

SENIORKURSUS STATA OG BIOSTATISTIK

Aarhus Universitet juni 2011

DAGENS TEMA: SAMMENLIGNINGER

FORMIDDAG: KONTINUERTE DATA

EFTERMIDDAG: KATEGORISKE DATA

16. juni 2011

Michael Væth

1

STATISTISK ANALYSE AF TO UAFHÆNGIGE STIKPRØVER FRA NORMALFORDELTE POPULATIONER

Eksempel. Fiskeolie tilskud og blodtryk hos gravide kvinder

Gravide kvinder blev randomiseret til en af to behandlingsgrupper. I den ene gruppe modtog kvinderne et tilskud af fiskeolie. Den anden gruppe var en kontrolgruppe.

Her: En sammenligning af de to behandlinger med ændring i diastolisk blodtryk, *difdia*, som udfald (respons variabel).

Vi har allerede set histogrammer af ændringens fordeling i hver af de to grupper. Disse plots antyder at den tilfældige variation kan beskrives med normalfordelinger.

I går brugte vi dette eksempel til at illustrere tankegangen bag statistiske test. Her repetition med fokus på implementering i Stata

16. juni 2011

Michael Væth

2

Standard analysen af dette problem bygger på følgende **statistiske model**, som specificerer de antagelser som forudsættes opfyldte

I hver gruppe: En stikprøve af uafhængige observationer (her: en differens mellem to blodtryk) fra en normalfordeling med ukendte parametre (middelværdi og spredning):

Gruppe	Middelv.	Varians
Kontrol	μ_1	σ^2
Fiskeolie	μ_2	σ^2

De to sæt observationer er indbyrdes uafhængige.

Obs: Den tilfældige variation (variansen eller spredningen) antages at være den **samme** i de to grupper. Denne antagelses rimelighed skal (også) vurderes.

Analysens formål er at kvantificere forskellen mellem den forventede ændring af blodtrykket i de to grupper og vurdere om denne ændring er væsentligt forskellig fra 0.

16. juni 2011

Michael Væth

3

Modellens antagelser

- 1) **Uafhængighed** indenfor og mellem stikprøver
- 2) Stikprøver fra to populationer med **samme varians** (samme spredning)
- 3) Den tilfældige variation i hver gruppe kan beskrives med en **normalfordeling**

Bemærk: En normalfordeling er fuldstændigt karakteriseret ved middelværdi og varians, så hvis antagelserne er opfyldte og middelværdierne også er ens, kan vi konkludere at fordelingerne er ens

Kontrol af antagelser:

Ad 1) Overvejelserne afhænger af situationen.

Her: Er der kvinder som bidrager med flere graviditeter?

Ad 2) Ser estimatorne forskellige ud? Et formelt test af hypotesen om samme varians kan også beregnes (se nedenfor).

Ad 3) Se på histogrammer og Q-Q plots. Formelle test af normalitet har typisk ringe styrke.

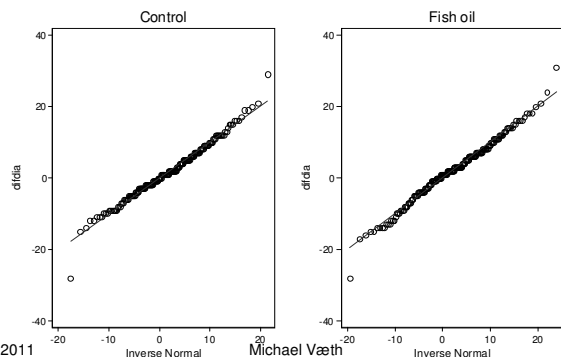
16. juni 2011

Michael Væth

4

Forudsætning 3: Stata

```
qnorm difdia if grp==1, title("Control")
    saving(q1,replace)
qnorm difdia if grp==2, title("Fish oil")
    saving(q2,replace)
graph combine q1.gph q2.gph
```



16. juni 2011

Michael Væth

5

Estimation

Basal idé: populationsværdierne (parametrene) estimeres af den tilsvarende stikprøvestørrelser. De to variansskøn samles til et fælles estimat (et vægtet gennemsnit af stikprøvernes varianser).

Stata udfører de basale beregninger med kommandoen
bysort grp: summarize difdia

-> grp = control					
variable	Obs	Mean	Std. Dev.	Min	Max
difdia	213	1.901408	7.528853	-28	29

-> grp = fish oil					
variable	Obs	Mean	Std. Dev.	Min	Max
difdia	217	2.193548	8.364904	-28	31

Kontrol: Middelværdi = 1.90

Fiskeolie: Middelværdi = 2.19

16. juni 2011

Michael Væth

6

Spredningerne er ret ens, så et fælles spredningsskøn (*pooled estimate*) synes naturligt. Hvordan beregnes det?

Statistisk teori: Det "bedste" fælles variansskøn beregnes som et vægtet gennemsnit af stikprøvernes varianser og det tilhørende spredningsskøn anvendes som fælles spredning.

Antal frihedsgrader ($f = n - 1$) anvendes som vægte ved beregning af de vægted gennemsnit:

$$s_{pooled}^2 = s_p^2 = \frac{f_1 s_1^2 + f_2 s_2^2}{f_1 + f_2} \quad \text{og} \quad s_{pooled} = s_p = \sqrt{s_p^2}$$

Stata's output fra t-test kommandoen indeholder ikke dette estimat, men det kan fås ved at skrive

```
quietly regress difdia grp
display e(rmse)
```

quietly foran en kommando medfører at output fra kommandoen ikke vises

Output: **7.9617662**, dvs. $s_p = 7.962$

tekst variablen group kan ikke bruges her

16. juni 2011

Michael Væth

7

Statistisk test:

Sammenligning af to uafhængige stikprøvers middelværdier

Den forventede ændring i diastolisk blodtryk er en smule højere i fiskeolie. **Er dette udtryk for en systematisk effekt?**

For at vurdere om forskellen kan forklares af den tilfældige variation testes hypotesen

$$H: \mu_1 = \mu_2$$

Dvs. den forventede værdi (middelværdien) er de samme i de to populationer.

Det sædvanlige ræsonnement: Antag at hypotesen er sand.

Den observerede forskel på -0.29 må da være et resultat af den tilfældige variation

Rimeligheden af denne forklaring vurderes ved at beregne en **p-værdi**, som er angiver sandsynligheden for at få et resultat som er mindst lige så ekstremt som det aktuelt observerede.

16. juni 2011

Michael Væth

8

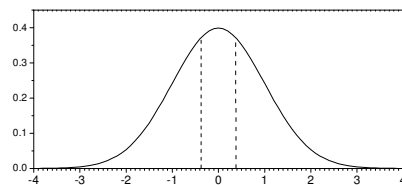
Statistisk test:

Sammenligning af to uafhængige stikprøvers middelværdier

I går: hypotesen kan testes ved at beregne

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -0.38$$

Teststørrelsen vurderes i en t-fordeling med $n_1 + n_2 - 2$ frihedsgrader. Det giver en p-værdi på 70%, så forskellen er ikke statistisk signifikant



16. juni 2011

Michael Væth

9

STATA: TWO SAMPLE t-TEST (ens varianser)

En Stata kommando giver alle resultater på én række

`ttest difdia , by(grp)`

s.d. i det samlede materiale
Ikke en pooled s.d.

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
control	213	1.901408	.5158685	7.528853	.8845197	2.918297
fish oil	217	2.193548	.5678467	8.364904	1.074318	3.312778
combined	430	2.048837	.3835675	7.953826	1.294932	2.802743
diff		-.2921399	.7679341		-1.801531	1.217252

Degrees of freedom: 428

Den testede hypotese

Ho: mean(control) - mean(fish oil) = diff = 0
Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
t = -0.3804 t = -0.3804 t = -0.3804
P < t = 0.3519 P > |t| = 0.7038 P > t = 0.6481

16. juni 2011

(to sider) p-værdi Michael Væth

10

Sammenligning af varianser: F-fordelingen

Analysen forudsatte samme varians (samme spredning) i de to populationer. For at evaluere denne antagelse kan vi teste hypotesen

$$H: \sigma_1^2 = \sigma_2^2$$

Et naturligt test er forholdet mellem de to variansskøn

$$F = \frac{s_1^2}{s_2^2}$$

Hvis hypotesen er sand vil man forvente en værdi tæt på 1. Både store og små værdier tyder på at varianserne er forskellige.

Statistisk teori: Fordelingen af to uafhængige variansestimater følger en F-fordeling if de tilsvarende populationsvarianser er ens (dvs hvis H er sand).

En F-fordeling er karakteriseret ved et par af frihedsgrader (frihedsgraderne for de to variansskøn). Ligesom for normalfordelingen, t-fordelingen og chi2-fordelingen findes der tabeller over F-fordelingen.

16. juni 2011

Michael Væth

11

STATA: SAMMENLIGNING AF TO VARIANSER

Stata's kommando `sdtest` kan bruges til at sammenligne to variansskøn.

`sdtest difdia , by(grp)`

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
control	213	1.901408	.5158685	7.528853	.8845197	2.918297
fish oil	217	2.193548	.5678467	8.364904	1.074318	3.312778
combined	430	2.048837	.3835675	7.953826	1.294932	2.802743

hypotesen

Ho: sd(control) = sd(fish oil)
F(212,216) observed = F_obs = 0.810
F(212,216) lower tail = F_L = F_obs = 0.810
F(212,216) upper tail = F_U = 1/F_obs = 1.234
Ha: sd(1) < sd(2) Ha: sd(1) != sd(2) Ha: sd(1) > sd(2)
P < F_obs = 0.0622 P < F_L + P > F_U = 0.1243 P > F_obs = 0.9378

16. juni 2011

(to sider) p-værdi Michael Væth

12

Sammenligning af middelværdier når varianserne er forskellige

Problem: Hvad skal vi gøre hvis varianserne ikke er ens?

Nogle muligheder:

1. Prøv at opnå varianshomogenitet ved at **transformere data** på passende vis, f.eks. ved at analysere de log-transformerede data.
2. Anvend et **approximativt t-test**, som ikke forudsætter ens varianser. Det approximative t-test har formen

$$t_{\text{approx}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Hvis hypotesen om ens middelværdier er sand vil fordelingen af denne teststørrelse kunne approksimeres med en passende t-fordeling. Antal frihedsgrader vil typisk være et ikke-heltal.

3. Use a **non-parametric test**, e.g. a Wilcoxon-Mann-Whitney test.

16. juni 2011

Michael Væth

13

STATA: TWO SAMPLE t-TEST (forskellige varianser)

Forslag 2: Stata beregner det approximative t-test med kommandoen

`ttest difdia , by(grp) unequal`

approximative
sikkerhedsgrenser

Group	Obs	Mea	Std. Err	Std. Dev.	[95% Conf. Interval]
control	213	1.901408	.5158685	7.528853	.8845197 2.918297
fish oil	217	2.193548	.5678467	8.364904	1.074318 3.312778
combined	430	2.048837	.3835675	7.953826	1.294932 2.802743
diff		-.2921399	.7671833		-1.800088 1.215808

Satterthwaite's degrees of freedom: 424.831 Antal frihedsgrader for det approximative t-test

Ho: mean(control) - mean(fish oil) = diff = 0
 Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 t = -0.3808 t = -0.3808 t = -0.3808
 P > |t| = 0.3518 P > |t| = 0.7035 P > t = 0.6482

16. juni 2011 (to sidedt) p-værdi

Michael Væth

Approximativt t-test

14

SAMMENLIGNING AF TO STIKPRØVER: IKKE-PARAMETRISKE METODER

Ikke-parametriske metoder eller **fordelingsfrie metoder**, er en gruppe af statistiske metoder som ikke forudsætter en bestemt parametriske fordeling i populationen

Fordele: Ikke-parametriske metoder bygger på færre og svagere forudsætninger og kan derfor anvendes i flere situationer.

Ulemper: Ikke-parametriske metoder er hovedsageligt statistiske test. Statistiske test kan derfor tillægges for stor betydning i forhold til andre dele af den statistiske analyse.

For en stor klasse af ikke-parametriske test vil teststørrelsen ikke afhænge af den observerede værdier, men kun af deres indbyrdes størrelsesorden (rangen). Disse metoder kan derfor også anvendes i situationer hvor udfaldet måles på en ordinalskala, f.eks. en komplikation som opgøres på en skala med trinene -, +, ++, or +++.

16. juni 2011

Michael Væth

15

TO UAFHÆNGIGE STIKPRØVER: WILCOXON-MANN-WHITNEY RANGSUM TEST

Data: To stikprøver med uafhængige observationer $N = n_1 + n_2$

Gruppe 1 $x_1, x_2, \dots, x_{n_1}$ fra en population med fordeling F_X

Gruppe 2 $y_1, y_2, \dots, y_{n_2}$ fra en population med fordeling F_Y

Hypotese: x 'erne og y 'erne er observationer fra den samme (kontinuerte) fordeling, dvs. $F_X = F_Y$. Alternative hypoteser af særlig interesse: y 'erne er forskudt mod større (eller mindre) værdier.

Teststørrelse (Wilcoxon's rangsum test)

T_1 = Sum of range i gruppe 1, eller

T_2 = Sum of range i gruppe 2

Et (to-sidet) test vil forkaste hypotesen for store eller små værdier af T_1 (eller T_2) **Obs:** De to teststørrelser er ækvivalente, da

$$T_2 = \frac{N(N+1)}{2} - T_1$$

16. juni 2011

Michael Væth

16

Wilcoxon's rangsum test (fortsat)

Hvis hypotesen er sand kan fordelingen af rangsummen bestemmes fuldstændigt vha. kombinatorik. Fordelingen er symmetrisk og der middelværdi og varians er givet ved

$$E(T_1) = n_1 \cdot (N+1)/2 \quad E(T_2) = n_2 \cdot (N+1)/2$$

$$Var(T_1) = Var(T_2) = n_1 \cdot n_2 \cdot (N+1)/12$$

Variationsformlen forudsætter at de observerede værdier alle er forskellige. Hvis dette ikke er tilfældet (**tied data**) bruges gennemsnitsrange (midranks) og rangsummens varians bliver mindre.

Rangsummens eksakte fordeling under hypotesen er tabelleret for små stikprøver. **Obs:** p-værdien bliver for stor hvis disse tabeller bruges med tied data.

P-værdien: Rangsummens fordeling approksimeres sædvanligvis ved at beregne en standardiseret teststørrelse, hvis fordeling approksimeres med en standard normalfordeling

$$z = \frac{T_1 - E(T_1)}{\sqrt{Var(T_1)}} \quad \text{eller} \quad z = \frac{|T_1 - E(T_1)| - 0.5}{\sqrt{Var(T_1)}}$$

16. juni 2011

Michael Væth

17

Rangsum test med Stata

Eksempel. Fiskeolie og diastolisk blodtryk hos gravide. I stedet for t-test kan vi sammenligne de to stikprøver med et Wilcoxon test. I Stata gøres det med

```
use fishoil.dta
ranksum difdia , by(grp)
```

Two-sample wilcoxon rank-sum (Mann-Whitney) test			
grp	obs	rank sum	expected
control	213	44953	45901.5
fish oil	217	47712	46763.5
combined	430	92665	92665

```
unadjusted variance 1660104.25
adjustment for ties -3237.25
```

```
adjusted variance 1656867.00
Ho: difdia(grp==control) = difdia(grp==fish oil)
z = -0.737
p = 0.4612
```

Stata beregner den standardiserede teststørrelse uden kontinuitetskorrektion

16. juni 2011

Michael Væth

18

TO STIKPRØVER MED PARREDE DATA

Eksempel: Måling af kropstemperatur

Stata filen **temp.dta** indeholder måling af 96 patienters kropstemperatur. Ved hjælp af Rektal Hg termometer, Oral Hg termometer og et elektronisk termometer (Craft-termometer).

```
list id hgrectal hgoral craft in 1/5
```

	id	hgrectal	hgoral	craft
1.	1	38.1	37.1	37.4
2.	2	38.1	37.8	37.8
3.	3	38.9	38.2	38.2
4.	4	38.4	38.2	38.1
5.	5	40.3	40	40.1

Data stammer fra en undersøgelse som blev iværksat for at evaluere det elektroniske termometer.

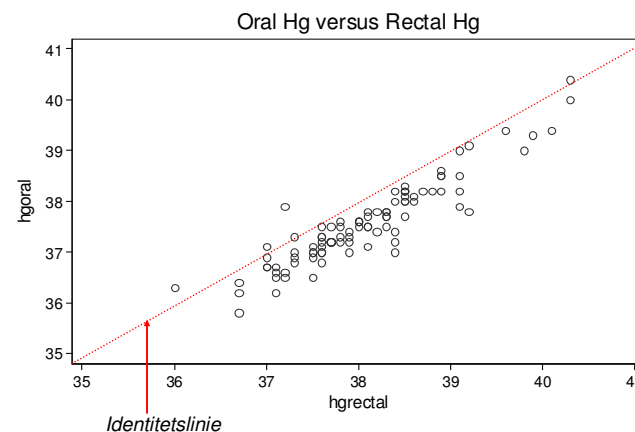
Her ser vi kun på de to sæt målinger med Hg termometre.

16. juni 2011

Michael Væth

19

Et plot af 94 par af målinger af oral og rektal temperatur (for to patienter manglede den rektale måling)



16. juni 2011

Michael Væth

20

Spørgsmål:

- *Hvad er forskellen mellem oral og rektal kropstemperatur?*
- *Hvor meget varierer forskellen fra patient til patient?*

Situationen ligner et to-stikprøve problem, **men** her er det de **samme** patienter som er målt ved begge metoder. Vi har ikke to uafhængige stikprøver, men **to parrede stikprøver**.

Den "oplagte" analyse

1. Beregn **differensen** mellem de to målinger for hver patient
2. Undersøg om differenserne kan betragtes som en stikprøve fra en normalfordeling og i givet fald beregn middeldifferens, differensernes spredning, sikkerhedsinterval for middelforskel, og et en-stikprøve t -test af hypotesen om at den forventede forskel er 0.

Hvis antagelse om normalfordelte differenser er urimelig, brug i stedet et ikke-parametrisk test.

16. juni 2011

Michael Væth

21

Bemærk:

Data kan bruges til at identificere **systematiske forskelle** mellem de to temperaturer, men gentagne målinger med samme metode er nødvendige hvis man vil udtale sig om metodernes **usikkerhed**.

Hvilke modelovervejelser leder til denne analyse?

For en given patient, lad

X = en rektal Hg temperatur måling

Y = en oral Hg temperature måling

$D = X - Y$ = forskellen mellem de to målinger (rektal – oral)

En temperatur måling kan betragtes som en sum af to komponenter

$X = A + E$

$Y = B + F$

16. juni 2011

Michael Væth

22

Her er

A = den "sande" rektale temperatur, dvs. den hypotetiske værdi vi ville få som gennemsnit, hvis "eksperimentet" blev gentaget mange gange.

B = den "sande" orale temperatur

E = **den tilfældige målefejl** for en rektal temperatur måling med et Hg termometer.

F = **den tilfældige målefejl** for en oral temperatur måling med et Hg termometer.

Differensen D kan derfor skrives som

$$D = (A - B) + (E - F)$$

Det **første led**, $A - B$, repræsenterer forskellen mellem den sande rektale og den sande orale temperatur for en given patient på en given dag.

Det **andet led**, $E - F$, er den samlede effekt af den tilfældige målefejl.

16. juni 2011

Michael Væth

23

I en patient population vil forskellen mellem de sande værdier, $A - B$, Variere mellem personer og også mellem dage for den samme person. Hvis δ angiver populationens middelforskel, haves

$$A - B = \delta + C,$$

hvor C repræsenterer **inter- og intra-individual variation** for forskellen mellem de sande temperaturer

Vi har derfor følgende dekomposition af forskellen mellem de to målinger

$$D = X - Y = \delta + (C + E - F) = \delta + U,$$

hvor δ er populationens middelværdi og U beskriver den tilfældige variation, som har tre komponenter, inter-individual variation af den sande differens, tilfældig rektal målefejl og tilfældig oral målefejl.

Den statistiske model for den "oplagte analyse" beskriver differenserne som en stikprøve fra en normalfordeling med middelværdi δ og varians σ^2

Obs: En model for differenserne.

Ingen forudsætninger om fordelingen af A og B i populationen

16. juni 2011

Michael Væth

24

Undersøge om antagelserne er rimelige:

- **Uafhængighed:** Hver patient bidrager med én differens, så uafhængighed synes rimelig.
- **Samme fordeling:**
 - Den **forventede forskel** mellem to målinger afhænger ikke af temperaturniveauet. Febrile patienter har ikke større forskel end patienter med normal temperatur. *Er det en rimelig antagelse?*
 - Den **tilfældige variation** af differenserne har samme størrelse for alle temperaturniveauer. *Er det en rimelig antagelse?*
- **Normal fordeling:** *Er det en rimelig antagelse?*

16. juni 2011

Michael Væth

25

Vigtige plots som man altid bør lave:

- Et plot af Y versus X
- Differensen mod gennemsnittet (eller mod summen)
- Histogram af differenserne
- Q-Q plot af differenserne

Betyder at kommandoen fortsætter på næste linie

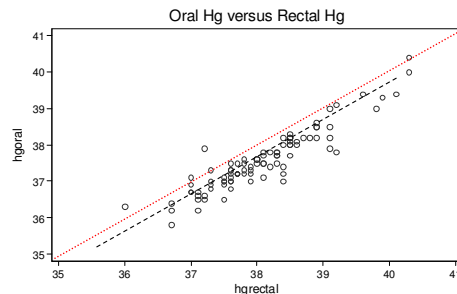
- ```
scatter hgoral hgrectal ,
 title("Oral Hg versus Rectal Hg")
 ylabel(35(1)41) xlabel(35(1)41)
 ///
```
- ```
gene difro = hgrectal-hgoral
gene hgmean=(hgrectal+hgoral)/2
scatter difro hgmean ,
      title("difference versus average")
      ylabel(-0.8(0.4)1.6) xlabel(37(1)41)
      ///
```
- ```
histogram difro ,
 title("Rectal Hg - Oral Hg")
 start(-0.8) width(0.2)
 ///
```
- ```
qnorm difro , title("Q-Q plot of difference")
      ///
```

16. juni 2011

Michael Væth

26

Plot a: Oral måling versus rektal måling



Kommentarer

Hvis der ikke er systematisk forskel på de to målinger vil punkterne fordele sig tilfældig omkring identitetslinien (den røde linie).

Den rektale temperatur er systematisk højere end den orale temperatur, men den systematiske forskel synes ikke at afhænge af temperaturrens størrelse.

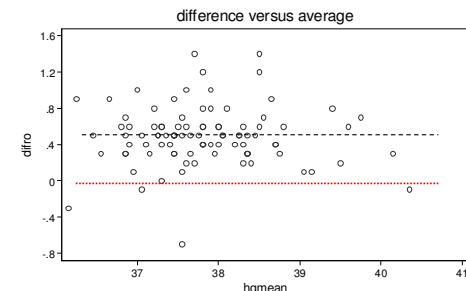
Punkternes variation omkring en linie med hældning 1 (den sorte linie) synes ikke at afhænge af temperaturrens størrelse.

16. juni 2011

Michael Væth

27

Plot b: Differens versus gennemsnit. Kaldes ofte et Bland-Altman plot



Kommentarer

Hvis der ikke er systematisk forskel på de to målinger vil punkterne fordele sig tilfældig omkring en vandret linie gennem 0 (den røde linie).

Den rektale temperatur er systematisk højere end den orale temperatur, men at den systematiske forskel synes ikke at afhænge af temperaturrens størrelse

Punkterne varierer omkring en vandret linie gennem (ca.) 0.5 (den sorte linie) og variationen er ret konstant omkring denne linie.

16. juni 2011

Michael Væth

28

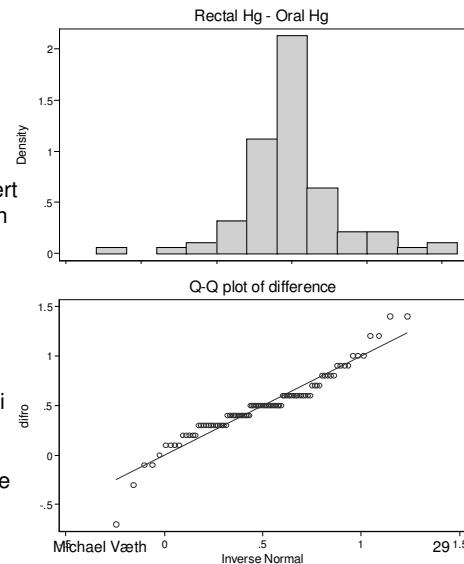
Plot c and d:
histogrammet og Q-Q
plottet for differenserne.

Kommentarer

Disse plots bruges primært
til at vurdere rimeligheden
af den normale fordeling

Her: Fordelingen er
nogenlunde symmetrisk
med beskedne afvigelser
fra normalfordelingen.
Det lidt S-formede forløb i
Q-Q plottet tyder på at
begge haler er for lange.
Det kunne tyde på enkelte
meget store afvigelser.

16. juni 2011



Analyse med Stata:

Et parret to-satiprøve problem kan analyseres direkte med kommandoen
ttest, dvs. uden først at genere en variabel med differenserne

ttest hgrectal=hgoral

== er også tilladt

variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
hgrectal	94	38.06808	.0852441	.8264723	37.89881	38.23736
hgoral	94	37.57447	.0855485	.8294235	37.40459	37.74435
diff	94	.4936168	.0330279	.3202179	.4280298	.5592038

gennemsnits differens

95% sikkerhedsgrænser for den forventede forskel

Ho: mean(hgrectal - hgoral) = mean(diff) = 0
 Ha: mean(diff) < 0 Ha: mean(diff) != 0 Ha: mean(diff) > 0
 t = 14.9454 t = 14.9454 t = 14.9454
 P > |t| = 1.0000 P > |t| = 0.000 P > t = 0.0000

p-værdi for hypotesen at den forventede forskel er 0

16. juni 2011

Michael Væth

30

Analyse (fortsat)

Fra outputtet fås følgende estimater

$$\hat{\delta} = 0.49 \quad (95\% \text{ sikkerhedsinterval: } 0.43 \quad 0.56)$$

$$\hat{\sigma} = 0.32 \quad (\text{i.e. } \hat{\sigma}^2 = 0.32^2 = 0.102)$$

95% sikkerhedsgrænser for σ kan også beregnes:

$$\sigma_L = 0.28 \quad \sigma_U = 0.37$$

Konklusion: Rektal Hg temperatur er i gennemsnit 0.5 grader højere
end oral Hg temperatur. Den tilfældige variation (målefejl, inter- og
intra-individuel) af differensen har en standard afvigelse på 0.32

Obs: Sikkerhedsintervallet for den forventede forskel beskriver ikke
overensstemmelsen mellem de to metoder. δ

Limits of agreement defineres som $\hat{\delta} \pm 2\hat{\sigma}$ (dvs. gns. diff. ± 2 s.d.).
Dette er et approx. 95% prædiktionsinterval. Grænserne er temmeligt
brede: -0.15 og 1.13. Det afspejler differensernes relative store variation

16. juni 2011

Michael Væth

31

PARREDE DATA:

ABSOLUTE ELLER RELATIVE ÆNDRING (FORSKEL)?

Data: en stikprøve på n par af observationer (x, y) .

Sådanne data optræder i mange situationer:

- Sammenligning af to målemetoder
- Sammenligning af en før-behandling måling og en efter-behandling måling
- Studier af inter-observer eller intra-observer variation

Problem: Adskiller x 'erne sig fra y 'erne? Hvis ja, hvordan?

Første idé: Se på differenser (den "oplagte" analyse ovenfor)

MEN skal vi se på absolut eller relativ forskel (ændring)?

$$\text{Absolut ændring} = y - x$$

$$\text{Relativ ændring} = \frac{y - x}{x} = \frac{y}{x} - 1$$

obs: den relative ændring
er blot forholdet "korrigeret"
så 0 svarer til ingen ændring

16. juni 2011

Michael Væth

32

Absolut eller relativ ændring?

Valget bør afspejle den systematiske variations struktur (en empirisk begrundelse).

Sammenhængen mellem de to sæt af målinger kan begrunde valget

To simple relationer mellem y og x

1. $y \approx x + a$
2. $y \approx x \cdot b$

I **situation 1** vil differenserne være relativt stabile.

Hvis variationen i data passer med denne sammenhæng vil en **absolut ændring** være det rigtige valg (fra et statistisk perspektiv).

I **situation 2** vil forholdene være relativt stabile.

Hvis variationen i data passer med denne sammenhæng vil en **relativ ændring** være det rigtige valg (fra et statistisk perspektiv).

16. juni 2011

Michael Væth

33

Det **parrede t-test** og det analoge ikke-parametriske er primært beregnet for **situation 1**, da disse test er baseret på differenserne.

Situation 2 klares sædvanligvis bedst ved at tage logaritmer:

$$y \approx x \cdot b \quad \text{betyder at} \quad \log(y) \approx \log(x) + \log(b),$$

så relationen mellem $\log(x)$ and $\log(y)$ er additiv (situation 1) med $a = \log(b)$

Opsummering: Hvis variationen i en serie af par (x, y) er **multiplikativ** (situation 2) vil den bedste løsning typisk være at beregne differenserne

$$\log(y) - \log(x) = \log\left(\frac{y}{x}\right)$$

og basere den statistiske analyse på disse differenser.

Dette svarer til at arbejde med forhold på en log-skala.

16. juni 2011

Michael Væth

34

Parede data: Differens af log-værdier versus relativ ændring

Hvorfor bruger man ikke den relative ændring i situation 2?

Eksempele

Patient	x	y	(y-x)/x	y/x	$\log(y) - \log(x)$
1	100	200	100%	2.00	0.693
2	200	100	-50%	0.50	-0.693
Gns.	150	150	25%	1.25	0

Patient 2 er blot en "omvendt version" af patient 1, så samlet set sker der ingenting.

- Gennemsnit af første måling (x) er magen til gennemsnit af 2. måling (y).
- Den relative ændring er i gennemsnit 25%: en svag stigning.
- Gennemsnittet af forholdene er større end 1: en svag stigning.
- Gennemsnittet af differenser af log-værdier = forholdet målt på en log-skala – er 0 svarende til ingen ændring: Rimeligt.

16. juni 2011

Michael Væth

35

Parede data: Differens af log-værdier versus relativ ændring

Det er bekvemt at se på forhold på en log-skala, da f.eks.

$$\log(y/x) = -\log(x/y)$$

The En relativ ændring har ikke denne egenskab

$$\frac{y-x}{x} \neq -\frac{x-y}{y}$$

Når forholdet y/x er tæt på 1 vil ($\log$ = naturlig logaritme)

$$\log(y) - \log(x) = \log\left\{\frac{y}{x}\right\} \approx \frac{y}{x} - 1 = \frac{y-x}{x}$$

Når forholdet er **tæt på 1** (e.g. within $\pm 20\%$) vil forskellen mellem de **naturlige logaritme-værdier** være cirka lig med den **relative ændring**.

Parrede data med en multiplikativ struktur (situation 2) vil typisk også have større variation for store værdier. Analyseres data på en log-skala vil derfor **stabilisere variationen** og kan betyde en simplere beskrivelse af den tilfældige variation (en variationskoefficient).

16. juni 2011

Michael Væth

36

Bemærk

Valget mellem metoder for normalfordelte data og ikke-parametriske metoder betragtes ofte som det vigtigste problem

MEN valget af et passende *mål for ændring* er ofte meget vigtigere end valget af teststørrelse.

Derfor: Lav altid *diagnostiske plots!* Bemærk dog at af og til er der ikke et klart svar, fordi

1. Hverken den oprindelige skala eller log-skalaen er fornuftig
2. Hvis variationen mellem parrene er beskeden, vil det være svært at skelne mellem en additiv struktur og en multiplikativ struktur i data. I sådanne situationer er resultaterne typisk meget ens for utransformerede og log-transformerede data.
3. Den transformation som giver additivitet er ikke den transformation som giver konstant variation.
4. Det hele er bare noget rod!

16. juni 2011

Michael Væth

37

TO PARREDE STIKPRØVER: WILCOXON'S SIGNED RANK TEST

Basalt problem: Analyse af parrede data uden at forudsætte normalfordelt variation.

Data: En stikprøve af n observationspar $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

Spørgsmål: Er x 'ernes fordeling forskellig fra y 'ernes fordeling?

Preliminære model overvejelser:

For et observationspar kan vi skrive

$$\begin{aligned}x &= \alpha + e_1 \\ y &= \alpha + \delta + e_2\end{aligned}$$

hvor α og $\alpha + \delta$ repræsenterer den forventede værdi af x og y , og hvor e_1 og e_2 er tilfældige fejl.

Antag: De tilfældige fejl er uafhængige mellem par og følger den samme fordeling

16. juni 2011

Michael Væth

38

Hvis fejllenede e_1 og e_2 følger den samme fordeling vil differensen

$$d = y - x$$

være **symmetrisk fordelt** med median (og middelværdi) δ .

Statistisk model: De n differenser $d_1, d_2, \dots, d_n$ betragtes som en tilfældig stikprøve fra en **symmetrisk** fordeling F med median δ .

Estimation: Populationens median estimeres med stikprøvens median

Hypotese: x 'erne og y 'erne har samme fordeling, dvs. $\delta = 0$.

Fortegntest (sign test)

Hvis medianen er 0, burde positive og negative differenser være lige hyppige, og antallet af positive differenser følger derfor en binomialfordeling med $p = 0.5$ (sml. møntkast). Hvis enkelte differenser er 0 reduceres stikprøvens størrelse med antallet af 0'er.

Stata: `signtest hgoral=craft`

16. juni 2011

Michael Væth

39

Wilcoxon's signed rank sum test

Fortegntestet ser kun på differensernes fortegn. Wilcoxon's signed rank test udnytter både differensernes fortegn og størrelse. Det giver et stærkere test.

Basal idé: Sortér differenserne efter størrelsen af de absolutte differenser (dvs. fortegnet ignoreres). Brug **summen af de positive differensers range** som teststørrelse.

Wilcoxon's signed rank-sum test

T_+ = sum af positive differensers range. Når differensernes ordnes efter størrelsen af den numeriske værdi.

Alternativt: brug T_- , som defineres på tilsvarende vis. De to teststørrelser er ækvivalente.

Hvis der ikke er ens differenser (*no ties*) og ingen 0'er, vil teststørrelsen have en symmetrisk fordeling og

$$0 \leq T_+ \leq n \cdot (n+1) / 2$$

$$E(T_+) = n \cdot (n+1) / 4 \quad \text{Var}(T_+) = n \cdot (n+1) \cdot (2n+1) / 24$$

16. juni 2011

Michael Væth

40

“Ties” og 0’er blandt differenserne

Gennemsnitrange (mid ranks) bruges hvis nogle differenser har samme værdi.

Differenser, som er lig 0 inkluderes ikke i teststørrelsen.

Stata bruger en variansformel, som er korrigeret for ties og 0’er. Nuller plejer man at tage højde for ved at ekskludere disse differenser og reducere stikprøvens størrelse tilsvarendeata.

P-værdien: Den eksakte fordeling er tabelleret for små stikprøver, men vil være konservative, hvis der er ties. Sædvanligvis bruges en approksimation med en normalfordeling for den standardiserede teststørrelse. Approksimationen er god selv hvis n er 10.

$$z = \frac{T_+ - E(T_+)}{\sqrt{\text{Var}(T_+)}} \quad \text{eller} \quad z = \frac{|T_+ - E(T_+)| - 0.5}{\sqrt{\text{Var}(T_+)}}$$

Stata beregner denne teststørrelse hvor variansen er korrigeret for ties og 0’er..

16. juni 2011

Michael Væth

41

Wilcoxon’s signed rank-sum test with Stata

Eksempel: Måling af kropstemperatur. De to temperaturmålinger kan sammenlignes med et Wilcoxon signed rank test.

`signrank hgrectal=hgoral`

Wilcoxon signed-rank test

sign	obs	sum ranks	expected
positive	89	4358.5	2232
negative	4	105.5	2232
zero	1	1	1
all	94	4465	4465

unadjusted variance 70323.75
adjustment for ties -247.75
adjustment for zeros -0.25
adjusted variance 70075.75

Ho: hgrectal = hgoral

z = 8.033

Prob > |z| = 0.0000

16. juni 2011

Michael Væth

42