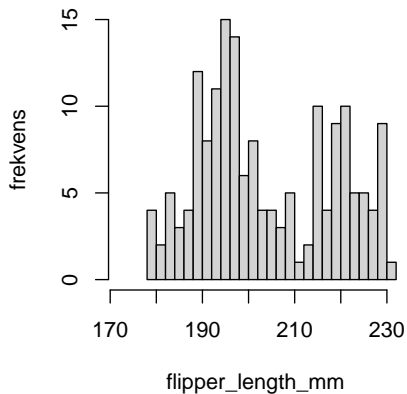
- ① Data inte är normalfördelade
  - Extra stort problem om vi inte har jämna gruppstorlekar.
- ② Vi har outliers.
- ③ Medelvärdet inte är “rätt” mått
  - För snedfördelningar kan medelvärdet ge en dålig beskrivning av vad som är “genomsnittligt”.

# Ensticksprovs-t-test: Normalfördelad?

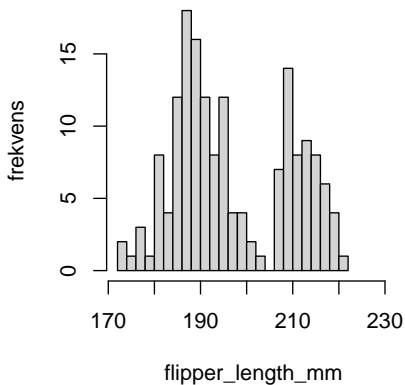


# Tvåsticksprovs-t-test: Normalfördelad?

## Histogram of males



## Histogram of females



# Wilcoxon-Mann-Whitney-testet

Våra data är

- ① Grupp 1:  $x_1, x_2, \dots, x_n$  av slumpvariabel  $X$
- ② Grupp 2:  $y_1, y_2, \dots, y_m$  av slumpvariabel  $Y$

Det tvåsticksprovs t-testet jämförar väntevärden av två grupper.

Hypoteser av **Wilcoxon-Mann-Whitney-testet** är

$$H_0 : P(X > Y) = P(X < Y)$$

$$H_1 : P(X > Y) \neq P(X < Y)$$

Ibland kallas testet **Wilcoxons rangsummetest** eller **Mann-Whitney test**.

# Wilcoxon-Mann-Whitney-testet

- Idé: rangordna observationerna!
- Våra data är  $x_1, x_2, \dots, x_n$  (Grupp 1) och  $y_1, y_2, \dots, y_m$  (Grupp 2). Totalt  $n + m$  observationer.
- För att utföra testet
  - 1 Sortera  $n + m$  observationer i stigande storleksordning
  - 2 Den minsta observationen får rang 1, den näst minsta rang 2,  $\dots$ , den största rang  $n + m$ .

Låt  $W_x$  vara summan av rangerna för  $x$ -observationer och  $W_y$  motsvarande summa för  $y$ -observationerna

- Om  $H_0$  stämmer:  $W_x$  och  $W_y$  borde vara ungefär lika stora
- Om  $H_1$  stämmer:  $W_x$  och  $W_y$  borde inte vara ungefär lika stora
- Stora skillnader mellan  $W_x$  och  $W_y$  ger ett lågt p-värde och får oss att tro att  $H_1$  stämmer

## Tvåstickprovs Mann-Whitney-Test

```
wilcox.test(flipper_length_mm ~ sex, data = penguins)

##
## Wilcoxon rank sum test with continuity correction
##
## data: flipper_length_mm by sex
## W = 9547, p-value = 9.011e-07
## alternative hypothesis: true location shift is not equal to 0

Female <- subset(penguins, sex == "female")
Male <- subset(penguins, sex == "male")
wilcox.test(Female$flipper_length_mm, Male$flipper_length_mm)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Female$flipper_length_mm and Male$flipper_length_mm
## W = 9547, p-value = 9.011e-07
## alternative hypothesis: true location shift is not equal to 0
```



## Mer om Hypoteser: Vara Försiktig

Ibland säger man att

$H_0$  :  $X$  och  $Y$  har samma fördelning

$H_1$  :  $X$  och  $Y$  har inte samma fördelning

eller

$H_0$  :  $X$  och  $Y$  har samma median

$H_1$  :  $X$  och  $Y$  har inte samma median

Båda är fel. Åtminstone inte 100% korrekta! Även R är fel!

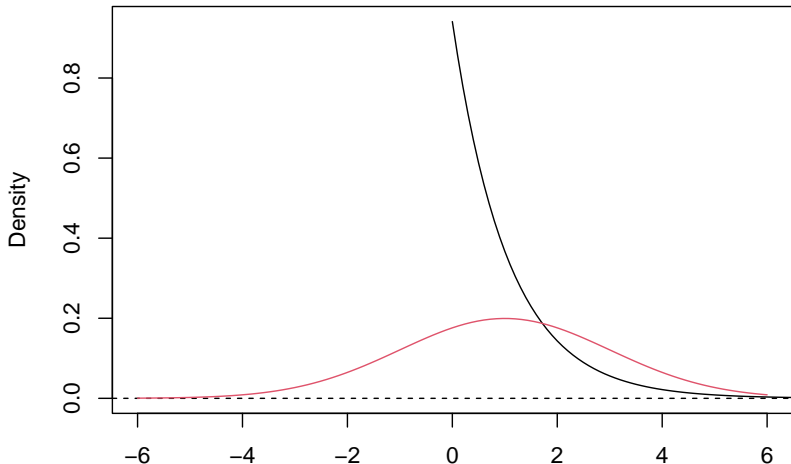
Om vi antar att

- 1 fördelningarna för  $X$  och  $Y$  har precis samma form
- 2 den eventuella skillnaden bara beror på en förskjutning

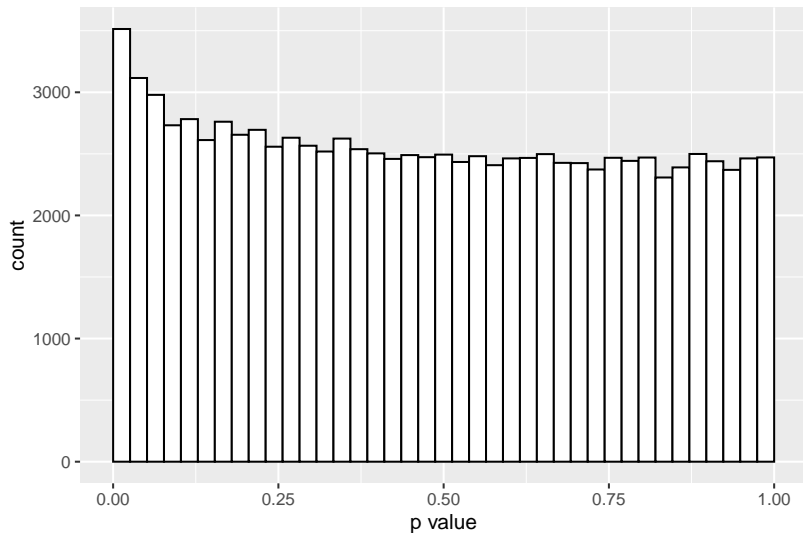
så förenklas hypoteserna till  $H_0$ : Samma median.

## Mer om Hypoteser: Vara Försiktig

Medianerna är 0.7367 (svart) och 1 (röd). Men  
 $P(X > Y) = P(X < Y) = 0.5$ .



# Mer om Hypoteser: Vara Försiktig



## Stickprov i par: Wilcoxon-Test

När vi har stickprov i par kan vi använda **Wilcoxons teckenrangtest**.

$H_0$  : Medianen i fördelningen av  $X - Y$  är 0

$H_1$  : Medianen i fördelningen av  $X - Y$  är inte 0

Med R:

```
wilcox.test(penguins$bill_length_mm, penguins$flipper_length_mm,  
            paired = TRUE)  
  
##  
## Wilcoxon signed rank test with continuity correction  
##  
## data:  penguins$bill_length_mm and penguins$flipper_length_mm  
## V = 0, p-value < 2.2e-16  
## alternative hypothesis: true location shift is not equal to 0
```

# Rangtest

- Påverkas inte av enstaka outliers
- Kräver inte normalfördelning
- Påverkas inte av monotona transformationer. Resultatet är detsamma för ursprungliga mätvärden och för logaritmerade mätvärden
- Ger inga konfidensintervall tyvärr

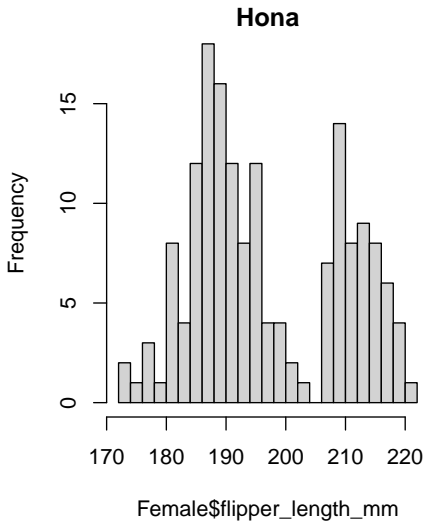
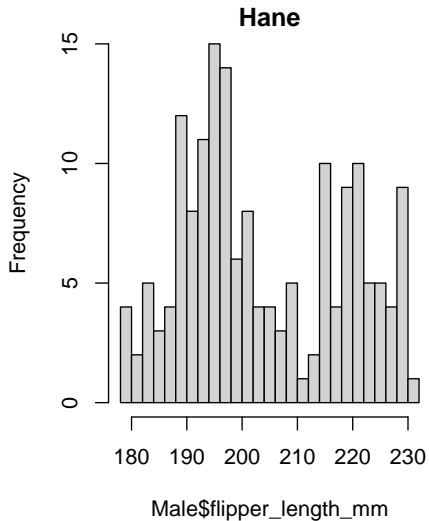
# Bootstrap

När t-testet fungerar dåligt är rangtest en möjlig lösning. Men vad ska vi göra om vi vill använda medelvärde och få konfidensintervall?

Vi kan använda **bootstrap** som kan kräva mycket datorkraft!

- 1 Dra  $n$  observationer från  $x_1, x_2, \dots, x_n$  med återläggning
- 2 Beräkna  $t$  för det nytt stickprovet (t-test)
- 3 Repetera Steg 1 och 2  $B$  gånger. Detta ger  $B$  stycken observationer  $t_1, t_2, \dots, t_B$  från fördelningen för  $t$
- 4 p-värdet är andelen  $t_i$  som är mer "extrema" än det  $t$  vi fick för våra ursprungliga data

# Normalfördelade?



## Vanligt tvåstickprovs t-test

```
# Vanligt t-test
t.test(Male$flipper_length_mm, Female$flipper_length_mm)

##
##  Welch Two Sample t-test
##
## data:  Male$flipper_length_mm and Female$flipper_length_mm
## t = 4.8079, df = 325.28, p-value = 2.336e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##    4.219821 10.064811
## sample estimates:
## mean of x mean of y
## 204.5060 197.3636
```



## R-kod för bootstrap-t-test (tvåstickprov)

```
# Bootstrap-t-test
library(MKinfer)
boot.t.test(Male$flipper_length_mm, Female$flipper_length_mm)

##
##  Bootstrap Welch Two Sample t-test
##
## data:  Male$flipper_length_mm and Female$flipper_length_mm
## bootstrap p-value < 2.2e-16
## bootstrap difference of means (SE) = 7.149401 (1.480879)
## 95 percent bootstrap percentile confidence interval:
##    4.231834 10.031861
##
## Results without bootstrap:
## t = 4.8079, df = 325.28, p-value = 2.336e-06
## alternative hypothesis: true difference in means is not equal to
## 95 percent confidence interval:
##    4.219821 10.064811
```

# Bootstrap

- Använder samma hypoteser som ett vanligt t-test
- Fungerar bättre än t-test om data inte är normalfördelade
- Fungerar bättre än t-test om vi inte har jämna gruppstorlekar
- Ger konfidensintervall

Men

- Är känsligt för outliers, men inte lika mycket som t-testet
- Kan ibland ta lång tid att köra, särskilt om man har mycket data

# Linjär regression: Modellförutsättningar

## ① Linearitet

- Om sambandet inte beskrivs tillräckligt bra av  $y = \alpha + \beta x$  så kommer modellen inte fungera särskilt bra.

## ② Lika varians

- Variationen i  $y$ -variabeln ska vara lika stora för alla värden på  $x$ . Lika varians krävs för konfidensintervall och p-värden!

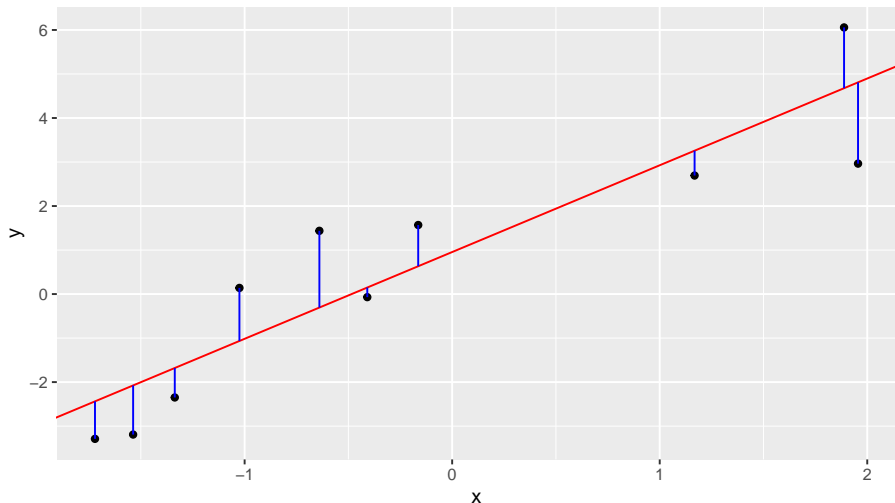
## ③ Normalfördelning

- Slumpavvikelserna  $\epsilon_i$  ska vara normalfördelade, vilket krävs för konfidensintervall och p-värden!

Bootstrap kan användas för att hantera problem 2 och 3!

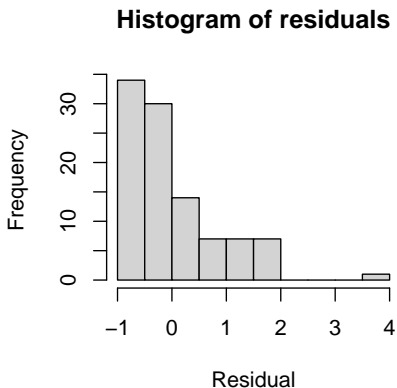
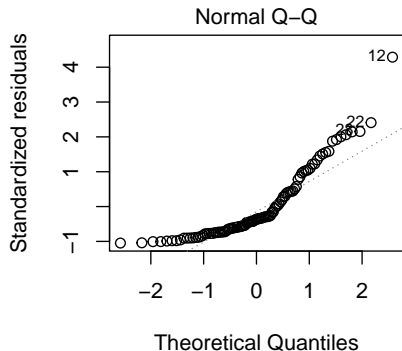
# Residualer

Modellens **residualer** är avstånden mellan datapunkterna och den anpassade linjen:



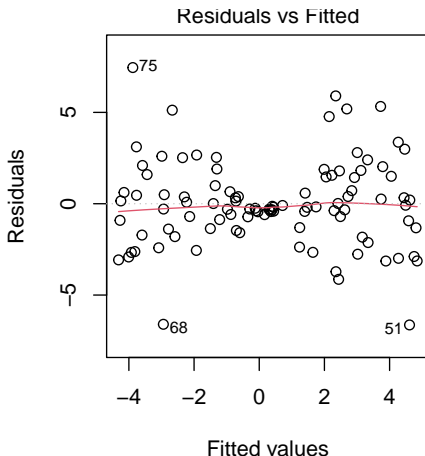
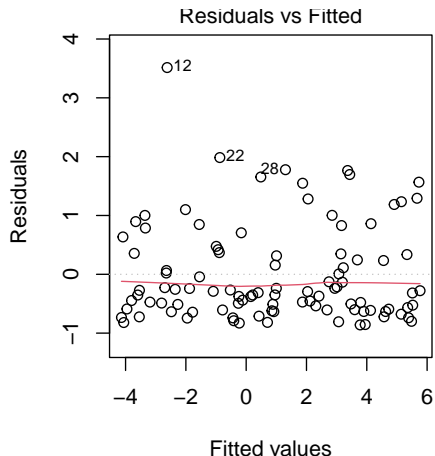
## Ingen normalfördelning

Om mätfeLEN  $\epsilon$  i modellen  $y = \alpha + \beta x + \epsilon$  inte är normalfördelade så blir heller inte residualerna det. Vi kan alltså utvärdera normalfördelningsantagandet genom att undersöka om residualerna är normalfördelade!



## Lika Varians?

Vi kan utvärdera antagandet genom att undersöka om residualerna beror på anpassat  $y$ !



## Konsekvenser

p-värden och konfidensintervall för  $\alpha$  och  $\beta$  kan beräknas med bootstrap utan att vi behöver normalfördelning och lika varians. Istället för `summary` använder vi `boot_summary` från paketet `boot.pval`:

```
gentoo <- subset(penguins, species == "Gentoo")
LM <- lm(bill_length_mm ~ bill_depth_mm, data = gentoo[-120,])
library(boot.pval)
boot_summary(LM)
```

##		Estimate	Lower.bound	Upper.bound	p.value
##	(Intercept)	17.229501	10.748219	23.976991	0
##	bill_depth_mm	2.020768	1.566116	2.456699	0

# method = "case" eller method = "residual"?

```
boot_summary {boot.pval}
```

R Documentation

## Summarising Regression Models Using the Bootstrap

### Description

Summaries for regression models, including "lm", "glm", "glm.nb", "nls", "rlm", "polr", and "merMod" ("lmer", "glmer") objects, using the bootstrap for p-values and confidence intervals.

### Usage

```
boot_summary(  
  model,  
  type = "perc",  
  method = NULL,  
  conf.level = 0.95,  
  R = 999,  
  coef = "raw",  
  pval_precision = NULL,  
  adjust.method = "none",  
  ...  
)
```

### Arguments

model	An object fitted using e.g. "lm", "glm", "glm.nb", "nls", "rlm", "polr", "lmer", or "glmer".
type	A vector of character strings representing the type of interval to base the test on. The value should be one of "norm", "basic", "stud", and "perc" (the default). "stud" is not supported for "lmer" and "glmer" models.
method	The method used for bootstrapping. For "lm" and "nls" objects use either "residual" (for resampling of scaled and centred residuals, the default) or "case" (for case resampling). For "glm" objects, use "case" (the default). For "merMod" objects (mixed models) use either "parametric" (the default) or "semiparametric".
conf.level	The confidence level for the confidence intervals. The default is 0.95.
R	The number of bootstrap replicates. The default is 999.



method = "case" eller method = "residual"?

method = "residual"

- kräver lika varians
- kräver att både  $y$  och  $\epsilon$  är slumpmässiga,
  - men  $x$  är inte slumpmassig p.g.a. forskningsplanen.
  - t.ex., vi ska jämföra koffeinhalt av Lindvalls Mörkrost och Lindvalls Brygg.

method = "case"

- kräver inte lika varians
- kräver att  $x$ ,  $y$  och  $\epsilon$  är slumpmässiga

# Sammanfattning

## ① Rangtest

- Påverkas inte av enstaka outliers
- Kräver inte normalfördelning
- Påverkas inte av monotona transformationer
- Wilcoxon-Mann-Whitney test
- Ger inga konfidensintervall

## ② Bootstrap

- Fungerar bättre än t-test om data inte är normalfördelade
- Fungerar bättre än t-test om vi inte har jämna gruppstorlekar
- Är känsligt för outliers, men inte lika mycket som t-testet
- Ger konfidensintervall
- Kan användas vid linjär regression
- Kan ibland ta lite tid att köra, särskilt om man har mycket data