

# Statistik för Biologer

## F4: t-Test och Konfidensintervall

Shaobo Jin

Matematiska institutionen

# Medelvärden

I många studier är vi intresserade av medelvärdet av en grupp:

- Vad är medelvärdet för vikten på pingvinhanar?
- Hur många personer per dag, i genomsnitt, smittades med covid-19 under förra veckan?

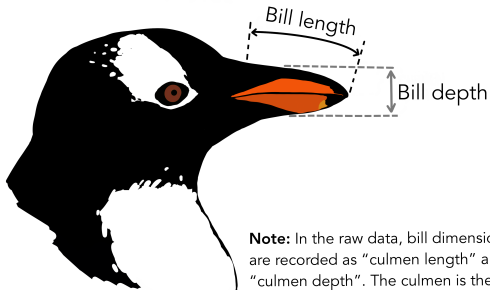
I många studier är vi också intresserade av medelvärden av två eller flera grupper:

- Hur stor är skillnaden i vikt mellan pingvinhonor och -hanar?

# Att Mäta Pingviner

Det finns tre arter av pingviner. Forskarna har mätt bland annat



- ① Näbbens längd (mm)
- ② Näbbens djup (mm)
- ③ Vingens längd (mm)
- ④ Vikt (g)
- ⑤ Kön (hona/hane)
- ⑥ Art (tre arter)



**Note:** In the raw data, bill dimensions are recorded as "culmen length" and "culmen depth". The culmen is the dorsal ridge atop the bill.

# Vikten av Gentoo

Encyclopedia of Life påstår att åsnepingviner i genomsnitt väger 5,9 kg.

 hur mycket väger en åsnepingvin 

Images Videos News Maps Books Flights Finance


All filters Tools SafeSei

---

About 1 090 results (0,68 seconds)

Gentoo penguin / Mass

5.9 kg



Gentoo penguin  
(Åsnepingvin)

Birds

The gentoo penguin is a penguin species in the genus *Pygoscelis*, most closely related to the Adélie penguin and the chinstrap penguin. The earliest scientific description was made in 1781 by Johann Reinhold Forster with a type locality in the Falkland Islands.


[Wikipedia](#)


**Scientific name:** *Pygoscelis papua*


**Conservation status:** Near Threatened (Population decreasing) [Encyclopedia of Life](#)

**Mass:** 5.9 kg [Encyclopedia of Life](#)

People also search for

 Chinstrap penguin  
3 – 5 kg

 Adélie penguin  
4.7 kg

 King penguin  
9.3 – 17 kg

Feedback · Sources include: Encyclopedia of Life

People also ask :

# Hypotesprövning

Vi vill undersöka om påståendet om åsnepingviners vikt stämmer med ett statistiskt hypotestest.

$H_0$  : den genomsnittliga vikten är 5.9 kg.

$H_1$  : den genomsnittliga vikten är inte 5.9 kg.

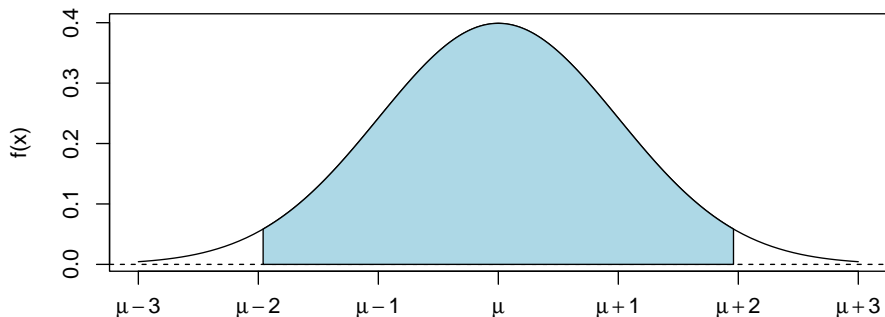
Om det är en stor skillnad mellan 5.9 kg och medelvikten i data tyder det på att  $H_1$  stämmer.

## En Idé

För normalfördelningen gäller att 95% av alla observationer hamnar inom två standardavvikelser från väntevärdet:

$$P(\mu - 1.96\sigma \leq X \leq \mu + 1.96\sigma) \approx 0.95$$

$$P\left(-1.96 \leq \frac{X - \mu}{\sigma} \leq 1.96\right) \approx 0.95.$$



# Centrala gränsvärdessatsen

Om  $X_1, X_2, \dots, X_n$  är:

- 1 oberoende slumpvariabler
- 2 som alla har samma fördelning, med  $E(X_i) = \mu$  och  $V(X_i) = \sigma^2$ ,  
så gäller att medelvärdet  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  är approximativt normalfördelat när  $n$  är tillräckligt stort.

Om  $n$  är tillräckligt stort så kommer alltså medelvärdet att ligga i spannet  $\mu \pm 2 \frac{\sigma}{\sqrt{n}}$  i 95% av alla studier!

# Differens

Vi kan ta ut åsnepingvinernas vikt från våra pingvindata:

```
library(palmerpenguins)
gentoo <- subset(penguins, species == "Gentoo")
```

Medelvärdet i våra data (na.rm tar bort saknat värden)

```
mean(gentoo$body_mass_g, na.rm = TRUE)

## [1] 5076.016
```

Skillnaden beror på mätskalan

```
mean(gentoo$body_mass_g, na.rm = TRUE) - 5900 # gram

## [1] -823.9837

mean(gentoo$body_mass_g/1000, na.rm = TRUE) - 5.9 # kg

## [1] -0.8239837
```



## Nytt mått på skillnad

$s/\sqrt{n}$  kallas för ett **medelfel**. Vi kan använda medelfelet  $s/\sqrt{n}$  för att skapa ett standardiserat mått på hur mycket medelvärdet avviker från  $\mu_0$ :

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

Beräkning med R:

```
# gram
xs <- mean(gentoo$body_mass_g, na.rm = TRUE)
s <- sd(gentoo$body_mass_g, na.rm = TRUE)
n <- sum(!is.na(gentoo$body_mass_g))
(xs - 5900) / (s / sqrt(n))

## [1] -18.12761
```

# Spelar Skalan Någon Roll?

```
# gram
xs <- mean(gentoo$body_mass_g, na.rm = TRUE)
s <- sd(gentoo$body_mass_g, na.rm = TRUE)
n <- sum(!is.na(gentoo$body_mass_g))
(xs - 5900) / (s / sqrt(n))

## [1] -18.12761
```

```
# kg
xs <- mean(gentoo$body_mass_g/1000, na.rm = TRUE)
s <- sd(gentoo$body_mass_g/1000, na.rm = TRUE)
(xs - 5.9) / (s / sqrt(n))

## [1] -18.12761
```

## Nytt mått på skillnad

- Måttet  $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$  mäter antalet standardfel som resultatet (medelvärde) avviker från nollhypotesen.
- Om nollhypotesen stämmer så kommer medelvärdet att ligga inom (approximativt) 2 standardfel från  $\mu_0$  i 95% av alla studier.
  - Beroende på hur stort  $n$  är så kommer medelvärdet i 95% av studierna ligga inom lite mer eller lite mindre än 2 standardfel.
- Här är avvikelserna mer än 18 standardfel!
- Vad är sannolikheten för en avvikelse som är minst så stor? (Det vill säga, vad är p-värdet?)

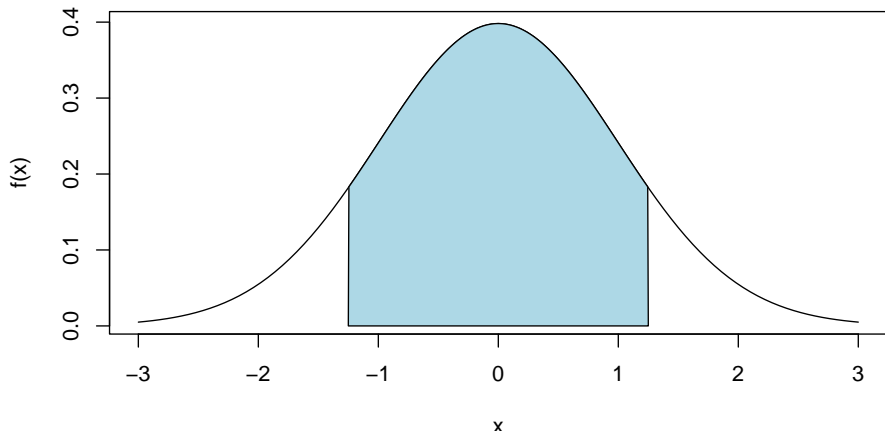
# Skål!



Guinness kemist W.S. Gosset brottades med det här problemet i början av 1900-talet. Han publicerade lösningen, som använder t-fördelningen, under pseudonymen “Student”. Testet kallas t-test eller Student’s t-test.

## Beräkning av p-värde

$H_0 : \mu = 5900\text{g}$  mot  $H_1 : \mu \neq 5900\text{g}$ . Ju mer den standardiserade differensen avviker från 0, desto extremare är resultatet. Om den standardiserade differensen t.ex är  $-1.25$  så är p-värdet = arean för de vita områdena under kurvan:



## t-test i R: dubbelsidig alternativhypotes

Mått i gram:  $H_0 : \mu = 5900\text{g}$  mot  $H_1 : \mu \neq 5900\text{g}$ .

```
t.test(gentoo$body_mass_g, mu = 5900)

##
##  One Sample t-test
##
## data:  gentoo$body_mass_g
## t = -18.128, df = 122, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 5900
## 95 percent confidence interval:
##  4986.034 5165.998
## sample estimates:
## mean of x
##  5076.016
```

## t-test i R: dubbelsidig alternativhypotes

Mått i kg:  $H_0 : \mu = 5.9\text{kg}$  mot  $H_1 : \mu \neq 5.9\text{kg}$ .

```
t.test(gentoo$body_mass_g / 1000, mu = 5.9)

##
##  One Sample t-test
##
## data:  gentoo$body_mass_g/1000
## t = -18.128, df = 122, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 5.9
## 95 percent confidence interval:
##  4.986034 5.165998
## sample estimates:
## mean of x
##  5.076016
```

## t-test i R: enkelsidig alternativhypotes

Mått i gram:  $H_0 : \mu \leq 5900\text{g}$  mot  $H_1 : \mu > 5900\text{g}$ .

```
t.test(gentoo$body_mass_g, mu = 5900,
       alternative = "greater")

##
##  One Sample t-test
##
## data:  gentoo$body_mass_g
## t = -18.128, df = 122, p-value = 1
## alternative hypothesis: true mean is greater than 5900
## 95 percent confidence interval:
##  5000.678      Inf
## sample estimates:
## mean of x
##  5076.016
```



## t-test i R: enkelsidig alternativhypotes

Mått i gram:  $H_0 : \mu \geq 5900\text{g}$  mot  $H_1 : \mu < 5900\text{g}$ .

```
t.test(gentoo$body_mass_g, mu = 5900,  
       alternative = "less")  
  
##  
## One Sample t-test  
##  
## data: gentoo$body_mass_g  
## t = -18.128, df = 122, p-value < 2.2e-16  
## alternative hypothesis: true mean is less than 5900  
## 95 percent confidence interval:  
##      -Inf 5151.355  
## sample estimates:  
## mean of x  
## 5076.016
```

# Konfidenstervall och test

- Ett konfidenstervall för väntevärdet täcker det sanna värdet på  $\mu$  i  $100(1 - \alpha)\%$  av alla studier.
- Hypotestest och konfidenstervall är olika sidor av samma mynt!
- Till varje test hör ett konfidenstervall, och vice versa.
- Ett  $1 - \alpha$  konfidenstervall för  $\mu$  innehåller de värden på  $\mu$  som inte skulle förkastas vid signifikansnivån  $\alpha$ .
- Ett värde  $\mu_0$  kan förkastas vid signifikansnivån  $\alpha$  om det inte ligger i  $1 - \alpha$  konfidenstervallet.

## Konfidensintervall för väntevärdet $\mu$

```
t.test(gentoo$body_mass_g, mu = 5900)

##
##  One Sample t-test
##
## data:  gentoo$body_mass_g
## t = -18.128, df = 122, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 5900
## 95 percent confidence interval:
##  4986.034 5165.998
## sample estimates:
## mean of x
##  5076.016
```

Detta brukar skrivas som ett intervall: (4986.304, 5165.998) eller  $4986.304 < \mu < 5165.998$ .

## Konfidsensintervall för väntevärdet $\mu$

Spelar  $\mu =$  i `t.test()` någon roll?

```
t.test(gentoo$body_mass_g, mu = 0)

##
##  One Sample t-test
##
## data:  gentoo$body_mass_g
## t = 111.67, df = 122, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  4986.034 5165.998
## sample estimates:
## mean of x
##  5076.016
```

# Konfidsensintervall

Konfidsensintervallet kan beräknas för hand som:

$$\bar{x} \pm t_{\alpha/2}^{(n-1)} \frac{s}{\sqrt{n}}$$

där  $t_{\alpha/2}^{(n-1)}$  anger inom hur många medelfel från  $\mu$  som medelvärdet hamnar i  $100(1 - \alpha)\%$  av alla studier.

$t_{\alpha/2}^{(n-1)}$  är den  $\alpha/2$ -kvantilen för t-fördelningen med  $n - 1$  frihetsgrader, dvs

$$P\left(T \geq t_{\alpha/2}^{(n-1)}\right) = \frac{\alpha}{2},$$

där  $T$  är t-fördelad med  $n - 1$  frihetsgrader

## Konfidsensintervall för väntevärdet $\mu$

Vi kan kontrollera konfidsensgraden  $1 - \alpha$ .

```
t.test(gentoo$body_mass_g, conf.level = 0.99) # alpha = 0.01

##
##  One Sample t-test
##
## data:  gentoo$body_mass_g
## t = 111.67, df = 122, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  4957.074 5194.959
## sample estimates:
## mean of x
##  5076.016
```

## t-kvantiler

Kvantilerna för ett 95% konfidenzintervall ( $\alpha = 0.05$ ,  $\alpha/2 = 0.025$ ) beror på stickprovsstorleken  $n$ :

Stickprovsstorlek $n$	$t_{0.025}^{(n-1)}$
5	2.776
10	2.262
20	2.093
50	2.01
100	1.984
1000	1.962
$\infty$	1.96

Om slumpvariabel  $X \sim N(0, 1)$  har vi

$$P(X \leq 1.96) = 0.975,$$

$$P(X \geq 1.96) = 0.025.$$

# Förutsättningar!

- ① Vi har inte outliers
- ② Normalfördelade data!
  - ① Resultaten som fås via t-fördelningen är **exakta** om data är normalfördelade, oavsett värdet på  $n$ .
  - ② Resultaten som fås via t-fördelningen är **approximativa** om data inte är normalfördelade men  $n$  är stort.
  - ③ Ni ska studera olika sätt att undersöka om data är normalfördelade i Datalab S3.
- ③  $X_1, X_2, \dots, X_n$  är oberoende slumpvariabler



# Strunta i Dem

- 1 Kursboken beskriver en metod för att utföra t-test som använder sig av olika tabeller. Ingen gör någonsin så nuförtiden! Använd R istället.
- 2 Boken nämner också **z-testet**. Använd t-testet istället.

# Forskningsfrågor om pingviner: skillnader

I många studier är vi också intresserade av skillnaderna mellan två eller flera grupper:

- Hur stor är skillnaden i vikt mellan pingvinhonor och -hanar?
- Finns storleksskillnader mellan könen för respektive art?
- Söker könen föda på samma sätt?

# Skilnader i vikt

- Vi har mätt hanarnas vikter:  $z_1, z_2, \dots, z_n$  och beräknat medelvärdet  $\bar{z} = \sum_{i=1}^n z_i/n$ .
- Vi har mätt honornas vikter:  $x_1, x_2, \dots, x_m$  och beräknat medelvärdet  $\bar{x} = \sum_{i=1}^m x_i/m$ .
- $n$  och  $m$  behöver inte vara samma!
- Differensen är  $\bar{z} - \bar{x}$ .

Hur kan vi få ett standardiserat mått på differens, som inte beror på vilken viktenhet vi använder?

## Standardiserade differens

När vi har två stickprov blir den standardiserade differensen:

$$t = \frac{\bar{z} - \bar{x}}{\sqrt{s_z^2/n + s_x^2/m}}$$

Om nollhypotesen  $H_0 : \mu_z = \mu_x$  är sann så är  $t$  t-fördelad, vilket gör att vi kan beräkna p-värdet.

## t-Test för två stickprov med R: variant 1

Om vi har det uppmätta värdet och grupptillhörighet som variabler i en tabell (data frame):

```
t.test(body_mass_g ~ sex, data = gentoo)

##
##  Welch Two Sample t-test
##
## data:  body_mass_g by sex
## t = -14.761, df = 116.64, p-value < 2.2e-16
## alternative hypothesis: true difference in means between groups is not equal to 0
## 95 percent confidence interval:
##   -913.1130 -697.0763
## sample estimates:
## mean in group female    mean in group male
##           4679.741           5484.836
```

## t-Test för två stickprov med R: variant 2

Om vi har data i två vektorer:

```
females <- subset(gentoo, sex == "female")
males <- subset(gentoo, sex == "male")
t.test(males$body_mass_g, females$body_mass_g)

##
##  Welch Two Sample t-test
##
## data:  males$body_mass_g and females$body_mass_g
## t = 14.761, df = 116.64, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  697.0763 913.1130
## sample estimates:
## mean of x mean of y
##  5484.836  4679.741
```

# Olika t-test

Vi har nu sett två sorters t-test:

- ① En grupp: ensticksprovs-t-test
- ② Två grupper: tvåsticksprovs-t-test

De används för olika frågeställningar, men går under samma namn eftersom de bygger på samma matematiska modell: t-fördelningen.

## Variant med lika standardavvikelse

I en del situationer har man anledning att anta att standardavvikelsen är densamma i båda grupperna.

- Exempel: variationen beror till största delen på mätfel, som beter sig på samma sätt för båda grupperna.

I sådana fall kan man beräkna den standardiserade differensen på ett lite annat sätt, vilket påverkar p-värdet.

Varianten där vi inte antar att grupperna har samma standardavvikelse är den vi oftast ska använda! Den kallas **Welch t-test** som är approximativt t-fördelat.



## Variant med lika standardavvikelse

```
t.test(body_mass_g ~ sex, data = gentoo, var.equal = TRUE)

##
##  Two Sample t-test
##
## data:  body_mass_g by sex
## t = -14.722, df = 117, p-value < 2.2e-16
## alternative hypothesis: true difference in means between groups is not equal to 0
## 95 percent confidence interval:
##  -913.4008 -696.7886
## sample estimates:
## mean in group female      mean in group male
##           4679.741           5484.836
```

# Bästa sättet att söva älgar

Veterinärer vid Kolmården ville utvärdera olika metoder för att söva älgar. Man vill testa följande hypoteser:

- $H_0$ : älgens puls är densamma vid sövningens start och slut
- $H_1$ : älgens puls är inte densamma vid sövningens start och slut

## Puls före och efter sövning

Namn	Före	Efter
Flash	43	42
Frostar	46	60
Hagel	43	47
Hjördis	29	36
Linus	42	30
Rosa Linda	38	30
Helga	46	107

Vi kan inte testa om det finns en skillnad genom ett vanligt tvåstickprovs-t-test eftersom vi har gjort flera mätningarna på samma individer!

## Puls före och efter sövning

Vi kan beräkna skillnaden Slut – Start för varje individ! Varje individ blir sin egen kontroll:

Namn	Före	Efter	Differens
Flash	43	42	-1
Frostar	46	60	14
Hagel	43	47	4
Hjördis	29	36	7
Linus	42	30	-12
Rosa Linda	38	30	-8
Helga	46	107	-61

# Puls före och efter sövning

Man vill testa följande hypoteser:

- $H_0$ : älgens puls är densamma vid sövningens start och slut
  - Är samma sak som  $H_0$ : väntevärdet för differensen är 0
- $H_1$ : älgens puls är inte densamma vid sövningens start och slut
  - Är samma sak som  $H_1$ : väntevärdet för differensen är inte 0

# Test

Vi kan utvärdera hypoteserna med ett t-test för differensen:

```
differens <- c(-1, 14, 4, 7, -12, -8, 61)
t.test(differens, mu = 0)

##
##  One Sample t-test
##
## data:  differens
## t = 1.0047, df = 6, p-value = 0.3538
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -13.32870  31.90013
## sample estimates:
## mean of x
##  9.285714
```

# Test

Testet kallas t-test vid stickprov i par. Vi kan utföra det direkt utan att behöva beräkna differensen:

```
start <- c(43, 46, 43, 29, 42, 38, 46)
slut <- c(42, 60, 47, 36, 30, 30, 107)
t.test(slut, start, mu = 0, paired = TRUE)

##
## Paired t-test
##
## data: slut and start
## t = 1.0047, df = 6, p-value = 0.3538
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -13.32870 31.90013
## sample estimates:
## mean difference
```

# Outliers och icke-normalitet

Här har vi en outlier!

Namn	Före	Efter	Differens
Flash	43	42	-1
Frostar	46	60	14
Hagel	43	47	4
Hjördis	29	36	7
Linus	42	30	-12
Rosa Linda	38	30	-8
<b>Helga</b>	<b>46</b>	<b>107</b>	<b>-61</b>

Orsaken kan vara

- mätfel?
- Viktigt specialfall?
- eller flera.



# Måste vi vara orolig?

t-test fungerar ofta inte som tänkt om vi har outliers (testet får fel signifikansnivå).

Möjliga lösningar (icke-parametriska metoder):

- Bootstrap-test: jämför medelvärdet efter och före (t-test utan normalfördelning)
- Rangtest: jämför medianen efter och före, utan normalfördelning (vidareutveckling av teckentestet)

# Sammanfattning

- ① Hypoteser om medelvärden kan testas med t-test
  - ① En grupp: ensticksprovs-t-test
  - ② Två grupper: tvåsticksprovs-t-test
  - ③ Två mätningar på samma individer: t-test för stickprov i par
- ② Förutsättningar!