Test af andel i populationer

Test af andel i en population

(se bogen: 8.1.2)

Example 8.3 About 13% of the population is left-handed. A biologist suspects that the scientific community is not like the general population in terms of handedness. He will conduct a study, and if the *P*-value from his test is less than 5%, he will conclude that the evidence supports his theory. He queries 200 scientists and finds that 36, or 18%, are left-handed. Does this data support the biologist's theory?

Test af andel i en population

(se bogen: 8.1.2)

Example 8.3 About 13% of the population is left-handed. A biologist suspects that the scientific community is not like the general population in terms of handedness. He will conduct a study, and if the *P*-value from his test is less than 5%, he will conclude that the evidence supports his theory. He queries 200 scientists and finds that 36, or 18%, are left-handed. Does this data support the biologist's theory?

ukendt parameter

Model: Antal $X \sim \text{Binom}(n, p)$ hvor n er kendt.

Nulhypotese: $H_0 : p = p_0$. $= 0.13$ i eksemplet

Test: Teststørrelse: X. p-værdi: beregn binomialsandsynlighed, at X er mindst så ekstrem som den observerede værdi x. Brug $p = p_0$.

i eksemplet
$x = 36$
$n = 200$

$$P(X \le x) = \sum_{k=0}^{x} \binom{n}{k} p_0^k (1-p_0)^{n-k}.$$

$H_A : p < p_0$	p-værdi $= P(X \le x)$
$H_A : p > p_0$	p-værdi $= P(X \ge x) = 1 - P(X \le x-1)$
$\boxed{H_A : p \ne p_0}$	p-værdi $= 2 \cdot \min(P(X \le x), P(X \ge x))$ ← biologens valg

i R : $\mathbb{P}(X \le x) =$ pbinom (x, n, prob = p_0)

her: pbinom (36, 200, 0.13)

$\mathbb{P}(X \ge x) = 1 - \underbrace{\mathbb{P}(X < x)}$ $x \in$ udf. rum

$= 1 - \mathbb{P}(X \leq x-1)$
$= 1 - \text{pbinom}(35, 200, 0.13)$

```
p1 <- pbinom(36, 200, 0.13)
p2 <- 1-pbinom(35, 200, 0.13)
2 * min(p1, p2)
```

Fun fact:

når teststørrelsen T er kontinuert, så gælder, at under H_0

~~Tobs~~ p-value: $\mathbb{P}(T > t_{obs}$
$< t_{obs}$

hvis vi gentoges mange samples fra H_0: $\Rightarrow$ forskellig p-værdi hver gang

$\Rightarrow$ p-værdi af random sample $\sim U[0,1]$

Bootstrap t-tests

Bootstrap t-test

(se bogen: 8.2)

Formål: test $H_0 : \theta = \theta_0$.

Teststørrelse:

$$T = \frac{\hat{\theta} - \theta_0}{\widehat{\mathrm{SE}}(\hat{\theta})}.$$

Problem: Stikprøvefordeling af $\hat{\theta}$ er ikke normal $\implies$ T har ikke en t-fordeling.

Bootstrap-idé: Erstat stikprøvefordeling af T ved bootstrapfordelingen af

teoretisk fordeling

$$\boxed{T^* = \frac{\hat{\theta}^* - \hat{\theta}}{\widehat{\mathrm{SE}}(\hat{\theta}^*)}.}$$

vi får bootstrap fordeling

$\downarrow$

estimer $\mathbb{P}(T \leq t_{obs})$

$\mathbb{P}(T \geq t_{obs})$

Bootstrap t-test

(se bogen: 8.2)

Formål: test $H_0 : \theta = \theta_0$.

Teststørrelse:

$$T = \frac{\hat{\theta} - \theta_0}{\widehat{\mathrm{SE}}(\hat{\theta})}.$$

Problem: Stikprøvefordeling af $\hat{\theta}$ er ikke normal $\implies$ T har ikke en t-fordeling.

Bootstrap-idé: Erstat stikprøvefordeling af T ved bootstrapfordelingen af

$$T^* = \frac{\hat{\theta}^* - \hat{\theta}}{\widehat{\mathrm{SE}}(\hat{\theta}^*)}.$$

Det praktiske:

- Simuler n_{boot} bootstrap værdier $T_1^*, \ldots T_{n_{\text{boot}}}^*$.
- Bootstrap p-værdi beregnes som i permutationstest, p-værdi =andel af bootstrap T^* som er mindst så ekstrem som t (obs. t indregnet), e.g. p-værdi = $(\sum_{i=1}^{n_{\text{boot}}} \mathbf{1}(T_i^* \geq t) \ + 1)/(n_{\text{boot}} + 1)$, når $H_A : \theta > \theta_0$.

bedst

til middelværdi af pop.

Bootstrap T tests anbefales

temmeligt

- frem for almindelig T-test, når stikprøvefordeling af $\hat{\theta}$ er skæv.
- frem for permutationstests til sammenligning af middelværdier, når variansen er meget forskelligt. Når varianserne ikke er vildt forskelligt, foretrækkes permutationstests frem for bootstrap-t-test.
- stikprøven bør ikke være alt for lille ($n = 30$ er fint).

permutationstest

tester, om fordelinger es éns

Eksempel

Thursday, 12 March 2020 09.48

Eksempel 8.4: Arsenforurening af grundvand i Bangladesh

Situation: der anbefales en arsenindhold på maks 100 μ g/l i vand til vanding i landbrug.

Det videnskabelige problem: Er grundvandet i Bangladesh OK?

Hypoteser: $H_0 : \mu = 100$ versus $H_A : \mu > 100$.

? Ville en miljøaktivist vælge samme alternativ hypotese?

Eksempel 8.4: Arsenforurening af grundvand i Bangladesh

```
t.test(Bangladesh$Arsenic, mu = 100, alt = "greater")
t = 1.3988, df = 270, p-value = 0.08151
...
```

To obtain the *P*-value from the bootstrap *t* approach:

```
Arsenic <- Bangladesh$Arsenic
observedT <- t.test(Arsenic, mu = 100)$statistic
xbar <- mean(Arsenic)
n <- length(Arsenic)
Tstar <- numeric(N)
for (i in 1:N)
{
  bootx <- sample(Arsenic, n, replace = TRUE)
  Tstar[i] <- (mean(bootx)- xbar)/(sd(bootx)/sqrt(n))
}
hist(Tstar)
abline(v = observedT)

(sum(Tstar >= observedT)+1)/(N+1)
```

notebook

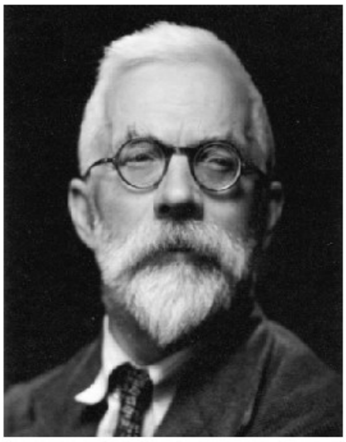
Hypotesetests: to forskellige tilgangsmåder

Tests skal hjælpe os at vurdere, om en observerede stikprøve kan forklares som udfald af en nulmodel, dvs, ved en nulhypotese.

P-værdier: Tilgang som blev gjort populær af Ronald Aylmer Fisher

Vurdering af hypoteser: R.A. Fisher vs. J. Neyman og E.S. Pearson

Sir R.A. Fishers's filosofi



Fisher (og mange andre):

> *Valider nulhypotesen ved hjælp af p-værdien.*

p-værdien angiver sandsynligheden for, at se et eksperimentelt resultat som er mindst lige så ekstrem som det observerede. Hvis p-værdien er lille, burde vi undersøge sagen nærmere.

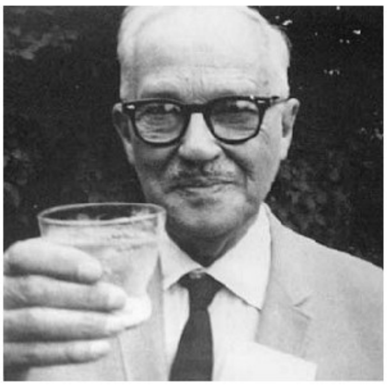
☞ H_A spiller ingen rolle i sig selv – vi bruger den kun for at fastlægge hvad "ekstrem"betyder.

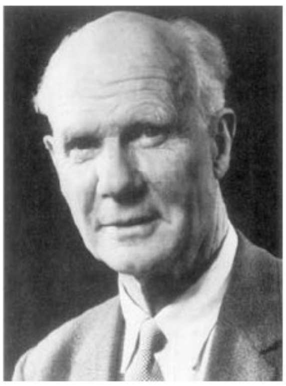
stikprøve $\underline{x}$ $\longrightarrow$ p værdi $\in [0,1]$

Jerzy Neyman og Egon Pearson prøvede at kaste det hele i en matematisk formalismus og endte med en ny tilgang.

Vurdering af hypoteser: R.A. Fisher vs. J. Neyman og E.S. Pearson

Neyman-Pearson hypotesetests





Jerzy Neyman(1894–1981), Egon Pearson (1895–1980)

Filosofi: Der antages, at enten H_0 eller H_A er sand.

kodeiva

Jerzy Neyman(1894–1981), Egon Pearson (1895–1980)

Filosofi: Der antages, at enten H_0 eller H_A er sand.
Valg mellem to hypoteser H_0 eller H_A skal afgøres vha. test.
Kontroller sandsynlighederne for fejlagtig beslutning.

kodering

$$\text{stikprøve } \underline{x} \longrightarrow \text{beslutning} \begin{cases} \text{forkast } H_0 & 0 \\ \text{ikj forkast } H_0 & 1 \end{cases}$$

Type1 og type 2 fejl

Hvilke situationer er muligt?

Sandheden	Beslutning: forkaste ikke H_0	forkaste H_0
H_0 er sandt	✓	type 1 fejl
H_A er sandt	type 2 fejl	✓

H_0 er sandt

Sandsynligheder vi ønsker at kontrollere:

$P(\text{type 1 fejl}) = P(H_0 \text{ forkastes} \mid H_0 \text{ er sandt})$

$P(\text{type 2 fejl}) = P(H_0 \text{ forkastes } \underline{\text{ej}} \mid \underline{H_A} \text{ er sandt})$

afhænger af den specielle valg af H_A

Eksempel for beregning af type 1 fejl

Eksempel 8.10 og 8.12: Normalfordelingsmodel med kendt varians

Example 8.10 We return to math SAT scores for 2009 college-bound seniors (scores are distributed $N(515, 116^2)$). Suppose local educators in a certain city want to know how their students compare with the national average. The educators have also decided that if their students average much lower than the national average of 515 points, then they will request more money from the city council for new teaching reforms. The educators will obtain a random sample of scores and test $H_0\colon\ \mu = 515$ versus $H_A\colon\ \mu < 515$, where μ denotes the mean math SAT score in their city. If their sample size is 100 and they decide that a sample average of 505 or less is their criterion for making the funding request, what is the probability they make a type I error? Assume that the standard deviation of scores in the city is $\sigma = 116$.

Data: 100 SAT test resultater (Scholastic Aptitude Test), $x_1, \ldots, x_{100}$.

Formål: Undersøg, om det gennemsnitlige resultat som førhen låg på 515 pts er blevet dårligere.

Model: $X_1, \ldots X_{100} \overset{\text{i.i.d.}}{\sim} N(\mu, 116^2)$.

Hypoteser: $H_0 : \mu = 515$ vs. $H_A : \mu < 515$.

Test: vi beslutter os for H_0 hvis $\bar{x} > 505$ og for H_A hvis $\bar{x} \le 505$.

Eksempel 8.10: Sandsynlighed for type 1 fejl

Test: vi beslutter os for H_0 hvis $\bar{x} > 505$ og forkaster H_0 hvis $\bar{x} \leq 505$.

$$
\begin{aligned}
P(\text{type 1 fejl}) &= P(H_0 \text{ forkastes} \mid H_0 \text{ er sand}) \\
&= P(\bar{X} \leq 505 \mid X_1, \ldots \overset{\text{i.i.d.}}{\sim} N(515, 116^2)) \\
&= P(\bar{X} \leq 505 \mid \bar{X} \sim N(515, \frac{116^2}{100})) \\
&= P(Z \leq \underbrace{\frac{505 - 515}{116/\sqrt{100}}}_{=-0.8621} \mid Z \sim N(0,1)) \qquad Z = \frac{\bar{X} - 515}{116/\sqrt{100}} \\
&= \Phi(-.8621) = 0.1943
\end{aligned}
$$

Kontrol af type 1 fejl ssh vha kritisk område

Neyman-Pearson signifikanstest

Mål: konstruer test således, at ssh for type 1 fejl kontrolleres.

Princippet af signifikanstests iflg Neyman og Pearson:
Byg et test der forkaster H_0, alt efter hvordan den observerede værdi af teststørrelse T ser ud.

Toleransgrænsne for type 1 fejl ssh α fastlægges i forvejen.

Definition 8.2:
Signifikansniveauet α er den største sandsyndlighed for type 1 fejl som vi synes er acceptabel.

Definition 8.3:
Betragt en test af H_0 mod H_A til signifikansniveau α, hvor der bruges en teststørrelse T. Mængden af alle værdier af T som medfører at H_0 forkastes, betegnes som kritisk område $\mathcal{R}$ af T. Grænserne af $\mathcal{R}$ betegnes som kritiske værdier.

- Teststørrelse $T(\underline{x}) \in \mathbb{R}$ $\quad T: \mathcal{X} \longmapsto \mathbb{R}$
- kritisk område $\mathcal{R}$
- beslutning: $T(\underline{x}) \in \mathcal{R} \iff$ forkaste H_0
  $T(\underline{x}) \notin \mathcal{R} \iff$ H_0 forkastes ej

ensided test $\quad$ kritisk værdi

$\mathcal{R} = (-\infty, C]$ $\quad$ hvis små værdier af T hentyder, at H_A må være sandt

$\mathcal{R} = [C, \infty)$ $\quad$ hvis store værdier af T ses hyppigt under H_A

tosided test

$\mathcal{R} = (-\infty, C_1] \cup [C_2, \infty)$

Eksempel 8.10 og 8.12: Normalfordelingsmodel med kendt varians

H_0

Example 8.10 We return to math SAT scores for 2009 college-bound seniors (scores are distributed $N(515, 116^2)$). Suppose local educators in a certain city want to know how their students compare with the national average. The educators have also decided that if their students average much lower than the national average of 515 points, then they will request more money from the city council for new teaching reforms. The educators will obtain a random sample of scores and test $H_0: \mu = 515$ versus $H_A: \mu < 515$, where μ denotes the mean math SAT score in their city. If their sample size is 100 and they decide that a sample average of 505 or less is their criterion for making the funding request, what is the probability they make a type I error? Assume that the standard deviation of scores in the city is $\sigma = 116$.

Data: 100 SAT test resultater (Scholastic Aptitude Test), $\boxed{x_1, \ldots, x_{100}}$. Stikprøve

Formål: Undersøg, om det gennemsnitlige resultat som førhen låg på 515 pts er blevet dårligere.

kendt varians

Model: $X_1, \ldots X_{100} \overset{\text{i.i.d.}}{\sim} N(\mu, 116^2)$.

Hypoteser: $H_0: \mu = 515$ vs. $H_A: \mu < 515$.

$R = (-\infty, 505)$

Test: vi beslutter os for H_0 hvis $\bar{x} \geq 505$ og for H_A hvis $\bar{x} < 505$.

$T(\underline{x}) = \bar{x}$

Eksempel 8.10: Sandsynlighed for type 1 fejl

Test: vi beslutter os for H_0 hvis $\bar{x} \geq 505$ og for H_A hvis $\bar{x} < 505$.

obs!

$$
\begin{aligned}
P(\text{type 1 fejl}) &= P(H_0 \text{ forkastes} \mid H_0 \text{ er sand}) \\
&= P(\bar{X} < 505 \mid X_1, \ldots \overset{\text{i.i.d.}}{\sim} N(515, 116^2)) \\
&= P(\bar{X} < 505 \mid \bar{X} \sim N(515, \frac{116^2}{100})) \\
&= P(Z < \underbrace{\frac{505 - 515}{116/\sqrt{100}}}_{=-0.8621} \mid Z \sim N(0,1)) \qquad Z = \frac{\bar{X} - 515}{116/\sqrt{100}} \\
&= \Phi(-.8621) = 0.1943
\end{aligned}
$$

i R
pnorm(-0.8621)

☞ Hvordan kan sandsynligheden for type 1 fejl kontrolleres, dvs. skulle grænsen (som var 505) vælges således at $P(\text{type 1 fejl}) = \alpha$?

signifikans niveau

Kontrol af type 1 fejl i normalfordelingsmodellen

(se bogen: Ex. 8.12)

Model: $X_1, \dots X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$.

Hypoteser: $H_0 : \mu = \mu_0$ vs. $H_A : \mu < \mu_0$.

Test: Forkast H_0, hvis $\bar{X} \leq C$.

Opgaven: Find grænse C således, at

$$P(\bar{X} \leq C \mid \bar{X} \sim N(\mu_0, \sigma^2/n)) \leq \alpha$$

Løsning:

$$C = \mu_0 + q \cdot \frac{\sigma}{\sqrt{n}}, \quad q = \Phi^{-1}(\alpha)$$

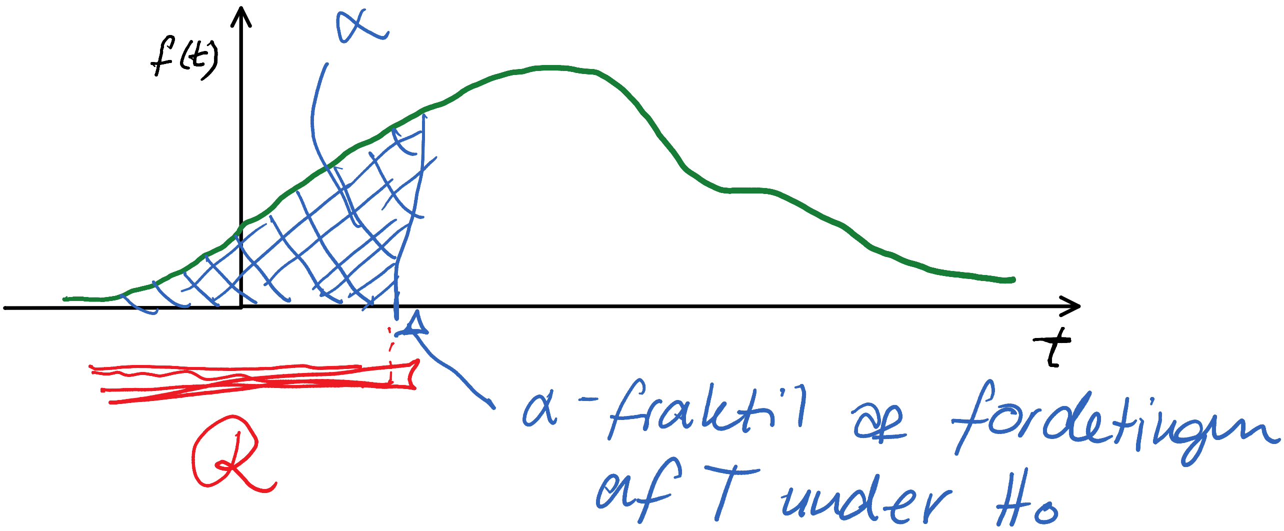
I eksemplet med SAT test (Example 8.10 og 8.12): der ønskes $\alpha = 0.1$. Kritisk værdi:

$$C = \mu_0 + \Phi^{-1}(\alpha) \cdot \frac{\sigma}{\sqrt{n}} = 515 + (-1.2816) \cdot \frac{116}{10} = 500.13.$$

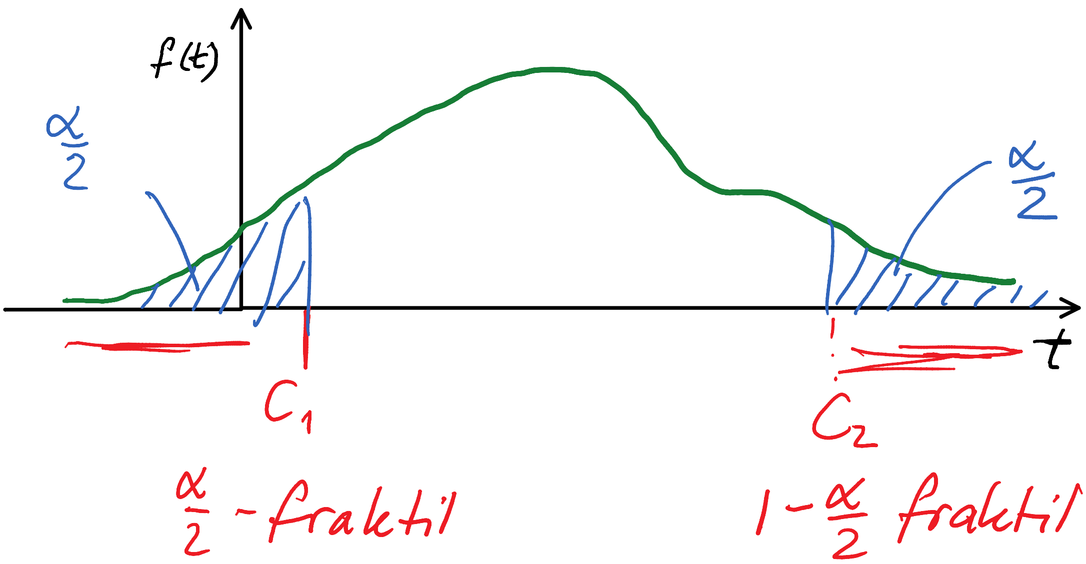
Beregning af kritiske værdier

Ingredienser:

- fordeling af teststørrelse T under H_0
- kriterier, hvornår T anses som ekstrem
  - små
  - store
  - både store og små
- signifikansniveau α



tosidet test



☐ Type I fejl, signifikansniveau, kritiske værdier: eksempel med diskret fordeling

Eksempel 8.13:
En terning mistænkes ikke at være fair. Specielt antal af ⚀

Eksperiment: 60 kast med terningenen, bestem antal X af ⚀

Model: udgangene af de enkelte kast er uafhængige og identisk fordelt.

$X \sim Bin(n=60, p)$

Nulhypotese: $H_0: p = 1/6 \xrightarrow{\text{under } H_0} EX = 10$

Alternativ $H_A: p \neq 1/6$

Tosided test: Forkast nulhypotesen, hvis den observerede X er enten meget lav eller meget høj.

$\leq C_1 \quad X \geq C_2 \qquad \mathcal{R} = (-\infty, C_1] \cup [C_2, \infty)$

Signifikansniveau: $\alpha = 0.05$

$P(X \in \mathcal{R} \mid H_0) \leq \alpha$

$= \underbrace{P(X \leq C_1)}_{\leq \frac{\alpha}{2} \; 0.025} + \underbrace{P(X \geq C_2)}_{\leq \frac{\alpha}{2} \; 0.025} \leq \alpha$

$\boxed{P(X \leq C_1)}$: pbinom(C_1, size = 60, prob = 1/6) ≤ 0.025

$\boxed{P(X \geq C_2)} = 1 - P(X < C_2) = 1 - P(X \leq C_2 - 1)$
$= 1 -$ pbinom ($C_2 - 1$, 60, 1/6) ≤ 0.025

Table 8.2 Cumulative probabilities for Binom(60, 1/6).

C_1	0	1	2	3	4	5
$P(X \leq C_1)$	1.77×10^{-5}	2.307×10^{-4}	0.0015	0.0063	0.0202	0.0512
C_2	~~16~~ 15	~~17~~ 16	~~18~~ 17	19	20	~~21~~ 20
$P(X \geq C_2)$	0.0648	0.0338	0.0164	0.0074	0.0031	0.0012

$P(X \leq C_1) \leq 0.025 \Rightarrow C_1 \leq 4$

$P(X \leq C_1) \leq 0.025 \Rightarrow C_1 \leq 4$

$P(X \geq C_2) \leq 0.025 \Rightarrow C_2 \geq 17$

$\mathcal{R} = \{0, 1, 2, 3, 4, 17, 18, 19, \dots, 60\}$

$\mathbb{P}(X \in \mathcal{R} \mid H_0) = 0.0202 + 0.0164 = 0.0366 \leq 0.05$

Definition

et test til det nominelle signifikansniveau α kaldes konservativ, når $\mathbb{P}(\text{type I fejl}) < \alpha$
liberal $> \alpha$

Type II fejl og styrke af tests (8.4.2)

Definition 8.4 *Power* is the probability of correctly rejecting a false null hypothesis.

$$1 - \beta = P(\text{Reject } H_0 \mid H_A \text{ true}). \qquad ||$$

ssh for type 2 fejl

$\beta = P($ Ho forkastes ej $\mid H_A$ er sandt

Sammensat hypotese: det alternative parameterrum indeholder mange elementer...

fx: $H_0: \mu = \mu_0$ $H_A: \mu \neq \mu_0 \Rightarrow$ fx $\mu = \mu_0 + 0.1$
i modellen $N(\mu, \sigma^2)$ $\mu = \mu_0 - 13$

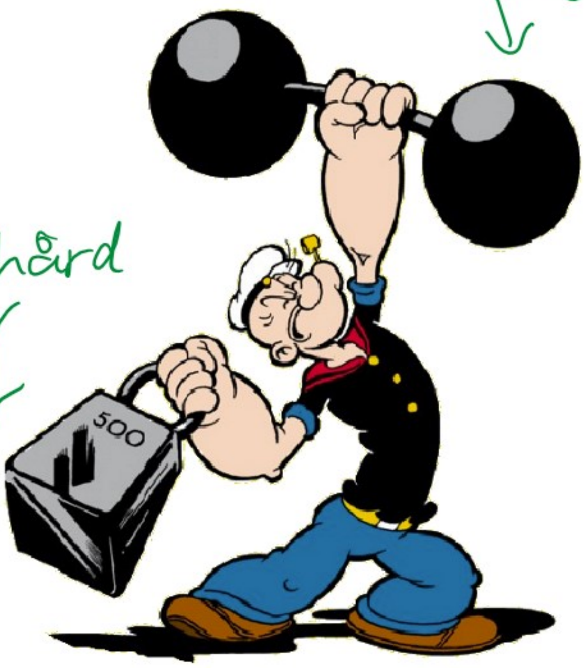
Styrke afhænger af den specielle H_A

nem

fx $\mu - 13$

hård

$\mu + 0.1$



Example 8.15 Suppose the heights of 2-year-old girls are normally distributed with a mean of 30 in. and a standard deviation of 6 in. Let μ denote the true mean heights of these girls. A researcher wishes to test

$$H_0 : \mu = 30 \quad \text{versus} \quad H_A : \mu < 30.$$

She plans to obtain a random sample of 30 girls and measure their heights using $\alpha = 0.05$ for a significance level. What is the probability of correctly rejecting

$$H_0 : \mu = 30 \quad \text{versus} \quad H_A : \mu < 30.$$

She plans to obtain a random sample of 30 girls and measure their heights using $\alpha = 0.05$ for a significance level. What is the probability of correctly rejecting the null hypothesis if the mean height of 2-year-old girls in this community is actually 27 in.?

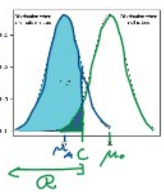
$X_1, \dots, X_{30} \sim N(\mu, \sigma^2 = 6^2)$

$H_0: \mu = 30$ vs $H_A\ \mu < 30$

testes ved hjælp $\bar{X} = \frac{1}{30}\sum_{i=1}^{30} X_i$, $\alpha = 0.05$

? hvor stor er styrken når $\mu = 27$?

① find det kritiske område



$R = (-\infty, C]$, hvor

$$C = \mu_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(\alpha)$$

$$= 30 + \frac{6}{\sqrt{30}} \cdot (-1.645)$$

$$= \underline{\underline{28.2}}$$

② beregn $1-\beta$ = Styrken ved $\mu = 27$

$P(H_0 \text{ forkastes} \mid \text{sande } \mu = 27, \sigma = 6, X_i \sim N(27, 6^2))$

$$= P\left(\bar{X} \leq 28.2 \mid \bar{X} \sim N\left(27, \frac{6^2}{30}\right)\right)$$

$$= \dots \ \underline{\underline{0.8632}}$$