

# **Biostatistik og epidemiologi**

Modul 4, Holdtime 1 og 2

Maja Hermansen

Epidemiologi, Biostatistik og Biodemografi (EBB), SDU

# Holdtimer

14 holdtimer (7 gange af 2 timer med 15 minutters pause)

- BSEH 1+2: d. 28/4 kl. 10-12 (U27A) – Datatyper, prædiktionsinterval og konfidensinterval
- BSEH 3+4: d. 4/5 kl. 8-10 (U27A) – Uparret t-test, hypotesebegrebet og p-værdi
- BSHE 5+6: d. 7/5 kl. 10-12 (U9) – Betinget/Ubetinget sandsynlighed
- BSHE 7+8: d. 8/5 kl. 8-10 (U27A) – Tværsnitsstudier
- BSHE 9+10: d. 12/5 kl. 8-10 (U9) – Kohortestudier
- BSHE 11+12: d. 13/5 kl. 10-12 (U230) – Case-kontrol studier
- BSHE 13+14: d. 20/5 kl. 10-12 (U28A) – Opsamling og tidligere eksamensopgave

# I dag

Population vs.  
Stikprøve

Datatyper

Illustration af data

Normalfordeling

Gennemsnit,  
spredning og  
standard error

Prædiktionsinterval

Konfidensinterval

# Opgave 1

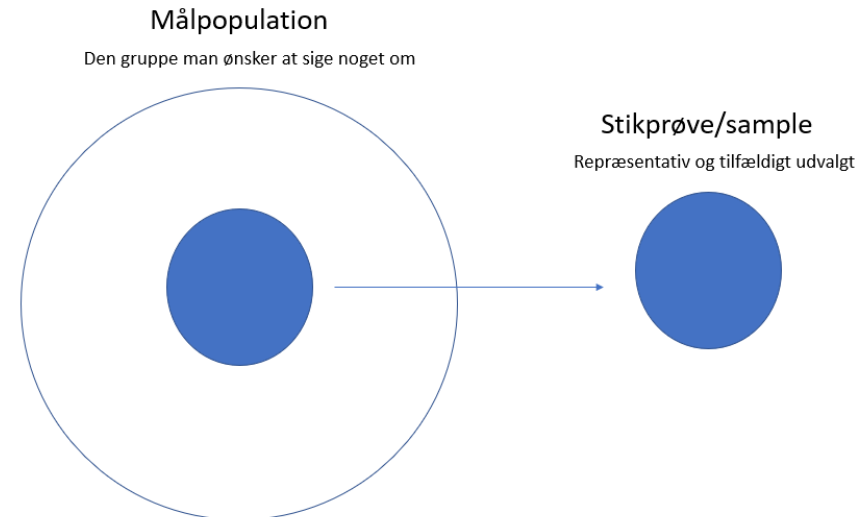
## 1) Hvad er en population og en stikprøve?

### Population

- En (ofte større) gruppe af ”individer” som er emnet for et (statistisk) studie
  - For medicinere/biomek: ”individer” er typisk mennesker
  - ”Individer” kan også være andet: hofteproteser, hospitaler, praksisenheder, ...
  - Den gruppe, vi ønsker at undersøge

### Stikprøve

- En delmængde af populationen
  - Tilfældigt udtræk fra populationen



# Opgave 1

**2) Hvis der fra samme population tages to tilfældige stikprøver, hvor man estimerer middelværdi i begge stikprøver, vil man så få præcis samme middelværdi eller spredning i de to stikprøver?**

Nej. Det kaldes sampling-variation/stikprøvevariationen. Det er denne variation, som standard error beskriver

- 1. stikprøve: giver målingerne  $x_1, x_2, \dots, x_n$
- 2. stikprøve: giver målingerne  $y_1, y_2, \dots, y_n$
- Da målingerne ved de to stikprøver vil være forskellige, vil de to gennemsnit  $\bar{x}$  og  $\bar{y}$  også være forskellige (næsten altid), men de skulle gerne minde om hinanden
- Det samme gælder for de to spredninger

# Opgave 1

## 3) Hvordan kan man bruge et 95% CI til at konkludere på middelværdien $\mu$ i hele populationen ud fra en stikprøve?

Når data er normalfordelte, kan vi bruge stikprøvens gennemsnit som bedste estimat for den sande middelværdi i populationen. Dvs. bedste bud på middelværdien ( $\mu$ ) i populationen er stikprøvens gennemsnit  $\bar{x}$ .

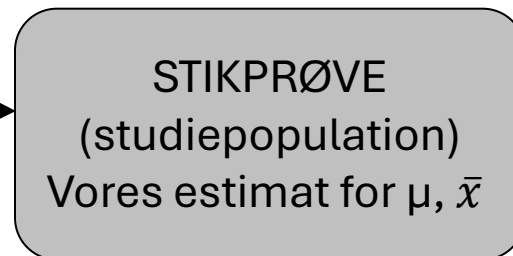
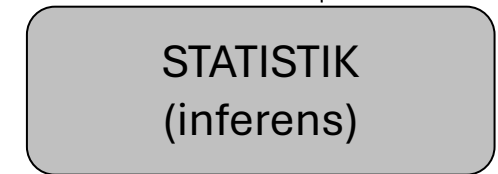
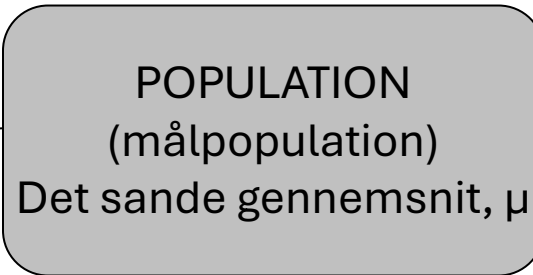
Pga. stikprøvevariationen vil man ud fra en stikprøve kun have en idé om, hvad middelværdien er ude i populationen, og man skal bruge statistik for at have evidens omkring middelværdien.

For normalfordelte data er 95% CI givet ved  $95\% \text{ CI} = [\bar{x} \pm t' \times se]$

Med 95% sandsynlighed vil populationens middelværdi ( $\mu$ ) ligge i 95%-konfidensintervallet, altså i intervallet  $[\bar{x} - t' \cdot se; \bar{x} + t' \cdot se]$ , hvor  $t'$  er det tosidede 5%-punkt i  $t$ -fordelingen med  $n - 1$  frihedsgrader

s.e afhænger af SD og antal,  $n$ . Jo større  $n$ , desto mindre bliver SE og dermed smallere CI, dvs. jo mere præcist vil man kunne udtale sig om den sande middelværdi i populationen.

Diabetikere



Diabetikere i Odense

*Epidemiologi:*  
Anvendes på  
populationen til  
udtagelse af  
repræsentativ  
stikprøve ved  
studiedesign

*Biostatistik:*  
Anvendes på  
stikprøven til at sige  
noget om  
populationen

# Datatyper

## Kategoriske data (grupperede):

- Drikker du øl? ← BINÆR  
Ja/Nej
- Hvilket mærke drikker du? ← KATEGORISK (NOMINAL)  
Ingen rækkefølge (flere end 2)
- Hvilken alkoholprocent indtager du? ← ORDNET KATEGORISK (ORDINAL)  
Grupperet – rækkefølge af alkohol styrke

## Numeriske data (tal):

- Hvor mange hele øl drak du i lørdags? ← DISKRET  
Antal uden decimaler
- Hvad var din promille? ← KONTINUERT  
Noget der kan måles og antage alle tal



# Datatyper

Valg af den rette statistiske metode dikteres delvist af de typer af data der indgår i analysen.

- **Numeriske variabel:** kvantitativ variabel, d.v.s. som i sin natur/sit væsen kvantificerer via en talværdi
  - **Kontinuert variabel** ("flydende" mål): tid, vægt, ...
    - Variabel som i princippet kan måles med uendelig præcision (variabel på en kontinuert skala)
  - **Diskret variabel:** antal søskende, antal bakterier, ...
    - Variabel der antager et begrænset antal (diskrete) værdier, typisk hele tal
- **Kategorisk variabel:** kvalitativ variabel, d.v.s. som grupperer (forskellige) observationer
  - **Binær variabel:** Død/ i live, syg/rask, mand/kvinde, ...
    - Variabel der inddeler i 2 grupper.
  - **Kategorisk variabel:** Etnicitet, fødested, ...
    - Variabel med "≥ 3" grupper.
  - **Ordnet kategorisk variabel:** Socialklasse, grupperet fødselsvægt, ...
    - "≥ 3" grupper med en ordning.

# Opgave 2

Angiv hvilken datatype nedenstående udfald (outcomes) er:

Outcome	Kategoriske data			Numerisk data	
	Binær	Kategorisk	Ordnet kategorisk	Diskret	Kontinuert
Antal fødsler				X	
Hårfarve		X			
Blodtryk					X
Køn (biologisk ved fødsel)†	X				
Kønsidentitet		X			
Socialt klassetrin			X		
BMI			X	← X	
Karakter			(X)	X	
Syg/ikke syg	X				

†Denne variabel benyttes ved statistiske analyser..

# Opgave 3

70 kvinder i et studie får taget en blodprøve, der måler hæmoglobin. Nedenstående tabel viser de forskellige niveauer af hæmoglobin.

1) Find median samt nedre og øvre kvartil (hhv Q1 og Q3). Beregn også IQR og range.

Hæmoglobin-niveau i g/100ml blandt 70 kvinder

8.8	9.3	9.4	9.7	10.2	10.2	10.3
10.4	10.4	10.5	10.6	10.6	10.7	10.8
10.8	10.9	10.9	10.9	11.0	11.0	11.1
11.1	11.2	11.2	11.3	11.4	11.4	11.4
11.5	11.6	11.6	11.7	11.7	11.8	11.8
11.9	11.9	12.0	12.0	12.1	12.1	12.1
12.2	12.3	12.5	12.5	12.7	12.9	12.9
12.9	12.9	13.0	13.1	13.1	13.2	13.3
13.3	13.4	13.4	13.5	13.5	13.6	13.7
13.7	14.1	14.6	14.6	14.7	14.9	15.1

## Median

(Number of values is odd)

8 11 15 18 24 30 31



8 11 15 **18** 24 30 31

## Median

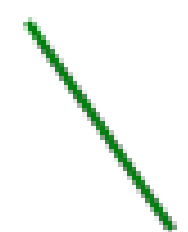
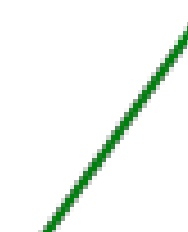
(Number of values is even)

11 15 18 24 30 31



11 15 18 **21** 24 30 31

18 19 20 **21** 22 23 24



# Opgave 3

Hæmoglobin-niveau i g/100ml blandt 70 kvinder

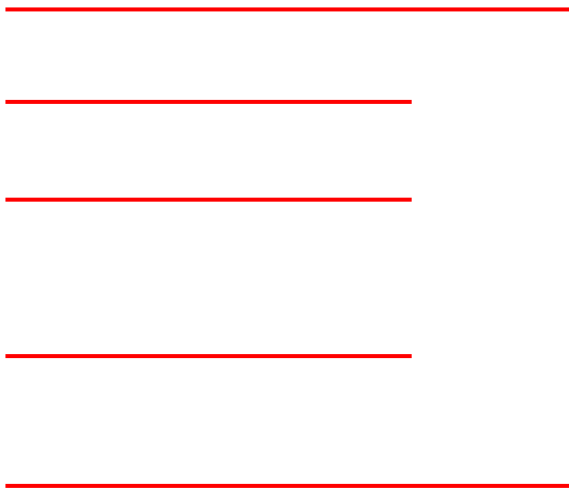
8.8	9.3	9.4	9.7	10.2	10.2	10.3	
10.4	10.4	10.5	10.6	10.6	10.7	10.8	
10.8	10.9	10.9	10.9	11.0	11.0	11.1	Q1
11.1	11.2	11.2	11.3	11.4	11.4	11.4	
11.5	11.6	11.6	11.7	11.7	11.8	11.8	
11.9	11.9	12.0	12.0	12.1	12.1	12.1	Median (11.8+11.9)/2
12.2	12.3	12.5	12.5	12.7	12.9	12.9	
12.9	12.9	13.0	13.1	13.1	13.2	13.3	Q3
13.3	13.4	13.4	13.5	13.5	13.6	13.7	
13.7	14.1	14.6	14.6	14.7	14.9	15.1	

IQR  $Q3 - Q1 = 13.1 - 10.9 = 2.2$

Range Min til max =  $15.1 - 8.8 = 6.3$

# Opgave 3

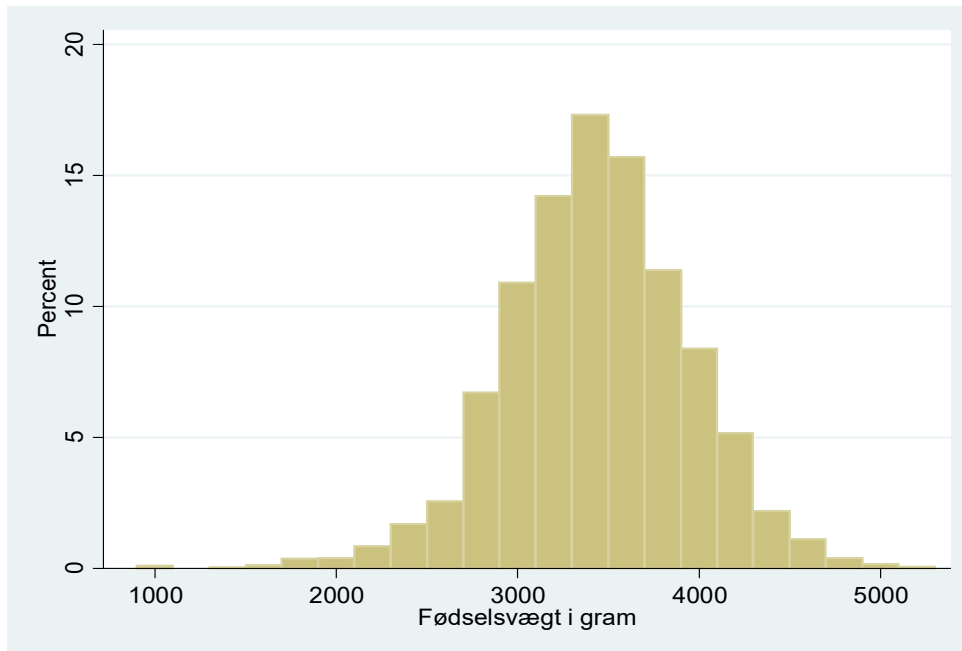
2) Her ser du et boxplot over alderen på forsøgsdeltagerne fra spørgsmål 2.  
Aflæs median, Q1, Q3, min og max



# Opgave 4

Et studie undersøgte 3978 amerikanske børns fødselsvægt.

Ud fra data på de 3978 børns vægte blev der beregnet et gennemsnit,  $\bar{x} = 3437$  gram og en standard afvigelse,  $SD = 507$  gram.



## Databoks

$$n = 3.978$$

$$\bar{x} = 3.437$$

$$SD = 507$$

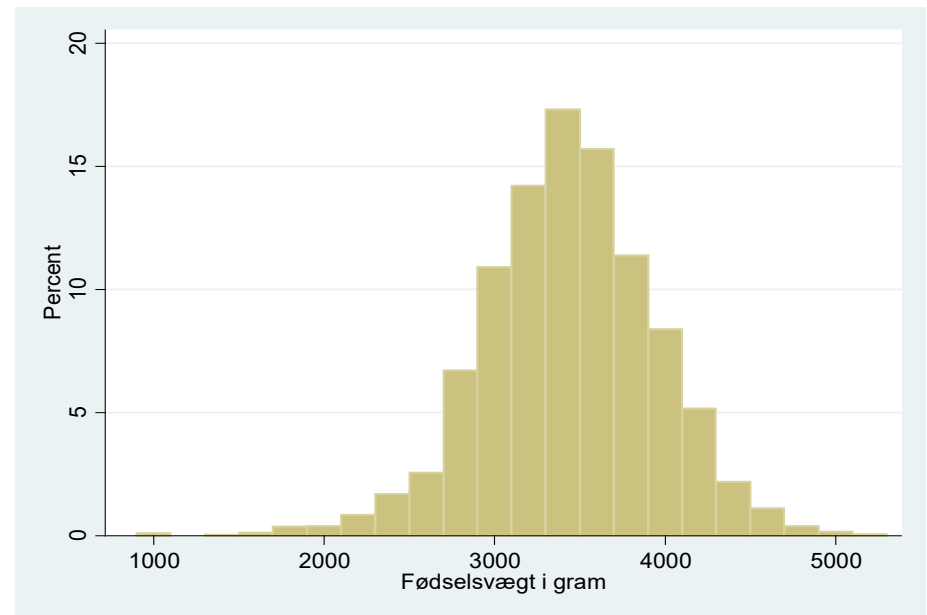
# Opgave 4

## Databoks

$$n = 3.978$$

$$\bar{x} = 3.437$$

$$SD = 507$$



**1) Hvilken type data er fødselsvægte? Karakteriser desuden ovenstående fordeling.**

Datatype: numerisk, kontinuerte data

Fordeling: fødselsvægtene er normalfordelte, klokkeformet fordeling

**2) Beregn 95% prædiktionsintervallet (=referenceintervallet) af børnenes vægte**

$$95\% \text{ PI} = [\bar{x} \pm 1,96 * SD]$$

$$95\% \text{ PI} = [\bar{x} \pm 1,96 * SD] = [3437 \pm 1,96 * 507] = [2443 \text{ gram}; 4431 \text{ gram}]$$



# Opgave 4

## 3) Hvad er fortolkningen af dette 95% prædiktionsinterval?

Forventning: 95% af nye målinger er indeholdt i dette interval.

Så i eksemplet med fødselsvægte og  $PI = [2443; 4431]$  gram?

Et amerikansk barn vil med 95% sandsynlighed have en fødselsvægt mellem 2443 og 4431 gram.

Anvendelse: et sådan interval bruges tit i lægers/fagpersons vurdering af den enkelte patients helbred. Et referenceinterval.

Eksempel med blodtryk: for raske unge mænd er systolisk normalt 100-140 og diastolisk normalt 60-90. En tilfældig mand i denne alder med værdier over eller under vil ligge uden for normalen.

# Opgave 4

**Børn, der vejer under 2500 g, siges at have lav fødselsvægt. Man kan under tiden finde det relevant at kategorisere børn ud fra om de har en lav fødselsvægt eller ej.**

**4) Hvilken datatype repræsenterer denne kategorisering?**

Lav fødselsvægt er en binær variabel, kategorisk.

Numerisk data kan konverteres til kategorisk, men IKKE den anden vej rundt

# Opgave 4

## 5) Beregn *standard error*

$$\text{s.e.} = SD/\sqrt{n} = 507/\sqrt{3978} = 507/63,07 = 8,04$$

## 6) Hvad er fortolkningen af *standard error*?

Standard error er lig spredningen i stikprøvefordelingen af gennemsnittet.

Hvis man tager 100 lige store stikprøver og beregner 100 gennemsnit, vil de gennemsnit være normalfordelte om den sande middelværdi i populationen. Gennemsnittene vil typisk ligge tættere på den sande middelværdi end de oprindelige enkelte målinger. Dermed har gennemsnittene en mindre spredning end målingerne i den bagvedliggende population. Og det er spredningen i gennemsnittenes fordeling der kaldes standard error, eller s.e.

s.e. er lig spredningen på gennemsnittene (s.e.= SD af gennemsnittene)

s.e. er et mål for, hvor præcist stikprøvens gennemsnittet estimerer den sande middelværdi

s.e. er bestemt ud fra stikprøvens størrelse og standard afvigelsen (SD) i den enkelte stikprøve.

### Databoks

$$n = 3978$$

$$\bar{x} = 3437$$

$$SD = 507$$

$$s.e. = \frac{SD}{\sqrt{n}}$$

# Opgave 4

Lad  $\mu$  betegne middelfødselsvægten i populationen af børn til amerikanske mødre.

7) Beregn et 95% konfidensinterval (CI) for børnenes middelvægt  $\mu$

Antal frihedsgrader (d.f.) =  $n-1 = 3977$

$$95\% CI = [\bar{x} \pm t' \times se]$$

Opslag i tabel A3 (t-distribution)  $\rightarrow$  5%-punktet

## Databoks

$$n = 3978$$

$$\bar{x} = 3437$$

$$SD = 507$$

$$s.e. = \frac{SD}{\sqrt{n}} = 8,04$$

$$95\% CI = [\bar{x} \pm t' \times s.e.]$$

**Table A3 Percentage points of the  $t$  distribution**Adapted from Table 7 of White *et al.* (1979) with permission of the authors and publishers.

d.f.	One-sided $P$ -value								
	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
	Two-sided $P$ -value								
	0.5	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001
1	1.00	3.08	6.31	12.71	31.82	63.66	127.32	318.31	636.62
2	0.82	1.89	2.92	4.30	6.96	9.92	14.09	22.33	31.60
3	0.76	1.64	2.35	3.18	4.54	5.84	7.45	10.21	12.92
4	0.74	1.53	2.13	2.78	3.75	4.60	5.60	7.17	8.61
5	0.73	1.48	2.02	2.57	3.36	4.03	4.77	5.89	6.87
6	0.72	1.44	1.94	2.45	3.14	3.71	4.32	5.21	5.96
7	0.71	1.42	1.90	2.36	3.00	3.50	4.03	4.78	5.41
8	0.71	1.40	1.86	2.31	2.90	3.36	3.83	4.50	5.04
9	0.70	1.38	1.83	2.26	2.82	3.25	3.69	4.30	4.78
10	0.70	1.37	1.81	2.23	2.76	3.17	3.58	4.14	4.59
11	0.70	1.36	1.80	2.20	2.72	3.11	3.50	4.02	4.44
12	0.70	1.36	1.78	2.18	2.68	3.06	3.43	3.93	4.32
13	0.69	1.35	1.77	2.16	2.65	3.01	3.37	3.85	4.22
14	0.69	1.34	1.76	2.14	2.62	2.98	3.33	3.79	4.14
15	0.69	1.34	1.75	2.13	2.60	2.95	3.29	3.73	4.07
16	0.69	1.34	1.75	2.12	2.58	2.92	3.25	3.69	4.02
17	0.69	1.33	1.74	2.11	2.57	2.90	3.22	3.65	3.96
18	0.69	1.33	1.73	2.10	2.55	2.88	3.20	3.61	3.92
19	0.69	1.33	1.73	2.09	2.54	2.86	3.17	3.58	3.88
20	0.69	1.32	1.72	2.09	2.53	2.84	3.15	3.55	3.85
21	0.69	1.32	1.72	2.08	2.52	2.83	3.14	3.53	3.82
22	0.69	1.32	1.71	2.07	2.51	2.82	3.12	3.50	3.79
23	0.68	1.32	1.71	2.07	2.50	2.81	3.10	3.48	3.77
24	0.68	1.32	1.71	2.06	2.49	2.80	3.09	3.47	3.74
25	0.68	1.32	1.71	2.06	2.48	2.79	3.08	3.45	3.72
26	0.68	1.32	1.71	2.06	2.48	2.78	3.07	3.44	3.71
27	0.68	1.31	1.70	2.05	2.47	2.77	3.06	3.42	3.69
28	0.68	1.31	1.70	2.05	2.47	2.76	3.05	3.41	3.67
29	0.68	1.31	1.70	2.04	2.46	2.76	3.04	3.40	3.66
30	0.68	1.31	1.70	2.04	2.46	2.75	3.03	3.38	3.65
40	0.68	1.30	1.68	2.02	2.42	2.70	2.97	3.31	3.55
60	0.68	1.30	1.67	2.00	2.39	2.66	2.92	3.23	3.46
120	0.68	1.29	1.66	1.98	2.36	2.62	2.86	3.16	3.37
$\infty$	0.67	1.28	1.65	1.96	2.33	2.58	2.81	3.09	3.29

# Opgave 4

Lad  $\mu$  betegne middelfødselsvægten i populationen af børn til amerikanske mødre.

**7) Beregn et 95% konfidensinterval (CI) for børnenes middelvægt  $\mu$**

Antal frihedsgrader (d.f.) =  $n-1 = 3977$

$$95\% CI = [\bar{x} \pm t' \times se]$$

Opslag i tabel A3 (t-distribution)  $\rightarrow$  5%-punktet

$$t' = 1,96$$

$$95\% CI = [\bar{x} \pm t' \times se] = [3437 \pm 1,96 \times 8,04] = [3421 \text{ g}; 3453 \text{ g}]$$

## Databoks

$$n = 3978$$

$$\bar{x} = 3437$$

$$SD = 507$$

$$s.e. = \frac{SD}{\sqrt{n}} = 8,04$$

$$95\% CI = [\bar{x} \pm t' \times s.e.]$$

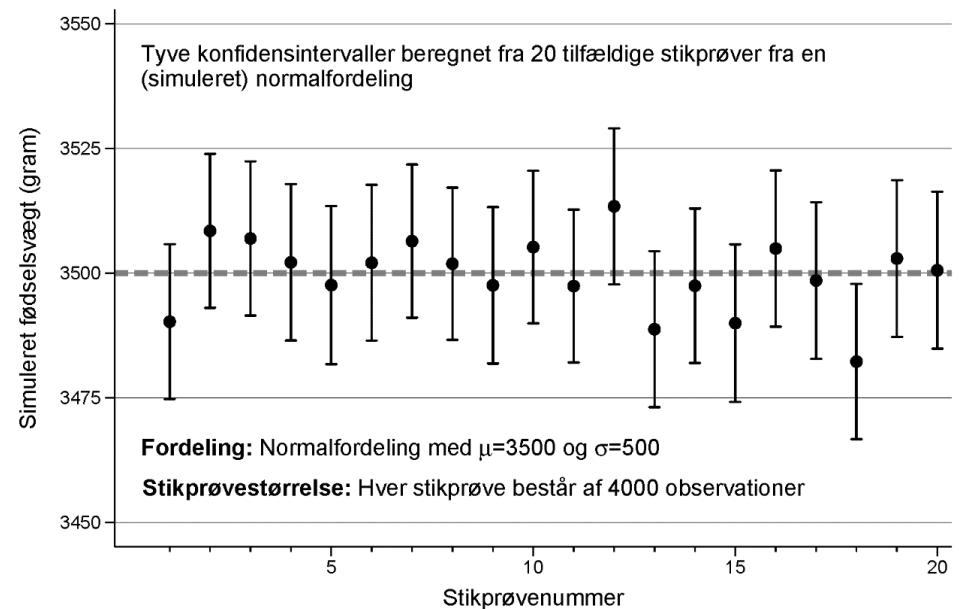
# Opgave 4

## 8) Hvad er fortolkningen af dette 95% CI?

Den sande middelværdi for amerikanske børns fødselsvægt ligger med 95% sandsynlighed mellem 3421 og 3453 gram.

Hvis man tager 20 tilfældige stikprøver af samme størrelse fra samme population, vil 19 af de 20 forskellige 95% konfidensintervaller indeholde den ukendte, sande middelværdi for populationen.

Det fortolker vi som: med 95% sandsynlighed indeholder konfidensintervallet den sande (faste, men ukendte) værdi.



# Opgave 4

Antag nu, at du kun har data på 27 børns fødselsvægte. De 27 børn har samme gennemsnitlig fødselsvægt og standardafvigelse som ovenfor

## 9) Beregn en ny standard error

$$s.e. = SD/\sqrt{n} = 507/\sqrt{27} = 507/5,20 = 97,57$$

## 10) Beregn et nyt 95% CI for børnenes middelfødselsvægt.

Frihedsgrader,  $n-1 = 26$ ;  $t'$  i tabel A3

### Databoks

$$n = 27$$

$$\bar{x} = 3437$$

$$SD = 507$$

$$s.e. = \frac{SD}{\sqrt{n}}$$

$$95\% CI = [\bar{x} \pm t' \times s.e.]$$



**Table A3 Percentage points of the  $t$  distribution**Adapted from Table 7 of White *et al.* (1979) with permission of the authors and publishers.

d.f.	One-sided $P$ -value								
	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
	Two-sided $P$ -value								
	0.5	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001
1	1.00	3.08	6.31	12.71	31.82	63.66	127.32	318.31	636.62
2	0.82	1.89	2.92	4.30	6.96	9.92	14.09	22.33	31.60
3	0.76	1.64	2.35	3.18	4.54	5.84	7.45	10.21	12.92
4	0.74	1.53	2.13	2.78	3.75	4.60	5.60	7.17	8.61
5	0.73	1.48	2.02	2.57	3.36	4.03	4.77	5.89	6.87
6	0.72	1.44	1.94	2.45	3.14	3.71	4.32	5.21	5.96
7	0.71	1.42	1.90	2.36	3.00	3.50	4.03	4.78	5.41
8	0.71	1.40	1.86	2.31	2.90	3.36	3.83	4.50	5.04
9	0.70	1.38	1.83	2.26	2.82	3.25	3.69	4.30	4.78
10	0.70	1.37	1.81	2.23	2.76	3.17	3.58	4.14	4.59
11	0.70	1.36	1.80	2.20	2.72	3.11	3.50	4.02	4.44
12	0.70	1.36	1.78	2.18	2.68	3.06	3.43	3.93	4.32
13	0.69	1.35	1.77	2.16	2.65	3.01	3.37	3.85	4.22
14	0.69	1.34	1.76	2.14	2.62	2.98	3.33	3.79	4.14
15	0.69	1.34	1.75	2.13	2.60	2.95	3.29	3.73	4.07
16	0.69	1.34	1.75	2.12	2.58	2.92	3.25	3.69	4.02
17	0.69	1.33	1.74	2.11	2.57	2.90	3.22	3.65	3.96
18	0.69	1.33	1.73	2.10	2.55	2.88	3.20	3.61	3.92
19	0.69	1.33	1.73	2.09	2.54	2.86	3.17	3.58	3.88
20	0.69	1.32	1.72	2.09	2.53	2.84	3.15	3.55	3.85
21	0.69	1.32	1.72	2.08	2.52	2.83	3.14	3.53	3.82
22	0.69	1.32	1.71	2.07	2.51	2.82	3.12	3.50	3.79
23	0.68	1.32	1.71	2.07	2.50	2.81	3.10	3.48	3.77
24	0.68	1.32	1.71	2.06	2.49	2.80	3.09	3.47	3.74
25	0.68	1.32	1.71	2.06	2.48	2.79	3.08	3.45	3.72
26	0.68	1.32	1.71	2.06	2.48	2.78	3.07	3.44	3.71
27	0.68	1.31	1.70	2.05	2.47	2.77	3.06	3.42	3.69
28	0.68	1.31	1.70	2.05	2.47	2.76	3.05	3.41	3.67
29	0.68	1.31	1.70	2.04	2.46	2.76	3.04	3.40	3.66
30	0.68	1.31	1.70	2.04	2.46	2.75	3.03	3.38	3.65
40	0.68	1.30	1.68	2.02	2.42	2.70	2.97	3.31	3.55
60	0.68	1.30	1.67	2.00	2.39	2.66	2.92	3.23	3.46
120	0.68	1.29	1.66	1.98	2.36	2.62	2.86	3.16	3.37
$\infty$	0.67	1.28	1.65	1.96	2.33	2.58	2.81	3.09	3.29

# Opgave 4

Antag nu, at du kun har data på 27 børns fødselsvægte. De 27 børn har samme gennemsnitlig fødselsvægt og standardafvigelse som ovenfor

## 9) Beregn en ny standard error

$$\text{s.e.} = SD/\sqrt{n} = 507/\sqrt{27} = 507/5,20 = 97,57$$

## 10) Beregn et nyt 95% CI for børnenes middelfødselsvægt.

Frihedsgrader,  $n-1 = 26$ ;  $t'$  i tabel A3

$$t' = 2,06$$

$$95\% \text{ CI} = [\bar{x} \pm 2,06 * \text{s.e.}] = [3437 \pm 2,06 * 97,57] = [3236 \text{ gram}; 3638 \text{ gram}]$$

### Databoks

$$n = 27$$

$$\bar{x} = 3437$$

$$SD = 507$$

$$\text{s.e.} = \frac{SD}{\sqrt{n}}$$

$$95\% \text{ CI} = [\bar{x} \pm t' \times \text{s.e.}]$$

# Opgave 4

Scenarie 1 (n=3978): 95% CI =  $[\bar{x} \pm 1,96 * \text{s.e.}] = [3437 \pm 1,96 * 8,04] = [3421 \text{ g}; 3453 \text{ g}]$

Scenarie 2 (n=27): 95% CI =  $[\bar{x} \pm 2,06 * \text{s.e.}] = [3437 \pm 2,06 * 97,57] = [3236 \text{ g}; 3638 \text{ g}]$

Dvs. noget bredere interval end før = højere usikkerhed på estimatet. Med et lille antal (n) i stikprøven, bliver vi altså dårligere til at estimere den sande middelværdi i populationen.

Vi kan ikke udtale os med så stor sikkerhed om den ukendte, sande middelværdi i populationen.

# Opgave 5

**1) Hvad forudsættes om blodtryksdata, hvis vi efterfølgende beregner prædiktionsintervallet og konfidensintervallet?**

At data er normalfordelte, så vi kan beskrive data ud fra  $\bar{x}$  og SD, så vi kan beregne PI og CI.

# Opgave 5

## 2) Hvornår bruges spredning og standard error i udregning af henholdsvis prædiktionsintervallet og konfidensintervallet?

Spredning, standard deviation, bruges til at beregne prædiktionsintervallet (tabel A1)

$$95\% PI = [\bar{x} \pm 1.96 \times SD] \text{ eller } [\bar{x} \pm t' \times SD]$$

SD fortæller, hvor målingerne ligger i forhold til  $\bar{x}$  i stikprøven. Med 95% sikkerhed vil målingerne ligge indenfor  $[\bar{x} \pm 1.96 \times SD]$

Standard error bruges til at beregne konfidensintervallet (tabel A3)

$$95\% CI = [\bar{x} \pm 1.96 \times s.e] - \text{approksimativt}$$

$$95\% CI = [\bar{x} \pm t' \times s.e] - \text{eksakt}$$

s.e. fortæller om usikkerheden på af vores estimat  $\bar{x}$  af den sande middelværdi i populationen. Vi estimerer, at den sande middelværdi ligger i intervallet  $[\bar{x} \pm t' \times s.e]$  med 95% sikkerhed.

# Opgave 5

## 3) Hvilket interval benyttes til at undersøge, om en tilfældig patient ligger uden for "normalområdet"?

Prædiktionsintervallet, referenceintervallet. et normalt systolisk blodtryk på 100-140 og diastolisk blodtryk på 60-90 for mænd i alderen på 40-50 år er fundet ved at kigge på raske mænd i denne aldersgruppe, og det er derefter vurderet, at her ligger størstedelen af observationerne. Så hvis en tilfældig 40-50-årig mand får målt sit blodtryk, vil det med 95% sandsynlighed ligge i prædiktionsintervallet. Hvis blodtrykket ligger udenfor denne 'normal', bør man reagere på denne måling.

## 4) Hvilket interval benyttes til, at vurdere, hvor den sande middelværdi i hele populationen ligger?

Konfidensintervallet: med 95% sandsynlighed vil intervallet indeholde den sande middelværdi i populationen. En anden stikprøve giver et andet interval. Sagt på en anden måde vil 19 ud af 20 konfidensintervaller indsamlet på tilsvarende måde indeholde den sande middelværdi i populationen. Jo større stikprøve, jo smallere CI.

# Opsummering

- Datatyper
- Begrebet ”stikprøvevariation”
- Estimering af middelværdien  $\mu$  i én population: (punkt)estimatet  $\bar{x}$  og 95%-CI( $\mu$ )
- Beregning og fortolkning af 95%-konfidensinterval for middelværdien i én normalfordelt stikprøve
  - Eksakt formel
  - Approksimativ formel
- Standard error (s.e.): beregning og fortolkning
- Beregning og fortolkning af 95%-prædiktionsinterval i én normalfordelt stikprøve