Kapitel 7
Forskelle mellem centraltendenser

Peter Tibert Stoltze
stat@peterstoltze.dk

Elementær statistik
F2011

Indledning

1. z-test for ukorrelerede data
2. t-test for ukorrelerede data med ens varianser
3. t-test for ukorrelerede data med uens varianser
4. z-test for korrelerede data
5. t-test for korrelerede data
6. Mann-Whitney U-test for ukorrelerede data
7. Wilcoxons rangtest for korrelerede data

Parametriske tests

- Vi starter med at opstille nulhypotesen H_0 og en relevant alternativhypotese H_1
- Vi beregner en teststørrelse med kendt fordeling på baggrund af data
- Ved opslag i tabel omsættes teststørrelsen til en signifikanssandsynlighed kaldet p
- For n mindre end 30 foretrækkes t-fordelingen, og ellers z-fordelingen
- På baggrund af p konkluderes det, om H_0 kan antages eller bør forkastes

Ikke-parametriske tests

- De ikke-parametriske tests benyttes når data er målt på en ordinalskala, altså når de kan ordnes i en rækkefølge men differensen mellem observationer ikke tillægges nogen speciel betydning
- Et centralt begreb er rangværdi, der udtrykker placeringen når observationerne ordnes i rækefølge
- Som ved de parametriske tests opstilles der et hypotesesæt og der beregnes en teststørrelse, men omsætningen til en signifikanssandsynlighed p må ske ved hjælp af testspecifikke tabeller

Parametriske tests for ukorrelerede data

- Disse tests er aktuelle når data er
  - målt på en ratio- eller intervalskala
  - ukorrelerede, dvs. hænger ikke naturligt sammen i par
- Når $n \geq 30$ kan vi anvende et z-test
- Når $n < 30$ anvender vi et t-test, idet vi dog først skal teste, om der kan antages ens varians

1. z-test for ukorrelerede data: Læsescore

- Vi starter med z-testet for ukorrelerede data på det eksempel, der omhandler (opdigtede) læsescores for 30 piger og 30 drenge i tredje klasse
- Vi kan opstille følgende summariske beskrivelse af data:

Køn	i	n_i	$\bar{x}_i$	s_i
piger	1	30	59,83	16,58
drenge	2	30	51,90	19,24

1. z-test for ukorrelerede data: Hypoteser

- Første trin er at opstille en relevant hypotese. I dette tilfælde kan vi se, at stikprøvegennemsnittet er højst for pigerne, hvilket vi vælger at bruge til at formulere et retningsbestemt alternativ:

$$\begin{aligned} H_0 &: \quad \mu_{piger} = \mu_{drenge} \\ H_1 &: \quad \mu_{piger} > \mu_{drenge} \end{aligned}$$

- Vi vender tilbage til betydningen af alternativhypotesen H_1 ift. bestemmelsen af signifikanssandsynligheden p

1. z-test for ukorrelerede data: Teststørrelse

- Teststørrelsen z beregnes som forskellen på de to stikprøvegennemsnit normeret med spredningen på denne
- Forskellen beregnes let til

$$\bar{x}_1 - \bar{x}_2 = 59,83 - 51,90 = 7,93$$

- Variansen på forskellen beregnes som

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{16,58^2}{30} + \frac{19,24^2}{30}} = 4,636$$

- Endelig kan teststørrelsen z beregnes som

$$z = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{7,93}{4,636} = 1,711$$

1. z-test for ukorrelerede data: Signifikanssandsynlighed

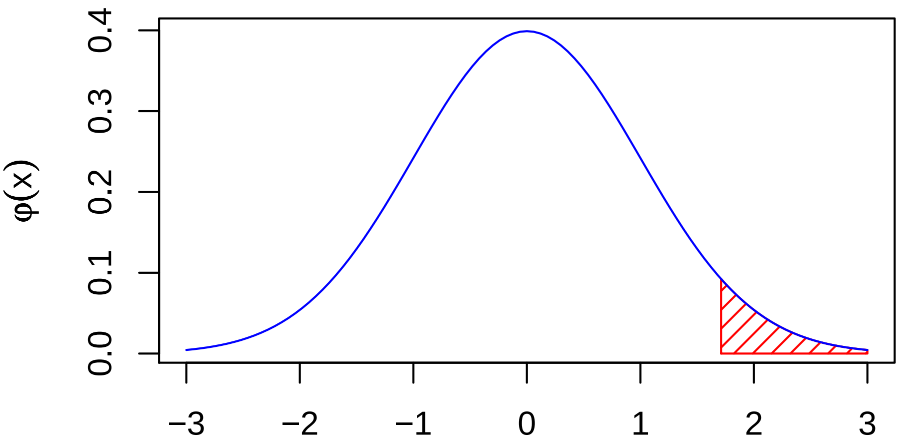
- Nu skal teststørrelsen z omsættes til en signifikans-sandsynlighed p, idet vi tager hensyn til vores sæt af hypoteser

- Alternativhypotesen er $\mu_{piger} > \mu_{drenge}$, så vi skal finde sandsynligheden for en teststørrelse på 1,711 eller mere i standardnormalfordelingen:

$$\begin{aligned} p = \mathrm{P}(Z > 1,711) &= 1 - \mathrm{P}(Z < 1,711) \\ &= 1 - \Phi(1,711) \\ &= 1 - 0,956 = 0,044 \end{aligned}$$

1. z-test for ukorrelerede data: Signifikanssandsynlighed

- Sandsynlighed kan også illustreres grafisk ved hjælp af det skraverede areal i følgende figur:



- Sandsynligheden for en teststørrelsen på mere end 1,711 er altså 4,4% under den antagelse at H_0 er sand

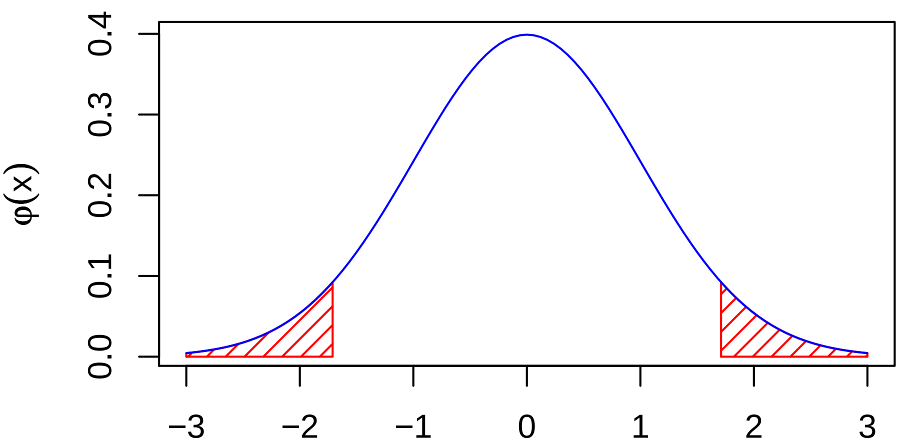
1. z-test for ukorrelerede data: Konklusion

- Med signifikanssandsynligheden $p = 0,044$ vælger vi at forkaste vores nulhypotese, $\mu_{piger} = \mu_{drenge}$, og i stedet overgå til alternativhypotesen, $\mu_{piger} > \mu_{drenge}$
- Vi skal dog være opmærksomme på, at selvom $\mu_{piger} = \mu_{drenge}$ faktisk er sandt, så vil vi have en sandsynlighed på 4,4% for at observere en forskel $x_1 - x_2$ på 7,93 eller mere
- Signifikanssandsynligheden p udtrykker altså også risikoen for at lave en fejl af Type I

1. z-test for ukorrelerede data: Tosidet alternativ

- Med alternativhypotesen $\mu_{piger} \neq \mu_{drenge}$ får vi

$$p = 2 \cdot \mathrm{P}(Z > |z|) = 2(1 - \Phi(z)) = 2(1 - 0,956) = 0,088$$

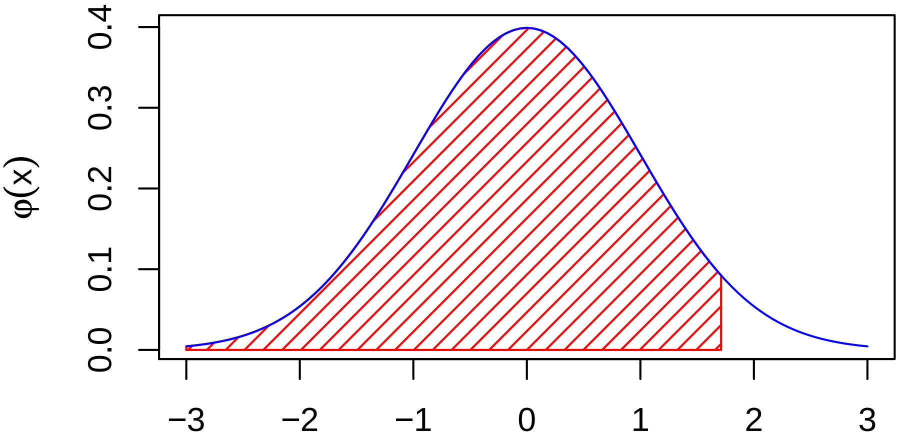


- Dette er det dobbelte af den oprindelige p så det bliver sværere at afvise H_0 — og grunden til at et retningsbestemt alternativ ofte foretrækkes

1. z-test for ukorrelerede data: Forkert alternativ

- Med alternativhypotesen $\mu_{piger} < \mu_{drenge}$ får vi

$$p = \mathrm{P}(Z < z) = \Phi(z) = 0,956$$



- Dette er meget højt fordi nulhypotesen er meget attraktiv sammenlignet med alternativet, der jo er i direkte modstrid med vores observationer

Sammenhæng mellem de parametriske tests

- De forskellige tests
  - z-test for ukorrelerede data
  - t-test for ukorrelerede data med ens varianser
  - t-test for ukorrelerede data med uens varianser
  - z-test for korrelerede data
  - t-test for korrelerede data

  er alle bygget op over samme ide:

$$\text{teststørrelse} = \frac{\text{forskel}}{\text{spredning på forskel}}$$

- Signifikanssandsynligheden p beregnes ud fra teststørrelsen (med kendt fordeling) og de opstillede hypoteser

Præ 2 og 3: F-test for ens varians

- Eksempel i Tabel 7.2
- Starter med at beregne stikprøvevarianserne $s_1^2 = 3,0$ og $s_2^2 = 2,2$ idet $n_1 = 4$ og $n_2 = 5$
- Dernæst beregnes F som

$$F = \frac{s_{max}^2}{s_{min}^2} = \frac{3,0}{2,2} = 1,36$$

  der følger en F-fordelingen med 4-1 $=$ 3 frihedsgrader i tæller og 5-1 $=$ 4 frihedsgrader i nævner
- Vi finder kritisk værdi for $\alpha = 0,05$ til 6,59 hvilket betyder at $p > 0,05$ og at H_0 $(s_1^2 = s_2^2)$ derfor ikke kan forkastes

2. t-test for ukorrelerede data med ens varians

- Samme princip som i z-testet men nu beregnes teststørrelsen t som

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2(1/n_1 + 1/n_2)}}$$

  hvor

$$s^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

- t-værdien vurderes i en t-fordeling med antal frihedsgrader svarende til $n_1 + n_2 - 2$

3. t-test for ukorrelerede data med uens varians

- Nu beregnes teststørrelsen t som

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

- t-værdien vurderes i en t-fordeling med antal frihedsgrader svarende til

$$df = \frac{(n_1 - 1)(n_2 - 1)}{(n_2 - 1)c^2 + (n_1 - 1)(1 - c)^2}$$

hvor

$$c = \frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2}$$

Parametriske tests for korrelerede data

- Disse tests er aktuelle når data er
  - målt på en ratio- eller intervalskala
  - korrelerede, dvs. hænger naturligt sammen i par
- Når $n \geq 30$ kan vi anvende et z-test
- Når $n < 30$ anvender vi et t-test
- I begge tilfælde regnes på differenserne $d_i = x_i - y_i$

4. z-test for korrelerede data

- Først bestemmes spredningen på differenserne som

$$s_d^2 = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$$

- Herefter beregnes teststørrelsen z som

$$z = \frac{\bar{d}}{s_d/\sqrt{n}}$$

- Signifikanssandsynligheden p bestemmes ved opslag i Tabel A

5. t-test for korrelerede data

- Når $n < 30$ benyttes et t-test i stedet for z-testet
- Teststørrelsen t beregnes præcist som ved z-testet, men p bestemmes ved opslag i en t-fordeling med $n - 1$ frihedsgrader

t-test med regneark

- Regnearksfunktionen `ttest(...)` kræver fire argumenter:
  - Første datavektor
  - Anden datavektor
  - Formulering af alternativhypotese (1=ensidig, 2=tosidig)
  - Testtype (1=korreleret, 2=ukorreleret men med ens varians, 3=ukorreleret med uens varians)
- Regnearksfunktionen `ztest(...)` testet tosidet for en bestemt middelværdi og kræver tre argumenter:
  - Datavektor
  - Middelværdi under nulhypotesen
  - Eventuelt kendt spredning

  og er altså ikke den samme z-test som vi kender fra kompendiet...

De ikke-parametriske test

- Mann-Whitney U-testet svarer til t- og z-test for ukorrelerede data
- Wilcoxons rangtest svarer til t- og z-test for korrelerede data
- Begge test benytter rangværdier, der udtrykker placeringen af en observation i en ordnet stikprøve - hvis stikprøven f.eks. er

$$X = \{7, 9, 13, 5\}$$

  har vi følgende rangværdier

$$R(7) = 2, \quad R(9) = 3, \quad R(13) = 4, \quad R(5) = 1$$

Mann-Whitney U-test

- Svarer til t- og z-test for ukorrelerede data
- Fælles rangordning af observationer i to stikprøver (1 og 2)
- Dernæst summeres antallet af 1'ere før hver 2'er og omvendt, og det mindste tal anvendes som teststørrelse
- Optælling kan erstattes af følgende formler:

$$U = \min(U_1, U_2)$$

hvor

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad \text{og} \quad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

og R_i er summen af rangværdierne for observationerne i stikprøve i

Mann-Whitney — eksempel

Stikprøve	Observationer
1	3,6,6,7
2	1,3,3,4,5

Stikprøve 1		Stikprøve 2	
Score	Rang	Score	Rang
3	3	1	1
6	7,5	3	3
6	7,5	3	3
7	9	4	5
		5	6
R_1	27	R_2	18

$$U_1 = 4 \cdot 5 + \frac{4(4+1)}{2} - 27 = 3$$

$$U_2 = 4 \cdot 5 + \frac{5(5+1)}{2} - 18 = 17$$

- Teststørrelsen kan nu bestemmes som $U = \min(3, 17) = 3$
- Kritiske værdier fra Tabel D idet $(n_1, n_2) = (4, 5)$:

α (enkelt)	Kritisk værdi
0,050	2
0,025	1
0,010	0
0,001	n/a

- Vi kan altså ikke afvise, at stikprøve 1 og 2 er trukket fra ens populationer

Mann-Whitney U for store stikprøver

- For n_1 eller n_2 større end 20 erstattes opslag af kritisk værdi i Tabel D med beregning af z-værdi, der sammenlignes med standardnormalfordelingen (Tabel A):

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Wilcoxons rangtest

- Svarer til t- og z-test for korrelerede data
- Vurderer sum af rangværdier for numeriske ændringer fordelt til stigninger $T_{(+)}$ og fald $T_{(-)}$
- For små stikprøver (højst 15 gyldige sammenligninger) benyttes Tabel E med kritiske værdier
- For større stikprøver benyttes en skaleret teststørrelse og opslag i normalfordeling

Wilcoxon — eksempel

Person	Før	Efter	d	$R(\lvert d \rvert)$	$R_{(+)}(\lvert d \rvert)$	$R_{(-)}(\lvert d \rvert)$
A	9	4	-5	4,5		4,5
B	8	2	-6	6,5		6,5
C	7	9	$+2$	1,5	1,5	
D	8	3	-5	4,5		4,5
E	10	3	-7	8		8
F	7	7	0	.		
G	6	10	$+4$	3	3	
H	9	1	-8	9		9
I	5	3	-2	1,5		1,5
J	8	2	-6	6,5		6,5
					$T_{(+)} = 4,5$	$T_{(-)} = 40,5$

Wilcoxon — eksempel *(fortsat)*

- I eksemplet er altså teststørrelsen

$$T = \min(T_{(+)}, T_{(-)}) = \min(4,5; 40,5) = 4,5$$

- Kritiske værdier er givet i Tabel E, og for $n = 9$ (vi udelod en observation med $d = 0$) findes følgende

α (énsidet)	0,050	0,010	0,001
T_{krit}	8	3	n/a

- Vi kan altså konkludere, at efter-scoren er signifikant lavere end før-scoren ($0,01 < p < 0,05$, $n = 9$)

Wilcoxon for store stikprøver

- For større stikprøver ($n > 15$) skaleres teststørrelsen til en z-værdi, der sammenlignes med standardnormalfordelingen (Tabel A):

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$