Grundläggande matematisk statistik

Hypotestest
Del II: Några parametriska test

Uwe Menzel, 2017
uwe.menzel@matstat.org
www.matstat.org

Z-Test

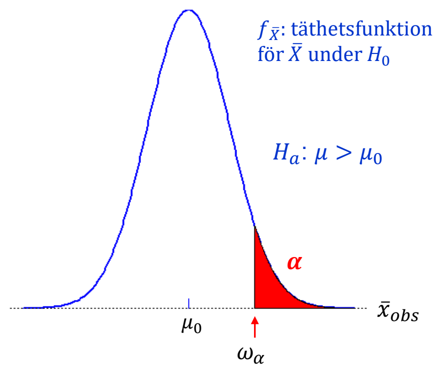
- testar en hypotes för väntevärdet μ i en normalfördelad population
- Antagandet: standardavvikelsen σ är känd.
- Nollhypotes: $H_0\colon \mu = \mu_0$ (μ_0 = hypotetiskt värde)
- Alternativa Hypoteser:
  - $H_a\colon \mu > \mu_0$ (ensidigt test)
  - $H_a\colon \mu < \mu_0$ (ensidigt test)
  - $H_a\colon \mu \neq \mu_0$ (tvåsidigt test)
- stickprov → $\bar{x}_{obs}$ (observation av skattningen $\bar{X}$ för μ)

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$$ skattning för μ

$$X_i \sim N(\mu, \sigma)$$
$$x = (x_1, x_2, \dots x_n)$$

Z-Test för H_a: $\mu > \mu_0$

Nollhypotesen (H_0) förkastas om en observation $\bar{x}_{obs}$ av skattningen $\bar{X}$ överskrider ett kritiskt värde ω_α i riktning mot den alternativa hypotesen (bild), alltså för $\bar{x}_{obs} > \omega_\alpha$. Det kritiska värdet ω_α beräknas utifrån den på förhand fastlagda signifikansnivån α (= sannolikhet för felet typ I; rött markerade area = värdet av α).



Z-Test för H_a: $\mu > \mu_0$

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

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

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

För H_a: $\mu > \mu_0$ och signifikansnivån α kan det kritiska värdet ω_α för testvariabeln $\bar{X}$ beräknas med hjälp av följande ekvation (se föreläsning **F11**):

$$\boxed{P\left(\underbrace{\bar{X} > \omega_\alpha}_{H_0 \text{ förkastas}} \mid H_0\ sann\right) = \alpha}$$

fel typ I

ekvation för att beräkna ω_α för givet α

Ekvationen betyder: sannolikheten att H_0 förkastas (trots att den är sann), ska vara α (= sannolikhet för fel typ I). Genom att omforma i parentesen får vi:

$$P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{\omega_\alpha - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha$$

Under H_0 är den vänstra termen i parentesen $N(0,1)$-fördelad och vi kan beteckna den med Z. Betingningen "H_0 *sann*" är därmed inräknad.

Z-Test för $H_a: \mu > \mu_0$

$$P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > \frac{\omega_\alpha-\mu_0}{\sigma/\sqrt{n}}\right) = \alpha$$

Under H_0 är den vänstra termen i parentesen $N(0,1)$-fördelad, den betecknas därför med Z.

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

Allmänt gäller:

$P(Z > \lambda_\alpha) = \alpha$ Definition för kvantiler för $N(0,1)$

Jämförelse av de sista båda uttrycken ger $\lambda_\alpha = \dfrac{\omega_\alpha-\mu_0}{\sigma/\sqrt{n}}$

$$\Rightarrow \quad \omega_\alpha = \mu_0 + \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

Kritiskt värde för $\bar{X}$ vid Z-test med $H_a: \mu > \mu_0$;
H_0 förkastas om $\bar{x}_{obs} > \omega_\alpha$

Nollhypotesen H_0 förkastas alltså för $\bar{x}_{obs} > \mu_0 + \lambda_\alpha \cdot \dfrac{\sigma}{\sqrt{n}}$

www.matstat.org

Z-Test för $H_a: \mu > \mu_0$

Nollhypotesen H_0 förkastas alltså för $\bar{x}_{obs} > \mu_0 + \lambda_\alpha \cdot \dfrac{\sigma}{\sqrt{n}}$

eller för $\dfrac{\bar{x}_{obs}-\mu_0}{\sigma/\sqrt{n}} > \lambda_\alpha$

$$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$$

Vanligtvis används inte $\bar{X}$, utan Z som testvariabel. En observation av Z kan betecknas med z_{obs}:

$$z_{obs} = \frac{\bar{x}_{obs}-\mu_0}{\sigma/\sqrt{n}}$$

observation av testvariabeln $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

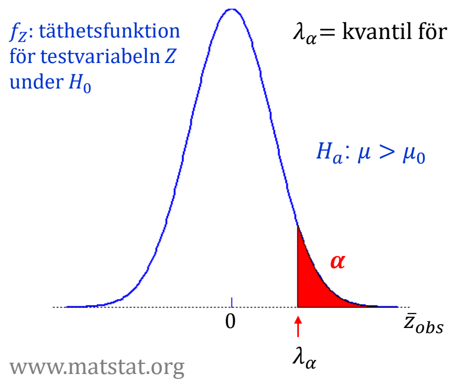
$\boxed{\text{Nollhypotesen förkastas därmed om } z_{obs} > \lambda_\alpha}$

www.matstat.org

Z-Test för $H_a: \mu > \mu_0$

$$\boxed{z_{obs} = \frac{\bar{x}_{obs} - \mu_0}{\sigma/\sqrt{n}}}$$ observation av testvariabeln $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

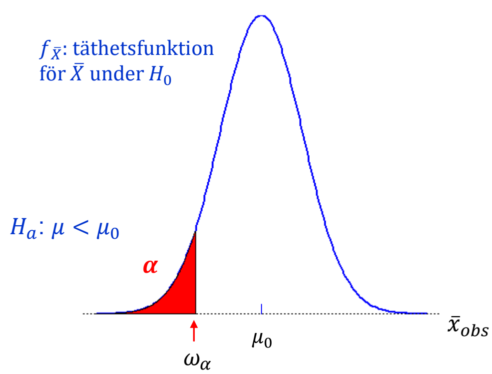
Nollhypotesen förkastas om $z_{obs} > \lambda_\alpha$



www.matstat.org

Z-Test för $H_a: \mu < \mu_0$

Nollhypotesen (H_0) förkastas om en observation $\bar{x}_{obs}$ av skattningen $\bar{X}$ överskrider ett kritiskt värde ω_α i riktning mot den alternativa hypotesen (bild), alltså för $\bar{x}_{obs} < \omega_\alpha$. Det kritiska värdet ω_α beräknas utifrån den på förhand fastlagda signifikansnivån α (= sannolikhet för felet typ I).



www.matstat.org

Z-Test för H_a: $\mu < \mu_0$

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

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

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

För H_a: $\mu < \mu_0$ och signifikansnivån α kan det kritiska värdet ω_α för testvariabeln $\bar{X}$ beräknas med hjälp av följande ekvation (se föreläsning **F11**) :

$$\boxed{P\left(\underbrace{\bar{X} < \omega_\alpha}_{H_0\ \text{förkastas}} \mid H_0\ sann\right) = \alpha}$$

ekvation för att beräkna ω_α för givet α

fel typ I

Ekvationen betyder: sannolikheten att H_0 förkastas (trots att den är sann), ska vara α (= sannolikhet för fel typ I). Genom att omforma i parentesen får vi:

$$P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < \frac{\omega_\alpha - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha$$

Under H_0 är den vänstra termen i parantesen $N(0,1)$-fördelad och vi kan beteckna den med Z. Betingningen " H_0 *sann*" är därmed inräknad.

uwe.menzel@matstat.org

Z-Test för H_a: $\mu < \mu_0$

$$P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < \frac{\omega_\alpha - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha$$

Under H_0 är den vänstra termen i parentesen $N(0,1)$-fördelad, den betecknas därför med Z.

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

Allmänt gäller:

$$P(Z < -\lambda_\alpha) = \alpha$$

följer från definitionen för kvantiler för $N(0,1)$

Jämförelse av de sista båda uttrycken ger $-\lambda_\alpha = \dfrac{\omega_\alpha - \mu_0}{\sigma/\sqrt{n}}$

$$\Rightarrow \omega_\alpha = \mu_0 - \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

Kritiskt värde för $\bar{X}$ vid Z-testet med H_a: $\mu < \mu_0$;
H_0 förkastas för $\bar{x}_{obs} < \omega_\alpha$

Nollhypotesen H_0 förkastas alltså för $\bar{x}_{obs} < \mu_0 - \lambda_\alpha \cdot \dfrac{\sigma}{\sqrt{n}}$

Z-Test für $H_a: \mu < \mu_0$

Nollhypotesen H_0 förkastas alltså för $\bar{x}_{obs} < \mu_0 - \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}}$

eller för $\frac{\bar{x}_{obs} - \mu_0}{\sigma/\sqrt{n}} < -\lambda_\alpha$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Vanligtvis används inte $\bar{X}$, utan Z som testvariabel. En observation av Z kan betecknas med z_{obs} :

$$z_{obs} = \frac{\bar{x}_{obs} - \mu_0}{\sigma/\sqrt{n}}$$

observation av testvariabeln $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

Nollhypotesen förkastas därmed om $z_{obs} < -\lambda_\alpha$

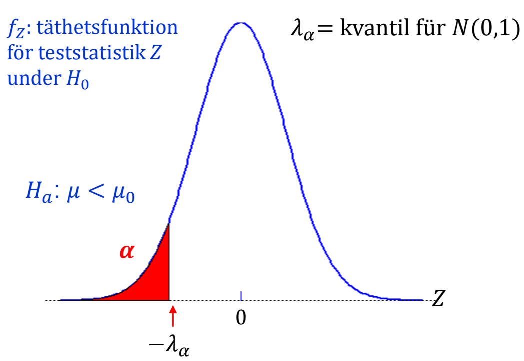
www.matstat.org

Z-Test för $H_a: \mu < \mu_0$

$$\boxed{z_{obs} = \frac{\bar{x}_{obs} - \mu_0}{\sigma/\sqrt{n}}}$$

observation av testvariabeln $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

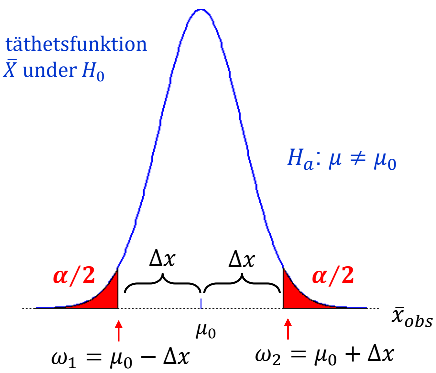
Nollhypotesen förkastas om $z_{obs} < -\lambda_\alpha$



www.matstat.org

Z-Test för $H_a: \mu \neq \mu_0$

Nollhypotesen (H_0) förkastas om en observation $\bar{x}_{obs}$ av $\bar{X}$ överskrider ett av de kritiska värdena $\omega_{2,1} = \mu_0 \pm \Delta x$, alltså för $|\bar{x}_{obs} - \mu_0| > \Delta x$ (bild). De kritiska värdena $\omega_{2,1}$ beräknas utifrån den på förhand fastlagda signifikansnivån α ($=$ sannolikhet för felet typ I).



Z-Test für $H_a: \mu \neq \mu_0$

Om nollhypotesen $\mu = \mu_0$ är sann gäller det att $X_i \sim N(\mu_0, \sigma)$ 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

För H_a: $\mu \neq \mu_0$ och signifikansnivån α kan de kritiska värdena $\omega_{2,1}$ för testvariabeln $\bar{X}$ beräknas med hjälp av följande ekvation (se föreläsning **F11**)

$$\boxed{P\left(\underbrace{|\bar{X} - \mu_0| > \Delta x}_{H_0 \text{ wird verworfen}} \mid H_0\ sann\right) = \alpha}$$

Fehler Typ I

ekvation för att beräkna
$\omega_{2,1} = \mu_0 \pm \Delta x$ för givet α

Ovanstående formel anger sannolikheten att $\bar{X}$ blir antingen större än $\omega_2 = \mu_0 + \Delta x$ eller mindre än $\omega_1 = \mu_0 - \Delta x$. På grund av täthetsfunktionens symmetri är denna sannolikhet dubbelt så stor som sannolikheten att $\bar{X}$ blir större än ω_2 (se bild ovan). Vi kan därför förenkla ovanstående ekvation →

Z-Test för $H_a: \mu \neq \mu_0$

$$P\left(|\bar{X} - \mu_0| > \Delta x \mid H_0 \; sann\right) = \alpha$$

ekvation för att beräkna $\omega_{2,1} = \mu_0 \pm \Delta x$ för givet α

$$2 \cdot P\left(\bar{X} > \mu_0 + \Delta x \mid H_0 \; sann\right) = \alpha$$

pga. symmetri

$$P\left(\bar{X} > \mu_0 + \Delta x \mid H_0 \; sann\right) = \alpha/2$$

omforma i parentesen →

$$P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{\Delta x}{\sigma/\sqrt{n}}\right) = \alpha/2$$

Under H_0 är den vänstra termen i parentesen $N(0,1)$-fördelad → vi kan beteckna den med Z. Betingningen "H_0 *sann*" är därmed inräknad.

$$P\left(Z > \frac{\Delta x}{\sigma/\sqrt{n}}\right) = \alpha/2 \qquad Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

Allmänt gäller: $P\left(Z > \lambda_{\alpha/2}\right) = \alpha/2$ genom jämförelse får vi →

$$\lambda_{\alpha/2} = \frac{\Delta x}{\sigma/\sqrt{n}} \Rightarrow \Delta x = \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Nollhypotesen H_0 förkastas alltså för $|\bar{x}_{obs} - \mu_0| > \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

www.matstat.org

Z-Test för $H_a: \mu \neq \mu_0$

Nollhypotesen H_0 förkastas alltså för $|\bar{x}_{obs} - \mu_0| > \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

alltså för $\frac{|\bar{x}_{obs} - \mu_0|}{\sigma/\sqrt{n}} > \lambda_{\alpha/2}$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Vanligtvis används inte $\bar{X}$, utan Z som testvariabel. En observation av Z kan betecknas med z_{obs} :

$$z_{obs} = \frac{\bar{x}_{obs} - \mu_0}{\sigma/\sqrt{n}}$$

observation av testvariabeln $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Nollhypotesen förkastas därmed om $|z_{obs}| > \lambda_{\alpha/2}$

www.matstat.org

Z-Test för $H_a: \mu \neq \mu_0$

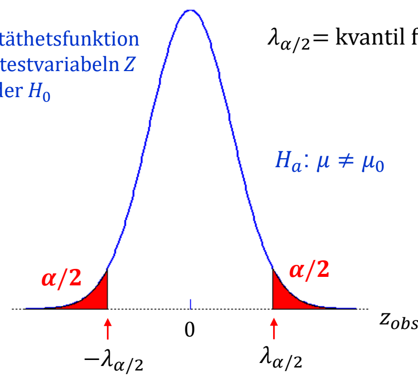
$$\boxed{z_{obs} = \frac{\bar{x}_{obs} - \mu_0}{\sigma/\sqrt{n}}}$$

observation av testvariabeln $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

Nollhypotesen förkastas om $|z_{obs}| > \lambda_{\alpha/2}$

f_Z: täthetsfunktion för testvariabeln Z under H_0

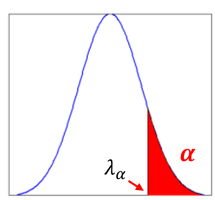
$\lambda_{\alpha/2}$= kvantil für $N(0,1)$



www.matstat.org

Sammanfattning: Z-Test

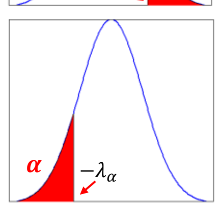
H_0	H_a	Test
$\mu = \mu_0$	$\mu > \mu_0$	ensidigt



förkasta H_0 om $z_{obs} > \lambda_\alpha$

$$\bar{x}_{obs} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

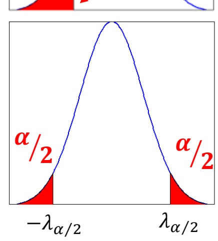
H_0	H_a	Test
$\mu = \mu_0$	$\mu < \mu_0$	ensidigt



förkasta H_0 om $z_{obs} < -\lambda_\alpha$

$$z_{obs} = \frac{\bar{x}_{obs} - \mu_0}{\sigma/\sqrt{n}}$$

H_0	H_a	Test
$\mu = \mu_0$	$\mu \neq \mu_0$	ensidigt



förkasta H_0 om $|z_{obs}| > \lambda_{\alpha/2}$

www.matstat.org

t-Test för ett stickprov

- testar en hypotes för väntevärdet μ i en normalfördelad population
- Antagandet: standardavvikelsen σ är okänd (skillnad till Z-testet!)
- Nollhypotes: $H_0\colon \mu = \mu_0$ (μ_0 = hypotetiskt värde)
- Alternativa hypoteser:
  - $H_a\colon \mu > \mu_0$ (ensidigt test)
  - $H_a\colon \mu < \mu_0$ (ensidigt test)
  - $H_a\colon \mu \neq \mu_0$ (tvåsidigt test)
- stickprov → $\bar{x}_{obs}$ (observation av skattningen $\bar{X}$ för μ)

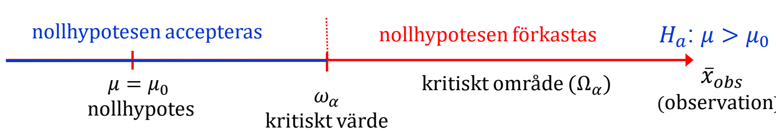
$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \qquad\qquad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

skattning för μ skattning för σ^2

www.matstat.org

t-test för $H_a\colon \mu > \mu_0$

- Nollhypotesen (H_0) förkastas om en observation $\bar{x}_{obs}$ av skattningen $\bar{X}$ överskrider ett kritiskt värde ω_α i riktning mot den alternativa hypotesen. För $H_a\colon \mu > \mu_0$ är detta fallet om $\bar{x}_{obs} \gg \mu_0$.
- Vi vet: En observation kan med en viss sannolikhet överskrida det kritiska värdet i riktning mot den alternativa hypotesen, **även om** nollhypotesen är sann (fel typ I).
- Det kritiska värdet ω_α justeras så att sannolikheten för felet typ I antar ett på förhand fastlagd värde α (signifikansnivå).



www.matstat.org

t-test för H_a: $\mu > \mu_0$

Om nollhypotesen $\mu = \mu_0$ är sann gäller det att $X_i \sim N(\mu_0, \sigma)$ 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 tillämpas som testvariabel. (det går inte att beräkna ett numeriskt värde för en observation av Z). Vi vet dock (föreläsning **F9**) att vi får en t-fördelad variabel om σ ersätts med S i det ovanstående uttrycket:

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

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

Fördelningen för T är helt känd; en observation av T kan beräknas utifrån stickprovet (μ_0 är den hypotetiska parametern, alltså också känd). En observation av T ska i det följande betecknas med t_{obs} :

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

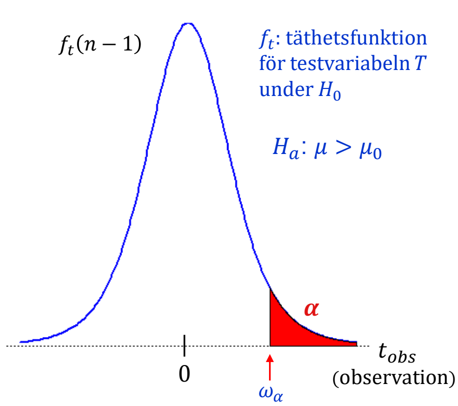
observation av T

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

uwe.menzel@matstat.org

t-Test för H_a: $\mu > \mu_0$

t-fördelningen är symmetrisk runt noll (bild). Om nollhypotesen är sann borde observationen t_{obs} inte ligga för långt från nollpunkten, för då borde $\bar{x}_{obs} \approx \mu_0$. Vi är dock benägna att acceptera den alternativa hypotesen om $\bar{x}_{obs} \gg \mu_0$. I detta fall antar t_{obs} stora positiva värden. När slutligen $t_{obs} > \omega_\alpha$ förkastas H_0 till förmån för den alternativa hypotesen.



f_t: täthetsfunktion för testvariabeln T under H_0

H_a: $\mu > \mu_0$

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

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

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

ω_α: kritiskt värde

t-Test för H_a: $\mu > \mu_0$

För H_a: $\mu > \mu_0$ och en på förhand fastlagd signifikansnivå α kan det kritiska värdet ω_α för testvariabeln T beräknas genom att kräva att sannolikheten för felet 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 givet α (= sannolikhet för felet typ I)

Under H_0 är termen på vänster sida i parentesen $t(n-1)$-fördelad, vi kan alltså skriva (betingningen "H_0 *sann*" är därmed inräknad):
(uttrycket är då och först då t-fördelat om H_0 är sann, alltså om $\mu = \mu_0$)

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

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

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

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

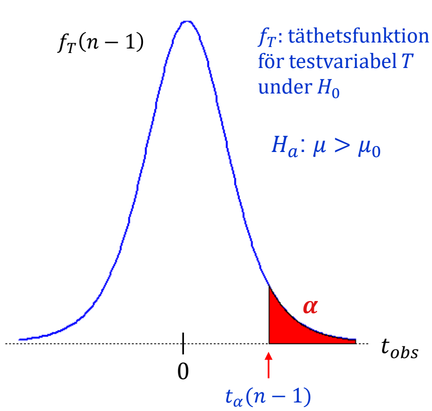
www.matstat.org

t-Test för H_a: $\mu > \mu_0$

$$\boxed{\text{Nollhypotesen förkastas om } \; t_{obs} > t_\alpha(n-1)}$$

Det kritiska området för signifikansnivå α betecknas med Ω_α :

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



f_T: täthetsfunktion för testvariabel T under H_0

H_a: $\mu > \mu_0$

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

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

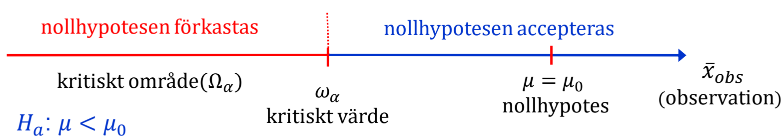
H_0 förkastas om observationen t_{obs} hamnar i det kritiska området (röd).

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

www.matstat.org

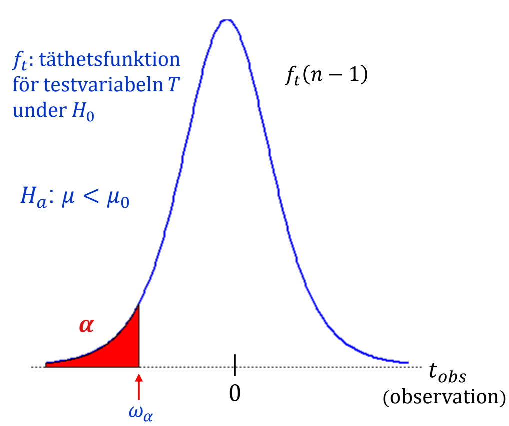
t-Test för $H_a: \mu < \mu_0$

- Nollhypotesen (H_0) förkastas om en observation $\bar{x}_{obs}$ av skattningen $\bar{X}$ överskrider ett kritiskt värde ω_α i riktning mot den alternativa hypotesen. För $H_a: \mu < \mu_0$ är detta fallet om $\bar{x}_{obs} \ll \mu_0$.
- Vi vet: En observation kan med en viss sannolikhet överskrida det kritiska värdet i riktning mot den alternativa hypotesen, **även om** nollhypotesen är sann (fel typ I).
- Det kritiska värdet ω_α justeras så att sannolikheten för felet typ I antar ett på förhand fastlagd värde α (signifikansnivå).



t-Test för $H_a: \mu < \mu_0$

Om nollhypotesen är sann borde observationen t_{obs} inte ligga för långt från nollpunkten, för då borde $\bar{x}_{obs} \approx \mu_0$. Är dock $\bar{x}_{obs} \ll \mu_0$ antar t_{obs} stora negativa värden, och vi blir benägna att acceptera den alternativa hypotesen. När sluttligen $t_{obs} < \omega_\alpha$ förkastas H_0 till förmån för 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

t-Test för H_a: $\mu < \mu_0$

För H_a: $\mu < \mu_0$ och en på förhand fastlagd signifikansnivå α kan det kritiska värdet ω_α för testvariabeln T beräknas genom att kräva att sannolikheten för felet 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 givet α (= sannolikhet för felet typ I)

Under H_0 är termen på vänster sida i parentesen $t(n-1)$-fördelad, vi kan alltså skriva (betingningen "H_0 *sann*" är därmed inräknad):

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

Allmänt gäller : $P(T < -t_\alpha(n-1)) = \alpha$

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

Genom att jämföra de sista båda uttrycken får vi $\omega_\alpha = -t_\alpha(n-1)$.
Nollhypotesen H_0 förkastas alltså om en observation av T (= t_{obs}) blir mindre än $-t_\alpha(n-1)$.

www.matstat.org

t-Test för H_a: $\mu < \mu_0$

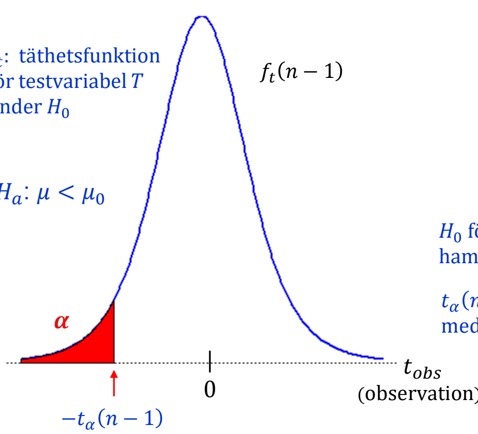
$$\boxed{\text{Nollhypotesen } H_0 \text{ förkastas om } t_{obs} < -t_\alpha(n-1)}$$

Det kritiska området för signifikansnivå α betecknas med Ω_α :

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

f_t: täthetsfunktion för testvariabel T under H_0

H_a: $\mu < \mu_0$



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

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

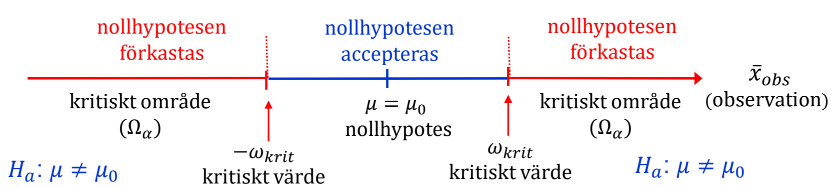
H_0 förkastas om observationen t_{obs} hamnar i det kritiska området (röd).

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

www.matstat.org

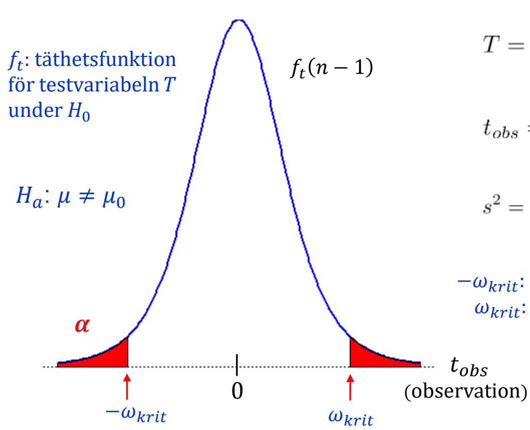
t-Test för $H_a: \mu \neq \mu_0$

- Nollhypotesen (H_0) förkastas om en observation $\bar{x}_{obs}$ av skattningen $\bar{X}$ överskrider ett kritiskt värde ω_α i riktning mot den alternativa hypotesen För $H_a: \mu \neq \mu_0$ är detta fallet om $\bar{x}_{obs} \ll \mu_0$ eller $\bar{x}_{obs} \gg \mu_0$.
- En observation kan med en viss sannolikhet överskrida det kritiska värdet i riktning mot den alternativa hypotesen, **även om** nollhypotesen är sann (fel typ I).
- Det kritiska värdet ω_α justeras så att sannolikheten för felet typ I antar ett på förhand fastlagd värde α (signifikansnivå).



t-Test för $H_a: \mu \neq \mu_0$

Under H_0 borde det gälla att $\bar{x}_{obs} \approx \mu_0$. Är dock $\bar{x}_{obs} \ll \mu_0$ eller $\bar{x}_{obs} \gg \mu_0$ antar t_{obs} stora negativa eller positiva värden. I så fall är vi benägna att acceptera den alternativa hypotesen. Om $t_{obs} < -\omega_{krit}$ eller $t_{obs} > \omega_{krit}$ förkastas H_0 till förmån för den alternativa hypotesen (pga. t-fördelningens symmetri ligger de kritiska värdena symmetriskt runt noll).



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

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

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

$-\omega_{krit}$: undre kritiskt värde
ω_{krit}: övre kritiskt värde

t-Test för $H_a: \mu \neq \mu_0$

För H_a: $\mu \neq \mu_0$ och en på förhand fastlagd signifikansnivå α kan de kritiska värdena $\pm\omega_{krit}$ för testvariabeln T beräknas genom att kräva att sannolikheten för felet typ I ska vara lika med α:

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

fel typ I

ekvation för att beräkna ω_{krit} för givet α (= sannolikhet för felet typ I)

Pga. symmetrin (se bild ovan) är denna sannolikhet dubbelt så stor som sannolikheten att testvariabeln överskrider det övre kritiska värdet:

$$2 \cdot P\left(\frac{\bar{X}-\mu_0}{S/\sqrt{n}} > \omega_{krit} \mid H_0\ wahr\right) = \alpha$$

Under H_0 är termen på vänster sida i parentesen $t(n-1)$-fördelad, vi kan alltså skriva (betingningen "$H_0\ sann$" är därmed inräknad):

$$P\left(T > \omega_{krit}\right) = \alpha/2 \quad \text{med} \quad T \sim t(n-1)$$

www.matstat.org

t-Test för $H_a: \mu \neq \mu_0$

$$P\left(T > \omega_{krit}\right) = \alpha/2 \quad \text{med} \quad T \sim t(n-1)$$

$$\text{Allmänt gäller}: \; P\left(T > t_{\alpha/2}(n-1)\right) = \alpha/2$$

$t_{\alpha/2}(n-1)$ = kvantil för t-fördelningen för $f = n-1$

Genom att jämföra de sista båda uttrycken får vi för det övre kritiska värdet $\omega_{krit} = t_{\alpha/2}(n-1)$.

H_0 förkastas alltså om $t_{obs} < -t_{\alpha/2}(n-1)$ eller om $t_{obs} > +t_{\alpha/2}(n-1)$, alltså om $|t_{obs}| > t_{\alpha/2}(n-1)$

$$\boxed{\text{Nollhypotesen förkastas om} \quad |t_{obs}| > t_{\alpha/2}(n-1)}$$

Det kritiska området för signifikansnivån α betecknas med Ω_α :

$$\Omega_\alpha = \left\{|t_{obs}| > t_{\alpha/2}(n-1)\right\}$$

www.matstat.org

t-Test för H_a: $\mu \neq \mu_0$

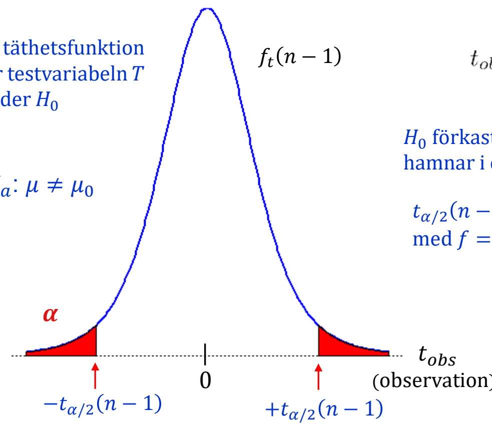
Nollhypotesen förkastas om $|t_{obs}| > t_{\alpha/2}(n-1)$

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

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

f_t: täthetsfunktion för testvariabeln T under H_0

H_a: $\mu \neq \mu_0$



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

$t_{\alpha/2}(n-1)$ = kvantil för t-fördelningen med $f = n-1$ för signifikansnivå α

www.matstat.org

Sammanfattning: t-Test

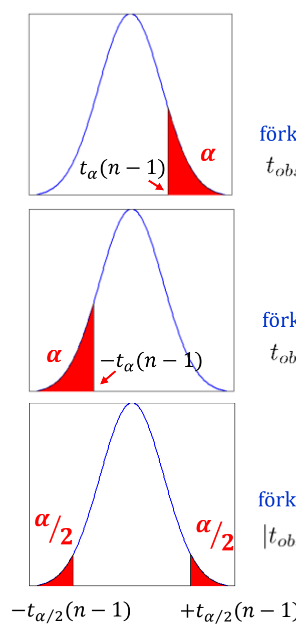
H_0	H_a	test
$\mu = \mu_0$	$\mu > \mu_0$	ensidigt

$$\bar{x}_{obs} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

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

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



förkasta H_0 om $t_{obs} > t_\alpha(n-1)$

förkasta H_0 om $t_{obs} < -t_\alpha(n-1)$

förkasta H_0 om $|t_{obs}| > t_{\alpha/2}(n-1)$

www.matstat.org

t-Test, exempel: Vita blodceller

Antalet vita blodceller per ml blod hos friska vuxna är normalfördelad med $\mu_0 = 7500$ (mätt hos miljontals människor, kan därför anses som sanna populationsparameter)

Fråga: Har astronauter samma genomsnittliga koncentration av vita blodceller?

Nollhypotes H_0: $\mu = \mu_0 = 7500$

Alternativ hypotes H_a: $\mu \neq \mu_0$ (vi har ingen aning åt vilket håll μ kan avvika)

Signifikansnivå $\alpha = 0.05$

Stickprov: 7130, 6845, 7055, 7235, 7200, 7450, 7750, 7950, 7340, 7150

$$\bar{x}_{obs} = 7310.5 \ ; \ s = 330.1$$

Observation för testvariabel: $$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}} = \frac{7310.5 - 7500}{330.1/\sqrt{10}} = \underline{\underline{-1.815}}$$

t-Test, exempel: Vita blodceller

Observation för testvariabel : $$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}} = \frac{7310.5 - 7500}{330.1/\sqrt{10}} = \underline{\underline{-1.815}}$$

Kritiskt område: $\Omega_\alpha = \left\{ |t_{obs}| > t_{\alpha/2}(n-1) \right\}$ för tvåsidigt test

$$t_{\alpha/2}(n-1) = t_{0.025}(9) = 2.262$$

Jämför testvariabeln med det kritiska värdet:

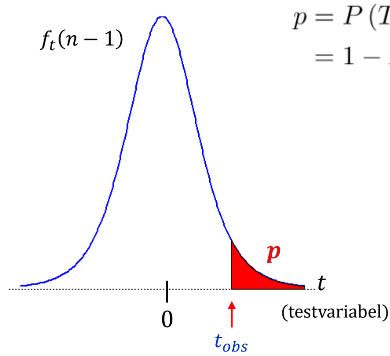
$$|t_{obs}| = 1.815 < 2.262 = t_{0.025}(9)$$

Testvariabelns observation t_{obs} ligger **inte** i det kritiska området. Nollhypotesen kan **inte** förkastas. Vi kan inte påstå att astronauter har en koncentration av vita blodceller som avviker från "jordpopulationen". (Att förkasta nollhypotesen kan dock vara möjligt med ett större stickprov ...)

Beräkning av p-värdet, H_a: $\mu > \mu_0$

$X_i \sim N(\mu, \sigma)$ σ okänd $H_0: \mu = \mu_0$ $H_a: \mu > \mu_0$

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



Under H_0 gäller $T \sim t(n-1)$

$$p = P(T > t_{obs}) = 1 - P(T \le t_{obs})$$
$$= 1 - F_{t(n-1)}(t_{obs}) = F_{t(n-1)}(-t_{obs})$$

(se appendixet)

$F_{t(n-1)}(t)$ = fördelningsfunktion för Student's t, för $n-1$ frihetsgrader

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

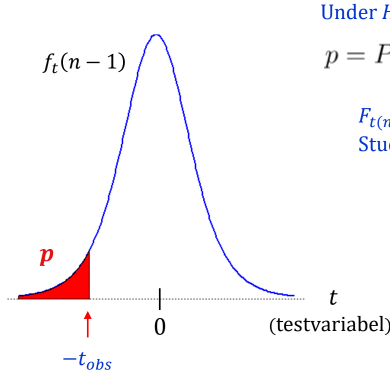
$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}}$$

www.matstat.org

Beräkning av p-värdet, H_a: $\mu < \mu_0$

$X_i \sim N(\mu, \sigma)$ σ okänd $H_0: \mu = \mu_0$ $H_a: \mu < \mu_0$

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



Under H_0 gäller $T \sim t(n-1)$

$$p = P(T \le t_{obs}) = F_{t(n-1)}(t_{obs})$$

$F_{t(n-1)}(t)$ = fördelningsfunktion för Student's t, för $n-1$ frihetsgrader

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

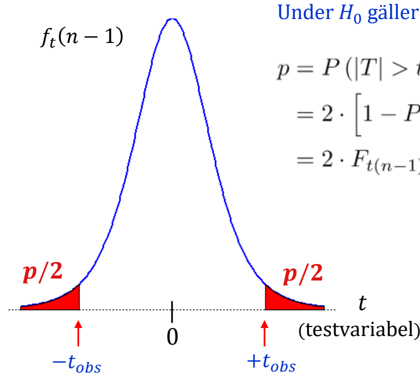
$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}}$$

www.matstat.org

Beräkning av p-värdet, H_a: $\mu \neq \mu_0$

$X_i \sim N(\mu, \sigma)$ σ okänd $H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$

p-värdet är lika med sannolikheten att erhålla ett utfall av testvariabeln minst så extremt som det faktiskt observerade, givet att nollhypotesen är sann.



Under H_0 gäller $T \sim t(n-1)$

positiv!

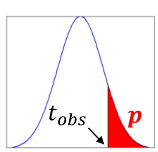
$$p = P(|T| > t_{obs}) = 2 \cdot P(T > |t_{obs}|)$$
$$= 2 \cdot \left[1 - P(T \leq |t_{obs}|)\right] = 2 \cdot \left[1 - F_{t(n-1)}(|t_{obs}|)\right)$$
$$= 2 \cdot F_{t(n-1)}(-|t_{obs}|)$$

$F_{t(n-1)}(t)$ = fördelningsfunktion för Student's t, för $n-1$ frihetsgrader

www.matstat.org

t-Test med R

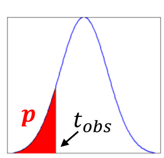
```
?t.test       # hjälp
x = c(32.2, 32, 30.4, 31, 31.2, 31.2, 30.3, 29.6, 30.5, 30.8)
```



hypotes $1-\alpha$

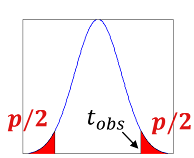
$H_a: \mu > \mu_0$

```
t.test(x, alternative = "greater", mu = 30, conf.level = 0.95)
```



$H_a: \mu < \mu_0$

```
t.test(x, alternative = ”less", mu = 30, conf.level = 0.95)
```



$H_a: \mu \neq \mu_0$

```
t.test(x, alternative = ”two.sided", mu = 30, conf.level = 0.95)
```

www.matstat.org

t-Test för två stickprov
("Two-sample t-test")

- testar om två oberoende, normalfördelade populationer uppvisar en viss hypotetisk skillnad $\Delta\mu$ mellan sina väntevärden
  - (mest testas om $\Delta\mu = 0$)
- Nollhypotes : H_0: $\mu_x - \mu_y = \Delta\mu_0$
  - $\Delta\mu_0$ mest 0, alltså H_0: $\mu_x = \mu_y$
- två stickprov → $\bar{x}_{obs}$; $\bar{y}_{obs}$; s_x ; s_y (observationer av punktskattningar)

$X_i \sim N(\mu_x, \sigma_x)$ σ_x okänd $x = (x_1, x_2, \dots x_n)$ stickprov 1

$Y_i \sim N(\mu_y, \sigma_y)$ σ_y okänd $y = (y_1, y_2, \dots y_m)$ stickprov 2

$$s_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (y_i - \bar{y})^2 \qquad s_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (x_i - \bar{x})^2$$

punktskattning för σ_y^2 punktskattning för σ_x^2

www.matstat.org

Kritiska områden, fall a)

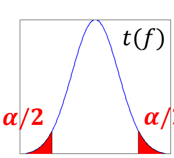
Fall a) Standardavvikelser okända, men lika $\boldsymbol{\sigma_x = \sigma_y = \sigma}$
(detta antagande hade vi redan för intervallskattning, föreläsning **F9**).

testvariabel (under H_0)
$$T = \frac{\bar{X} - \bar{Y} - \Delta\mu_0}{S_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t(n_x + n_y - 2)$$
$f = n_x + n_y - 2$ frihetsgrader

observation
$$t_{obs} = \frac{\bar{x}_{obs} - \bar{y}_{obs}}{s_p \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$
för H_0: $\Delta\mu_0 = 0$ ($\mu_x = \mu_y$)

$$s_p = \sqrt{\frac{(n_x - 1) \cdot s_x^2 + (n_y - 1) \cdot s_y^2}{(n_x - 1) + (n_y - 1)}}$$
"pooled standard deviation"

$$T \sim t(f) \Rightarrow P\left(|T| > t_{\alpha/2}(f)\right) = \alpha$$



Kritiska områden för tvåsidigt test, signifikansnivå α:

$$\Omega_\alpha = \left\{ |t_{obs}| > t_{\alpha/2}(f) \right\} \quad \text{för} \quad H_a : \Delta\mu \neq 0$$

(gör ensidiga test analogt, se formelsamling på matstat.org)

www.matstat.org

Kritiska områden, fall b)

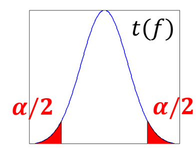
Fall b) Standardavvikelser okända och olika $\boldsymbol{\sigma_x \neq \sigma_y}$
(Welch-Test, Smith-Satterthwaite-Test)

testvariabel (under H_0) $$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} \sim t(f)$$ für H_0: $\Delta\mu_0 = 0$ $(\mu_x = \mu_y)$

observation $$t_{obs} = \frac{\bar{x}_{obs} - \bar{y}_{obs}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$ för H_0: $\Delta\mu_0 = 0$ $(\mu_x = \mu_y)$

$$f = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{\left(\frac{s_x^2}{n_x}\right)^2}{n_x - 1} + \frac{\left(\frac{s_y^2}{n_y}\right)^2}{n_x - 1}}$$ frihetsgrader, avrundas (ner) om inte heltal

$$T \sim t(f) \Rightarrow P\left(|T| > t_{\alpha/2}(f)\right) = \alpha$$



Kritiska områden(tvåsidigt), signifikansnivå α:

$$\Omega_\alpha = \left\{|t_{obs}| > t_{\alpha/2}(f)\right\} \quad \text{för} \quad H_a: \Delta\mu \neq 0$$

www.matstat.org

t-Test för två parade stickprov

Person	A	B	C	D	E	F	G	H
före	78.1	66.9	74.3	72.5	90.9	78.3	68.4	72.5
efter	79.2	67.0	77.1	73.3	92.0	78.1	68.4	72.9

Modell för parade stickprov (repetition från föreläsning **F9**):

$$\left.\begin{array}{l} X_i \sim N(\mu_i, \sigma_x) \\ Y_i \sim N(\mu_i + \Delta\mu, \sigma_y) \end{array}\right\}$$ systematisk förskjutning $\Delta\mu$ mellan båda grupper

$Z_i = Y_i - X_i$ beräknas

$Z_i \sim N(\Delta\mu, \sigma_z)$ fördelning för Z_i

$$\bar{Z} \sim N\left(\Delta\mu, \frac{\sigma_z}{\sqrt{n}}\right)$$ fördelning för medelvärdet av Z_i

$$\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2 - 2 \cdot \underbrace{\rho \cdot \sigma_x \sigma_y}}$$ standardavvikelse för σ_z (okänd)

$$C(X, Y) = \rho \cdot \sigma_x \sigma_y$$ kovarians mellan X och Y

t-Test för två parade stickprov

- Nollhypotes: H_0: $\Delta\mu = \Delta\mu_0$
  - $\Delta\mu_0$ mest 0, alltså H_0: $\Delta\mu = 0$ (ingen skillnad)

$$\bar{Z} \sim N\left(\Delta\mu, \frac{\sigma_z}{\sqrt{n}}\right)$$ fördelning för medelvärdet av Z_i

$$\frac{\bar{Z} - \Delta\mu_0}{\sigma_z/\sqrt{n}} \sim N(0,1)$$ σ_z är dock okänd, därför $\sigma_z \rightarrow S_z$ som förut

$$T = \frac{\bar{Z} - \Delta\mu_0}{S_z/\sqrt{n}} \sim t(n-1)$$ testvariabel med $S_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$

observation $t_{obs} = \frac{\bar{z}_{obs}}{s_z/\sqrt{n}}$ för $\Delta\mu_0 = 0$ $\Rightarrow$ H_0: $\Delta\mu = 0$

$$\bar{z}_{obs} = \frac{1}{n}\sum_{i=1}^{n} z_i \qquad s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z}_{obs})^2$$

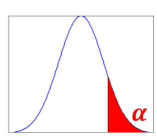
t-Test för två parade stickprov

Testvariabel $T = \frac{\bar{Z} - \Delta\mu_0}{S_z/\sqrt{n}} \sim t(n-1)$ observation H_a: $\mu > \mu_0$

Som vanligt förkastar vi nollhypotesen om en observation av testvariabeln T ligger för långt i riktning mot den alternativa hypotesen, alltså om den överskrider kvantilen $t_\alpha(n-1)$ resp. $t_{\alpha/2}(n-1)$ i riktning mot den alternativa hypotesen:

H_a: $\mu > \mu_0$ $P\left(T > t_\alpha(n-1)\right) = \alpha$

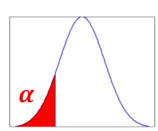
krit. område $\Omega_\alpha = \{t_{obs} > t_\alpha(n-1)\}$



förkasta H_0 om $t_{obs} > t_\alpha(n-1)$ signifikansnivå α

H_a: $\mu < \mu_0$ $P\left(T < -t_\alpha(n-1)\right) = \alpha$

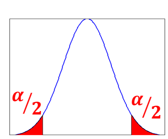
krit. område $\Omega_\alpha = \{t_{obs} < -t_\alpha(n-1)\}$



förkasta H_0 om $t_{obs} < -t_\alpha(n-1)$

H_a: $\mu \neq \mu_0$ $P\left(|T| > t_{\alpha/2}(n-1)\right) = \alpha$

krit. område $\Omega_\alpha = \left\{|t_{obs}| > t_{\alpha/2}(n-1)\right\}$



förkasta H_0 om $|t_{obs}| > t_{\alpha/2}(n-1)$

Sammanfattning: t-Test för två parade stickprov

$$X_i \sim N(\mu_i, \sigma_x)$$
$$Y_i \sim N(\mu_i + \Delta\mu, \sigma_y)$$

$\sigma_{x,y}$ okända $\quad H_0$: $\Delta\mu = 0$

testvariabel: $t_{obs} = \dfrac{\bar{z}_{obs}}{s_z/\sqrt{n}}$

$$z_i = y_i - x_i \qquad \bar{z}_{obs} = \frac{1}{n}\sum_{i=1}^{n} z_i$$
$$s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z}_{obs})^2$$

- Nollhypotes H_0: $\Delta\mu = 0$ (eller $\Delta\mu = \Delta\mu_0$)
- Alternativ hypotes H_a: $\Delta\mu \neq 0$ eller $\Delta\mu > 0$ eller $\Delta\mu < 0$
- signifikansnivå bestäms, t. ex. $\alpha = 0.05$
- observation av testvariabeln beräknas $t_{obs} = \cdots$
- H_0 förkastas om t_{obs} ligger i det kritiska området Ω_α.

H_a: $\mu > \mu_0 \quad \Omega_\alpha = \{t_{obs} > t_\alpha(n-1)\}$

H_a: $\mu < \mu_0 \quad \Omega_\alpha = \{t_{obs} < -t_\alpha(n-1)\}$

H_a: $\mu \neq \mu_0 \quad \Omega_\alpha = \{|t_{obs}| > t_{\alpha/2}(n-1)\}$

t-Test med R

hypotes $\quad 1-\alpha$

Ett stickprov, H_0: $\mu = \mu_0$; H_a: $\mu \neq \mu_0$
t.test(x, alternative = "two.sided", mu = 7500, conf.level = 0.95)

Två stickprov, lika varianser, H_0: $\mu_x = \mu_y$; H_a: $\mu_x \neq \mu_y$
t.test(x, y, alternative = "two.sided", mu = 0, var.equal = TRUE, conf.level = 0.95)

Två stickprov, olika varianser, H_0: $\mu_x = \mu_y$; H_a: $\mu_x \neq \mu_y$
t.test(x, y, alternative = "two.sided", mu = 0, var.equal = FALSE, conf.level = 0.95)

Två parade stickprov, H_0: $\Delta\mu = 0$; H_a: $\Delta\mu \neq 0$
t.test(x1, y1, alternative = "two.sided", paired = TRUE , mu = 0, conf.level = 0.95)

Är X, Y normalfördelade?:

hist(x, col="red")   # histogram, borde ungefär se ut som en normalfördelning
qqnorm(x); qqline(x, col = ”red”)   # punkterna borde ungefär ligga på den röda linjen
normalfördelning avslås med: ad.test (library(nortest) ; ks.test ; shapiro.test

Appendix

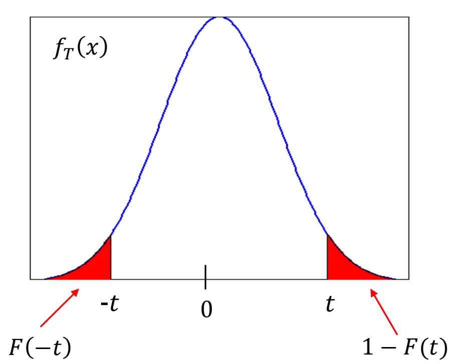
Hypotestest, del 2

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

Symmetri och fördelningsfunktion

Om täthetsfunktionen är symmetrisk kring noll [t. ex. N, $t(f)$] gäller för fördelningsfunktionen att::

$$F(-t) = 1 - F(t)$$

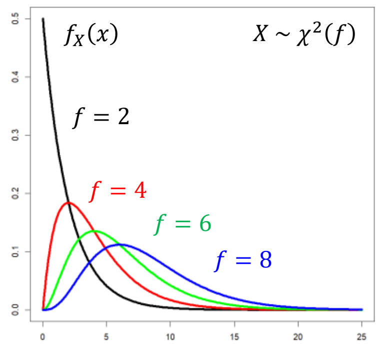


Chi-kvadrat-fördelningen

$$W = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{\sigma^2} = \frac{(n-1)\cdot S^2}{\sigma^2} \sim \chi^2(n-1)$$

Detta gäller allmänt om X_i:erna är oberoende, normalfördelade slumpvariabler med standardavvikelsen σ.

$$X_i \sim N(\mu, \sigma)$$



Täthetsfunktionen för χ^2 för olika frihetsgrader f.
Observationerna X_i måste komma från en normalfördelning.

www.matstat.org

t-fördelningen

(Students fördelning)

William Gosset
("Student")

Antaganden: $Z \sim N(0,1)$

$W \sim \chi^2(\nu)$

Om slumpvariablerna Z och W är oberoende och har ovanstående fördelningar har följande kvot en så kallad t-fördelning:

$$\boxed{\frac{Z}{\sqrt{\frac{W}{\nu}}} \sim t(\nu)}$$

t-fördelning med ν frihetsgrader

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1 + \frac{t^2}{\nu}\right)^{-(\frac{\nu+1}{2})}$$

täthetsfunktion

www.matstat.org

Students t-fördelning

- $Z \sim N(0,1)$ standardiserad normalfördelning
- $W \sim \chi^2(\nu)$ Chi-kvadrat-fördelning, låt $\nu = \text{n} - 1$ (frihetsgrader)
- Z och W oberoende

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \qquad W = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\frac{Z}{\sqrt{\frac{W}{\nu}}} = \frac{\frac{(\bar{X}-\mu_0)\cdot\sqrt{n}}{\sigma}}{\sqrt{\frac{(n-1)\cdot S^2}{\sigma^2\cdot(n-1)}}} = \frac{(\bar{X}-\mu_0)\cdot\sqrt{n}}{S} = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$$

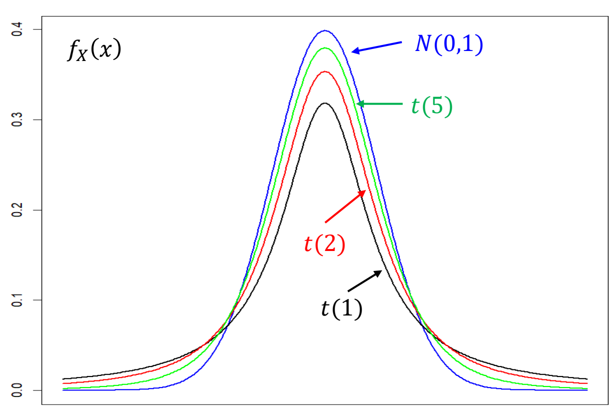
Vänster sida är $t(n-1)$ – fördelad, alltså måste detta även gäller för höger sida:

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

Slumpvariablerna X_i måste vara normalfördelade och oberoende.

Students t-fördelning

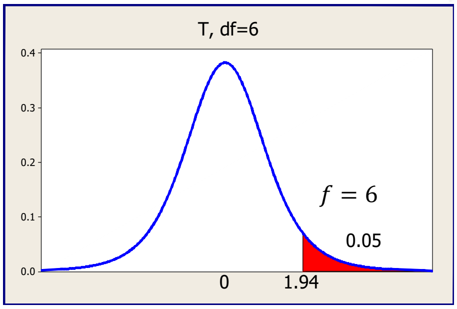
Täthetsfunktion för t-fördelningen för olika frihetsgrader och jämförelse med $N(0,1)$



$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-(\frac{\nu+1}{2})}$$

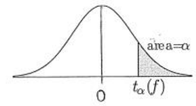
Täthetsfunktion
ν = antalet frihetsgrader

Kvantiler för t-fördelningen



t-fördelningen är lite bredare än N, men ju större n blir, desto mera liknar de varandra (se sista råden i tabellen).

$$P\left(X > t_\alpha(f)\right) = \alpha$$
$$X \in t(f)$$



f \ α	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	1.89	2.92	4.30	6.96	9.92	22.33	31.60
3	1.64	2.35	3.18	4.54	5.84	10.21	12.92
4	1.53	2.13	2.78	3.75	4.60	7.17	8.61
5	1.48	2.02	2.57	3.36	4.03	5.89	6.87
6	1.44	1.94	2.45	3.14	3.71	5.21	5.96
7	1.41	1.89	2.36	3.00	3.50	4.79	5.41
8	1.40	1.86	2.31	2.90	3.36	4.50	5.04
9	1.38	1.83	2.26	2.82	3.25	4.30	4.78
10	1.37	1.81	2.23	2.76	3.17	4.14	4.59
11	1.36	1.80	2.20	2.72	3.11	4.02	4.44
12	1.36	1.78	2.18	2.68	3.05	3.93	4.32
13	1.35	1.77	2.16	2.65	3.01	3.85	4.22
14	1.35	1.76	2.14	2.62	2.98	3.79	4.14
15	1.34	1.75	2.13	2.60	2.95	3.73	4.07
16	1.34	1.75	2.12	2.58	2.92	3.69	4.01
17	1.33	1.74	2.11	2.57	2.90	3.65	3.97
18	1.33	1.73	2.10	2.55	2.88	3.61	3.92
19	1.33	1.73	2.09	2.54	2.86	3.58	3.88
20	1.33	1.72	2.09	2.53	2.85	3.55	3.85
21	1.32	1.72	2.08	2.52	2.83	3.53	3.82
22	1.32	1.72	2.07	2.51	2.82	3.50	3.79
23	1.32	1.71	2.07	2.50	2.81	3.48	3.77
24	1.32	1.71	2.06	2.49	2.80	3.47	3.75
25	1.32	1.71	2.06	2.49	2.79	3.45	3.73
26	1.31	1.71	2.06	2.48	2.78	3.43	3.71
27	1.31	1.70	2.05	2.47	2.77	3.42	3.69
28	1.31	1.70	2.05	2.47	2.76	3.41	3.67
29	1.31	1.70	2.05	2.46	2.76	3.40	3.66
30	1.31	1.70	2.04	2.46	2.75	3.39	3.65
40	1.30	1.68	2.02	2.42	2.70	3.31	3.55
60	1.30	1.67	2.00	2.39	2.66	3.23	3.46
120	1.29	1.66	1.98	2.36	2.62	3.16	3.37
∞	1.28	1.64	1.96	2.33	2.58	3.09	3.29 $= \lambda_\alpha$

www.matstat.org

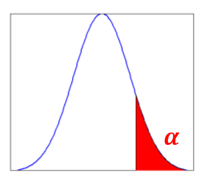
t-Test, Kritiska områden

Observation av testvariabel

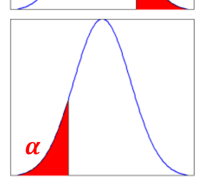
$$t_{obs} = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}}$$

H_0: $\mu = \mu_0$

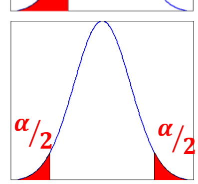
H_a	test	kritiskt område
$\mu > \mu_0$	ensidigt	$\Omega_\alpha = \{t > t_\alpha(n-1)\}$



H_a	test	kritiskt område
$\mu < \mu_0$	ensidigt	$\Omega_\alpha = \{t < -t_\alpha(n-1)\}$



H_a	test	kritiskt område
$\mu \neq \mu_0$	tvåsidigt	$\Omega_\alpha = \left\{\lvert t\rvert > t_{\alpha/2}(n-1)\right\}$



www.matstat.org