Utvalgsfordelinger (Kapittel 5)

Oversikt pensum, fortid og fremtid

- Eksplorativ data-analyse (Kap 1, 2)
- Hvordan produsere data (Kap 3)
- Sannsynlighetsteori (Kap 4)
- Utvalgsfordelinger til observatorer (Kap 5): I DAG
- Introduksjon til inferens om ukjente parametre (Kap 6)
  - Konfidensintervaller
  - Hypotesetesting
- Inferens på
  - Forventning (Kap 7)
  - Regresjonsmodeller (Kap 10,11)

Observator

- En observator er en funksjon av data for mange individer, for eksempel
  - Gjennomsnitt
  - Andel
  - Stigningstall i regresjonslinje
- En observator er en tilfeldig variabel
  - Har en sannsynlighetsfordeling som kalles observatorfordeling
  - Beskriver hvordan observatoren varierer når utvalget/eksperimentet repeteres mange ganger
- Individuelle data: Populasjonsfordeling

Eksempel: Populasjonsfordelinger og utvalgsfordelinger

- Høyde kvinne N(64.5,2.5)-fordelt
  - N(64.5,2.5): Populasjonsfordeling
- Gjennomsnitt av 100 kvinner N(64.5,0.25)-fordelt
  - $\bar{x}$ observator
  - N(64.5,0.25): Observatorfordeling til $\bar{x}$
  - 64.5 er forventningen (μ), en parameter som er et fast tall, har ingen fordeling

Observatorer

- Vi skal i dag se på observatorene:
  - Gjennomsnitt
  - Antall/andeler
- Er ute etter egenskaper til observatorer
  - Forventning
  - Spredning
  - Sannsynlighetsfordeling

Utvalgsfordeling: Gir svaret på hva som ville skje dersom vi så på mange utvalg med størrelse *n* fra den samme populasjonen

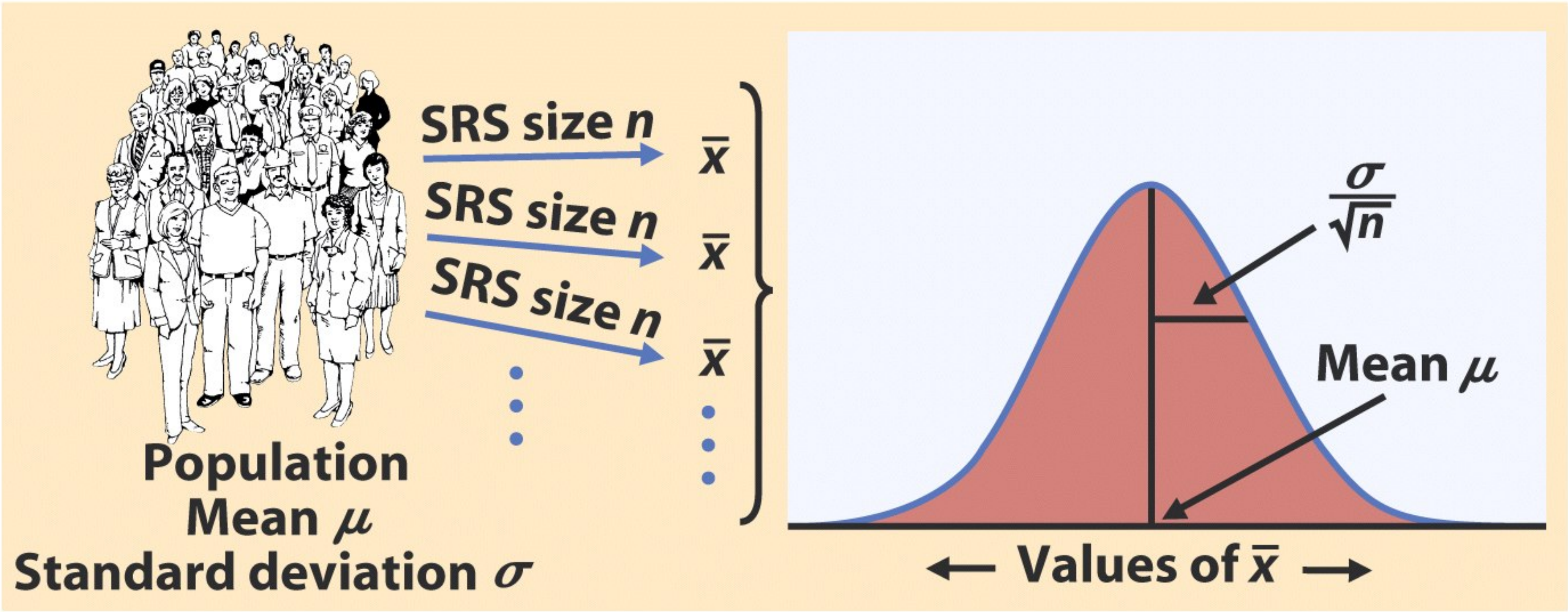


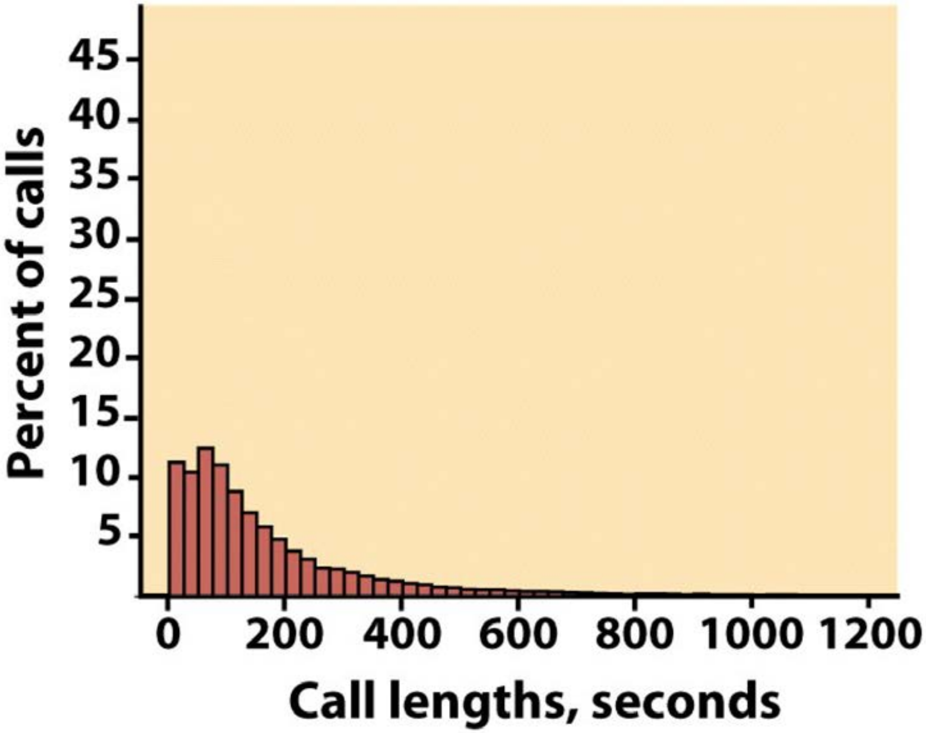
Figure 5-12
Introduction to the Practice of Statistics, Fifth Edition

Kap. 5.1: Utvalgsfordelingen for gjennomsnitt

- $\bar{x}$: kontinuerlig tilfeldig variabel
- Gjenomsnitt er mindre variable enn individuelle observasjoner
- Gjennomsnitt mer normalfordelte enn individuelle observasjoner

Histogram over lengden av ca 3000 oppringninger (ekstreme uteliggere er fjernet)

Histogram over gjennomsnittlig lengde på oppringninger



Kvantilplott for gjennomsnittlig lengde på oppringninger (500 utvalg av størrelse 80 fra populasjonen på 3000)

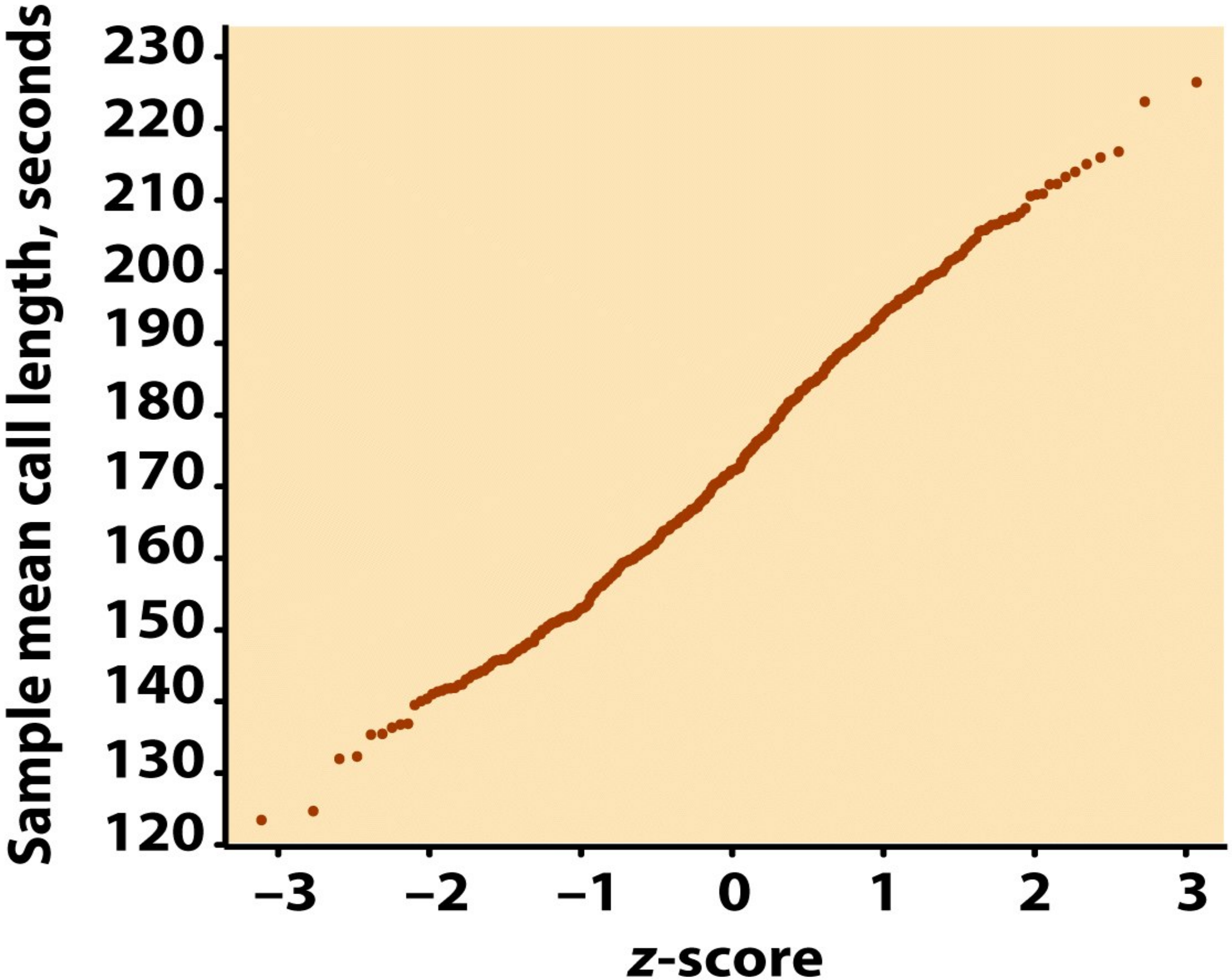


Figure 5-9
Introduction to the Practice of Statistics, Fifth Edition

Forventning til gjennomsnitt

- Har et SRS av størrelse *n* fra en populasjon
- Observerer variabelen X for hvert individ i utvalget, som har forventning μ og standardavik σ (populasjonsparametre)
- De *n* observasjonene er verdier av *n* stokastiske variable $X_1, X_2, \ldots, X_n$
- $\bar{x} = (x_1 + x_2 + \ldots + x_n)/n = S/n$ (S er summen av alle x_i -ene)
- Forventning til x_i er μ
- $\mu_S =$
- $\mu_{\bar{x}} =$

Varians og standardavvik til gjennomsnitt

- $\bar{x}=(x_1+x_2+...+x_n)/n=S/n$
- Varians til x_i er σ^2
- $\sigma_S{}^2=$
- $\sigma^2{}_{\bar{x}}=$
- $\sigma_{\bar{x}}=$

MEAN AND STANDARD DEVIATION OF A SAMPLE MEAN

Let $\bar{x}$ be the mean of an SRS of size n from a population having mean μ and standard deviation σ. The mean and standard deviation of $\bar{x}$ are

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

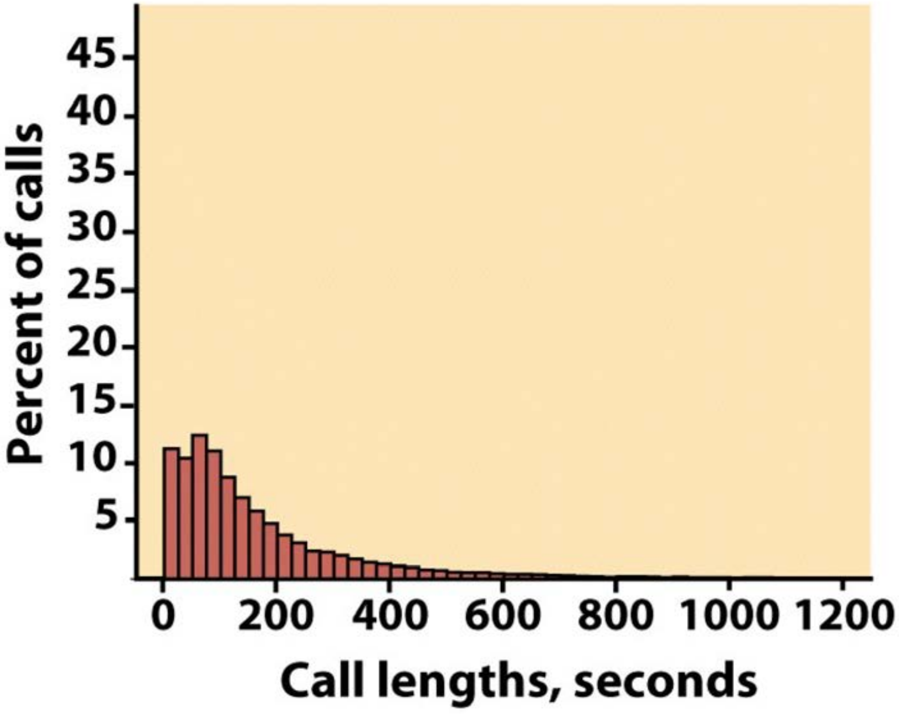
Definition, pg 361
Introduction to the Practice of Statistics, Fifth Edition

Fordeling til gjennomsnitt

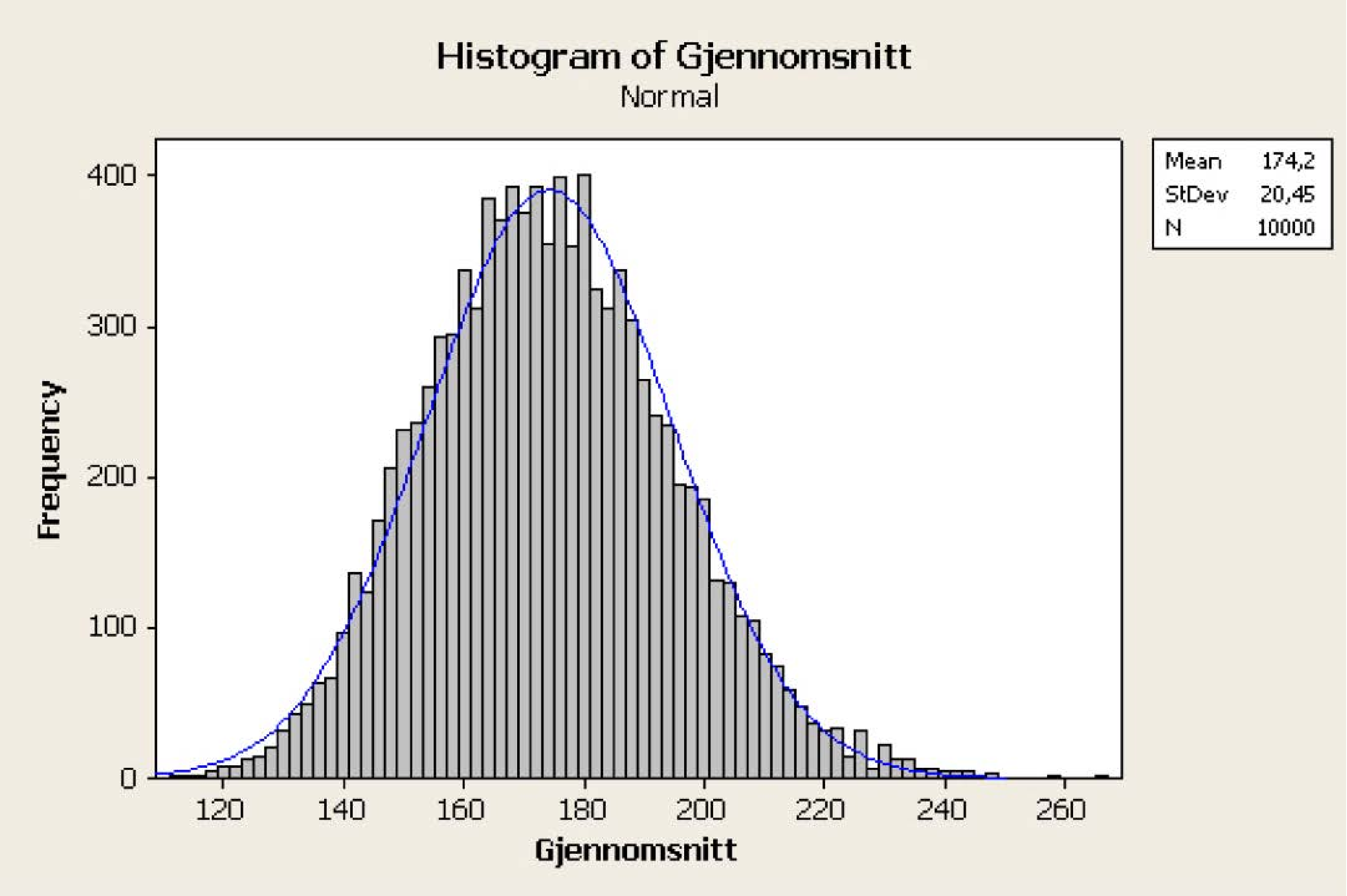
- Vi har beskrevet senter (forventningen) og spredning (standardavviket) til fordelingen til gjennomsnittet for et utvalg på størrelse *n*, men ikke formen
- Normalfordelte variable
  - Anta X_i er uavhengige og $N(\mu,\sigma)$-fordelte
  - Da er $\bar{x}$ $N(\mu,\sigma/\sqrt{n})$-fordelt
- Sentralgrenseteoremet for SRS av størrelse n
  - Forventning til X_i er μ
  - Standardavvik til X_i er σ
  - Anta n stor
  - Da er $\bar{x}$ tilnærmet $N(\mu,\sigma/\sqrt{n})$-fordelt

Telefonoppringninger

- 31492 oppringninger til en banks kundesenter
- Fordelingen til de individuelle observasjonene er langt fra normal
- Repeter 10000 ganger:
  - Trekk enkelt tilfeldig utvalg av størrelse n=80
  - Beregn $\bar{x}$
- Lag histogram av de 10000 $\bar{x}$'er



Histogram over gjennomsnittet av de 10000 utvalgene av størrelse 80



Eksponensiell fordeling

- Mye brukt for levetider
  - Lyspærer
  - Elektroniske komponenter
  - Individer
- Langt fra normal

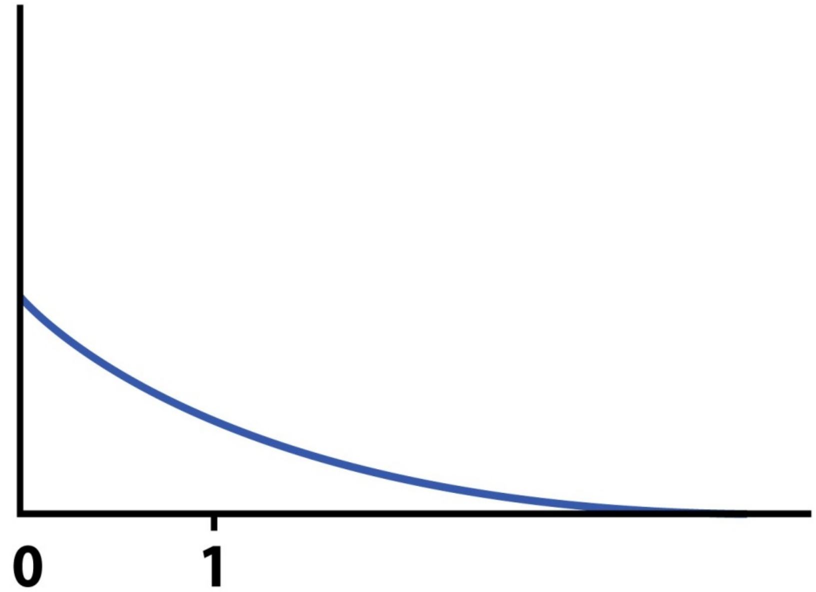


Figure 5-10a
Introduction to the Practice of Statistics, Fifth Edition

a) Fordelingen til gj.sn av 1 observasjon fra en eksponensialfordelt populasjon
b) Fordelingen til gj.sn av 2 observasjoner fra en eksponensialfordelt populasjon
c) Fordelingen til gj.sn av 10 observasjon fra en eksponensialfordelt populasjon
d) Fordelingen til gj.sn av 25 observasjoner fra en eksponensialfordelt populasjon

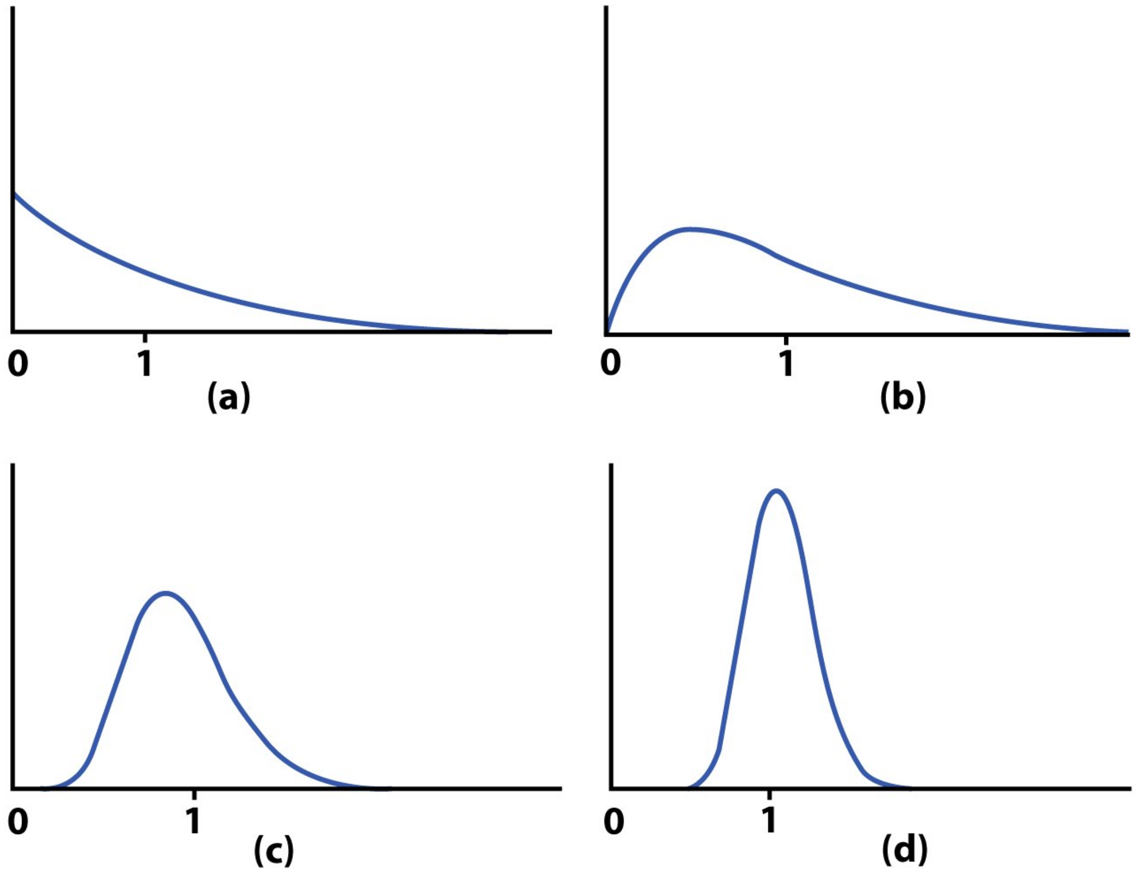


Figure 5-10
Introduction to the Practice of Statistics, Fifth Edition

Eksponensiell - eksempel

- X: tid tekniker bruker for vedlikehold av system
- Eksponensiell, μ=1 time,σ=1 time
- 70 enheter, $P(\bar{x}>50$ min)=$P(\bar{x}>0.83$ timer)=?
- Sentralgrenseteoremet: $\bar{x}$ tilnærmet $N(1,1/\sqrt{70})=N(1,0.12)$
- $P(\bar{x}>0.83)=$

- Eksakt: 0.9294

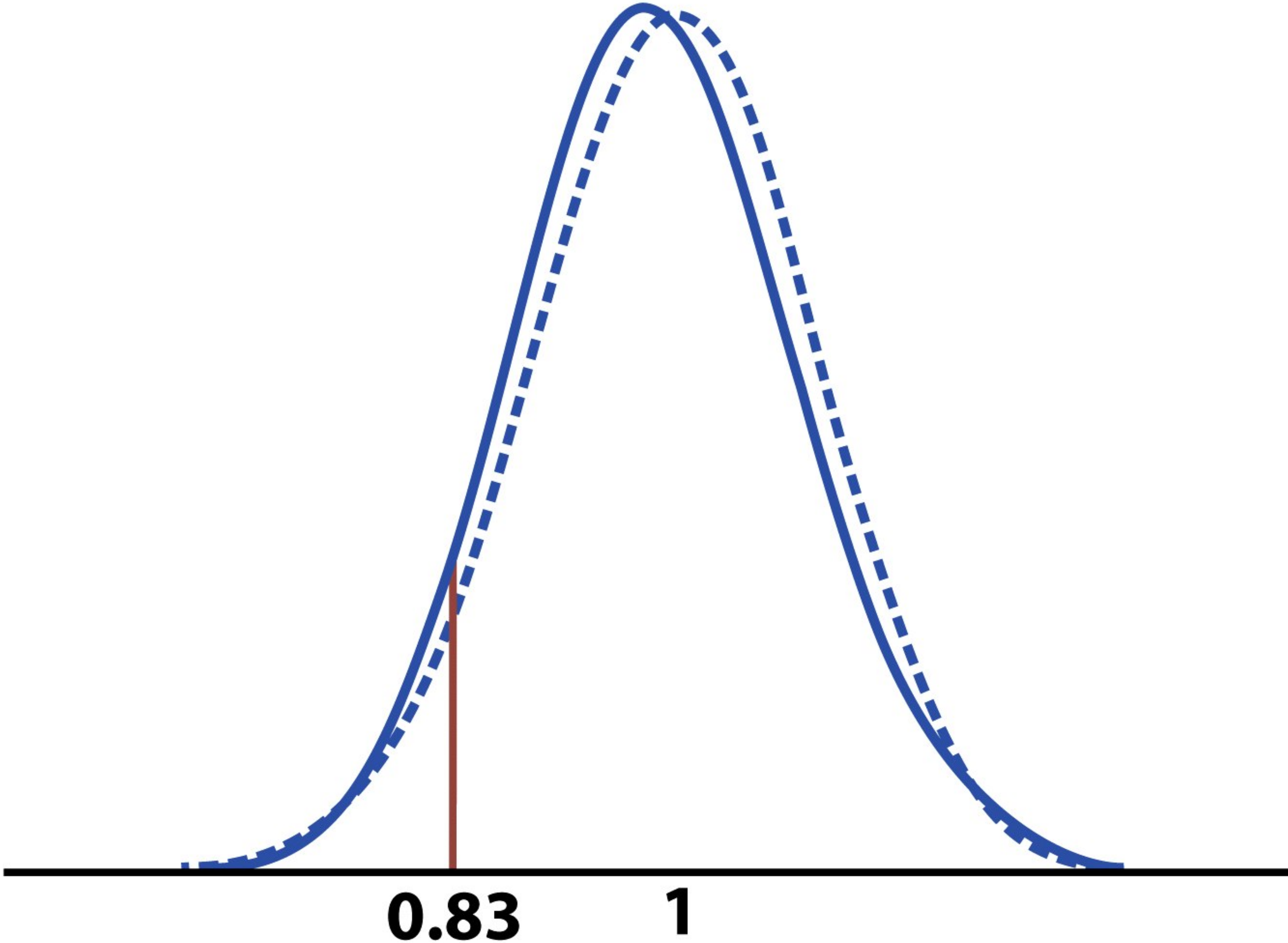


Figure 5-11
Introduction to the Practice of Statistics, Fifth Edition

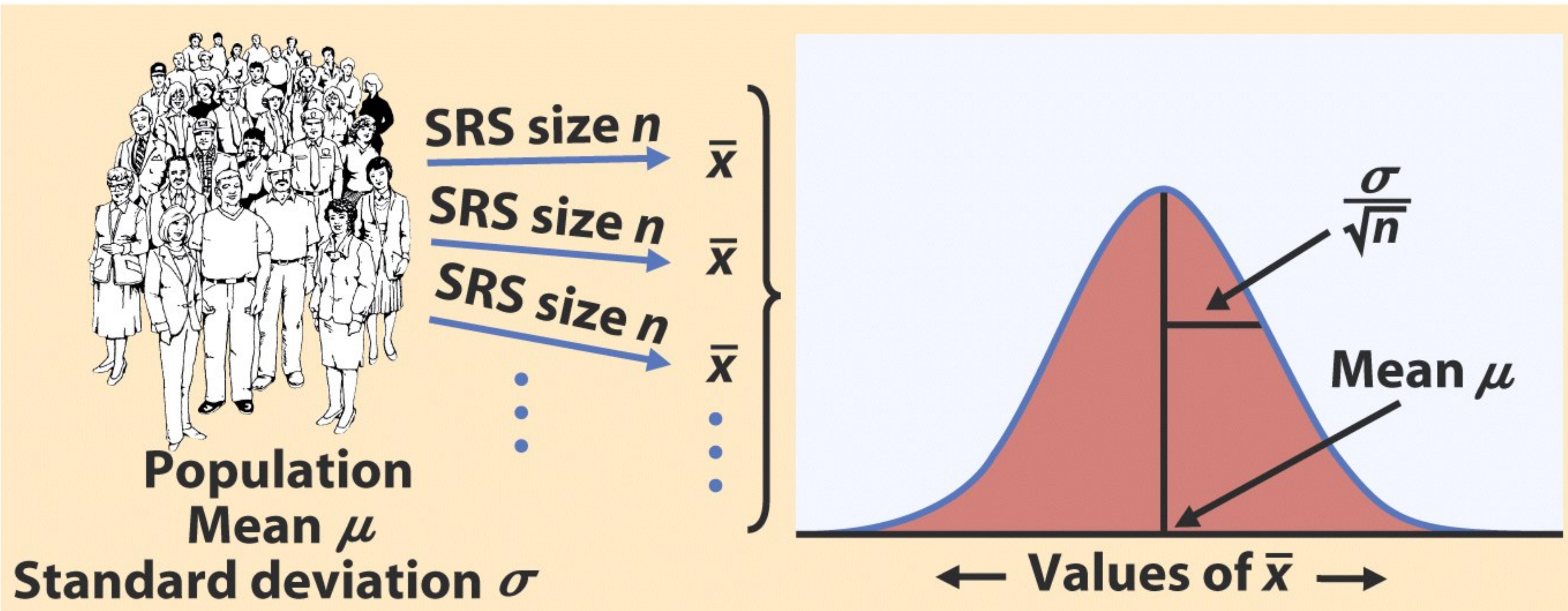


Figure 5-12
Introduction to the Practice of Statistics, Fifth Edition

Fordeling til en lineærkombinasjon av normalfordelte variabler

- At gjennomsnittet x av observerte verdier av *n* uavhengige normalfordelte variabler $X_1, X_2, \ldots, X_n$ også er normalfordelt er et spesialtilfelle av en generell regel:
  - X er $N(\mu_X, \sigma_X)$-fordelt og Y er $N(\mu_Y, \sigma_Y)$-fordelt
  - X og Y er uavhengige (dvs korrelasjon=0)
  - Da er $Z = aX+bY$ (der a og b er faste konstanter) også normalfordelt med
    - forventning $\mu_Z = a\mu_X + b\mu_Y$
    - standardavvik $\sigma_Z = \sqrt{(a^2\sigma_X{}^2 + b^2\sigma_Y{}^2)}$