
Ilkka Mellin

Tilastolliset menetelmät

Osa 3: Tilastolliset testit

Testit järjestysasteikollisille muuttujille

Testit järjestysasteikollisille muuttujille

>> **Järjestysasteikollisten muuttujien testit**

Merkitesti

Wilcoxonin rankitesti

Mannin ja Whitneyyn testi

Wilcoxonin rankisummatesti

Testit järjestysasteikollisille muuttujille 1/2

- Tarkastelemme seuraavia testejä (jatkuville) *järjestysasteikollisille* muuttujille:
 - **Merkitesti**
 - **Wilcoxonin rankitesti**
 - **Mannin ja Whitneyyn testi eli Wilcoxonin rankisummatesti**
- On syytä huomata, että testejä saa – ja on usein myös järkevää – käyttää *välimatka-* ja *suhdeasteikollisille muuttujille*.
- Mitta-asteikot: ks. lukua **Tilastollisten aineistojen kerääminen ja mittaaminen**.

Testit järjestysasteikollisille muuttujille 2/2

- Kaikki käsiteltävät testit ovat *ei-parametrisia* eli *jakaumista riippumattomia*, millä tarkoitetaan sitä, että testien yleiset hypoteesit eivät määrittele tarkasti perusjoukon jakaumaa.
- *Merkkitesti* ja *Wilcoxonin rankitesti* ovat luonteeltaan **yhden otoksen testejä**, mutta niitä voidaan soveltaa myös **parivertailuasetelmissä**.
- *Mannin ja Whitneyyn testi* eli *Wilcoxonin rankisummatesti* on luonteeltaan **kahden otoksen testi**.
- Kaikissa käsiteltävissä testeissä testataan tarkemmin määrittelemättömän todennäköisyysjakauman **sijainti-parametria (mediaania)** koskevia hypoteeseja

Testit järjestysasteikollisille muuttujille

Järjestysasteikollisten muuttujien testit

>> Merkkitesti

Wilcoxonin rankitesti

Mannin ja Whitneyyn testi

Wilcoxonin rankisummatesti

Testausasetelma

- Olkoon X_1, X_2, \dots, X_n yksinkertainen satunnaisotos perusjoukosta S , jonka jakauma on *symmetrinen*.
- Asetetaan jakauman *mediaanille* Me nollahypoteesi
$$H_0 : Me = Me_0$$
- Testausongelma:
Ovatko havainnot *sopuinnussa* nollahypoteesin H_0 kanssa?
- Ongelman eräänä ratkaisuna on **merkitesti**, joka vastaa parametristen testausasetelmien *yhden otoksen t-testiä*.

Testisuureet

- Määritellään *erotukset*

$$D_i = X_i - Me_0, i = 1, 2, \dots, n'$$

ja olkoon

$$n \leq n'$$

niiden erotusten D_i lukumäärä, jotka ovat $\neq 0$.

- Jos *nollahypoteesi* H_0 pätee, positiivisten ja negatiivisten erotusten on *jakauduttava suunnilleen tasan*.
- Määritellään **testisuureet** S^- ja S^+ :

S^- = negatiivisten erotusten $D_i = X_i - Me_0$ lukumäärä

S^+ = positiivisten erotusten $D_i = X_i - Me_0$ lukumäärä

Merkkitesti

Testisuureiden S^- ja S^+ ominaisuudet

- (i) $S^- + S^+ = n$
- (ii) Jos *nollahypoteesi* H_0 pätee, testisuureet S^- ja S^+ noudattavat *binomijakaumaa* $\text{Bin}(n, q)$ parametrein n ja $q = 1/2$:
 $S^- \sim \text{Bin}(n, 1/2)$
 $S^+ \sim \text{Bin}(n, 1/2)$
- (iii) Jos *nollahypoteesi* H_0 pätee,
 $E(S^-) = E(S^+) = nq = \frac{1}{2}n$
- (iv) Jos *nollahypoteesi* H_0 pätee,
 $D^2(S^-) = D^2(S^+) = nq(1 - q) = \frac{1}{4}n$

Merkkitesti

Eksakti testi

- Testisuureiden S^- ja S^+ jakaumat on taulukoitu ja monet tietokoneohjelmat laskevat testin p -arvoja.
- Merkkitestin p -arvot määrätään seuraavilla kaavoilla, joissa s^- (s^+) on testisuureen S^- (S^+) havaittu arvo:
 - (i) *Vaihtoehtoinen hypoteesi* $H_1 : Me > Me_0$
Testin p -arvo: $p = \Pr(S^+ > s^+)$
 - (ii) *Vaihtoehtoinen hypoteesi* $H_1 : Me < Me_0$
Testin p -arvo: $p = \Pr(S^- < s^-)$
 - (iii) *Vaihtoehtoinen hypoteesi*: $H_1 : Me \neq Me_0$
Testin p -arvo: $p = 2 \times \min\{\Pr(S^+ > s^+), \Pr(S^- < s^-)\}$

Standardoitu S-testisuure ja sen jakauma 1/2

- Jos *nollahypoteesi* H_0 pätee,

$$E(S^-) = E(S^+) = \frac{1}{2}n$$

$$D^2(S^-) = D^2(S^+) = \frac{1}{4}n$$

- Määritellään **testisuure**

$$z = \frac{S^* - E(S^*)}{D(S^*)}$$

jossa $S^* = S^-$ tai S^+ .

Standardoitu S-testisuure ja sen jakauma 2/2

- Jos *nollahypoteesi* H_0 pätee, niin testisuure

$$z = \frac{S^* - E(S^*)}{D(S^*)}$$

noudattaa suurissa otoksissa *approksimatiivisesti standardoitua normaalijakaumaa* $N(0, 1)$:

$$z \sim_a N(0,1)$$

- Approksimaatio on tavallisesti *riittävän hyvä*, jos $n > 20$.
- *Pienissä otoksissa* nojataan testisuureen S^* *tarkkaan jakaumaan*.

Asymptoottinen testi

- Testisuureen

$$z = \frac{S^* - E(S^*)}{D(S^*)}$$

normaaliarvo = 0, koska *nollahypoteesin* H_0 *pätiessä*

$$E(z) = 0$$

- Siten itseisarvoltaan *suuret* testisuureen z arvot viittaavat siihen, että *nollahypoteesi* H_0 *ei päde*.
- *Nollahypoteesi* H_0 *hylätään*, jos testin p -arvo on *kyllin pieni*.
- Hylkäysalueen valinta ja p -arvon määrittäminen:
ks. lukua **Tilastollinen testaus**.

Kommentteja

- Merkkitesti voidaan tulkita *yhden otoksen t -testin ei-parametriseksi vastineeksi*.
- Merkkitestissä *ei tehdä*
 - toisin kuin yhden otoksen t -testissä –
mitään *oletuksia perusjoukon jakauman tyypistä*.
- Merkkitestin testisuureen arvo *ei riipu havaintoarvoista*, vaan ainoastaan niiden keskinäisestä *järjestyksestä*.

Merkkitestin soveltaminen parivertailuasetelmiin 1/2

- Merkkitestiä voidaan soveltaa *parivertailuasetelmiin*, joissa havainnot muodostuvat toisistaan *riippumattomista mittauspareista*

$$(X_i, Y_i), i = 1, 2, \dots, n'$$

- Oletetaan, että X - ja Y -mittausten jakaumat ovat muuten samat, mutta niiden *mediaaneilla* (*sijaintiparametreilla*) saattaa olla eri arvot.
- Määritellään havaintojen X_i ja Y_i erotukset

$$D_i = X_i - Y_i, i = 1, 2, \dots, n'$$

ja olkoon n niiden erotusten D_i lukumäärä, jotka ovat $\neq 0$.

Merkkitestin soveltaminen parivertailuasetelmiin 2/2

- Tehdään oletus, että erotusten

$$D_i = X_i - Y_i, i = 1, 2, \dots, n$$

jakauma on *symmetrinen*.

- Määritellään **testisuureet** S^- ja S^+ erotuksille D_i kuten edellä.
- Olkoon

$$Me_D$$

erotusten $D_i = X_i - Y_i, i = 1, 2, \dots, n$ *mediaani*.

- Tällöin *nollahypoteesin*

$$H_0 : Me_D = 0$$

testaamiseen voidaan soveltaa **merkkitestiä**.

Testit järjestysasteikollisille muuttujille

Järjestysasteikollisten muuttujien testit

Merkkitesti

>> Wilcoxonin rankitesti

Mannin ja Whitneyyn testi

Wilcoxonin rankisummatesti

Wilcoxonin rankitesti

Testausasetelma

- Olkoon X_1, X_2, \dots, X_n yksinkertainen satunnaisotos perusjoukosta S , jonka jakauma on *symmetrinen*.
- Asetetaan jakauman *mediaanille* Me nollahypoteesi
$$H_0 : Me = Me_0$$
- Testausongelma:
Ovatko havainnot *sopuinnussa* nollahypoteesin H_0 kanssa?
- Ongelman eräänä ratkaisuna on **Wilcoxonin rankitesti**, joka vastaa parametristen testausasetelmien *yhden otoksen t-testiä*.

Wilcoxonin rankitesti

Testisuure 1/2

- Olkoon

$$|D_i| = |X_i - Me_0|, i = 1, 2, \dots, n'$$

ja olkoon

$$n \leq n'$$

niiden erotusten D_i lukumäärä, jotka ovat $\neq 0$.

- Olkoot

$$Z_1, Z_2, \dots, Z_n$$

itseisarvot $|D_i|$ järjestettyinä *suuruusjärjestykseen* pienimmästä suurimpaan ja olkoon

$$R(Z_i) = \text{itseisarvon } Z_i \text{ järjestysnumero eli ranki,} \\ i = 1, 2, \dots, n$$

Wilcoxonin rankitesti

Testisuure 2/2

- Määritellään **testisuure**

$$W^- = \sum_{D_i < 0} R(Z_i)$$

- W^- on niiden *rankien summa*, joita vastaavat erotukset

$$D_i = X_i - Me_0 < 0$$

- Määritellään **testisuure**

$$W^+ = \sum_{D_i > 0} R(Z_i)$$

- W^+ on niiden *rankien summa*, joita vastaavat erotukset

$$D_i = X_i - Me_0 > 0$$

Wilcoxonin rankitesti

Testisuureiden W^- ja W^+ ominaisuudet

(i) $W^- + W^+ = \frac{1}{2}n(n+1)$

(ii) Jos *nollahypoteesi* H_0 pätee,

$$E(W^-) = E(W^+) = \frac{1}{4}n(n+1)$$

(iii) Jos *nollahypoteesi* H_0 pätee,

$$D^2(W^-) = D^2(W^+) = \frac{1}{24}n(n+1)(2n+1)$$

Wilcoxonin rankitesti

Eksakti testi

- Testisuureiden W^- ja W^+ jakaumat on taulukoitu ja monet tietokoneohjelmat laskevat testin p -arvoja.
- Wilcoxonin rankitestin p -arvot määrätään seuraavilla kaavoilla, joissa w^- ja w^+ ovat testisuureiden W^- ja W^+ havaitut arvot:
 - (i) *Vaihtoehtoinen hypoteesi* $H_1 : Me > Me_0$
Testin p -arvo: $p = \Pr(W^+ > w^+)$
 - (ii) *Vaihtoehtoinen hypoteesi* $H_1 : Me < Me_0$
Testin p -arvo: $p = \Pr(W^- < w^-)$
 - (iii) *Vaihtoehtoinen hypoteesi* $H_1 : Me \neq Me_0$
Testin p -arvo: $p = 2 \times \min\{\Pr(W^+ > w^+), \Pr(W^- < w^-)\}$

Standardoitu W -testisuure ja sen jakauma 1/2

- Jos *nollahypoteesi* H_0 pätee,

$$E(W^-) = E(W^+) = \frac{1}{4}n(n+1)$$

$$D^2(W^-) = D^2(W^+) = \frac{1}{24}n(n+1)(2n+1)$$

- Määritellään **testisuure**

$$z = \frac{W^* - E(W^*)}{D(W^*)}$$

jossa $W^* = W^-$ tai W^+ .

Standardoitu W -testisuure ja sen jakauma 2/2

- Jos *nollahypoteesi* H_0 pätee, niin testisuure

$$z = \frac{W^* - E(W^*)}{D(W^*)}$$

noudattaa suurissa otoksissa *approksimatiivisesti standardoitua normaalijakaumaa* $N(0, 1)$:

$$z \sim_a N(0,1)$$

- Approksimaatio on tavallisesti *riittävän hyvä*, jos $n > 20$.
- *Pienissä otoksissa* nojataan testisuureen W^* *tarkkaan jakaumaan*.

Asymptoottinen testi

- Testisuureen

$$z = \frac{W^* - E(W^*)}{D(W^*)}$$

normaaliarvo = 0, koska *nollahypoteesin* H_0 *pätiessä*

$$E(z) = 0$$

- Siten itseisarvoltaan *suuret* testisuureen z arvot viittaavat siihen, että *nollahypoteesi* H_0 *ei päde*.
- *Nollahypoteesi* H_0 *hylätään*, jos testin p -arvo on *kyllin pieni*.
- Hylkäysalueen valinta ja p -arvon määrittäminen:
ks. lukua **Tilastollinen testaus**.

Wilcoxonin rankitesti

Kommentteja

- Wilcoxonin rankitesti voidaan tulkita *yhden otoksen t -testin ei-parametriseksi vastineeksi*.
- Wilcoxonin rankitestissä *ei tehdä*
– toisin kuin yhden otoksen t -testissä –
mitään oletuksia perusjoukon jakauman tyypistä.
- Wilcoxonin rankitestin testisuureen arvo ei riipu *havaintoarvoista*, vaan ainoastaan niiden keskinäisestä *järjestyksestä*.
- Wilcoxonin rankitesti käyttää merkkitestistä *enemmän informaatiota havaintojen järjestyksestä*.
- Wilcoxonin rankitesti on *voimakkaampi* kuin merkkitesti.

Wilcoxonin rankitestin soveltaminen parivertailuasetelmiin 1/2

- Wilcoxonin rankitestiä voidaan soveltaa *parivertailuasetelmiin*, joissa havainnot muodostuvat toisistaan *riippumattomista mittauspareista*

$$(X_i, Y_i), i = 1, 2, \dots, n'$$

- Oletetaan, että X - ja Y -mittausten jakaumat ovat muuten samat, mutta niiden *mediaaneilla* (*sijaintiparametreilla*) *saattaa olla eri arvot*.

- Määritellään havaintojen X_i ja Y_i erotukset

$$D_i = X_i - Y_i, i = 1, 2, \dots, n'$$

ja olkoon n niiden erotusten D_i lukumäärä, jotka ovat $\neq 0$.

Wilcoxonin rankitestin soveltaminen parivertailuasetelmiin 2/2

- Oletetaan, että erotusten

$$D_i = X_i - Y_i, i = 1, 2, \dots, n$$

jakauma on *symmetrinen*.

- Määritellään **testisuureet** W^- ja W^+ erotuksille D_i kuten edellä.
- Olkoon

$$Me_D$$

erotusten $D_i = X_i - Y_i, i = 1, 2, \dots, n$ *mediaani*.

- Tällöin *nollahypoteesin*

$$H_0 : Me_D = 0$$

testaamiseen voidaan soveltaa **Wilcoxonin rankitestiä**.

Testit järjestysasteikollisille muuttujille

Järjestysasteikollisten muuttujien testit

Merkkitesti

Wilcoxonin rankitesti

>> Mannin ja Whitneyyn testi

Wilcoxonin rankisummatesti

Mannin ja Whitneyyn testi

Testausasetelma 1/2

- Oletetaan, että

$$X_1, X_2, \dots, X_n$$

ovat *riippumattomia havaintoja* satunnaismuuttujan X jakaumasta perusjoukossa S_1 (otos 1).

- Oletetaan, että

$$Y_1, Y_2, \dots, Y_m$$

ovat *riippumattomia havaintoja* satunnaismuuttujan Y jakaumasta perusjoukossa S_2 (otos 2).

- Olkoot otokset lisäksi toisistaan *riippumattomia*.

Testausasetelma 2/2

- Oletetaan, että satunnaismuuttujat X ja Y noudattavat muuten *samaa jakaumaa*, mutta niiden *mediaanit (sijaintiparametrit)* saattavat *erota toisistaan*.
- Asetetaan *nollahypoteesi*, että satunnaismuuttujilla X ja Y on sama *mediaani (sijaintiparametri)*.
- Testausongelma:
Ovatko havainnot *sopuosoinnussa* nollahypoteesin H_0 kanssa?
- Ongelman eräänä ratkaisuna on **Mannin ja Whitneyyn testi**, joka vastaa parametristen testausasetelmien *kahden riippumattoman otoksen t-testiä*.

Mannin ja Whitneyyn testi

Yleinen hypoteesi

- *Yleinen hypoteesi* H :
 - (1) Havainnot $X_i \sim F_X, i = 1, 2, \dots, n$
 - (2) Havainnot $Y_j \sim F_Y, j = 1, 2, \dots, m$
 - (3) Jakaumat F_X ja F_Y ovat muuten samat, mutta niiden *mediaanit (sijaintiparametrit)* saattavat *erota toisistaan*.
 - (4) Havainnot X_i ja Y_j ovat *riippumattomia* kaikille i ja j
- Huomautus:
 - Oletus (3) sisältää *kolme riippumattomuusoletusta*:
 - Havainnot ovat riippumattomia otoksien 1 ja 2 *sisällä*.
 - Havainnot ovat riippumattomia otoksien 1 ja 2 *välillä*.

Nollahypoteesi ja vaihtoehtoinen hypoteesi

- *Nollahypoteesi* H_0 :

$$H_0 : F_X = F_Y$$

- *Vaihtoehtoinen hypoteesi* H_1 :

$$H_1 : F_X \neq F_Y$$

Testin idea

- *Yhdistetään X- ja Y-havainnot yhdeksi otokseksi ja järjestetään yhdistetyn otoksen havainnot suuruusjärjestykseen pienimmästä suurimpaan.*
- *Tarkastellaan miten X- ja Y-havainnot seuraavat yhdistetyssä otoksessa toisiaan.*
- *Jos kaikki X-havainnot (Y-havainnot) edeltävät kaikkia Y-havaintoja (X-havaintoja), ei ole uskottavaa, että nollahypoteesi H_0 pätee.*
- *Jos satunnaismuuttujat X ja Y noudattavat samaa jakaumaa, on ilmeistä, että X- ja Y-havaintojen on sekoituttava sopivasti toisiinsa.*
- *Mannin ja Whitneyyn testisuure mittaa tätä sekoittumista.*

Testisuure U_1 – muoto 1

- Määritellään satunnaismuuttujat

$$D_{ij}^{(1)} = \begin{cases} 1, & \text{jos } X_i < Y_j \\ 0, & \text{jos } X_i > Y_j \end{cases}$$
$$i = 1, 2, \dots, n, j = 1, 2, \dots, m$$

ja testisuure

$$U_1 = \sum_{i=1}^n \sum_{j=1}^m D_{ij}^{(1)}$$

Testisuure U_1 – muoto 2

- Määritellään satunnaismuuttujat

$R(X_i)$ = havainnon X_i järjestysnumero eli *ranki*
yhdistetyssä otoksessa
 $i = 1, 2, \dots, n$

ja testisuure

$$U_1 = nm + \frac{1}{2}n(n+1) - \sum_{i=1}^n R(X_i)$$

- Testisuureen U_1 muodot 1 ja 2 ovat *ekvivalentteja*.

Testisuureen U_1 ominaisuudet

- Testisuureen U_1 arvo ei riipu X - ja Y -havaintoarvojen *suuruudesta*, vaan ainoastaan niiden keskinäisestä *järjestyksestä*.

- Aina pätee

$$0 \leq U_1 \leq nm$$

ja erityisesti

$$U_1 = 0, \text{ jos } X_i > Y_j \text{ kaikille } i \text{ ja } j$$

$$U_1 = nm, \text{ jos } X_i < Y_j \text{ kaikille } i \text{ ja } j$$

Testisuure U_2 – muoto 1

- Määritellään satunnaismuuttujat

$$D_{ji}^{(2)} = \begin{cases} 1, & \text{jos } Y_j < X_i \\ 0, & \text{jos } Y_j > X_i \end{cases}$$
$$j = 1, 2, \dots, m, i = 1, 2, \dots, n$$

ja testisuure

$$U_2 = \sum_{j=1}^m \sum_{i=1}^n D_{ji}^{(2)}$$

Testisuure U_2 – muoto 2

- Määritellään satunnaismuuttujat

$$R(Y_j) = \text{havainnon } Y_j \text{ järjestysnumero eli ranki}$$

yhdistetyssä otoksessa

$$j = 1, 2, \dots, m$$

ja testisuure

$$U_2 = nm + \frac{1}{2}m(m+1) - \sum_{j=1}^m R(Y_j)$$

- Testisuureen U_2 muodot 1 ja 2 ovat *ekvivalentteja*.

Testisuureen U_2 ominaisuudet

- Testisuureen U_2 arvo ei riipu X - ja Y -havaintoarvojen *suuruudesta*, vaan ainoastaan niiden keskinäisestä *järjestyksestä*.

- Aina pätee

$$0 \leq U_2 \leq nm$$

ja erityisesti

$$U_2 = 0, \text{ jos } Y_j > X_i \text{ kaikille } i \text{ ja } j$$

$$U_2 = nm, \text{ jos } Y_j < X_i \text{ kaikille } i \text{ ja } j$$

Mannin ja Whitneyyn testi

Testisuureiden U_1 ja U_2 ominaisuudet

(i) $U_1 + U_2 = nm$

(ii) Jos *nollahypoteesi* H_0 pätee,

$$E(U_1) = E(U_2) = \frac{1}{2}nm$$

(iii) Jos *nollahypoteesi* H_0 pätee,

$$D^2(U_1) = D^2(U_2) = \frac{1}{12}nm(n + m + 1)$$

Standardoitu U_1 -testisuure ja sen jakauma

- Jos *nollahypoteesi* H_0 pätee, niin standardoitu satunnaismuuttuja

$$z_1 = \frac{U_1 - E(U_1)}{D(U_1)}$$

noudattaa suurissa otoksissa *approksimatiivisesti standardoitua normaalijakaumaa*:

$$z_1 \sim_a N(0,1)$$

- Approksimaatio on tavallisesti *riittävän hyvä*, jos $n > 10$ ja $m > 10$.
- *Pienissä otoksissa* nojataan testisuureen U_1 *tarkkaan jakaumaan*.

Asymptoottinen testi – muoto 1

- **Testisuureen**

$$z_1 = \frac{U_1 - E(U_1)}{D(U_1)}$$

normaaliarvo = 0, koska *nollahypoteesin* H_0 *pätiessä*

$$E(z_1) = 0$$

- Siten itseisarvoltaan *suuret* testisuureen z_1 arvot viittaavat siihen, että *nollahypoteesi* H_0 *ei päde*.
- *Nollahypoteesi* H_0 *hylätään*, jos testin p -arvo on *kyllin pieni*.
- Hylkäysalueen valinta ja p -arvon määrittäminen:
ks. lukua **Tilastollinen testaus**.

Standardoitu U_2 -testisuure ja sen jakauma

- Jos *nollahypoteesi* H_0 pätee, niin standardoitu satunnaismuuttuja

$$z_2 = \frac{U_2 - E(U_2)}{D(U_2)}$$

noudattaa suurissa otoksissa *approksimatiivisesti standardoitua normaalijakaumaa*:

$$z_2 \sim_a N(0,1)$$

- Approksimaatio on tavallisesti *riittävän hyvä*, jos $n > 10$ ja $m > 10$.
- *Pienissä otoksissa* nojataan testisuureen U_2 *tarkkaan jakaumaan*.

Asymptoottinen testi – muoto 2

- **Testisuureen**

$$z_2 = \frac{U_2 - E(U_2)}{D(U_2)}$$

normaaliarvo = 0, koska *nollahypoteesin* H_0 *pätiessä*

$$E(z_2) = 0$$

- Siten itseisarvoltaan *suuret* testisuureen z_2 arvot viittaavat siihen, että *nollahypoteesi* H_0 *ei päde*.
- *Nollahypoteesi* H_0 *hylätään*, jos testin p -arvo on *kyllin pieni*.
- Hylkäysalueen valinta ja p -arvon määrittäminen:
ks. lukua **Tilastollinen testaus**.

Mannin ja Whitneyyn testi

Kommentteja 1/2

- Mannin ja Whitneyyn testi voidaan tulkita *kahden riippumattoman otoksen t -testin ei-parametriseksi vastineeksi*.
- Mannin ja Whitneyyn testissä *ei tehdä*
 - toisin kuin kahden riippumattoman otoksen t -testissä – *mitään oletuksia perusjoukkojen jakaumasta*.
- Mannin ja Whitneyyn testisuureiden arvo ei riipu muuttujien X ja Y arvoista, vaan ainoastaan niiden keskinäisestä *järjestyksestä*.

Mannin ja Whitneyyn testi

Kommentteja 2/2

- Jos havainnot *ovat normaalijakautuneita*, Mannin ja Whitneyyn testi *ei ole yhtä voimakas* kuin kahden riippumattoman otoksen *t*-testi.
- Jos havainnot *eivät ole normaalijakautuneita*, Mannin ja Whitneyyn testi *saattaa olla paljon voimakkaampi* kuin kahden riippumattoman otoksen *t*-testi.
- Mannin ja Whitneyyn testi *on vartenotettava vaihtoehto kahden riippumattoman otoksen t-testille*, jos otoskoot *eivät ole kovin isoja* ja *perusjoukot eivät ole normaalijakautuneita*.

Testit järjestysasteikollisille muuttujille

Järjestysasteikollisten muuttujien testit

Merkkitesti

Wilcoxonin rankitesti

Mannin ja Whitneyyn testi

>> Wilcoxonin rankisummatesti

Wilcoxonin rankisummatesti

Wilcoxonin rankisummatesti ja Mannin ja Whitneyyn testi

- *Wilcoxonin rankisummatesti* perustuu Mannin ja Whitneyyn testisuureiden muodoissa 2 esiintyviin havaintojen *rankisummiin* eli *järjestyslukujen summiin*.
- Wilcoxonin rankisummatesti on *ekvivalentti* Mannin ja Whitneyyn testin kanssa.

Testisuure T_1

- Määritellään satunnaismuuttujat

$R(X_i)$ = havainnon X_i järjestysnumero eli *ranki*
yhdistetyssä otoksessa
 $i = 1, 2, \dots, n$

ja testisuure

$$T_1 = \sum_{i=1}^n R(X_i)$$

Testisuure T_2

- Määritellään satunnaismuuttujat

$R(Y_j)$ = havainnon Y_j järjestysnumero eli *ranki*
yhdistetyssä otoksessa
 $j = 1, 2, \dots, m$

ja testisuure

$$T_2 = \sum_{j=1}^m R(Y_j)$$

Wilcoxonin rankisummatesti

Testisuureiden T_1 ja T_2 ominaisuudet

(i) $T_1 + T_2 = \frac{1}{2}(n + m)(n + m + 1)$

(ii) Jos *nollahypoteesi* H_0 pätee,

$$E(T_1) = \frac{1}{2}n(n + m + 1)$$

$$E(T_2) = \frac{1}{2}m(n + m + 1)$$

(iii) Jos *nollahypoteesi* H_0 pätee,

$$D^2(T_1) = D^2(T_2) = \frac{1}{12}nm(n + m + 1)$$

Standardoitu T_1 -testisuure ja sen jakauma

- Jos *nollahypoteesi* H_0 pätee, niin standardoitu satunnaismuuttuja

$$z_1 = \frac{T_1 - E(T_1)}{D(T_1)}$$

noudattaa suurissa otoksissa *approksimatiivisesti standardoitua normaalijakaumaa*:

$$z_1 \sim_a N(0,1)$$

- Approksimaatio on tavallisesti *riittävän hyvä*, jos $n > 10$ ja $m > 10$.
- *Pienissä otoksissa* nojataan testisuureen T_1 *tarkkaan jakaumaan*.

Asymptoottinen testi – muoto 1

- **Testisuureen**

$$z_1 = \frac{T_1 - E(T_1)}{D(T_1)}$$

normaaliarvo = 0, koska *nollahypoteesin* H_0 *pätiessä*

$$E(z_1) = 0$$

- Siten itseisarvoltaan *suuret* testisuureen z_1 arvot viittaavat siihen, että *nollahypoteesi* H_0 *ei päde*.
- *Nollahypoteesi* H_0 *hylätään*, jos testin p -arvo on *kyllin pieni*.
- Hylkäysalueen valinta ja p -arvon määrittäminen:
ks. lukua **Tilastollinen testaus**.

Standardoitu T_2 -testisuure ja sen jakauma

- Jos *nollahypoteesi* H_0 pätee, niin standardoitu satunnaismuuttuja

$$z_2 = \frac{T_2 - E(T_2)}{D(T_2)}$$

noudattaa suurissa otoksissa *approksimatiivisesti standardoitua normaalijakaumaa*:

$$z_2 \sim_a N(0,1)$$

- Approksimaatio on tavallisesti *riittävän hyvä*, jos $n > 10$ ja $m > 10$.
- *Pienissä otoksissa* nojataan testisuureen T_2 tarkkaan jakaumaan.

Asymptoottinen testi – muoto 2

- **Testisuureen**

$$z_2 = \frac{T_2 - E(T_2)}{D(T_2)}$$

normaaliarvo = 0, koska *nollahypoteesin* H_0 *pätiessä*

$$E(z_2) = 0$$

- Siten itseisarvoltaan *suuret* testisuureen z_2 arvot viittaavat siihen, että *nollahypoteesi* H_0 *ei päde*.
- *Nollahypoteesi* H_0 *hylätään*, jos testin p -arvo on *kyllin pieni*.
- Hylkäysalueen valinta ja p -arvon määrittäminen:
ks. lukua **Tilastollinen testaus**.