Grundläggande matematisk statistik

Hypotestest
Del I: Introduktion

Uwe Menzel, 2018
uwe.menzel@matstat.org
www.matstat.org

Allmänt om hypotestest
Nollhypotes och alternativ hypotes

- **Mål med (parametriskt) test:** att avgöra, om en fördelningsparameter ($\mu, \Delta\mu, \sigma, \lambda, \ldots$) kan ha ett förmodat (hypotetiskt) värde.
  - Vi kan t. ex. förmoda att väntevärdet i en normalfördelad population är $\mu = 20$.
- Hypotesen för parametern kallas nollhypotes (H_0)
  - nollhypotesen kan alltså t. ex. vara $\boldsymbol{H_0:\ \mu = 20}$.
- Nollhypotesen ställs mot en alternativ hypotes (H_a) för parametern.
  - den alternativa hypotesen kan t. ex. vara $\boldsymbol{H_a: \mu > 20}$.

Allmänt om hypotestest
Exempel för hypoteser

- $X \sim N(\mu, \sigma)$ (det är känt att normalfördelning föreligger)
  - nollhypotes H_0: $\mu = 20$ (väntevärdet är 20)
  - alternativ hypotes H_a: $\mu \neq 20$ (väntevärdet är **inte** 20)
- $X \sim N(\mu, \sigma)$
  - nollhypotes H_0: $\mu = 20$ (väntevärdet är 20)
  - alternativ hypotes H_a: $\mu > 20$ (väntevärdet är **större** än 20)
- $X \sim N(\mu, \sigma)$
  - nollhypotes H_0: $\mu = 20$ (väntevärdet är 20)
  - alternativ hypotes H_a: $\mu < 20$ (väntevärdet är **mindre** än 20)
- $X \sim Bin(n, p)$ (det är känt att binomialfördelning föreligger)
  - nollhypotes H_0: $p = {}^1/_2$ (sannolikheten ”att lyckas” är ${}^1/_2$)
  - alternativ hypotes H_a: $p > {}^1/_2$ (sannolikheten ”att lyckas” är **större** än ${}^1/_2$

www.matstat.org

Allmänt om hypotestest
Stickprov och testvariabel

- För att avgöra vilken hypotes som stämmer tas ett slumpmässigt stickprov från populationen.
- Från stickprovet beräknas ett numeriskt värde - en observation av en så kallad testvariabel (eller teststatistik, ”statistic” i engelsk litteratur)
  - Testvariabeln kan t. ex. vara: det (standardiserade) **medelvärdet för stickprovet** eller antalet ”lyckade” Bernoulli-försök (för binomialfördelning) – i allmänhet en skattning för parametern eller en storhet som härleds från skattningen.
- Det kollas nu om det observerade värdet för testvariabeln står i motsats till nollhypotesen. Om så är fallet, förkastas nollhypotesen och den alternativa hypotesen accepteras.
  - En motsats finns t. ex. när det så kallade **p-värdet** är litet eller när testvariabeln överskrider ett **kritiskt värde**. Mer om det senare …
- Att nollhypotesen förkastas är i allmänhet ett önskat resultat!
  - I enlighet därmed väljs i allmänhet också nollhypotesen. Vill man t. ex. visa att två populationer har olika väntevärden, skulle man välja H_0: $\Delta\mu = \mu_1 - \mu_2 = 0$ och sedan försöka att motbevisa detta.

www.matstat.org

Allmänt om hypotestest
Att förkasta nollhypotesen

Det finns olika metoder för att pröva om nollhypotesen borde förkastas.

1. Jämförelse av testvariabelns stickprovsvärde med ett kritiskt värde:

- Nollhypotesen förkastas när testvariabelns värde överskrider ett kritiskt värde i riktning mot den alternativa hypotesen.

2. Jämförelse av testvariabelns värde med ett konfidensintervall:

- Nollhypotesen förkastas när det hypotetiska värdet för parametern ligger utanför konfidensintervallets gränser för parametern.

- **3.** Jämförelse av p-värdet med signifikansnivån:
  - Nollhypotesen förkastas om p-värdet är mindre än ett förut bestämt värde – den så kallade signifikansnivån.

Alla tre metoder leder till samma slutsatser.

www.matstat.org

1. Jämförelse av testvariabelns värde med ett kritiskt värde

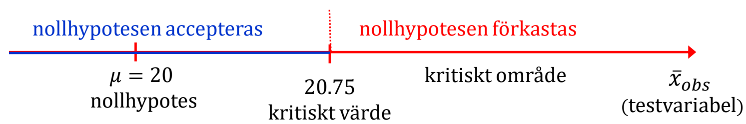
Exempel ”torktid”:

- Ett bolag tillverkar färg och påstår att torktiden bara är 20 minuter.
- Konsumentverket tvivlar på detta påstående och misstänker att torktiden i verkligheten kunde vara längre.
- Det antas att torktiden är normalfördelad: $X \sim N(\mu, \sigma)$
  - nollhypotes H_0: $\mu = 20$
  - alternativ hypotes: H_a: $\mu > 20$
- Ett stickprov ska testa bolagets påstående:
  - det målas 36 brädor ($n = 36$)
  - torktiderna mäts (torktiderna = slumpvariabler X_i)
  - testvariabel: $\bar{X}_{obs} = {}^{1}/_{36} \cdot \sum_{i=1}^{36} X_i$ (genomsnittlig torktid i stickprovet = skattning för μ)
- För vilket värde av testvariabeln ($\bar{X}_{obs}$) förkastas H_0 ?
  - → kritiskt värde måste fastslås

www.matstat.org

Val av det kritiska värdet

- Nollhypotesen ska förkastas för $\bar{x}_{obs} \geq \omega_{krit} = 20.75$ - ett sådant resultat stödjer den alternativa hypotesen H_a: $\mu > 20$
- Valet av det kritiska värdet verkar vara godtyckligt – vi ska senare se hur det kritiska värdet ω_{krit} **beräknas** på grundval av felrisken.
- Det kritiska värdet ω_{krit} definerar det kritiska området – om testvariabeln hamnar i detta område förkastas H_0.



OBS!: Formuleringen"nollhypotesen accepteras" betyder **inte** att denna därmed är bevisad ! Det kan bara konstateras att det aktuella stickprovet inte motsätter sig nollhypotesen och att nollhypotesen därför inte kan förkastas.

www.matstat.org

Feltyper

Feltyper: När ett kritiskt värde defineras är det ofrånkomligt att två typer av fel riskeras:

Fel typ I: det är möjligt att stickprovet ger $\bar{x}_{obs} > 20.75$ **även om** nollhypotesen $\mu = 20$ är korrekt:

$\rightarrow$ Nollhypotesen förkastas trots att den är sann.

Fel typ II: det är möjligt att stickprovet ger $\bar{x}_{obs} < 20.75$ **även om** nollhypotesen $\mu = 20$ är falsk:

$\rightarrow$ Nollhypotesen accepteras trots att den är falsk.

Dessa fel är oundvikliga därför att vi fattar våra beslut på grundval av ett (eller flera) stickprov som styrs av slumpen.

.

	H_0 accepteras	H_0 förkastas
H_0 är sann	korrekt	Fel typ I
H_0 är falsk	Fel typ II	korrekt

www.matstat.org

Sannolikheten att göra ett fel typ I: α

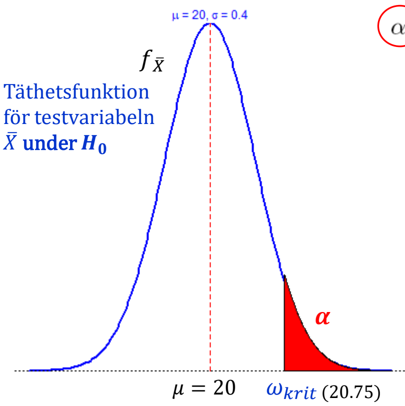
Fel typ I: Nollhypotesen förkastas trots att den är sann.

Antaganden: normalfördelning, $\sigma = 2.4$ känd. Under H_0 gäller $\boldsymbol{\mu = 20}$.

$$X_i \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \bar{X} \sim N(\underline{\underline{20}}, 0.4) \quad \text{(om } H_0 \text{ sann)}$$

$f_{\bar{X}}$

Täthetsfunktion för testvariabeln $\bar{X}$ **under** $\boldsymbol{H_0}$



$$\begin{aligned}
\alpha &= P(\text{Fel typ I}) \\
&= P(H_0 \text{ förkastas} \mid H_0 \text{ sann}) \\
&= P(\bar{X} > 20.75 \mid \mu = 20) \\
&= 1 - P(\bar{X} \leq 20.75 \mid \mu = 20) \\
&= 1 - F_{\bar{X}}(20.75) \quad \text{för } \mu = 20 \\
&= 1 - \Phi\left(\frac{20.75 - 20}{0.4}\right) \\
&= 1 - \Phi(1.875) = 0.03
\end{aligned}$$

3% risk för fel typ I

Man kan förminska α genom att **förstora** ω_{krit}.

uwe.menzel@matstat.org

Sannolikheten att göra ett fel typ II: β

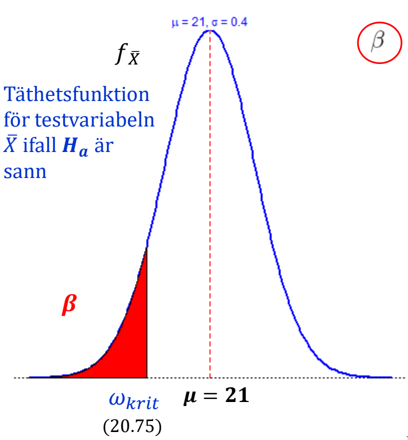
Fel typ II: Nollhypotesen accepteras trots att den är falsk.

Antaganden: $\sigma = 2.4$ känd; H_0 falskt, t. ex. $\boldsymbol{\mu = 21}$

$$X_i \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \bar{X} \sim N(\underline{\underline{21}}, 0.4) \quad (H_0 \text{ falsk})$$

$f_{\bar{X}}$

Täthetsfunktion för testvariabeln $\bar{X}$ ifall $\boldsymbol{H_a}$ är sann



$$\begin{aligned}
\beta &= P(\text{Fel typ II}) \\
&= P(H_0 \text{ accepteras} \mid H_0 \text{ falsk}) \\
&= P(\bar{X} \leq 20.75 \mid \mu = 21) \\
&= F_{\bar{X}}(20.75) \quad \text{för } \mu = 21 \\
&= \Phi\left(\frac{20.75 - 21}{0.4}\right) \\
&= \Phi(-0.625) \\
&= 1 - \Phi(0.625) = 0.266
\end{aligned}$$

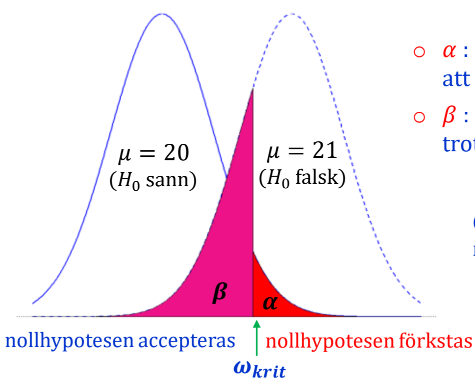
ca. 27% risk för fel typ II

Man kan förminska β genom att **förminska** ω_{krit}.

uwe.menzel@matstat.org

Kompromiss mellan α och β

- Man kan förminska α (fel typ I) genom att **förstora** ω_{krit}.
- Man kan förminska β (fel typ II) genom att **förminska** ω_{krit}.
- → Kompromiss mellan α och β behövs!



- α : Fel typ I : H_0 förkastas trots att den är sann
- β : Fel typ II : H_0 accepteras trots att den är falsk

Om ω_{krit} blir större, blir α mindre, men β större

uwe.menzel@matstat.org

Justering av sannolikheten för fel typ I (α)

- I praktiken fastslås α före testet.
- Genom detta bestämmer användaren hur stort felet av typ I ska vara.
- Värdet α kallas signifikansnivå.
- Ofta väljer man $\alpha = 0.05$ eller $\alpha = 0.01$.
- När α är fastlagt kan ω_{krit} beräknas på följande sätt:

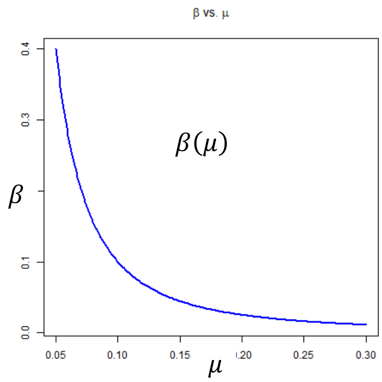
$$
\begin{aligned}
\alpha &= P\,(H_0 \text{ förkastas} \mid H_0 \text{ sann}) && \text{fel typ I} \\
&= P\left(\bar{X} > \omega_{krit} \mid \mu = 20\right) && \omega_{krit} \text{ måste bestämmas} \\
&= 1 - P\left(\bar{X} \leq \omega_{krit} \mid \mu = 20\right) && \text{allt under } H_0\text{: } \mu = 20 \\
&= 1 - F_{\bar{X}}\left(\omega_{krit}\right) && F_{\bar{X}} = \text{fördelningsfunktion för } \bar{X} \\
&= 1 - \Phi\left(\frac{\omega_{krit} - 20}{0.4}\right) && \text{Ekvation för att beräkna } \omega_{krit} \text{ (med tabell över } \Phi\text{), } \alpha \text{ är förbestämd}
\end{aligned}
$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

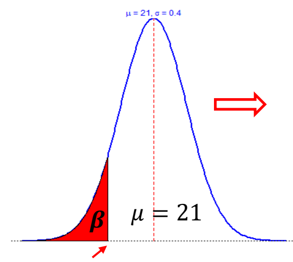
uwe.menzel@matstat.org

Justering av sannolikheten för fel typ II (β) ?

- Fel typ II (β) kan inte justeras på det sättet.
- I exemplet räknades för H_a med $\mu = 21$. Detta värde känner vi dock inte till (den alternativa hypotesen var H_a: $\mu > 20$).
- Man kan beräkna en funktion $\beta(\mu)$ i ett visst intervall:



Om μ blir större förskjuts fördelningen mot höger. Om samma kritiska värde bibehålls blir β mindre:



krit. värde

www.matstat.org

Valet av signifikansnivån α

- I praktiken fastslås signifikansnivån α före testet.
- Det kritiska värdet ω_{krit} beräknas sedan under H_0.
- Ofta väljs $\alpha = 0.05$ eller $\alpha = 0.01$.
  - $\alpha = 0.05$: "i det långa loppet" förkastas H_0 i 1 av 20 test felaktigt
  - $\alpha = 0.01$: " i det långa loppet" förkastas H_0 i 1 av 100 test felaktigt

Konsekvenser av valet av α :

... för exemplet " torktid":

Ett mindre α, t. ex. $\alpha = 0.01$ har följande konsekvenser:

- ω_{krit} förskjuts till högre värden (20.75 → 20.93, se nere)
- det blir mera osannolikt att nollhypotesen förkastas
- bättre för bolaget – eventuellt sämre för konsumenterna

$$\alpha = 0.01 = 1 - \Phi\left(\frac{\omega_{krit} - 20}{0.4}\right) = 1 - \Phi(x)$$
$$x = \Phi^{-1}(0.99) = 2.326$$
$$\omega_{krit} = x \cdot 0.4 + 20 = \underline{\underline{20.93}}$$

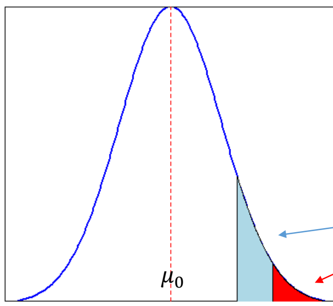
uwe.menzel@matstat.org

Valet av signifikansnivån α

Ett mindre $\boldsymbol{\alpha}$ betyder:

- Risken för felet typ I förminskas, risken för felet typ II (β) förhöjs.
- Man kan vara säkrare på att H_0 verkligen är falsk när den förkastas.

Täthetsfunktion för testvariabeln $\bar{X}$ **under $\boldsymbol{H_0}$**



om $\alpha = \alpha_2$ förkastas H_0 först när $\bar{x}_{obs}$ avviker ännu mer från μ_0 (i riktning mot H_a).

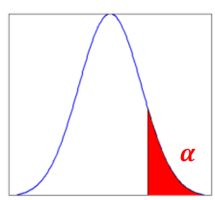
$\alpha_1 = 0.05$

$\alpha_2 = 0.01$

www.matstat.org

Olika alternativa hypoteser (H_a)

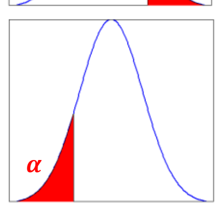
H_0	H_a	test
$\mu = \mu_0$	$\mu > \mu_0$	ensidigt



förkasta H_0 för $\bar{x}_{obs} > \omega_\alpha$

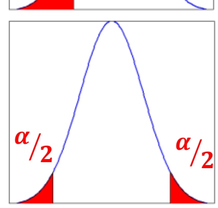
ω_α = kritiskt värde

H_0	H_a	test
$\mu = \mu_0$	$\mu < \mu_0$	ensidigt



förkasta H_0 för $\bar{x}_{obs} < \omega_\alpha$

H_0	H_a	test
$\mu = \mu_0$	$\mu \neq \mu_0$	tvåsidigt



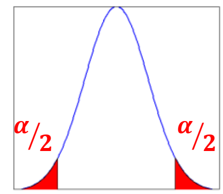
förkasta H_0 för $\bar{x}_{obs} < -\omega_{\alpha/2}$ eller $\bar{x}_{obs} > +\omega_{\alpha/2}$ (för symmetrisk täthetsfunktion)

www.matstat.org

Olika alternativa hypoteser

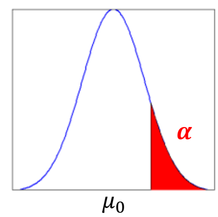
Exempel, tvåsidigt test:

- Ett byggelement måste ha diametern 30 cm, det får varken vara för stort eller för litet
- H_0: $\mu = 30$; förkasta H_0 om
  - $\bar{x}_{obs} > 30.5$ eller
  - $\bar{x}_{obs} < 29.5$



Exempel, ensidigt test:

- En gammal maskin producerar i genomsnitt 2500 komponenter per dag: $\mu_0 = 2500$ (normalfördelat).
- Köpa en ny, dyrare maskin ?
  - bara när μ är större för den nya maskinen!
  - välj H_a: $\mu > \mu_0$
  - köp bara om H_0 kan förkastas!



Beräkning av det kritiska värdet utav α
när σ är okänt

Om nollhypotesen H_0: $\mu = \mu_0$ är sann gäller $X_i \sim N(\mu_0, \sigma)$ och därför:

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right) \Longrightarrow Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

en $N(0{,}1)$-variabel betecknas ofta med Z

Standardavvikelsen σ är dock okänd, därför kan Z inte användas som testvariabel. (Ett numeriskt värde för en observation av Z kan inte beräknas). Vi vet dock (föreläsning **F9**) att vi får en t-fördelad variabel, om σ i uttrycket ovan ersätts med skattningen S:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$$

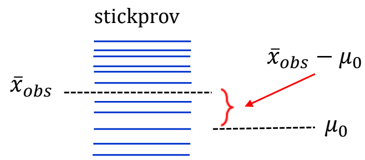
$t(n-1)$: Students t–fördelning för $n-1$ frihetsgrader

Testvariabeln T är indikativ för frågan om populationens väntevärde kan ha det hypotetiska värdet μ_0 eller inte. Avviker det observerade medelvärdet $\bar{x}_{obs}$ mycket från μ_0 och är samtidigt stickprovets observerade standardavvikelse s inte alltför stor, då antar en observation av T stora positiva eller negativa värden. Ligger dessa värden in riktning mot den alternativa hypotesen H_a borde vi vara benägna att acceptera H_a.

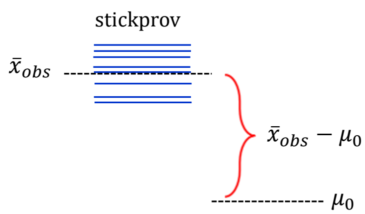
Testvariabeln T är indikativ för testet

En observation för T ska betecknas med t_{obs} :

$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}}$$



Det observerade medelvärdet $\bar{x}_{obs}$ avviker inte mycket från μ_0 , dessutom är stickprovets standardavvikelse s stor. En observation för T antar inte alltför stora positiva eller negativa värden. **Nollhypotesen $\boldsymbol{\mu = \mu_0}$ förkastas inte.**



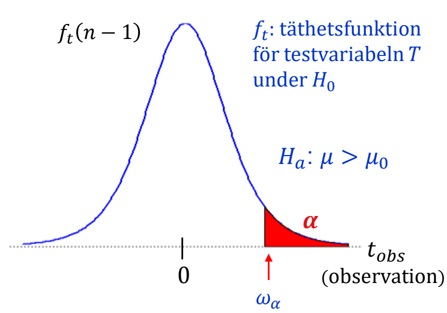
Det observerade medelvärdet $\bar{x}_{obs}$ avviker mycket från μ_0, dessutom är stickprovets standardavvikelse s liten. En observation för T antar stora positiva eller negativa värden. Om dessa värden ligger i riktning mot den alternativa hypotesen, kan **nollhypotesen $\boldsymbol{\mu = \mu_0}$ förkastas**

www.matstat.org

Beräkning av det kritiska värdet utav α

för H_a: $\mu > \mu_0$

t-fördelningen är symmetrisk runt noll. Om nollhypotesen är sann borde en observation t_{obs} inte ligger långt ifrån nollpunkten, för då är $\bar{x}_{obs} \approx \mu_0$. Vi är t.ex. benägna att acceptera den alternativa hypotesen H_a: $\mu > \mu_0$, om $\bar{x}_{obs} \gg \mu_0$ och om s samtidigt inte är alltför stort. I det här fallet antar t_{obs} ett stort positivt värde. När slutligen $t_{obs} > \omega_\alpha$ förkastas H_0 till förmån av den alternativa hypotesen.



$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t\,(n-1)$$

$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

ω_α: kritiskt värde

www.matstat.org

Kritiskt värde för H_a: $\mu > \mu_0$

För H_a: $\mu > \mu_0$ och en förut bestämd signifikansnivå α kan det kritiska värdet ω_α för testvariabeln T beräknas genom att kräva att sannolikheten för ett fel typ I ska vara lika med α:

$$\boxed{P\left(\underbrace{\frac{\bar{X}-\mu_0}{S/\sqrt{n}} > \omega_\alpha}_{H_0 \text{ förkastas}} \mid H_0 \; sann\right) = \alpha}$$

fel typ I

ekvation för att beräkna ω_α för förbestämd α (= sannolikhet för fel typ I)

Under H_0 är den vänstra termen i parantesen $t(n-1)$-fördelad, vi kan alltså skriva (betingningen "| H_0 *sann*" är därmed inräknad):
(uttrycket är bara t-fördelat om H_0 är sann, alltså när $\mu = \mu_0$)

$$P\left(T > \omega_\alpha\right) = \alpha \quad \text{med} \quad T \sim t(n-1)$$

Allmänt gäller: $P\left(T > t_\alpha(n-1)\right) = \alpha$

$t_\alpha(n-1)$ = kvantil för t-fördelningen med $f = n-1$ och signifikansnivå α

Genom att jämföra de sista uttrycken får vi det kritiska värdet $\omega_\alpha = t_\alpha(n-1)$.
H_0 förkastas alltså ifall en observation för T blir större än $t_\alpha(n-1)$.

www.matstat.org

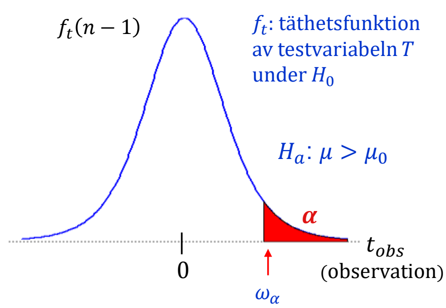
Kritiskt värde för H_a: $\mu > \mu_0$

Nollhypotesen förkastas om $t_{obs} > t_\alpha(n-1)$. Dessa värden utgör det kritiska området för t_{obs} för signifikansnivån α. Den kritiska regionen betecknas med Ω_α :

$$\Omega_\alpha = \{t_{obs} > t_\alpha(n-1)\}$$

$$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t\,(n-1)$$

$$t_{obs} = \frac{\bar{x}_{obs}-\mu_0}{s/\sqrt{n}}$$



f_t: täthetsfunktion av testvariabeln T under H_0

H_a: $\mu > \mu_0$

H_0 förkastas om observationen t_{obs} hamnar i det kritiska området (röd).

$t_\alpha(n-1)$ = kvantil för t-fördelningen med $f = n-1$ och signifikansnivå α

www.matstat.org

Kritiskt värde för $\bar{X}$ för H_a: $\mu > \mu_0$

Vanligtvis utförs testet med testvariabeln T. Med hjälp av det ovan beräknade kritiska värdet för t_{obs} kann man dock enkelt härleda ett kritiskt värde även för $\bar{x}_{obs}$. Vi hade sett att nollhypotesen förkastas ifall $t_{obs} > t_\alpha(n-1)$, alltså om

$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}} > t_\alpha(n-1)$$

H_0 förkastas alltså om $\quad \bar{x}_{obs} > \mu_0 + t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$

Det kritiska värdet med hänsyn till $\bar{X}$ är därför:

$$\omega_{\bar{X}} = \mu_0 + t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$$

Nollhypotesen förkastas alltså om $\bar{x}_{obs} > \omega_{\bar{X}}$.

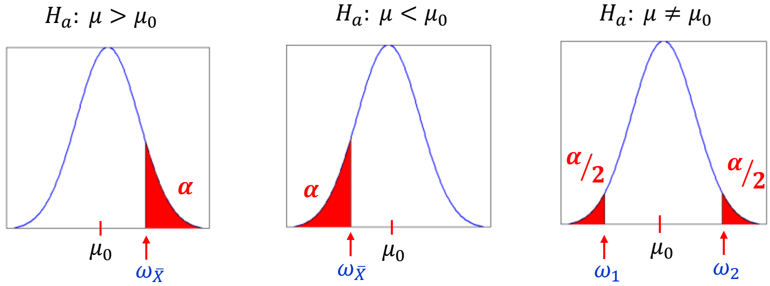
De kritiska värdena för t_{obs} resp. $\bar{x}_{obs}$ för de andra alternativa hypoteserna kan härledas analogt (se föreläsning **F12**). I följande listas de kritiska områdena för $\bar{x}_{obs}$:

www.matstat.org

De kritiska värdena för $\bar{x}_{obs}$ för olika alternativhypoteser

Lokaliseringen av det kritiska området (röd) för $\bar{x}_{obs}$ beror på valet av den alternativa hypotesen.

H_a: $\mu > \mu_0$ | H_a: $\mu < \mu_0$ | H_a: $\mu \neq \mu_0$



$\omega_{\bar{X}} = \mu_0 + t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$ | $\omega_{\bar{X}} = \mu_0 - t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$ | $\omega_1 = \mu_0 - t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$

$\omega_2 = \mu_0 + t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$

ensidiga test

tvåsidigt test

uwe.menzel@matstat.org

Sammanfattning
Jämförelse av testvariabeln med ett kritiskt värde

1. Definera nollhypotesen (H_0) och den alternativa hypotesen (H_a)
2. Slå fast signifikansnivån $\alpha = P(fel\ typ\ I \mid H_0)$
3. Bestäm testvariabeln ($\bar{X}_{obs}, t_{obs}, \ldots$) och det kritiska värdet ω_{krit} för den. Fördelningen för testvariabeln under H_0 måste vara känd.
4. Beräkna testvariabelns stickprovsvärde ($\bar{x}_{obs}, t_{obs}, \ldots$). .
5. Förkasta nollhypotesen H_0 på signifikansnivå α om observationen av testvariabeln ligger i det kritiska området, annars ”acceptera” H_0

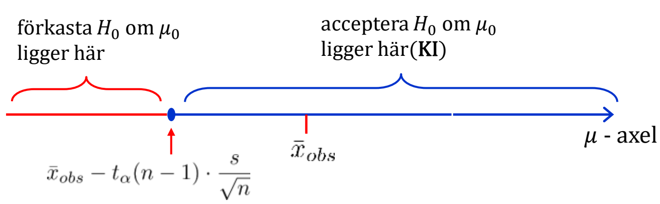
www.matstat.org

2. Jämförelse testvariabel ⇄ konfidensintervall

Ensidigt test $H_0\colon \mu = \mu_0 \quad H_a\colon \mu > \mu_0$

Vi hade: förkasta H_0 om $\bar{x}_{obs} > \mu_0 + t_\alpha(n-1)\cdot\frac{s}{\sqrt{n}}$

Det betyder: förkasta H_0 om $\mu_0 < \bar{x}_{obs} - t_\alpha(n-1)\cdot\frac{s}{\sqrt{n}}$



Kom ihåg: ensidigt konfidensintervall för μ (föreläsning **F9**):

$$I_\mu = \left(\bar{x}_{obs} - t_\alpha(n-1)\cdot\frac{s}{\sqrt{n}}\ ;\ +\infty\right)$$

detta är precis det blå markerade området ovan

Nollhypotesen förkastas **inte** (”accepteras”), om den hypotetiska parametern ligger innanför konfidensintervallets gränser.

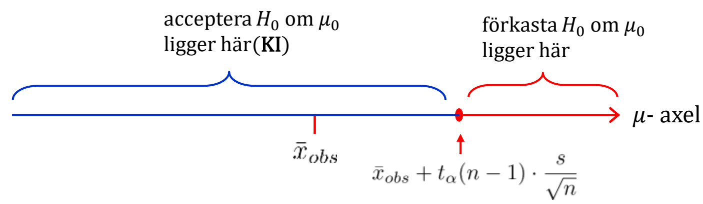
uwe.menzel@matstat.org

Jämförelse testvariabel ⇄ konfidensintervall

Ensidigt test H_0: $\mu = \mu_0$ H_a: $\mu < \mu_0$

Vi hade: förkasta H_0 om $\bar{x}_{obs} < \mu_0 - t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$

Det betyder : förkasta H_0 om $\mu_0 > \bar{x}_{obs} + t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}}$



Kom ihåg : ensidigt konfidensintervall för μ (föreläsning **F9**):

$$I_\mu = \left(-\infty \; ; \; \bar{x}_{obs} + t_\alpha(n-1) \cdot \frac{s}{\sqrt{n}} \right)$$

detta är precis det blå markerade området ovan

Nollhypotesen förkastas **inte** ("accepteras"), om den hypotetiska parametern ligger innanför konfidensintervallets gränser.

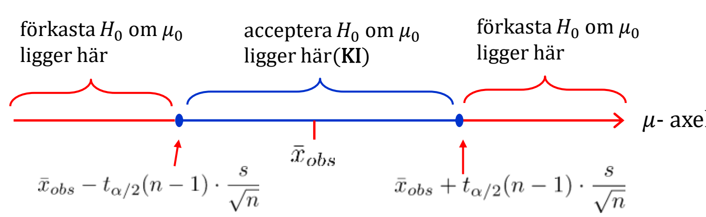
uwe.menzel@matstat.org

Jämförelse testvariabel ⇄ konfidensintervall

Tvåsidigt test H_0: $\mu = \mu_0$ H_a: $\mu \neq \mu_0$

förkasta H_0 om $\bar{x}_{obs} < \mu_0 - t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$ eller $\bar{x}_{obs} > \mu_0 + t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$

→ förkasta H_0 om $\mu_0 > \bar{x}_{obs} + t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$ eller $\mu_0 < \bar{x}_{obs} - t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$



Kom ihåg : tvåsidigt konfidensintervall för μ (föreläsning **F9**):

$$I_\mu = \bar{x}_{obs} \pm t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$$

detta är precis det blå markerade området ovan

Nollhypotesen förkastas **inte** ("accepteras"), om den hypotetiska parametern ligger innanför konfidensintervallets gränser.

uwe.menzel@matstat.org

Sammanfattning
Jämförelse av testvariabeln med konfidensintervallet

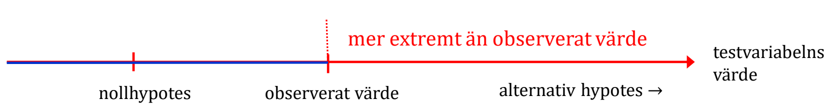
1. Beräkna ett konfidensintervall (KI) med konfidensgrad $1 - \alpha$ för parametern θ som ska testas på grundval av ett stickprov.
2. Acceptera $H_0: \theta = \theta_0$ på signifikansnivån α om den hypotetiska parametern θ_0 ligger innanför konfidensintervallets gränser.
3. Förkasta $H_0: \theta = \theta_0$ på signifikansnivå α om den hypotetiska parametern θ_0 ligger utanför detta konfidensintervall,

- Testvariabelmetoden och konfidensmetoden leder till samma slutsatser.

3. Jämförelse p-värde ⇄ signifikansnivå

Definition p-värde: p-värdet är sannolikheten av en erhållen observation, eller ännu mer extrema observationer, givet att nollhypotesen är sann.

Observationen är här värdet av testvariabeln som erhållits. Ett värde av testvariabeln anses som ”mer extremt” om det ligger ”längre bort från nollhypotesen”, gentemot den alternativa hypotesen; det är mindre sannolikt än $\bar{x}_{obs}$ om H_0 är sann.



Föt att kunna beräkna p - värdet måste man välja en tesvariabel med känd fördelning under H_0 - skattnigen för parametern eller ett uttryck som är härlett från en sådan skattning, t. ex. den standardiserade skattningen.

p-värdet för normalfördelningen

Exempel 1: Det ska testas om väntevärdet av en normalfördelad population är $\mu = 20$ eller större:

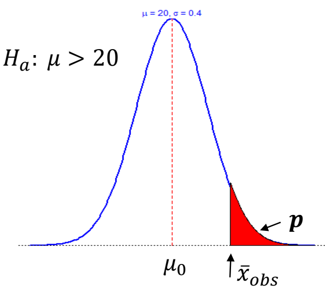
Hypoteser: H_0: $\mu = \mu_0 = 20$; H_a: $\mu > 20$

Testvariabel: skattning för μ, medelvärdet $\bar{X}$

Fördelning av testvariabeln "under H_0": $X_i \sim N(\mu_0, \sigma) \Rightarrow \bar{X} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$

För enkelhetens skull antar vi att standardavvikelsen σ är känd. Därmed är fördelningen av $\bar{X}$ under H_0 helt känd.

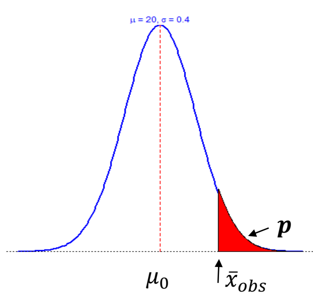
H_a: $\mu > 20$



Låt den blåa kurvan vara täthetsfunktionen för $\bar{X}$ under antagandet att nollhypotesen är sann: $f_{\bar{X}}(x \mid H_0)$. Låt det observerade värdet av testvariabeln vara $\bar{x}_{obs}$. Därmed är arean av det röd markerade området lika med p-värdet = sannolikhet av observationen $\bar{x}_{obs}$ eller ännu mer extrema värden i riktning mot den alternativa hypotesen.

www.matstat.org

p-värdet för normalfördelningen



Låt den blåa kurvan vara täthetsfunktionen för $\bar{X}$ under antagandet att H_0 är sann: $f_{\bar{X}}(x \mid H_0)$. Låt det observerade värdet av testvariabeln vara $\bar{x}_{obs}$. Därmed är arean av det röd markerade området lika med p-värdet

$$P_{wert} = \int_{\bar{x}_{obs}}^{\infty} f_{\bar{X}}(x \mid H_0)\, dx$$

för H_a: $\mu > 20$

Eftersom integralen inte alltid är lätt att beräkna är det oftast bättre att använda det standardiserade medelvärdet som testvariabel, mer om det senare …

Ett mycket litet p-värde betyder:

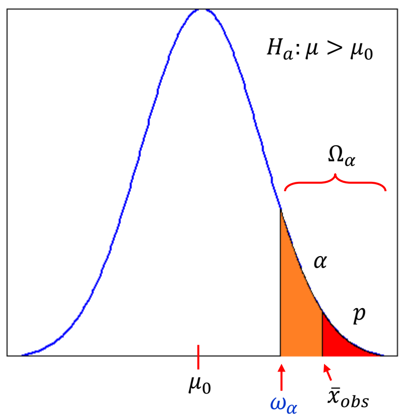
- Det observerade värdet av testvariabeln (= "observation") är osannolikt om man antar att nollhypotesen H_0 är sann.
- Nollhypotesen förkastas därför.

www.matstat.org

p-värdet och kritiskt område

- Om observationen $\bar{x}_{obs}$ ligger i det kritiska området för signifikansnivån α är p–värdet mindre än α och tvärtom.
- Nollhypotesen förkastas alltså för $p < \alpha$ (på signifikansnivå α).

Täthetsfunktion för testvariabeln $\bar{X}$ **under H_0**



$\alpha = P(fel\ typ\ I)$

Ω_α = kritiskt område

ω_α = kritiskt värde

röd area = p-värde

$\boxed{H_0 \text{ förkastas om } p < \alpha}$

www.matstat.org

p-värdet för binomialfördelningen

Exempel 2: Vi har en behållare med flera tärningar. De flesta tärningarna är rättvisa, men det finns några som är förfalskade och har mer än en sida med sexan. Vi drar slumpmässigt en tärning och vill med hjälp av 10 tärningskast – utan att titta närmare på tärningen! – ta reda på om vi har dragit en förfalskad tärning.

Slumpvariabel X : Ögontalet vid en kast. Låt $p = P(X = 6)$.

Hypoteser: $H_0 : p = 1/6$; $H_a : p > 1/6$ (nollhypotes = rättvis tärning)

Testvariabel (T): Antalet kastade sexor av totalt 10 kast. Under H_0 är T binomialfördelad med $T \sim Bin(10, 1/6)$, för att sannolikheten för att kasta en sexa med en rättvis tärning är $1/6$.

Stickprov: Vi antar att sexan kom upp 9 gånger, dvs. $T = 9$

p-värde: Observationen är $T = 9$. Ännu mer extremt i riktning mot H_a är $T = 10$. p-värdet är alltså enligt definitionen summan av sannolikheterna för $T = 9$ och $T = 10$.

www.matstat.org

p-värdet för binomialfördelningen

Hypoteser: $H_0: p = 1/6\,;\ H_a: p > 1/6$ (nollhypotes = rättvis tärning)

Testvariabel (T): $T \sim Bin(10, 1/6)$ under H_0

Stickprov: $T = 9$

p-värde: $p_{value} = P(T \geq 9 \mid H_0)$

Kom ihåg: $P(T = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$ om $T \sim Bin(n, p)$

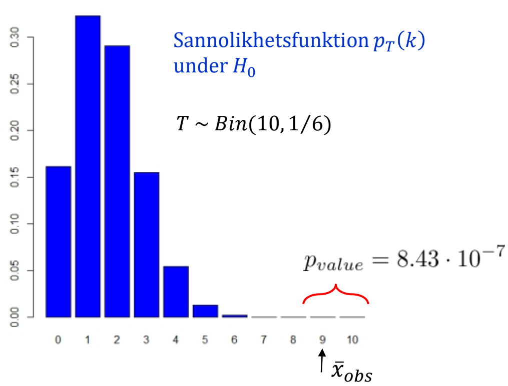
$$
\begin{aligned}
p_{value} &= P(T \geq 9 \mid H_0) = P(T = 9 \mid H_0) + P(T = 10 \mid H_0) \\
&= \binom{10}{9} \cdot (1/6)^9 \cdot (5/6)^1 + \binom{10}{10} \cdot (1/6)^{10} \cdot (5/6)^0 \\
&= 8.27 \cdot 10^{-7} + 1.65 \cdot 10^{-8} = \underline{\underline{8.43 \cdot 10^{-7}}}
\end{aligned}
$$

Facit: p-värdet är mycket litet. Utfallet vi fick är alltså mycket osannolikt om man antar att nollhypotesen stämmer. Nollhypotesen (rättvis tärning) förkastas därför – t. o. m. på signifikansnivån $\alpha = 10^{-6}$! (signifikansnivån slås dock fast **före** testet !)

Notera: Med ett sådant testresultat hade väl även ”sunt förnuft” tillbakavisat hypotesen att tärningen är rättvis.

www.matstat.org

p-värdet för binomialfördelningen



Sannolikhetsfunktion $p_T(k)$ under H_0

$T \sim Bin(10, 1/6)$

Facit: $p_{value} < \alpha = 10^{-6}$. Nollhypotesen kan t. o. m. förkastas på signifikansnivån $\alpha = 10^{-6}$. Risken för fel typ I är mycket liten.

www.matstat.org

p-värdet för normalfördelningen

Exempel 3: Ytor av vissa elektroniska komponenter måste ha ett kopparskikt som är 30 µm tjock. Skiktet får varken vara för tjockt eller för tunnt. Ytorna av $n = 10$ komponenter har mätts, medelvärdet blev $\bar{x}_{obs} = 30.91$.
Av lång erfarenhet vet man att tjockleken är normalfördelad med $\sigma = 0.788$ (vi kommer att räkna med okänd σ senare ...).

Hypoteser: $H_0: \mu = 30$; $H_a: \mu \neq 30$

Testvariabel: skattning för μ, alltså $\bar{X}$

Om H_0 är sann gäller $\mu = \mu_0 = 30$, alltså:

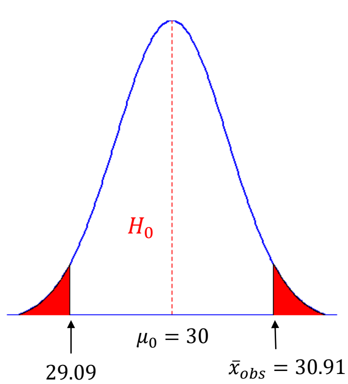
$$\bar{X} \sim N\left(30, \frac{0.788}{\sqrt{10}}\right) \Rightarrow \bar{X} \sim N(30, 0.249)$$

Den alternativa hypotesen $H_a: \mu \neq 30$ säger att μ får varken vara för stort eller för litet. Det kritiska området ligger på **båda** sidor av nollhypotesens täthetsfunktion.

www.matstat.org

p-värdet för normalfördelningen

- Alternativ hypotes: $H_a: \mu \neq 30$ $\rightarrow \mu$ får varken vara för stort eller för litet.
- p-värdet är nu summan av två areor (som har samma avstånd från μ_0 om täthetsfunktionen är symmetrisk runt μ_0).
- $\rightarrow$ tvåsidigt test



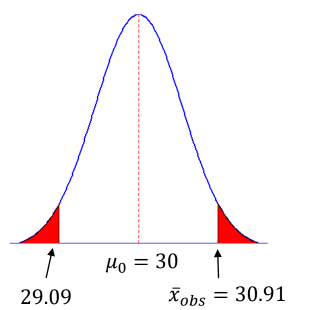
$H_0: \mu = 30$; $H_a: \mu \neq 30$

Tvåsidigt Test: båda röda områden är mer extrema än $\bar{x}_{obs}$!

uwe.menzel@matstat.org

p-värdet för normalfördelningen

$H_0: \mu = 30$; $H_a: \mu \neq 30$ (tvåsidigt test) $\bar{x}_{obs} = 30.91$



Tvåsidigt test : båda markerade områden är mer extrema än $\bar{x}_{obs}$!

För att beräkna p-värdet kan vi utnyttja symmetrin av täthetsfunktionen – båda röd markerade areor är lika stora:

$$
\begin{aligned}
p &= 2 \cdot P\left(\bar{X} > \bar{x}_{obs} \mid H_0\right) \\
&= 2 \cdot P\left(\bar{X} > 30.91 \mid H_0\right) \\
&= 2 \cdot \left[1 - P\left(\bar{X} \leq 30.91\right)\right] \quad H_0 \\
&= 2 \cdot \left[1 - F_{\bar{X}}\left(30.91\right)\right] \\
&= 2 \cdot \left[1 - \Phi\left(\frac{30.91 - 30}{0.249}\right)\right] \\
&= 2 \cdot \left[1 - \Phi\left(3.65\right)\right] \\
&= 2 \cdot \left[1 - 0.99987\right] = \underline{\underline{2.6 \cdot 10^{-4}}}
\end{aligned}
$$

Facit: p-värdet är mindre än $10^{-3} \rightarrow H_0$ kan förkastas på signifikansnivån $\alpha = 10^{-3}$. Avvikelserna från den föreskrivna tjockleken är alltså för stora.

p-värdet för binomialfördelningen

Exempel 4: (Alm, Britton: **Stokastik**, Liber AB, Stockholm 2008)

En person påstår sig vara tillräckligt fingerfärdig för att kunna påverka vilken sida som kommer upp vid myntkast, så att han får ”krona” oftare än ”klave”, dvs. $p = P(krona) > {}^1/_2$. Vi är skeptiska och tror att $p = {}^1/_2$. För att pröva påståendet får han göra 10 kast och vi är beredda att tro honom om han får tillräckligt många ”kronor” vid försöket, mera exakt om p-värder är mindre än $\alpha = 0.01$. (dvs. vi godtar en sannolikheten 0.01 för ett fel typ I).

Hypoteser: H_0: $p = {}^1/_2$; $H_a: p > {}^1/_2$ (ensidigt test)

Testvariabel (T): Antalet kastade ”krona” $T \sim Bin(10, 1/2)$ under H_0

Stickprov: $T = 8$ ”krona” kastades 8 gånger

p-värde: $p_{value} = P\left(T \geq 8 \mid H_0\right)$

Signifikansnivå: $\alpha = 0.01$

p-värdet för binomialfördelningen

$T \sim Bin(10, 1/2)$ under H_0 $\quad$ p-värde: $\quad p_{value} = P(T \geq 8 \mid H_0)$

observation $\quad$ ännu mer extrema än observation

$$p_{value} = P(X \geq 8 \mid H_0) = p_X(8) + p_X(9) + p_X(10)$$
$$= \binom{10}{8}(1/2)^{10} + \binom{10}{9}(1/2)^{10} + \binom{10}{10}(1/2)^{10} = \underline{\underline{0.0986}}$$

Facit: Signifikansnivån α fastslogs med 0.01. Det beräknade p-värdet är dock större: $p > \alpha$. Nollhypotesen förkastas därför **inte** (”på signifikansnivån 0.01”). Vi tvivlar fortfarande på att denna person är utomordentligt fingerfärdig. Vi tror att det inte är särskilt osannolikt att få 8 av 10 ”kronor” på grund av ren slump.

Notera: Därmed är nollhypotesen dock **icke** ”bevisad”! Kanske räckte helt enkelt antalet kast inte – hade personen kastat 80 av 100 ”kronor”, hade resultatet väl varit annorlunda. Man kan förkasta nollhypotesen, men aldrig ”bevisa” den!

www.matstat.org

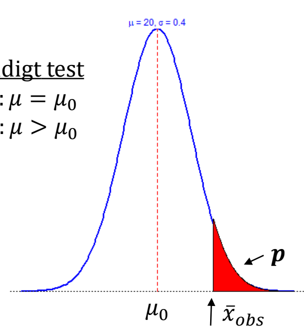
Sammanfattning: p-värde-metoden

1. Definiera nollhypotes (H_0) och alternativ hypotes (H_a)
2. Slå fast signifikansnivån $\alpha = P(fel\ typ\ I \mid H_0)$
3. Bestäm testvariabeln ($\bar{X}_{obs}, T, \ldots$). Testvariabelns fördelning under H_0 måste vara känd.
4. Beräkna p – värdet enligt definitionen.
5. Förkasta H_0 på signifikansnivån α om $p < \alpha$, annars ”acceptera” H_0

<u>ensidigt test</u>

H_0: $\mu = \mu_0$

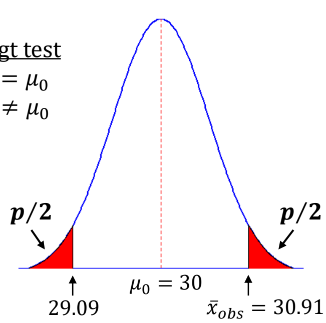
H_a: $\mu > \mu_0$



<u>tvåsidigt test</u>

H_0: $\mu = \mu_0$

H_a: $\mu \neq \mu_0$



www.matstat.org