# STK1110 Høsten 2025

## Ikke-parametriske metoder

Tilsvarer Avsnitt 14.1 og 14.2

# Motivasjon

# Motivasjon

- Metodene for å lage konfidensintervaller og teste hypoteser i Kap. 8, 9, 10, 11 og 12 er nesten utelukkende basert på en antakelse om normalfordelte data.

# Motivasjon

- Metodene for å lage konfidensintervaller og teste hypoteser i Kap. 8, 9, 10, 11 og 12 er nesten utelukkende basert på en antakelse om normalfordelte data.
- Så lenge antakelsen er tilnærmet oppfylt, vil konfidens- og signifikansnivåer typisk være omtrent som de skal.

# Motivasjon

- Metodene for å lage konfidensintervaller og teste hypoteser i Kap. 8, 9, 10, 11 og 12 er nesten utelukkende basert på en antakelse om normalfordelte data.
- Så lenge antakelsen er tilnærmet oppfylt, vil konfidens- og signifikansnivåer typisk være omtrent som de skal.
- Dersom den sanne fordelingen avviker mye fra normalfordelingen, kan imidlertid det reelle konfidensnivået være en del lavere, eller signifikansnivået være en del høyere enn ønsket.

# Motivasjon

- Metodene for å lage konfidensintervaller og teste hypoteser i Kap. 8, 9, 10, 11 og 12 er nesten utelukkende basert på en antakelse om normalfordelte data.
- Så lenge antakelsen er tilnærmet oppfylt, vil konfidens- og signifikansnivåer typisk være omtrent som de skal.
- Dersom den sanne fordelingen avviker mye fra normalfordelingen, kan imidlertid det reelle konfidensnivået være en del lavere, eller signifikansnivået være en del høyere enn ønsket.
- Et alternativ kan da være ikke-parametriske metoder som bootstrap persentilintervaller fra Kap. 8.5, eller fortegn- og Wilcoxon-testene vi skal se på nå.

# Konfidensintervall for $100p$-persentilen

# Konfidensintervall for $100p$-persentilen

- Anta at $x_1, \ldots, x_n$ er observasjoner av de stokastiske variablene $X_1, \ldots, X_n$, som er uavhengig identisk fordelt fra en kontinuerlig fordeling, og vi ønsker å si noe om det "typiske" nivået til $X_i$.

# Konfidensintervall for $100p$-persentilen

- Anta at $x_1, \ldots, x_n$ er observasjoner av de stokastiske variablene $X_1, \ldots, X_n$, som er uavhengig identisk fordelt fra en kontinuerlig fordeling, og vi ønsker å si noe om det "typiske" nivået til $X_i$.
- Dersom fordelingen til $X_i$ er nokså skjev, er medianen $\tilde{\mu}$ et mer naturlig mål å se på enn forventningen $\mu$.

# Konfidensintervall for $100p$-persentilen

- Anta at $x_1, \ldots, x_n$ er observasjoner av de stokastiske variablene $X_1, \ldots, X_n$, som er uavhengig identisk fordelt fra en kontinuerlig fordeling, og vi ønsker å si noe om det "typiske" nivået til $X_i$.
- Dersom fordelingen til $X_i$ er nokså skjev, er medianen $\tilde{\mu}$ et mer naturlig mål å se på enn forventningen $\mu$.
- Mer generelt skal vi finne en metode for å lage konfidensintervall for $100p$-persentilen $\eta_p$

# Konfidensintervall for $100p$-persentilen

- Anta at $x_1, \ldots, x_n$ er observasjoner av de stokastiske variablene $X_1, \ldots, X_n$, som er uavhengig identisk fordelt fra en kontinuerlig fordeling, og vi ønsker å si noe om det "typiske" nivået til $X_i$.
- Dersom fordelingen til $X_i$ er nokså skjev, er medianen $\tilde{\mu}$ et mer naturlig mål å se på enn forventningen $\mu$.
- Mer generelt skal vi finne en metode for å lage konfidensintervall for $100p$-persentilen $\eta_p$ $(0 < p < 1)$

# Konfidensintervall for $100p$-persentilen

- Anta at $x_1, \ldots, x_n$ er observasjoner av de stokastiske variablene $X_1, \ldots, X_n$, som er uavhengig identisk fordelt fra en kontinuerlig fordeling, og vi ønsker å si noe om det "typiske" nivået til $X_i$.
- Dersom fordelingen til $X_i$ er nokså skjev, er medianen $\tilde{\mu}$ et mer naturlig mål å se på enn forventningen $\mu$.
- Mer generelt skal vi finne en metode for å lage konfidensintervall for $100p$-persentilen $\eta_p$ $(0 < p < 1)$ i fordelingen til $X_i$

# Konfidensintervall for $100p$-persentilen

- Anta at $x_1, \ldots, x_n$ er observasjoner av de stokastiske variablene $X_1, \ldots, X_n$, som er uavhengig identisk fordelt fra en kontinuerlig fordeling, og vi ønsker å si noe om det "typiske" nivået til $X_i$.
- Dersom fordelingen til $X_i$ er nokså skjev, er medianen $\tilde{\mu}$ et mer naturlig mål å se på enn forventningen $\mu$.
- Mer generelt skal vi finne en metode for å lage konfidensintervall for $100p$-persentilen $\eta_p$ $(0 < p < 1)$ i fordelingen til $X_i$ ($\tilde{\mu}$ tilsvarer da $\eta_{0.5}$).

# Konfidensintervall for $100p$-persentilen

# Konfidensintervall for $100p$-persentilen

- La $Y_i = X_{(i)}$ være de ordnede verdiene, altså

# Konfidensintervall for $100p$-persentilen

- La $Y_i = X_{(i)}$ være de ordnede verdiene, altså

$$X_{(1)} \leq \cdots \leq X_{(n)}$$

# Konfidensintervall for $100p$-persentilen

- La $Y_i = X_{(i)}$ være de ordnede verdiene, altså

$$X_{(1)} \leq \cdots \leq X_{(n)}$$

slik at vi har en ordningsobservator

# Konfidensintervall for $100p$-persentilen

- La $Y_i = X_{(i)}$ være de ordnede verdiene, altså

$$X_{(1)} \leq \ldots \leq X_{(n)}$$

slik at vi har en ordningsobservator

$$Y_1 \leq \ldots \leq Y_n.$$

# Konfidensintervall for $100p$-persentilen

- La $Y_i = X_{(i)}$ være de ordnede verdiene, altså

$$X_{(1)} \leq \ldots \leq X_{(n)}$$

slik at vi har en ordningsobservator

$$Y_1 \leq \ldots \leq Y_n.$$

- Da $X_1, \ldots, X_n$ er kontinuerlige, har ingen av dem like verdier.

# Konfidensintervall for $100p$-persentilen

- La $Y_i = X_{(i)}$ være de ordnede verdiene, altså

$$X_{(1)} \leq \ldots \leq X_{(n)}$$

slik at vi har en ordningsobservator

$$Y_1 \leq \ldots \leq Y_n.$$

- Da $X_1, \ldots, X_n$ er kontinuerlige, har ingen av dem like verdier.
- For hvilke som helst heltall $r$ og $s$ med $1 \leq r < s \leq 1$ blir da

# Konfidensintervall for $100p$-persentilen

- La $Y_i = X_{(i)}$ være de ordnede verdiene, altså

$$X_{(1)} \leq \ldots \leq X_{(n)}$$

slik at vi har en ordningsobservator

$$Y_1 \leq \ldots \leq Y_n.$$

- Da $X_1, \ldots, X_n$ er kontinuerlige, har ingen av dem like verdier.
- For hvilke som helst heltall $r$ og $s$ med $1 \leq r < s \leq 1$ blir da

$$\mathrm{P}(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}.$$

# Konfidensintervall for $100p$-persentilen

- La $Y_i = X_{(i)}$ være de ordnede verdiene, altså

$$X_{(1)} \leq \ldots \leq X_{(n)}$$

slik at vi har en ordningsobservator

$$Y_1 \leq \ldots \leq Y_n.$$

- Da $X_1, \ldots, X_n$ er kontinuerlige, har ingen av dem like verdier.
- For hvilke som helst heltall $r$ og $s$ med $1 \leq r < s \leq 1$ blir da

$$\mathrm{P}(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}.$$

- Et $(1-\alpha) \cdot 100\%$ konfidensintervall for $\eta_p$

# Konfidensintervall for $100p$-persentilen

- La $Y_i = X_{(i)}$ være de ordnede verdiene, altså

$$X_{(1)} \leq \ldots \leq X_{(n)}$$

slik at vi har en ordningsobservator

$$Y_1 \leq \ldots \leq Y_n.$$

- Da $X_1, \ldots, X_n$ er kontinuerlige, har ingen av dem like verdier.
- For hvilke som helst heltall $r$ og $s$ med $1 \leq r < s \leq 1$ blir da

$$\mathrm{P}(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}.$$

- Et $(1-\alpha) \cdot 100\%$ konfidensintervall for $\eta_p$ oppnås da ved å finne $r$ og $s$

# Konfidensintervall for $100p$-persentilen

- La $Y_i = X_{(i)}$ være de ordnede verdiene, altså

$$X_{(1)} \leq \ldots \leq X_{(n)}$$

slik at vi har en ordningsobservator

$$Y_1 \leq \ldots \leq Y_n.$$

- Da $X_1, \ldots, X_n$ er kontinuerlige, har ingen av dem like verdier.
- For hvilke som helst heltall $r$ og $s$ med $1 \leq r < s \leq 1$ blir da

$$\mathrm{P}(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}.$$

- Et $(1-\alpha) \cdot 100\%$ konfidensintervall for $\eta_p$ oppnås da ved å finne $r$ og $s$ slik at $\mathrm{P}(Y_r \leq \eta_p \leq Y_s)$ kommer nærmest mulig $1-\alpha$ med

# Konfidensintervall for $100p$-persentilen

- La $Y_i = X_{(i)}$ være de ordnede verdiene, altså

$$X_{(1)} \leq \ldots \leq X_{(n)}$$

slik at vi har en ordningsobservator

$$Y_1 \leq \ldots \leq Y_n.$$

- Da $X_1, \ldots, X_n$ er kontinuerlige, har ingen av dem like verdier.
- For hvilke som helst heltall $r$ og $s$ med $1 \leq r < s \leq 1$ blir da

$$\mathrm{P}(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}.$$

- Et $(1-\alpha) \cdot 100\%$ konfidensintervall for $\eta_p$ oppnås da ved å finne $r$ og $s$ slik at $\mathrm{P}(Y_r \leq \eta_p \leq Y_s)$ kommer nærmest mulig $1-\alpha$ med

$$\mathrm{P}(Y_r \leq \eta_p \leq Y_s) \geq 1-\alpha.$$

# Konfidensintervall for $100p$-persentilen

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1 - \alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n, p)$,

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1 - \alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n, p)$, som vi kan finne med R.

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1 - \alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n, p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret,

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1 - \alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n, p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret, vil en som regel ikke få det eksakte konfidensnivået en ønsker,

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1 - \alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n, p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret, vil en som regel ikke få det eksakte konfidensnivået en ønsker, men det vil i alle fall ikke være lavere.

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1 - \alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n, p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret, vil en som regel ikke få det eksakte konfidensnivået en ønsker, men det vil i alle fall ikke være lavere.
- Alternativt, dersom $np > 10$ og $n(1 - p) > 10$,

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1 - \alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n, p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret, vil en som regel ikke få det eksakte konfidensnivået en ønsker, men det vil i alle fall ikke være lavere.
- Alternativt, dersom $np > 10$ og $n(1 - p) > 10$, kan en tilnærme den binomiske fordelingen med normalfordelingen $N(np, np(1 - p))$

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1 - \alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n, p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret, vil en som regel ikke få det eksakte konfidensnivået en ønsker, men det vil i alle fall ikke være lavere.
- Alternativt, dersom $np > 10$ og $n(1 - p) > 10$, kan en tilnærme den binomiske fordelingen med normalfordelingen $N(np, np(1 - p))$ med kontinuitetskorreksjon,

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1 - \alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n, p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret, vil en som regel ikke få det eksakte konfidensnivået en ønsker, men det vil i alle fall ikke være lavere.
- Alternativt, dersom $np > 10$ og $n(1 - p) > 10$, kan en tilnærme den binomiske fordelingen med normalfordelingen $N(np, np(1 - p))$ med kontinuitetskorreksjon, slik at $r$ og $s$ tilnærmes med

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1-\alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n,p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret, vil en som regel ikke få det eksakte konfidensnivået en ønsker, men det vil i alle fall ikke være lavere.
- Alternativt, dersom $np > 10$ og $n(1-p) > 10$, kan en tilnærme den binomiske fordelingen med normalfordelingen $N(np, np(1-p))$ med kontinuitetskorreksjon, slik at $r$ og $s$ tilnærmes med

$$r - \tfrac{1}{2} \approx np - z_{\alpha/2}\sqrt{np(1-p)}$$

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1-\alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n,p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret, vil en som regel ikke få det eksakte konfidensnivået en ønsker, men det vil i alle fall ikke være lavere.
- Alternativt, dersom $np > 10$ og $n(1-p) > 10$, kan en tilnærme den binomiske fordelingen med normalfordelingen $N(np, np(1-p))$ med kontinuitetskorreksjon, slik at $r$ og $s$ tilnærmes med

$$r - \tfrac{1}{2} \approx np - z_{\alpha/2}\sqrt{np(1-p)} \qquad \Rightarrow r \approx np + \tfrac{1}{2} - z_{\alpha/2}\sqrt{np(1-p)}$$

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1-\alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n,p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret, vil en som regel ikke få det eksakte konfidensnivået en ønsker, men det vil i alle fall ikke være lavere.
- Alternativt, dersom $np > 10$ og $n(1-p) > 10$, kan en tilnærme den binomiske fordelingen med normalfordelingen $N(np, np(1-p))$ med kontinuitetskorreksjon, slik at $r$ og $s$ tilnærmes med

$$\begin{aligned} r - \tfrac{1}{2} &\approx np - z_{\alpha/2}\sqrt{np(1-p)} \\ (s-1) + \tfrac{1}{2} &\approx np + z_{\alpha/2}\sqrt{np(1-p)} \end{aligned} \qquad \Rightarrow r \approx np + \tfrac{1}{2} - z_{\alpha/2}\sqrt{np(1-p)}$$

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1-\alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n,p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret, vil en som regel ikke få det eksakte konfidensnivået en ønsker, men det vil i alle fall ikke være lavere.
- Alternativt, dersom $np > 10$ og $n(1-p) > 10$, kan en tilnærme den binomiske fordelingen med normalfordelingen $N(np, np(1-p))$ med kontinuitetskorreksjon, slik at $r$ og $s$ tilnærmes med

$$\begin{aligned} r - \tfrac{1}{2} &\approx np - z_{\alpha/2}\sqrt{np(1-p)} && \Rightarrow r \approx np + \tfrac{1}{2} - z_{\alpha/2}\sqrt{np(1-p)} \\ (s-1) + \tfrac{1}{2} &\approx np + z_{\alpha/2}\sqrt{np(1-p)} && \Rightarrow s \approx np + \tfrac{1}{2} + z_{\alpha/2}\sqrt{np(1-p)}, \end{aligned}$$

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1-\alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n,p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret, vil en som regel ikke få det eksakte konfidensnivået en ønsker, men det vil i alle fall ikke være lavere.
- Alternativt, dersom $np > 10$ og $n(1-p) > 10$, kan en tilnærme den binomiske fordelingen med normalfordelingen $N(np, np(1-p))$ med kontinuitetskorreksjon, slik at $r$ og $s$ tilnærmes med

$$\begin{aligned} r - \tfrac{1}{2} &\approx np - z_{\alpha/2}\sqrt{np(1-p)} && \Rightarrow r \approx np + \tfrac{1}{2} - z_{\alpha/2}\sqrt{np(1-p)} \\ (s-1) + \tfrac{1}{2} &\approx np + z_{\alpha/2}\sqrt{np(1-p)} && \Rightarrow s \approx np + \tfrac{1}{2} + z_{\alpha/2}\sqrt{np(1-p)}, \end{aligned}$$

som så rundes opp til nærmeste heltall.

# Konfidensintervall for $100p$-persentilen

- Disse er henholdsvis øvre $(1-\alpha/2)$- og $\alpha/2$-kvantiler i den binomiske fordelingen $Bin(n,p)$, som vi kan finne med R.
- Da den binomiske fordelingen er diskret, vil en som regel ikke få det eksakte konfidensnivået en ønsker, men det vil i alle fall ikke være lavere.
- Alternativt, dersom $np > 10$ og $n(1-p) > 10$, kan en tilnærme den binomiske fordelingen med normalfordelingen $N(np, np(1-p))$ med kontinuitetskorreksjon, slik at $r$ og $s$ tilnærmes med

$$\begin{aligned} r - \tfrac{1}{2} &\approx np - z_{\alpha/2}\sqrt{np(1-p)} && \Rightarrow r \approx np + \tfrac{1}{2} - z_{\alpha/2}\sqrt{np(1-p)} \\ (s-1) + \tfrac{1}{2} &\approx np + z_{\alpha/2}\sqrt{np(1-p)} && \Rightarrow s \approx np + \tfrac{1}{2} + z_{\alpha/2}\sqrt{np(1-p)}, \end{aligned}$$

som så rundes opp til nærmeste heltall.

## Eksempel

Eks. 14.1 i boka.

# Utledning av $\mathrm{P}(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}$

# Utledning av $\mathrm{P}(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}$

For en hver gitt $k$ må vi ha

# Utledning av $\mathrm{P}(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}$

For en hver gitt $k$ må vi ha

$$\mathrm{P}(Y_k \leq \eta_p \leq Y_{k+1}) = \binom{n}{k} p^k (1-p)^{n-k}$$

# Utledning av $\mathrm{P}(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}$

For en hver gitt $k$ må vi ha

$$\mathrm{P}(Y_k \leq \eta_p \leq Y_{k+1}) = \binom{n}{k} p^k (1-p)^{n-k}$$

siden dette betyr at da er eksakt $k$ av $Y$-ene mindre enn $\eta_p$.

# Utledning av $P(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}$

For en hver gitt $k$ må vi ha

$$P(Y_k \leq \eta_p \leq Y_{k+1}) = \binom{n}{k} p^k (1-p)^{n-k}$$

siden dette betyr at da er eksakt $k$ av $Y$-ene mindre enn $\eta_p$.

Men begivenheten $Y_r \leq \eta_p \leq Y_s$ må være unionen over de disjunkte begivenhetene

# Utledning av $\mathrm{P}(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}$

For en hver gitt $k$ må vi ha

$$\mathrm{P}(Y_k \leq \eta_p \leq Y_{k+1}) = \binom{n}{k} p^k (1-p)^{n-k}$$

siden dette betyr at da er eksakt $k$ av $Y$-ene mindre enn $\eta_p$.

Men begivenheten $Y_r \leq \eta_p \leq Y_s$ må være unionen over de disjunkte begivenhetene

$$Y_r \leq \eta_p \leq Y_{r+1}$$

# Utledning av $P(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}$

For en hver gitt $k$ må vi ha

$$P(Y_k \leq \eta_p \leq Y_{k+1}) = \binom{n}{k} p^k (1-p)^{n-k}$$

siden dette betyr at da er eksakt $k$ av $Y$-ene mindre enn $\eta_p$.

Men begivenheten $Y_r \leq \eta_p \leq Y_s$ må være unionen over de disjunkte begivenhetene

$$Y_r \leq \eta_p \leq Y_{r+1}$$
$$Y_{r+1} \leq \eta_p \leq Y_{r+2}$$

# Utledning av $\mathrm{P}(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}$

For en hver gitt $k$ må vi ha

$$\mathrm{P}(Y_k \leq \eta_p \leq Y_{k+1}) = \binom{n}{k} p^k (1-p)^{n-k}$$

siden dette betyr at da er eksakt $k$ av $Y$-ene mindre enn $\eta_p$.

Men begivenheten $Y_r \leq \eta_p \leq Y_s$ må være unionen over de disjunkte begivenhetene

$$\begin{aligned} Y_r &\leq \eta_p \leq Y_{r+1} \\ Y_{r+1} &\leq \eta_p \leq Y_{r+2} \\ &\vdots \\ Y_{s-1} &\leq \eta_p \leq Y_s \end{aligned}$$

# Utledning av $P(Y_r \leq \eta_p \leq Y_s) = \sum_{k=r}^{s-1} \binom{n}{k} p^k (1-p)^{n-k}$

For en hver gitt $k$ må vi ha

$$P(Y_k \leq \eta_p \leq Y_{k+1}) = \binom{n}{k} p^k (1-p)^{n-k}$$

siden dette betyr at da er eksakt $k$ av $Y$-ene mindre enn $\eta_p$.

Men begivenheten $Y_r \leq \eta_p \leq Y_s$ må være unionen over de disjunkte begivenhetene

$$\begin{aligned} Y_r &\leq \eta_p \leq Y_{r+1} \\ Y_{r+1} &\leq \eta_p \leq Y_{r+2} \\ &\vdots \\ Y_{s-1} &\leq \eta_p \leq Y_s \end{aligned}$$

og dermed følger resultatet.

# Fortegntesten for medianen $\tilde{\mu}$

# Fortegntesten for medianen $\tilde{\mu}$

- Anta at vi nå ønsker å teste hypoteser knyttet til medianen $\tilde{\mu}$.

# Fortegntesten for medianen $\tilde{\mu}$

- Anta at vi nå ønsker å teste hypoteser knyttet til medianen $\tilde{\mu}$.
- Anta først at vi vil teste om

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

# Fortegntesten for medianen $\tilde{\mu}$

- Anta at vi nå ønsker å teste hypoteser knyttet til medianen $\tilde{\mu}$.
- Anta først at vi vil teste om

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

(der $\tilde{\mu}$ er faktisk og ukjent median og $\tilde{\mu}_0$ en gitt kandidat)

# Fortegntesten for medianen $\tilde{\mu}$

- Anta at vi nå ønsker å teste hypoteser knyttet til medianen $\tilde{\mu}$.
- Anta først at vi vil teste om

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

(der $\tilde{\mu}$ er faktisk og ukjent median og $\tilde{\mu}_0$ en gitt kandidat)

- La så $p = \mathrm{P}(X_i > \tilde{\mu}_0)$.

# Fortegntesten for medianen $\tilde{\mu}$

- Anta at vi nå ønsker å teste hypoteser knyttet til medianen $\tilde{\mu}$.
- Anta først at vi vil teste om

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

(der $\tilde{\mu}$ er faktisk og ukjent median og $\tilde{\mu}_0$ en gitt kandidat)

- La så $p = \mathrm{P}(X_i > \tilde{\mu}_0)$.
- Under $H_0$ er da

# Fortegntesten for medianen $\tilde{\mu}$

- Anta at vi nå ønsker å teste hypoteser knyttet til medianen $\tilde{\mu}$.
- Anta først at vi vil teste om

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

(der $\tilde{\mu}$ er faktisk og ukjent median og $\tilde{\mu}_0$ en gitt kandidat)

- La så $p = \mathrm{P}(X_i > \tilde{\mu}_0)$.
- Under $H_0$ er da

$$p = \mathrm{P}(X_i > \tilde{\mu}_0) \leq \mathrm{P}(X_i > \tilde{\mu}) = 0.5,$$

# Fortegntesten for medianen $\tilde{\mu}$

- Anta at vi nå ønsker å teste hypoteser knyttet til medianen $\tilde{\mu}$.
- Anta først at vi vil teste om

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

(der $\tilde{\mu}$ er faktisk og ukjent median og $\tilde{\mu}_0$ en gitt kandidat)

- La så $p = \mathrm{P}(X_i > \tilde{\mu}_0)$.
- Under $H_0$ er da

$$p = \mathrm{P}(X_i > \tilde{\mu}_0) \leq \mathrm{P}(X_i > \tilde{\mu}) = 0.5,$$

mens dersom $H_0$ ikke er sann, er

# Fortegntesten for medianen $\tilde{\mu}$

- Anta at vi nå ønsker å teste hypoteser knyttet til medianen $\tilde{\mu}$.
- Anta først at vi vil teste om

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

(der $\tilde{\mu}$ er faktisk og ukjent median og $\tilde{\mu}_0$ en gitt kandidat)

- La så $p = \mathrm{P}(X_i > \tilde{\mu}_0)$.
- Under $H_0$ er da

$$p = \mathrm{P}(X_i > \tilde{\mu}_0) \leq \mathrm{P}(X_i > \tilde{\mu}) = 0.5,$$

mens dersom $H_0$ ikke er sann, er

$$p = \mathrm{P}(X_i > \tilde{\mu}_0) > \mathrm{P}(X_i > \tilde{\mu}) = 0.5.$$

# Fortegntesten for medianen $\tilde{\mu}$

- Anta at vi nå ønsker å teste hypoteser knyttet til medianen $\tilde{\mu}$.
- Anta først at vi vil teste om

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

(der $\tilde{\mu}$ er faktisk og ukjent median og $\tilde{\mu}_0$ en gitt kandidat)

- La så $p = \mathrm{P}(X_i > \tilde{\mu}_0)$.
- Under $H_0$ er da

$$p = \mathrm{P}(X_i > \tilde{\mu}_0) \leq \mathrm{P}(X_i > \tilde{\mu}) = 0.5,$$

mens dersom $H_0$ ikke er sann, er

$$p = \mathrm{P}(X_i > \tilde{\mu}_0) > \mathrm{P}(X_i > \tilde{\mu}) = 0.5.$$

- Vi kan dermed reformulere hypotesene til

# Fortegntesten for medianen $\tilde{\mu}$

- Anta at vi nå ønsker å teste hypoteser knyttet til medianen $\tilde{\mu}$.
- Anta først at vi vil teste om

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

(der $\tilde{\mu}$ er faktisk og ukjent median og $\tilde{\mu}_0$ en gitt kandidat)

- La så $p = \mathrm{P}(X_i > \tilde{\mu}_0)$.
- Under $H_0$ er da

$$p = \mathrm{P}(X_i > \tilde{\mu}_0) \leq \mathrm{P}(X_i > \tilde{\mu}) = 0.5,$$

mens dersom $H_0$ ikke er sann, er

$$p = \mathrm{P}(X_i > \tilde{\mu}_0) > \mathrm{P}(X_i > \tilde{\mu}) = 0.5.$$

- Vi kan dermed reformulere hypotesene til

$$H_0 : p \leq 0.5 \text{ mot } \quad H_a : p > 0.5.$$

# Fortegntesten for medianen $\tilde{\mu}$

# Fortegntesten for medianen $\tilde{\mu}$

- Da kan vi bruke testprosedyrene for populasjonsandeler fra Avsnitt 9.3 med testobservatoren $T = \sum_{i=1}^{n} I(X_i > \tilde{\mu}_0)$, der $I(\cdot)$ er indikatorfunksjonen.

# Fortegntesten for medianen $\tilde{\mu}$

- Da kan vi bruke testprosedyrene for populasjonsandeler fra Avsnitt 9.3 med testobservatoren $T = \sum_{i=1}^{n} I(X_i > \tilde{\mu}_0)$, der $I(\cdot)$ er indikatorfunksjonen.
- Alternativt kan en se på størrelsene $X_i - \tilde{\mu}_0$, $i = 1, \ldots, n$.

# Fortegntesten for medianen $\tilde{\mu}$

- Da kan vi bruke testprosedyrene for populasjonsandeler fra Avsnitt 9.3 med testobservatoren $T = \sum_{i=1}^{n} I(X_i > \tilde{\mu}_0)$, der $I(\cdot)$ er indikatorfunksjonen.
- Alternativt kan en se på størrelsene $X_i - \tilde{\mu}_0$, $i = 1, \ldots, n$.
- Disse er positive når $X_i > \tilde{\mu}_0$, og negative ellers.

# Fortegntesten for medianen $\tilde{\mu}$

- Da kan vi bruke testprosedyrene for populasjonsandeler fra Avsnitt 9.3 med testobservatoren $T = \sum_{i=1}^{n} I(X_i > \tilde{\mu}_0)$, der $I(\cdot)$ er indikatorfunksjonen.
- Alternativt kan en se på størrelsene $X_i - \tilde{\mu}_0$, $i = 1, \ldots, n$.
- Disse er positive når $X_i > \tilde{\mu}_0$, og negative ellers.
- Testobservatoren $T$ over, er da det samme som antall $(X_i - \tilde{\mu}_0)$-er som er positive.

# Fortegntesten for medianen $\tilde{\mu}$

- Da kan vi bruke testprosedyrene for populasjonsandeler fra Avsnitt 9.3 med testobservatoren $T = \sum_{i=1}^{n} I(X_i > \tilde{\mu}_0)$, der $I(\cdot)$ er indikatorfunksjonen.
- Alternativt kan en se på størrelsene $X_i - \tilde{\mu}_0$, $i = 1, \ldots, n$.
- Disse er positive når $X_i > \tilde{\mu}_0$, og negative ellers.
- Testobservatoren $T$ over, er da det samme som antall $(X_i - \tilde{\mu}_0)$-er som er positive.
- En kan altså telle de positive fortegnene til $(X_i - \tilde{\mu}_0)$-ene, og hypotesetesten kalles derfor fortegntesten ("sign test").

# Fortegntesten for medianen $\tilde{\mu}$

- Da kan vi bruke testprosedyrene for populasjonsandeler fra Avsnitt 9.3 med testobservatoren $T = \sum_{i=1}^{n} I(X_i > \tilde{\mu}_0)$, der $I(\cdot)$ er indikatorfunksjonen.
- Alternativt kan en se på størrelsene $X_i - \tilde{\mu}_0$, $i = 1, \ldots, n$.
- Disse er positive når $X_i > \tilde{\mu}_0$, og negative ellers.
- Testobservatoren $T$ over, er da det samme som antall $(X_i - \tilde{\mu}_0)$-er som er positive.
- En kan altså telle de positive fortegnene til $(X_i - \tilde{\mu}_0)$-ene, og hypotesetesten kalles derfor fortegntesten ("sign test").

## Eksempel

Eks. 14.2 i boka.

# Fortegntesten for parvise data

# Fortegntesten for parvise data

- Fortegntesten kan også brukes på parvise data.

# Fortegntesten for parvise data

- Fortegntesten kan også brukes på parvise data.
- Anta at en har de parvise målingene $(X_1, Y_1)$, ...,$(X_n, Y_n)$, og la $D_i = X_i - Y_i$, $i = 1, \ldots, n$.

# Fortegntesten for parvise data

- Fortegntesten kan også brukes på parvise data.
- Anta at en har de parvise målingene $(X_1, Y_1)$, ...,$(X_n, Y_n)$, og la $D_i = X_i - Y_i$, $i = 1, \ldots, n$.
- Anta at vi vil teste om $X_i$ typisk er større enn eller lik $Y_i$, slik at $D_i$ typisk er $> 0$.

# Fortegntesten for parvise data

- Fortegntesten kan også brukes på parvise data.
- Anta at en har de parvise målingene $(X_1, Y_1)$, ...,$(X_n, Y_n)$, og la $D_i = X_i - Y_i$, $i = 1, \ldots, n$.
- Anta at vi vil teste om $X_i$ typisk er større enn eller lik $Y_i$, slik at $D_i$ typisk er $> 0$.
- La $\tilde{\mu}_D$ være medianen til $D_i$

# Fortegntesten for parvise data

- Fortegntesten kan også brukes på parvise data.
- Anta at en har de parvise målingene $(X_1, Y_1)$, ...,$(X_n, Y_n)$, og la $D_i = X_i - Y_i$, $i = 1, \ldots, n$.
- Anta at vi vil teste om $X_i$ typisk er større enn eller lik $Y_i$, slik at $D_i$ typisk er $> 0$.
- La $\tilde{\mu}_D$ være medianen til $D_i$
- Da kan vi teste

$$H_0 : \tilde{\mu}_D \leq 0 \text{ mot } H_a : \tilde{\mu}_D > 0.$$

# Fortegntesten for parvise data

- Fortegntesten kan også brukes på parvise data.
- Anta at en har de parvise målingene $(X_1, Y_1)$, ...,$(X_n, Y_n)$, og la $D_i = X_i - Y_i$, $i = 1, \ldots, n$.
- Anta at vi vil teste om $X_i$ typisk er større enn eller lik $Y_i$, slik at $D_i$ typisk er $> 0$.
- La $\tilde{\mu}_D$ være medianen til $D_i$
- Da kan vi teste

$$H_0 : \tilde{\mu}_D \leq 0 \text{ mot } H_a : \tilde{\mu}_D > 0.$$

- La nå $p = \mathrm{P}(X_i > Y_i) = \mathrm{P}(D_i > 0)$.

# Fortegntesten for parvise data

- Fortegntesten kan også brukes på parvise data.
- Anta at en har de parvise målingene $(X_1, Y_1)$, ...,$(X_n, Y_n)$, og la $D_i = X_i - Y_i$, $i = 1, \ldots, n$.
- Anta at vi vil teste om $X_i$ typisk er større enn eller lik $Y_i$, slik at $D_i$ typisk er $> 0$.
- La $\tilde{\mu}_D$ være medianen til $D_i$
- Da kan vi teste

$$H_0 : \tilde{\mu}_D \leq 0 \text{ mot } H_a : \tilde{\mu}_D > 0.$$

- La nå $p = \mathrm{P}(X_i > Y_i) = \mathrm{P}(D_i > 0)$. Da kan vi omformulere hypotesene til

# Fortegntesten for parvise data

- Fortegntesten kan også brukes på parvise data.
- Anta at en har de parvise målingene $(X_1, Y_1)$, ...,$(X_n, Y_n)$, og la $D_i = X_i - Y_i$, $i = 1, \ldots, n$.
- Anta at vi vil teste om $X_i$ typisk er større enn eller lik $Y_i$, slik at $D_i$ typisk er $> 0$.
- La $\tilde{\mu}_D$ være medianen til $D_i$
- Da kan vi teste

$$H_0 : \tilde{\mu}_D \leq 0 \text{ mot } H_a : \tilde{\mu}_D > 0.$$

- La nå $p = \mathrm{P}(X_i > Y_i) = \mathrm{P}(D_i > 0)$. Da kan vi omformulere hypotesene til

$$H_0 : p \leq 0.5 \text{ mot} \quad H_a : p > 0.5,$$

# Fortegntesten for parvise data

- Fortegntesten kan også brukes på parvise data.
- Anta at en har de parvise målingene $(X_1, Y_1)$, ...,$(X_n, Y_n)$, og la $D_i = X_i - Y_i$, $i = 1, \ldots, n$.
- Anta at vi vil teste om $X_i$ typisk er større enn eller lik $Y_i$, slik at $D_i$ typisk er $> 0$.
- La $\tilde{\mu}_D$ være medianen til $D_i$
- Da kan vi teste

$$H_0 : \tilde{\mu}_D \leq 0 \text{ mot } H_a : \tilde{\mu}_D > 0.$$

- La nå $p = \mathrm{P}(X_i > Y_i) = \mathrm{P}(D_i > 0)$. Da kan vi omformulere hypotesene til

$$H_0 : p \leq 0.5 \text{ mot } \quad H_a : p > 0.5,$$

og bruke tegn-testen for ett utvalg.

# Fortegntesten for parvise data

- Fortegntesten kan også brukes på parvise data.
- Anta at en har de parvise målingene $(X_1, Y_1)$, ...,$(X_n, Y_n)$, og la $D_i = X_i - Y_i$, $i = 1, \ldots, n$.
- Anta at vi vil teste om $X_i$ typisk er større enn eller lik $Y_i$, slik at $D_i$ typisk er $> 0$.
- La $\tilde{\mu}_D$ være medianen til $D_i$
- Da kan vi teste

$$H_0 : \tilde{\mu}_D \leq 0 \text{ mot } H_a : \tilde{\mu}_D > 0.$$

- La nå $p = \mathrm{P}(X_i > Y_i) = \mathrm{P}(D_i > 0)$. Da kan vi omformulere hypotesene til

$$H_0 : p \leq 0.5 \text{ mot } \quad H_a : p > 0.5,$$

og bruke tegn-testen for ett utvalg.

## Eksempel

Eks. 14.3 i boka.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**, kan vi lage en mer effektiv hypotesetest for medianen.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**, kan vi lage en mer effektiv hypotesetest for medianen.
- Når fordelingen er symmetrisk er forventningsverdien $\mu$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**, kan vi lage en mer effektiv hypotesetest for medianen.
- Når fordelingen er symmetrisk er forventningsverdien $\mu$ (så sant den eksisterer)

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**, kan vi lage en mer effektiv hypotesetest for medianen.
- Når fordelingen er symmetrisk er forventningsverdien $\mu$ (så sant den eksisterer) lik medianen,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**, kan vi lage en mer effektiv hypotesetest for medianen.
- Når fordelingen er symmetrisk er forventningsverdien $\mu$ (så sant den eksisterer) lik medianen, slik at dette også blir en test for forventningsverdien.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**, kan vi lage en mer effektiv hypotesetest for medianen.
- Når fordelingen er symmetrisk er forventningsverdien $\mu$ (så sant den eksisterer) lik medianen, slik at dette også blir en test for forventningsverdien.
- Anta at vi $x_1, \ldots, x_n$ er observasjoner

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**, kan vi lage en mer effektiv hypotesetest for medianen.
- Når fordelingen er symmetrisk er forventningsverdien $\mu$ (så sant den eksisterer) lik medianen, slik at dette også blir en test for forventningsverdien.
- Anta at vi $x_1, \ldots, x_n$ er observasjoner av de stokastiske variablene $X_1, \ldots, X_n$,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**, kan vi lage en mer effektiv hypotesetest for medianen.
- Når fordelingen er symmetrisk er forventningsverdien $\mu$ (så sant den eksisterer) lik medianen, slik at dette også blir en test for forventningsverdien.
- Anta at vi $x_1, \ldots, x_n$ er observasjoner av de stokastiske variablene $X_1, \ldots, X_n$, som er uif fra en kontinuerlig og symmetrisk fordeling,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**, kan vi lage en mer effektiv hypotesetest for medianen.
- Når fordelingen er symmetrisk er forventningsverdien $\mu$ (så sant den eksisterer) lik medianen, slik at dette også blir en test for forventningsverdien.
- Anta at vi $x_1, \ldots, x_n$ er observasjoner av de stokastiske variablene $X_1, \ldots, X_n$, som er uif fra en kontinuerlig og symmetrisk fordeling, og vi vil teste hypoteser av typen

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**, kan vi lage en mer effektiv hypotesetest for medianen.
- Når fordelingen er symmetrisk er forventningsverdien $\mu$ (så sant den eksisterer) lik medianen, slik at dette også blir en test for forventningsverdien.
- Anta at vi $x_1, \ldots, x_n$ er observasjoner av de stokastiske variablene $X_1, \ldots, X_n$, som er uif fra en kontinuerlig og symmetrisk fordeling, og vi vil teste hypoteser av typen
  1. $H_0 : \tilde{\mu} \leq \tilde{\mu}_0$ mot $H_a : \tilde{\mu} > \tilde{\mu}_0$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**, kan vi lage en mer effektiv hypotesetest for medianen.
- Når fordelingen er symmetrisk er forventningsverdien $\mu$ (så sant den eksisterer) lik medianen, slik at dette også blir en test for forventningsverdien.
- Anta at vi $x_1, \ldots, x_n$ er observasjoner av de stokastiske variablene $X_1, \ldots, X_n$, som er uif fra en kontinuerlig og symmetrisk fordeling, og vi vil teste hypoteser av typen
  1. $H_0 : \tilde{\mu} \leq \tilde{\mu}_0$ mot $H_a : \tilde{\mu} > \tilde{\mu}_0$
  2. $H_0 : \tilde{\mu} \geq \tilde{\mu}_0$ mot $H_a : \tilde{\mu} < \tilde{\mu}_0$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Ved å gjøre en ekstra antakelse om fordelingen til dataene, nemlig at den er **symmetrisk**, kan vi lage en mer effektiv hypotesetest for medianen.
- Når fordelingen er symmetrisk er forventningsverdien $\mu$ (så sant den eksisterer) lik medianen, slik at dette også blir en test for forventningsverdien.
- Anta at vi $x_1, \ldots, x_n$ er observasjoner av de stokastiske variablene $X_1, \ldots, X_n$, som er uif fra en kontinuerlig og symmetrisk fordeling, og vi vil teste hypoteser av typen
  1. $H_0 : \tilde{\mu} \leq \tilde{\mu}_0$ mot $H_a : \tilde{\mu} > \tilde{\mu}_0$
  2. $H_0 : \tilde{\mu} \geq \tilde{\mu}_0$ mot $H_a : \tilde{\mu} < \tilde{\mu}_0$
  3. $H_0 : \tilde{\mu} = \tilde{\mu}_0$ mot $H_a : \tilde{\mu} \neq \tilde{\mu}_0$.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- La $R_1, \ldots, R_n$ være rangene til $|X_1 - \tilde{\mu}_0|$, ...,$|X_n - \tilde{\mu}_0|$.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- La $R_1, \ldots, R_n$ være rangene til $|X_1 - \tilde{\mu}_0|$, ...,$|X_n - \tilde{\mu}_0|$.
- Testobservatoren vi skal bruke er

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- La $R_1, \ldots, R_n$ være rangene til $|X_1 - \tilde{\mu}_0|$, ...,$|X_n - \tilde{\mu}_0|$.
- Testobservatoren vi skal bruke er

$$S_+ = \sum_{i=1}^{n} I(X_i - \tilde{\mu}_0 > 0) \cdot R_i,$$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- La $R_1, \ldots, R_n$ være rangene til $|X_1 - \tilde{\mu}_0|$, ...,$|X_n - \tilde{\mu}_0|$.
- Testobservatoren vi skal bruke er

$$S_+ = \sum_{i=1}^{n} I(X_i - \tilde{\mu}_0 > 0) \cdot R_i,$$

altså summen av alle rangene som er assossiert med et positivt fortegn.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- La $R_1, \ldots, R_n$ være rangene til $|X_1 - \tilde{\mu}_0|$, ...,$|X_n - \tilde{\mu}_0|$.
- Testobservatoren vi skal bruke er

$$S_+ = \sum_{i=1}^{n} I(X_i - \tilde{\mu}_0 > 0) \cdot R_i,$$

  altså summen av alle rangene som er assossiert med et positivt fortegn.
- Anta først at vi vil teste

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- La $R_1, \ldots, R_n$ være rangene til $|X_1 - \tilde{\mu}_0|$, ...,$|X_n - \tilde{\mu}_0|$.
- Testobservatoren vi skal bruke er

$$S_+ = \sum_{i=1}^{n} I(X_i - \tilde{\mu}_0 > 0) \cdot R_i,$$

  altså summen av alle rangene som er assossiert med et positivt fortegn.
- Anta først at vi vil teste

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

- Dersom $H_0$ er sann, vil en forvente at mange av $(X_i - \tilde{\mu}_0)$-ene er negative,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- La $R_1, \ldots, R_n$ være rangene til $|X_1 - \tilde{\mu}_0|$, ...,$|X_n - \tilde{\mu}_0|$.
- Testobservatoren vi skal bruke er

$$S_+ = \sum_{i=1}^{n} I(X_i - \tilde{\mu}_0 > 0) \cdot R_i,$$

  altså summen av alle rangene som er assossiert med et positivt fortegn.
- Anta først at vi vil teste

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

- Dersom $H_0$ er sann, vil en forvente at mange av $(X_i - \tilde{\mu}_0)$-ene er negative, slik at mange av rangene $R_i$ er assossiert med et negativt fortegn,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- La $R_1, \ldots, R_n$ være rangene til $|X_1 - \tilde{\mu}_0|$, ...,$|X_n - \tilde{\mu}_0|$.
- Testobservatoren vi skal bruke er

$$S_+ = \sum_{i=1}^{n} I(X_i - \tilde{\mu}_0 > 0) \cdot R_i,$$

  altså summen av alle rangene som er assossiert med et positivt fortegn.
- Anta først at vi vil teste

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

- Dersom $H_0$ er sann, vil en forvente at mange av $(X_i - \tilde{\mu}_0)$-ene er negative, slik at mange av rangene $R_i$ er assossiert med et negativt fortegn, og at $S_+$ er liten.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- La $R_1, \ldots, R_n$ være rangene til $|X_1 - \tilde{\mu}_0|$, ...,$|X_n - \tilde{\mu}_0|$.
- Testobservatoren vi skal bruke er

$$S_+ = \sum_{i=1}^{n} I(X_i - \tilde{\mu}_0 > 0) \cdot R_i,$$

  altså summen av alle rangene som er assossiert med et positivt fortegn.
- Anta først at vi vil teste

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

- Dersom $H_0$ er sann, vil en forvente at mange av $(X_i - \tilde{\mu}_0)$-ene er negative, slik at mange av rangene $R_i$ er assossiert med et negativt fortegn, og at $S_+$ er liten.
- Dersom derimot $H_a$ er sann,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- La $R_1, \ldots, R_n$ være rangene til $|X_1 - \tilde{\mu}_0|$, ...,$|X_n - \tilde{\mu}_0|$.
- Testobservatoren vi skal bruke er

$$S_+ = \sum_{i=1}^{n} I(X_i - \tilde{\mu}_0 > 0) \cdot R_i,$$

  altså summen av alle rangene som er assossiert med et positivt fortegn.
- Anta først at vi vil teste

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

- Dersom $H_0$ er sann, vil en forvente at mange av $(X_i - \tilde{\mu}_0)$-ene er negative, slik at mange av rangene $R_i$ er assossiert med et negativt fortegn, og at $S_+$ er liten.
- Dersom derimot $H_a$ er sann, vil en forvente at mange av $(X_i - \tilde{\mu}_0)$-ene er positive,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- La $R_1, \ldots, R_n$ være rangene til $|X_1 - \tilde{\mu}_0|$, ...,$|X_n - \tilde{\mu}_0|$.
- Testobservatoren vi skal bruke er

$$S_+ = \sum_{i=1}^{n} I(X_i - \tilde{\mu}_0 > 0) \cdot R_i,$$

  altså summen av alle rangene som er assossiert med et positivt fortegn.
- Anta først at vi vil teste

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

- Dersom $H_0$ er sann, vil en forvente at mange av $(X_i - \tilde{\mu}_0)$-ene er negative, slik at mange av rangene $R_i$ er assossiert med et negativt fortegn, og at $S_+$ er liten.
- Dersom derimot $H_a$ er sann, vil en forvente at mange av $(X_i - \tilde{\mu}_0)$-ene er positive, slik at mange av rangene $R_i$ er assossiert med et positivt fortegn,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- La $R_1, \ldots, R_n$ være rangene til $|X_1 - \tilde{\mu}_0|$, ...,$|X_n - \tilde{\mu}_0|$.
- Testobservatoren vi skal bruke er

$$S_+ = \sum_{i=1}^{n} I(X_i - \tilde{\mu}_0 > 0) \cdot R_i,$$

  altså summen av alle rangene som er assossiert med et positivt fortegn.
- Anta først at vi vil teste

$$H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} > \tilde{\mu}_0.$$

- Dersom $H_0$ er sann, vil en forvente at mange av $(X_i - \tilde{\mu}_0)$-ene er negative, slik at mange av rangene $R_i$ er assossiert med et negativt fortegn, og at $S_+$ er liten.
- Dersom derimot $H_a$ er sann, vil en forvente at mange av $(X_i - \tilde{\mu}_0)$-ene er positive, slik at mange av rangene $R_i$ er assossiert med et positivt fortegn, og at $S_+$ er stor.

# “Wilcoxon signed rank test” for medianen $\tilde{\mu}$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Vi vil derfor forkaste $H_0$ ved signifikansnivå $\alpha$ for store verdier av $S_+$,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Vi vil derfor forkaste $H_0$ ved signifikansnivå $\alpha$ for store verdier av $S_+$, dvs. dersom $S_+ \geq c_1$,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Vi vil derfor forkaste $H_0$ ved signifikansnivå $\alpha$ for store verdier av $S_+$, dvs. dersom $S_+ \geq c_1$, der

$$\mathrm{P}(S_+ \geq c_1 | H_0 \text{ er sann }) \approx \alpha.$$

- Man kan regne på fordelingen til $S_+$ når $\tilde{\mu} = \tilde{\mu}_0$ med kombinatoriske argumenter

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Vi vil derfor forkaste $H_0$ ved signifikansnivå $\alpha$ for store verdier av $S_+$, dvs. dersom $S_+ \geq c_1$, der

$$\mathrm{P}(S_+ \geq c_1 | H_0 \text{ er sann }) \approx \alpha.$$

- Man kan regne på fordelingen til $S_+$ når $\tilde{\mu} = \tilde{\mu}_0$ med kombinatoriske argumenter (detaljene utelates)

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Vi vil derfor forkaste $H_0$ ved signifikansnivå $\alpha$ for store verdier av $S_+$, dvs. dersom $S_+ \geq c_1$, der

$$\mathrm{P}(S_+ \geq c_1 | H_0 \text{ er sann }) \approx \alpha.$$

- Man kan regne på fordelingen til $S_+$ når $\tilde{\mu} = \tilde{\mu}_0$ med kombinatoriske argumenter (detaljene utelates)
- Tilsvarende vil testen

$$H_0 : \tilde{\mu} \geq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} < \tilde{\mu}_0$$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Vi vil derfor forkaste $H_0$ ved signifikansnivå $\alpha$ for store verdier av $S_+$, dvs. dersom $S_+ \geq c_1$, der

$$\mathrm{P}(S_+ \geq c_1 | H_0 \text{ er sann }) \approx \alpha.$$

- Man kan regne på fordelingen til $S_+$ når $\tilde{\mu} = \tilde{\mu}_0$ med kombinatoriske argumenter (detaljene utelates)
- Tilsvarende vil testen

$$H_0 : \tilde{\mu} \geq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} < \tilde{\mu}_0$$

få signifikansnivå $\alpha$ dersom vi forkaster $H_0$ hvis $S_+ \leq c_2$,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Vi vil derfor forkaste $H_0$ ved signifikansnivå $\alpha$ for store verdier av $S_+$, dvs. dersom $S_+ \geq c_1$, der

$$\mathrm{P}(S_+ \geq c_1 | H_0 \text{ er sann }) \approx \alpha.$$

- Man kan regne på fordelingen til $S_+$ når $\tilde{\mu} = \tilde{\mu}_0$ med kombinatoriske argumenter (detaljene utelates)
- Tilsvarende vil testen

$$H_0 : \tilde{\mu} \geq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} < \tilde{\mu}_0$$

få signifikansnivå $\alpha$ dersom vi forkaster $H_0$ hvis $S_+ \leq c_2$, med

$$\mathrm{P}(S_+ \leq c_2 | H_0 \text{ er sann }) \approx \alpha.$$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Vi vil derfor forkaste $H_0$ ved signifikansnivå $\alpha$ for store verdier av $S_+$, dvs. dersom $S_+ \geq c_1$, der

$$\mathrm{P}(S_+ \geq c_1 | H_0 \text{ er sann }) \approx \alpha.$$

- Man kan regne på fordelingen til $S_+$ når $\tilde{\mu} = \tilde{\mu}_0$ med kombinatoriske argumenter (detaljene utelates)
- Tilsvarende vil testen

$$H_0 : \tilde{\mu} \geq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} < \tilde{\mu}_0$$

få signifikansnivå $\alpha$ dersom vi forkaster $H_0$ hvis $S_+ \leq c_2$, med

$$\mathrm{P}(S_+ \leq c_2 | H_0 \text{ er sann }) \approx \alpha.$$

- Da fordelingen til $X_i$ er symmetrisk om $\tilde{\mu}$,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Vi vil derfor forkaste $H_0$ ved signifikansnivå $\alpha$ for store verdier av $S_+$, dvs. dersom $S_+ \geq c_1$, der

$$\mathrm{P}(S_+ \geq c_1 | H_0 \text{ er sann }) \approx \alpha.$$

- Man kan regne på fordelingen til $S_+$ når $\tilde{\mu} = \tilde{\mu}_0$ med kombinatoriske argumenter (detaljene utelates)
- Tilsvarende vil testen

$$H_0 : \tilde{\mu} \geq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} < \tilde{\mu}_0$$

få signifikansnivå $\alpha$ dersom vi forkaster $H_0$ hvis $S_+ \leq c_2$, med

$$\mathrm{P}(S_+ \leq c_2 | H_0 \text{ er sann }) \approx \alpha.$$

- Da fordelingen til $X_i$ er symmetrisk om $\tilde{\mu}$, vil fordelingen til $S_+$, som går over verdiene $1,\ldots,n(n+1)/2$, være symmetrisk om midtpunktet når $\tilde{\mu} = \tilde{\mu}_0$,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Vi vil derfor forkaste $H_0$ ved signifikansnivå $\alpha$ for store verdier av $S_+$, dvs. dersom $S_+ \geq c_1$, der

$$\mathrm{P}(S_+ \geq c_1 | H_0 \text{ er sann }) \approx \alpha.$$

- Man kan regne på fordelingen til $S_+$ når $\tilde{\mu} = \tilde{\mu}_0$ med kombinatoriske argumenter (detaljene utelates)
- Tilsvarende vil testen

$$H_0 : \tilde{\mu} \geq \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} < \tilde{\mu}_0$$

få signifikansnivå $\alpha$ dersom vi forkaster $H_0$ hvis $S_+ \leq c_2$, med

$$\mathrm{P}(S_+ \leq c_2 | H_0 \text{ er sann }) \approx \alpha.$$

- Da fordelingen til $X_i$ er symmetrisk om $\tilde{\mu}$, vil fordelingen til $S_+$, som går over verdiene $1,\ldots,n(n+1)/2$, være symmetrisk om midtpunktet når $\tilde{\mu} = \tilde{\mu}_0$, slik at $c_2 = n(n+1)/2 - c_1$.

# “Wilcoxon signed rank test” for medianen $\tilde{\mu}$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- For tester av typen

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} \neq \tilde{\mu}_0$$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- For tester av typen

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} \neq \tilde{\mu}_0$$

vil vi forkaste $H_0$ ved signifikansnivå $\alpha$ dersom $S_+ \leq n(n+1)/2 - c$ **eller** $S_+ \geq c$,

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- For tester av typen

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} \neq \tilde{\mu}_0$$

vil vi forkaste $H_0$ ved signifikansnivå $\alpha$ dersom $S_+ \leq n(n+1)/2 - c$ **eller** $S_+ \geq c$, med

$$\mathrm{P}(S_+ \geq c | H_0 \text{ er sann }) \approx \alpha/2.$$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- For tester av typen

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} \neq \tilde{\mu}_0$$

vil vi forkaste $H_0$ ved signifikansnivå $\alpha$ dersom $S_+ \leq n(n+1)/2 - c$ **eller** $S_+ \geq c$, med

$$\mathrm{P}(S_+ \geq c | H_0 \text{ er sann }) \approx \alpha/2.$$

- Denne testen kalles "Wilcoxon signed rank test".

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- For tester av typen

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} \neq \tilde{\mu}_0$$

vil vi forkaste $H_0$ ved signifikansnivå $\alpha$ dersom $S_+ \leq n(n+1)/2 - c$ **eller** $S_+ \geq c$, med

$$\mathrm{P}(S_+ \geq c | H_0 \text{ er sann }) \approx \alpha/2.$$

- Denne testen kalles "Wilcoxon signed rank test".
- Den kan utføres i R med funksjonen `wilcox.test()`.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- For tester av typen

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} \neq \tilde{\mu}_0$$

vil vi forkaste $H_0$ ved signifikansnivå $\alpha$ dersom $S_+ \leq n(n+1)/2 - c$ **eller** $S_+ \geq c$, med

$$\mathrm{P}(S_+ \geq c | H_0 \text{ er sann }) \approx \alpha/2.$$

- Denne testen kalles "Wilcoxon signed rank test".
- Den kan utføres i R med funksjonen `wilcox.test()`.
- Når forventningsverdien $\mu$ eksisterer, er det også en test vedrørende denne, men testen er gyldig for medianen også når $\mu$ ikke eksisterer.

# “Wilcoxon signed rank test” for medianen $\tilde{\mu}$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- For store utvalg ($n > 20$), kan det vises at når $\tilde{\mu} = \tilde{\mu}_0$, er

$$S_+ \overset{tiln.}{\sim} N\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right),$$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- For store utvalg ($n > 20$), kan det vises at når $\tilde{\mu} = \tilde{\mu}_0$, er

$$S_+ \overset{\text{tiln.}}{\sim} N\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right),$$

slik at

$$Z = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \overset{\text{tiln.}}{\sim} N(0,1).$$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- For store utvalg ($n > 20$), kan det vises at når $\tilde{\mu} = \tilde{\mu}_0$, er

$$S_+ \overset{tiln.}{\sim} N\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right),$$

slik at

$$Z = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \overset{tiln.}{\sim} N(0,1).$$

- Da kan en bruke

$$c_1 = \frac{n(n+1)}{4} + z_\alpha \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- For store utvalg ($n > 20$), kan det vises at når $\tilde{\mu} = \tilde{\mu}_0$, er

$$S_+ \overset{tiln.}{\sim} N\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right),$$

slik at

$$Z = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \overset{tiln.}{\sim} N(0,1).$$

- Da kan en bruke

$$\begin{aligned} c_1 &= \frac{n(n+1)}{4} + z_\alpha \sqrt{\frac{n(n+1)(2n+1)}{24}} \\ c_2 &= \frac{n(n+1)}{4} - z_\alpha \sqrt{\frac{n(n+1)(2n+1)}{24}} \end{aligned}$$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- For store utvalg ($n > 20$), kan det vises at når $\tilde{\mu} = \tilde{\mu}_0$, er

$$S_+ \overset{tiln.}{\sim} N\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right),$$

slik at

$$Z = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \overset{tiln.}{\sim} N(0,1).$$

- Da kan en bruke

$$\begin{aligned} c_1 &= \frac{n(n+1)}{4} + z_\alpha \sqrt{\frac{n(n+1)(2n+1)}{24}} \\ c_2 &= \frac{n(n+1)}{4} - z_\alpha \sqrt{\frac{n(n+1)(2n+1)}{24}} \\ c &= \frac{n(n+1)}{4} + z_{\alpha/2} \sqrt{\frac{n(n+1)(2n+1)}{24}} \end{aligned}$$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- For store utvalg ($n > 20$), kan det vises at når $\tilde{\mu} = \tilde{\mu}_0$, er

$$S_+ \overset{\text{tiln.}}{\sim} N\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right),$$

slik at

$$Z = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \overset{\text{tiln.}}{\sim} N(0,1).$$

- Da kan en bruke

$$\begin{aligned} c_1 &= \frac{n(n+1)}{4} + z_\alpha \sqrt{\frac{n(n+1)(2n+1)}{24}} \\ c_2 &= \frac{n(n+1)}{4} - z_\alpha \sqrt{\frac{n(n+1)(2n+1)}{24}} \\ c &= \frac{n(n+1)}{4} + z_{\alpha/2} \sqrt{\frac{n(n+1)(2n+1)}{24}} \end{aligned}$$

i testene over.

# “Wilcoxon signed rank test” for medianen $\tilde{\mu}$

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Denne testen brukes typisk på parvise data $(X_1, Y_1)$, ...,$(X_n, Y_n)$.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Denne testen brukes typisk på parvise data $(X_1, Y_1)$, ...,$(X_n, Y_n)$.
- En lar da $D_i = X_i - Y_i$, $i = 1, \ldots, n$, og tester hypoteser knyttet til $\tilde{\mu}_D$ (eventuelt $\mu_D = \mathrm{E}(D_i)$) ved å bruke testen over på differansene.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Denne testen brukes typisk på parvise data $(X_1, Y_1)$, ...,$(X_n, Y_n)$.
- En lar da $D_i = X_i - Y_i$, $i = 1, \ldots, n$, og tester hypoteser knyttet til $\tilde{\mu}_D$ (eventuelt $\mu_D = \mathrm{E}(D_i)$) ved å bruke testen over på differansene.
- Selv om fordelingen til dataene er kontinuerlig, kan det skje at flere av observasjonene har nøyaktig samme verdi ("ties"),

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Denne testen brukes typisk på parvise data $(X_1, Y_1)$, ...,$(X_n, Y_n)$.
- En lar da $D_i = X_i - Y_i$, $i = 1, \ldots, n$, og tester hypoteser knyttet til $\tilde{\mu}_D$ (eventuelt $\mu_D = \mathrm{E}(D_i)$) ved å bruke testen over på differansene.
- Selv om fordelingen til dataene er kontinuerlig, kan det skje at flere av observasjonene har nøyaktig samme verdi ("ties"), f.eks. p.g.a. begrenset presisjon i måleapparatet.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Denne testen brukes typisk på parvise data $(X_1, Y_1)$, ...,$(X_n, Y_n)$.
- En lar da $D_i = X_i - Y_i$, $i = 1, \ldots, n$, og tester hypoteser knyttet til $\tilde{\mu}_D$ (eventuelt $\mu_D = \mathrm{E}(D_i)$) ved å bruke testen over på differansene.
- Selv om fordelingen til dataene er kontinuerlig, kan det skje at flere av observasjonene har nøyaktig samme verdi ("ties"), f.eks. p.g.a. begrenset presisjon i måleapparatet.
- Da lar en alle observasjoner med samme verdi få den gjennomsnittlige rangen til disse observasjoner.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Denne testen brukes typisk på parvise data $(X_1, Y_1)$, ...,$(X_n, Y_n)$.
- En lar da $D_i = X_i - Y_i$, $i = 1, \ldots, n$, og tester hypoteser knyttet til $\tilde{\mu}_D$ (eventuelt $\mu_D = \mathrm{E}(D_i)$) ved å bruke testen over på differansene.
- Selv om fordelingen til dataene er kontinuerlig, kan det skje at flere av observasjonene har nøyaktig samme verdi ("ties"), f.eks. p.g.a. begrenset presisjon i måleapparatet.
- Da lar en alle observasjoner med samme verdi få den gjennomsnittlige rangen til disse observasjoner.
- F.eks. hvis observasjonene med rang 2 og 3 har samme verdi, gis de begge rang 2.5.

# "Wilcoxon signed rank test" for medianen $\tilde{\mu}$

- Denne testen brukes typisk på parvise data $(X_1, Y_1)$, ...,$(X_n, Y_n)$.
- En lar da $D_i = X_i - Y_i$, $i = 1, \ldots, n$, og tester hypoteser knyttet til $\tilde{\mu}_D$ (eventuelt $\mu_D = \mathrm{E}(D_i)$) ved å bruke testen over på differansene.
- Selv om fordelingen til dataene er kontinuerlig, kan det skje at flere av observasjonene har nøyaktig samme verdi ("ties"), f.eks. p.g.a. begrenset presisjon i måleapparatet.
- Da lar en alle observasjoner med samme verdi få den gjennomsnittlige rangen til disse observasjoner.
- F.eks. hvis observasjonene med rang 2 og 3 har samme verdi, gis de begge rang 2.5.

## Eksempel

Eks. 14.4

# Hvilken test bør en velge?

# Hvilken test bør en velge?

- Hvilken test bør en velge mellom t-test, fortegntest eller "Wilcoxon signed rank test"?

# Hvilken test bør en velge?

- Hvilken test bør en velge mellom t-test, fortegntest eller "Wilcoxon signed rank test"?
- Dersom fordelingen til dataene er asymmetrisk, bør en bruke fortegntesten.

# Hvilken test bør en velge?

- Hvilken test bør en velge mellom t-test, fortegntest eller "Wilcoxon signed rank test"?
- Dersom fordelingen til dataene er asymmetrisk, bør en bruke fortegntesten.
- Når fordelingen er tilnærmet normal, er t-testen den beste, da den er noe mer asymptotisk effisient enn "signed rank"-testen og mye mer effisient enn fortegntesten.

# Hvilken test bør en velge?

- Hvilken test bør en velge mellom t-test, fortegntest eller "Wilcoxon signed rank test"?
- Dersom fordelingen til dataene er asymmetrisk, bør en bruke fortegntesten.
- Når fordelingen er tilnærmet normal, er t-testen den beste, da den er noe mer asymptotisk effisient enn "signed rank"-testen og mye mer effisient enn fortegntesten.
- Dersom fordelingen er symmetrisk, men nokså forskjellig fra normalfordelingen, vil "signed rank"-testen være best.

# Konfidensintervall for medianen $\tilde{\mu}$

# Konfidensintervall for medianen $\tilde{\mu}$

- En kan også bruke "Wilcoxon signed rank test" til å lage konfidensintervall for medianen $\tilde{\mu}$ (forventningsverdien $\mu$).

# Konfidensintervall for medianen $\tilde{\mu}$

- En kan også bruke "Wilcoxon signed rank test" til å lage konfidensintervall for medianen $\tilde{\mu}$ (forventningsverdien $\mu$).
- Da kan vi bruke at dersom vi tester

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} \neq \tilde{\mu}_0,$$

# Konfidensintervall for medianen $\tilde{\mu}$

- En kan også bruke "Wilcoxon signed rank test" til å lage konfidensintervall for medianen $\tilde{\mu}$ (forventningsverdien $\mu$).
- Da kan vi bruke at dersom vi tester

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} \neq \tilde{\mu}_0,$$

og $A$ er settet av alle verdier av $\tilde{\mu}_0$ som en **ikke** forkaster $H_0$ for ved signifikansnivå $\alpha$,

# Konfidensintervall for medianen $\tilde{\mu}$

- En kan også bruke "Wilcoxon signed rank test" til å lage konfidensintervall for medianen $\tilde{\mu}$ (forventningsverdien $\mu$).
- Da kan vi bruke at dersom vi tester

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} \neq \tilde{\mu}_0,$$

og $A$ er settet av alle verdier av $\tilde{\mu}_0$ som en **ikke** forkaster $H_0$ for ved signifikansnivå $\alpha$, så er $A$ et $(1-\alpha)\cdot 100\%$ konfidensintervall for $\tilde{\mu}$.

# Konfidensintervall for medianen $\tilde{\mu}$

- En kan også bruke "Wilcoxon signed rank test" til å lage konfidensintervall for medianen $\tilde{\mu}$ (forventningsverdien $\mu$).
- Da kan vi bruke at dersom vi tester

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} \neq \tilde{\mu}_0,$$

og $A$ er settet av alle verdier av $\tilde{\mu}_0$ som en **ikke** forkaster $H_0$ for ved signifikansnivå $\alpha$, så er $A$ et $(1-\alpha)\cdot 100\%$ konfidensintervall for $\tilde{\mu}$.

- Når vi skal lage konfidensintervall basert på "Wilcoxon signed rank test", er det enklere å spesifisere testobservatoren $S_+$ som følger:

# Konfidensintervall for medianen $\tilde{\mu}$

- En kan også bruke "Wilcoxon signed rank test" til å lage konfidensintervall for medianen $\tilde{\mu}$ (forventningsverdien $\mu$).
- Da kan vi bruke at dersom vi tester

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} \neq \tilde{\mu}_0,$$

og $A$ er settet av alle verdier av $\tilde{\mu}_0$ som en **ikke** forkaster $H_0$ for ved signifikansnivå $\alpha$, så er $A$ et $(1-\alpha)\cdot 100\%$ konfidensintervall for $\tilde{\mu}$.

- Når vi skal lage konfidensintervall basert på "Wilcoxon signed rank test", er det enklere å spesifisere testobservatoren $S_+$ som følger:

$$S_+ = \sum_{j=1}^{n} \sum_{i \leq j} I\left(\frac{X_i + X_j}{2} \geq \tilde{\mu}_0\right),$$

# Konfidensintervall for medianen $\tilde{\mu}$

- En kan også bruke "Wilcoxon signed rank test" til å lage konfidensintervall for medianen $\tilde{\mu}$ (forventningsverdien $\mu$).
- Da kan vi bruke at dersom vi tester

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 \text{ mot } H_a : \tilde{\mu} \neq \tilde{\mu}_0,$$

og $A$ er settet av alle verdier av $\tilde{\mu}_0$ som en **ikke** forkaster $H_0$ for ved signifikansnivå $\alpha$, så er $A$ et $(1-\alpha)\cdot 100\%$ konfidensintervall for $\tilde{\mu}$.

- Når vi skal lage konfidensintervall basert på "Wilcoxon signed rank test", er det enklere å spesifisere testobservatoren $S_+$ som følger:

$$S_+ = \sum_{j=1}^{n} \sum_{i \leq j} I\left(\frac{X_i + X_j}{2} \geq \tilde{\mu}_0\right),$$

der $\frac{X_i+X_j}{2}$ kalles Walsh-gjennomsnitt.

# Konfidensintervall for medianen $\tilde{\mu}$

# Konfidensintervall for medianen $\tilde{\mu}$

- La $\bar{x}_{(1)},\ldots,\bar{x}_{(n(n+1)/2)}$ være alle observerte Walsh-gjennomsnitt,

# Konfidensintervall for medianen $\tilde{\mu}$

- La $\bar{x}_{(1)},\ldots,\bar{x}_{(n(n+1)/2)}$ være alle observerte Walsh-gjennomsnitt, basert på observasjonene $x_1,\ldots,x_n$ av $X_1,\ldots,X_n$,

# Konfidensintervall for medianen $\tilde{\mu}$

- La $\bar{x}_{(1)},\ldots,\bar{x}_{(n(n+1)/2)}$ være alle observerte Walsh-gjennomsnitt, basert på observasjonene $x_1,\ldots,x_n$ av $X_1,\ldots,X_n$, ordnet i økende rekkefølge,

# Konfidensintervall for medianen $\tilde{\mu}$

- La $\bar{x}_{(1)},\ldots,\bar{x}_{(n(n+1)/2)}$ være alle observerte Walsh-gjennomsnitt, basert på observasjonene $x_1,\ldots,x_n$ av $X_1,\ldots,X_n$, ordnet i økende rekkefølge, altså

$$\bar{x}_{(1)} \leq \cdots \leq \bar{x}_{(n(n+1)/2)}.$$

# Konfidensintervall for medianen $\tilde{\mu}$

- La $\bar{x}_{(1)},\ldots,\bar{x}_{(n(n+1)/2)}$ være alle observerte Walsh-gjennomsnitt, basert på observasjonene $x_1,\ldots,x_n$ av $X_1,\ldots,X_n$, ordnet i økende rekkefølge, altså

$$\bar{x}_{(1)} \leq \ldots \leq \bar{x}_{(n(n+1)/2)}.$$

- Da er et $(1-\alpha)\cdot 100\%$ konfidensintervall for $\tilde{\mu}$ gitt ved

$$\left(\bar{x}_{(n(n+1)/2)-c+1}, \bar{x}_{(c)}\right),$$

# Konfidensintervall for medianen $\tilde{\mu}$

- La $\bar{x}_{(1)},\ldots,\bar{x}_{(n(n+1)/2)}$ være alle observerte Walsh-gjennomsnitt, basert på observasjonene $x_1,\ldots,x_n$ av $X_1,\ldots,X_n$, ordnet i økende rekkefølge, altså

$$\bar{x}_{(1)} \leq \ldots \leq \bar{x}_{(n(n+1)/2)}.$$

- Da er et $(1-\alpha)\cdot 100\%$ konfidensintervall for $\tilde{\mu}$ gitt ved

$$\left(\bar{x}_{(n(n+1)/2)-c+1}, \bar{x}_{(c)}\right),$$

der $c$ er slik at $H_0 : \tilde{\mu} = \tilde{\mu}_0$ forkastes til fordel for $H_a : \tilde{\mu} \neq \tilde{\mu}_0$ ved signifikansnivå $\alpha$ dersom $S_+ \geq c$ eller $S_+ \leq n(n+1)/2 - c$.

# Konfidensintervall for medianen $\tilde{\mu}$

- La $\bar{x}_{(1)},\ldots,\bar{x}_{(n(n+1)/2)}$ være alle observerte Walsh-gjennomsnitt, basert på observasjonene $x_1,\ldots,x_n$ av $X_1,\ldots,X_n$, ordnet i økende rekkefølge, altså

$$\bar{x}_{(1)} \leq \ldots \leq \bar{x}_{(n(n+1)/2)}.$$

- Da er et $(1-\alpha)\cdot 100\%$ konfidensintervall for $\tilde{\mu}$ gitt ved

$$\left(\bar{x}_{(n(n+1)/2)-c+1}, \bar{x}_{(c)}\right),$$

der $c$ er slik at $H_0 : \tilde{\mu} = \tilde{\mu}_0$ forkastes til fordel for $H_a : \tilde{\mu} \neq \tilde{\mu}_0$ ved signifikansnivå $\alpha$ dersom $S_+ \geq c$ eller $S_+ \leq n(n+1)/2 - c$.

- For store $n$ ($n > 20$), kan en bruke at $S_+$ er tilnærmet normalfordelt, slik at

# Konfidensintervall for medianen $\tilde{\mu}$

- La $\bar{x}_{(1)},\ldots,\bar{x}_{(n(n+1)/2)}$ være alle observerte Walsh-gjennomsnitt, basert på observasjonene $x_1,\ldots,x_n$ av $X_1,\ldots,X_n$, ordnet i økende rekkefølge, altså

$$\bar{x}_{(1)} \leq \ldots \leq \bar{x}_{(n(n+1)/2)}.$$

- Da er et $(1-\alpha)\cdot 100\%$ konfidensintervall for $\tilde{\mu}$ gitt ved

$$\left(\bar{x}_{(n(n+1)/2)-c+1}, \bar{x}_{(c)}\right),$$

der $c$ er slik at $H_0 : \tilde{\mu} = \tilde{\mu}_0$ forkastes til fordel for $H_a : \tilde{\mu} \neq \tilde{\mu}_0$ ved signifikansnivå $\alpha$ dersom $S_+ \geq c$ eller $S_+ \leq n(n+1)/2 - c$.

- For store $n$ ($n > 20$), kan en bruke at $S_+$ er tilnærmet normalfordelt, slik at

$$c \approx \frac{n(n+1)}{4} + z_{\alpha/2}\sqrt{\frac{n(n+1)(2n+1)}{24}}.$$

# Konfidensintervall for medianen $\tilde{\mu}$

# Konfidensintervall for medianen $\tilde{\mu}$

- Hvis $n$ er stor og fordelingen til $X_i$ er tilnærmet normal, vil Wilcoxon-intervallet være litt bredere enn et standard t-intervall.

# Konfidensintervall for medianen $\tilde{\mu}$

- Hvis $n$ er stor og fordelingen til $X_i$ er tilnærmet normal, vil Wilcoxon-intervallet være litt bredere enn et standard t-intervall.
- Dersom fordelingen avviker en del fra normalfordelingen, vil imidlertid Wilcoxon-intervallet typisk være kortere enn t-intervallet.

# To-utvalgs rang-test, 14.3, **NB. Ikke pensum**

# To-utvalgs rang-test, 14.3, **NB. Ikke pensum**

Hvis vi har to utvalg på hhv. $m$ og $n$ observasjoner

$$X_1, \ldots, X_m \sim f_X(x) \text{ og uavhengige}$$

# To-utvalgs rang-test, 14.3, **NB. Ikke pensum**

Hvis vi har to utvalg på hhv. $m$ og $n$ observasjoner

$$X_1, \ldots, X_m \sim f_X(x) \text{ og uavhengige}$$
$$Y_1, \ldots, Y_n \sim f_Y(y) \text{ og uavhengige}$$

# To-utvalgs rang-test, 14.3, **NB. Ikke pensum**

Hvis vi har to utvalg på hhv. $m$ og $n$ observasjoner

$$X_1, \ldots, X_m \sim f_X(x) \text{ og uavhengige}$$
$$Y_1, \ldots, Y_n \sim f_Y(y) \text{ og uavhengige}$$

slik at $f_X(x) = f_Y(x - \Delta_0)$ for en eller annen verdi $\Delta_0$

# To-utvalgs rang-test, 14.3, **NB. Ikke pensum**

Hvis vi har to utvalg på hhv. $m$ og $n$ observasjoner

$$X_1, \ldots, X_m \sim f_X(x) \text{ og uavhengige}$$
$$Y_1, \ldots, Y_n \sim f_Y(y) \text{ og uavhengige}$$

slik at $f_X(x) = f_Y(x - \Delta_0)$ for en eller annen verdi $\Delta_0$ kan vi teste

$$\mathrm{H}_0 : \Delta_0 = 0$$

mot ensidige og tosidig alternativ

# To-utvalgs rang-test, 14.3, **NB. Ikke pensum**

Hvis vi har to utvalg på hhv. $m$ og $n$ observasjoner

$$X_1, \ldots, X_m \sim f_X(x) \text{ og uavhengige}$$
$$Y_1, \ldots, Y_n \sim f_Y(y) \text{ og uavhengige}$$

slik at $f_X(x) = f_Y(x - \Delta_0)$ for en eller annen verdi $\Delta_0$ kan vi teste

$$\mathrm{H}_0 : \Delta_0 = 0$$

mot ensidige og tosidig alternativ basert på rangen til det sammenslåtte materialet $X_1, \ldots, X_m, Y_1, \ldots, Y_n$.

# To-utvalgs rang-test, 14.3, **NB. Ikke pensum**

Hvis vi har to utvalg på hhv. $m$ og $n$ observasjoner

$$X_1, \ldots, X_m \sim f_X(x) \text{ og uavhengige}$$
$$Y_1, \ldots, Y_n \sim f_Y(y) \text{ og uavhengige}$$

slik at $f_X(x) = f_Y(x - \Delta_0)$ for en eller annen verdi $\Delta_0$ kan vi teste

$$\mathrm{H}_0 : \Delta_0 = 0$$

mot ensidige og tosidig alternativ basert på rangen til det sammenslåtte materialet $X_1, \ldots, X_m, Y_1, \ldots, Y_n$. La

$$W = \text{ summen av ranger knyttet til } X_i\text{-ene}$$

Siden $\sum_{k=1}^{m+n} k = \frac{(m+n)(m+n+1)}{2}$ vil forventningen til $W$ under nullhypotesen være lik

$$\mathrm{E}(W) = m\frac{m+n+1}{2}$$

# To-utvalg-rang test, forts

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

samt at $W$ dessuten er tilnærmet normalfordelt under $H_0$.

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

samt at $W$ dessuten er tilnærmet normalfordelt under $H_0$. Dermed kan vi teste denne for de ulike alternativene basert på

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

samt at $W$ dessuten er tilnærmet normalfordelt under $H_0$. Dermed kan vi teste denne for de ulike alternativene basert på

$$Z = \frac{W - \text{E}(W)}{\sqrt{\text{Var}(W)}} \sim \text{N}(0,1) \text{ tilnærmet under } H_0$$

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

samt at $W$ dessuten er tilnærmet normalfordelt under $H_0$. Dermed kan vi teste denne for de ulike alternativene basert på

$$Z = \frac{W - \text{E}(W)}{\sqrt{\text{Var}(W)}} \sim \text{N}(0,1) \text{ tilnærmet under } H_0$$

og man forkaster hvis $Z$ er tilstrekkelig liten,

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

samt at $W$ dessuten er tilnærmet normalfordelt under $H_0$. Dermed kan vi teste denne for de ulike alternativene basert på

$$Z = \frac{W - \text{E}(W)}{\sqrt{\text{Var}(W)}} \sim \text{N}(0,1) \text{ tilnærmet under } H_0$$

og man forkaster hvis $Z$ er tilstrekkelig liten, tilstrekkelig stor

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

samt at $W$ dessuten er tilnærmet normalfordelt under $H_0$. Dermed kan vi teste denne for de ulike alternativene basert på

$$Z = \frac{W - \text{E}(W)}{\sqrt{\text{Var}(W)}} \sim \text{N}(0,1) \text{ tilnærmet under } H_0$$

og man forkaster hvis $Z$ er tilstrekkelig liten, tilstrekkelig stor eller $|Z|$ er tilstrekkelig stor

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

samt at $W$ dessuten er tilnærmet normalfordelt under $H_0$. Dermed kan vi teste denne for de ulike alternativene basert på

$$Z = \frac{W - \text{E}(W)}{\sqrt{\text{Var}(W)}} \sim \text{N}(0,1) \text{ tilnærmet under } H_0$$

og man forkaster hvis $Z$ er tilstrekkelig liten, tilstrekkelig stor eller $|Z|$ er tilstrekkelig stor avhengig av alternativet.

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

samt at $W$ dessuten er tilnærmet normalfordelt under $H_0$. Dermed kan vi teste denne for de ulike alternativene basert på

$$Z = \frac{W - \text{E}(W)}{\sqrt{\text{Var}(W)}} \sim \text{N}(0,1) \text{ tilnærmet under } H_0$$

og man forkaster hvis $Z$ er tilstrekkelig liten, tilstrekkelig stor eller $|Z|$ er tilstrekkelig stor avhengig av alternativet.

Man kan også utlede kombinatoriske argumenter for eksakt fordeling under $H_0$ og finne eksakte P-verdier.

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

samt at $W$ dessuten er tilnærmet normalfordelt under $H_0$. Dermed kan vi teste denne for de ulike alternativene basert på

$$Z = \frac{W - \text{E}(W)}{\sqrt{\text{Var}(W)}} \sim \text{N}(0,1) \text{ tilnærmet under } H_0$$

og man forkaster hvis $Z$ er tilstrekkelig liten, tilstrekkelig stor eller $|Z|$ er tilstrekkelig stor avhengig av alternativet.

Man kan også utlede kombinatoriske argumenter for eksakt fordeling under $H_0$ og finne eksakte P-verdier.

Dette kalles en to-utvalgs Wilcoxon-test

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

samt at $W$ dessuten er tilnærmet normalfordelt under $H_0$. Dermed kan vi teste denne for de ulike alternativene basert på

$$Z = \frac{W - \text{E}(W)}{\sqrt{\text{Var}(W)}} \sim \text{N}(0,1) \text{ tilnærmet under } H_0$$

og man forkaster hvis $Z$ er tilstrekkelig liten, tilstrekkelig stor eller $|Z|$ er tilstrekkelig stor avhengig av alternativet.

Man kan også utlede kombinatoriske argumenter for eksakt fordeling under $H_0$ og finne eksakte P-verdier.

Dette kalles en to-utvalgs Wilcoxon-test evt. en Mann-Whitney-test

# To-utvalg-rang test, forts

Man kan også vise at variansen til $W$ under nullhypotesen blir gitt ved

$$\text{Var}(W) = \frac{mn(m+n+1)}{12}$$

samt at $W$ dessuten er tilnærmet normalfordelt under $H_0$. Dermed kan vi teste denne for de ulike alternativene basert på

$$Z = \frac{W - \text{E}(W)}{\sqrt{\text{Var}(W)}} \sim \text{N}(0,1) \text{ tilnærmet under } H_0$$

og man forkaster hvis $Z$ er tilstrekkelig liten, tilstrekkelig stor eller $|Z|$ er tilstrekkelig stor avhengig av alternativet.

Man kan også utlede kombinatoriske argumenter for eksakt fordeling under $H_0$ og finne eksakte P-verdier.

Dette kalles en to-utvalgs Wilcoxon-test evt. en Mann-Whitney-test og kan også utføres i R-med kommandoen `wilcox.test`.