Peter Stoltze
6. marts 2009

Data Denne note gennemgår z-testet for ukorrelerede data med det eksempel, der er givet på siderne 47–50 i lærebogen. Data (opdigtede læsescores for 30 piger og 30 drenge i tredje klasse) er gengivet i følgende tabel:

Køn	Score
drenge	55 46 32 54 60 77 49 43 64 32
	24 44 59 79 89 19 21 48 57 52
	35 68 55 88 69 38 49 79 24 48
piger	22 63 40 78 64 55 45 88 33 53
	54 69 74 62 51 58 71 66 68 41
	67 51 38 42 63 71 84 95 55 71

Vi har allerede set hvordan vi kan summere data ved at beregne stikprøvegennemsnit og stikprøvestandardafvigelse for begge køn. Disse summariske mål for stikprøven er gengivet nedenfor:

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

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:

$$H_0: \quad \mu_{piger} = \mu_{drenge}$$
$$H_1: \quad \mu_{piger} > \mu_{drenge}$$

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

Teststørrelse Nu skal vi beregne en teststørrelse z, der består af forskellen mellem de to middelværdier normeret med spredningen på denne. Forskellen beregnes let til

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

Variansen på denne forskel findes som summen af de to varianser på middelværdierne, hvilket er en konsekvens af to generelle sætninger om variansberegninger:

$$\mathrm{var}(aX) = a^2 \cdot \mathrm{var}(X) \quad \text{og} \quad \mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) \ ,$$

hvor $a \in \mathbb{R}$ og X og Y er uafhængige i betydningen $\mathrm{cov}(X,Y) = 0$. I vores tilfælde har vi

$$\mathrm{var}(\bar{x}_1 - \bar{x}_2) = \mathrm{var}(\bar{x}_1) + (-1)^2\mathrm{var}(\bar{x}_2) = \mathrm{var}(\bar{x}_1) + \mathrm{var}(\bar{x}_2) \ ,$$

og kan altså beregne spredningen på forskellen i middelværdierne som

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

Teststørrelsen z kan nu beregnes som

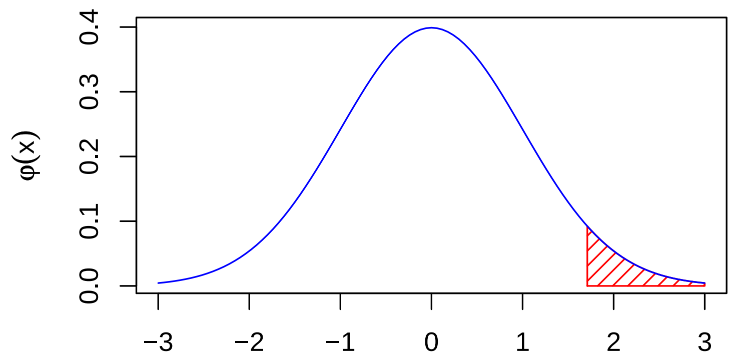
$$z = \frac{\bar{x}_1-\bar{x}_2}{s_{\bar{x}_1-\bar{x}_2}} = \frac{7,93}{4,636} = 1,711 \ .$$

Teststørrelsen følger under H_0 en standardnormalfordeling, $Z \sim N(0,1)$, idet de to middelværdier antages ens og derfor må forskellen være nul. Forskellen er desuden normeret med dens spredning så vi får enhedsvarians.

Signifikanssandsynlighed Nu skal vi omsætte teststørrelsen z til en signifikanssandsynlighed p, idet vi tager hensyn til vores sæt af hypoteser. Vores alternative hypotese er $\mu_{piger} > \mu_{drenge}$, og derfor skal vi finde sandsynligheden for en teststørrelse på 1,711 eller mere i standardnormalfordelingen. Denne beregnes som

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

Denne sandsynlighed kan også illustreres grafisk ved hjælp af det skraverede areal i følgende figur:



Konklusion På baggrund af signifikanssandsynligheden $p = 0,044$ kan vi tillade os 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.

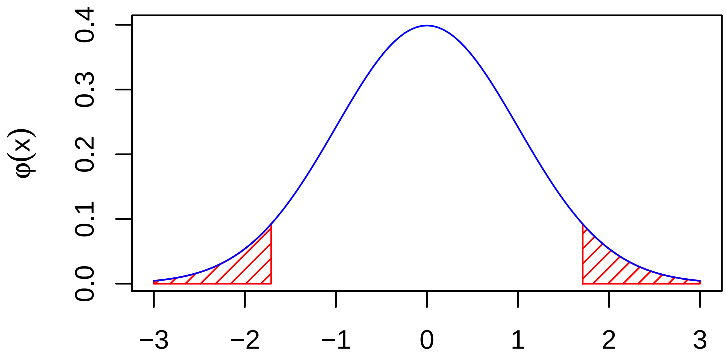
Andre alternativhypoteser Vi kan definere andre alternativhypoteser:

$$\begin{aligned} H_{1a}: &\quad \mu_{piger} \neq \mu_{drenge} \\ H_{1b}: &\quad \mu_{piger} < \mu_{drenge} \end{aligned}$$

Beregning af signifikanssandsynlighed bliver i det første tilfælde

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

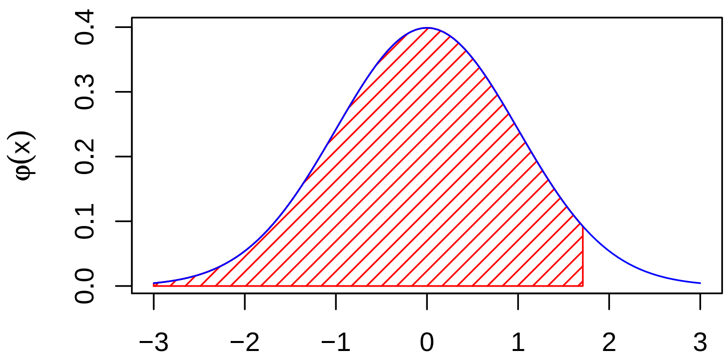
hvilket er det dobbelte af den oprindelige p. Betydningen er, at det bliver sværere at afvise H_0, hvilket typisk er det, vi er interesserede i. Sandsynligheden kan også illustreres grafisk ved hjælp af det skraverede areal i følgende figur:



Beregning af signifikanssandsynlighed bliver i det andet tilfælde

$$p_{1b} = \mathrm{P}(Z < z) = \Phi(z) = 0,956$$

hvilket er meget højt. Dette skyldes at nulhypotesen er meget attraktiv sammenlignet med alternativet, der jo er i direkte modstrid med vores observationer. Sandsynligheden kan illustreres grafisk ved hjælp af det skraverede areal i følgende figur:



Rækkefølge Havde vi skrevet hypoteserne med kønnene i den modsatte rækkefølge (hvilket naturligvis ville være uhøfligt), men med same retning, så bliver regningerne lidt anderledes med resultaterne de samme. Vi illustrerer det ved at skrive hypoteserne som følger:

$$H_0: \quad \mu_{drenge} = \mu_{piger}$$
$$H_1: \quad \mu_{drenge} < \mu_{piger}$$

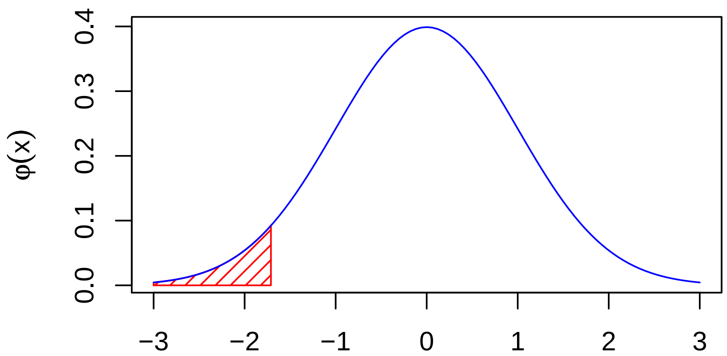
Teststørrelsen z beregnes nu som

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

Altså numerisk set samme størrelse som tidligere, men med modsat fortegn. Signifikanssandsynligheden findes som ved ovenstående H_{1b}, altså

$$p = \mathrm{P}(Z < z) = \Phi(z) = \Phi(-1,711) = 0,044 \ ,$$

hvilket selvfølgelig er det samme som med den modsatte rækkefølge. Grafisk findes sandsynligheden som vist i nedenstående figur.



Kritiske værdier Har vi ikke en avanceret regnemaskine eller et meget detaljeret tabelværk til rådighed, så kan vi i stedet benytte os af kritiske værdier. Kritiske værdier er fraktiler i den fordeling, vi bruger som model for teststørrelsen. I tabellen nedenfor er vist de kritiske værdier for et énsidet test med 5 forskellige signifikansniveauer α.

$\alpha_{ensidet}$	0.100	0.050	0.025	0.010	0.005
$\Phi^{-1}(1-\alpha)$	1.282	1.645	1.960	2.326	2.576
$t^{-1}_{1-\alpha;60}$	1.296	1.671	2.000	2.390	2.660
$t^{-1}_{1-\alpha;58}$	1.296	1.672	2.002	2.392	2.663

Med normalfordelingen som model for teststørrelsen (altså et z-test) kan vi aflæse, at signifikanssandsynligheden p er mindre end 5% men større end 2,5% da $1,645 < 1,711 < 1,960$. Vi får det samme interval for p hvis vi benytter en t-fordeling med 58 frihedsgrader $(30 + 30 - 2 = 58)$, og altså laver et t-test. I Tabel C (lærebogen side 122) kan vi kun finde en t-fordeling med 60 frihedsgrader, men vi kan se at konklusionen bliver den samme.

Med et direkte opslag af p (f.eks. med funktionerne `normfordeling()` og `tfordeling()` i Excel) kan vi se en lille forskel idet z-testet giver $p = 0,0435$ mens t-testet med 58 frihedsgrader giver $p = 0,0462$.