

TEST DI STUDENT

Vedremo per prima la situazione in cui si deve confrontare due medie di popolazioni gaussiane (test sulle differenze di medie). Abbiamo

$$X_1, \dots, X_{n_1} \text{ i.i.d. } \sim N(\mu_X, \sigma_X^2)$$

$$Y_1, \dots, Y_{n_2} \text{ i.i.d. } \sim N(\mu_Y, \sigma_Y^2)$$

supponiamo anche che $(X_i)_i$ e $(Y_i)_i$ siano indipendenti fra loro.

Modello statistico: $(P_\theta)_{\theta \in \Theta}$ t.c., sotto P_θ , i due campioni obbediscono alla distribuzione sopra, con $\theta = (\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2) \in \Theta = \mathbb{R}^2 \times (0, +\infty)^2$

Vogliamo testare un'ipotesi del tipo $\mu_X = \mu_Y$, contro un'ipotesi alternativa $\mu_X \neq \mu_Y$. Questo si scrive

$$\text{ipotesi: } \Theta_0 = \left\{ \theta = (\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2) \mid \mu_X = \mu_Y \right\} \in \Theta$$

$$\text{alternativa } \Theta_1 = \left\{ \theta \in \Theta \mid \mu_X \neq \mu_Y \right\}$$

o più semplicemente

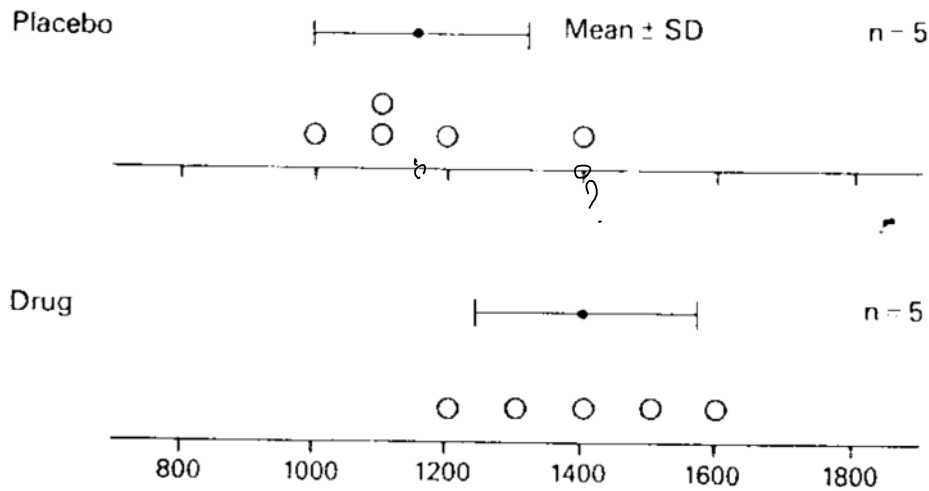
$$\text{ipotesi: } H_0: \mu_X = \mu_Y$$

$$\text{alternativa: } H_1: \mu_X \neq \mu_Y$$

idea: confrontare gli stime $\bar{X}$ e $\bar{Y}$

idea: confrontare gli statistiche $\bar{X}$ e $\bar{Y}$

esempio: test su un diuretico



idea: confrontare le medie facendone la differenza:

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}\right)$$

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{n_1}\right), \bar{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n_2}\right) \text{ ind.}$$

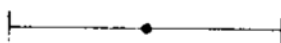
ipotesi semplificatrice: $\sigma_X^2 = \sigma_Y^2 (= \sigma^2)$

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

stima la ~~varianza~~ errore standard delle diff. di medie:
(dev. st.)

$$S_{\bar{X} - \bar{Y}} := \sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}} \quad \left(\neq \frac{S_X}{\sqrt{n_1}} + \frac{S_Y}{\sqrt{n_2}} \right)$$

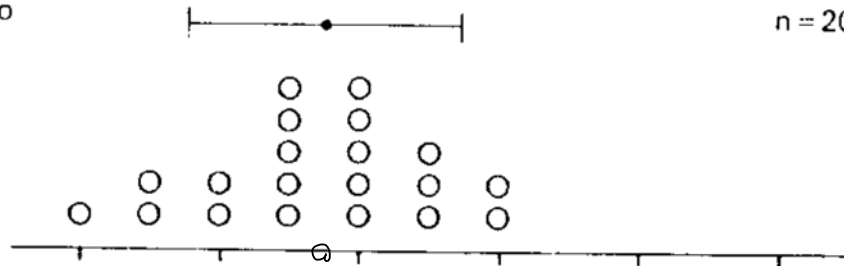
Placebo



n = 20

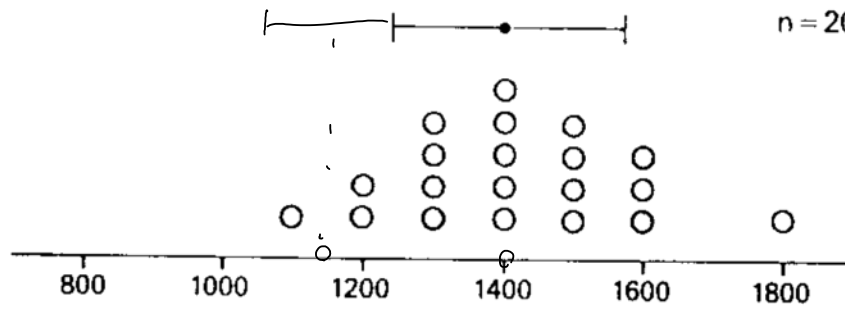
Placebo

n = 20



Drug

n = 20



Una volta calcolate $\bar{X} - \bar{Y}$ e $s_{\bar{X} - \bar{Y}}$, confrontiamole:

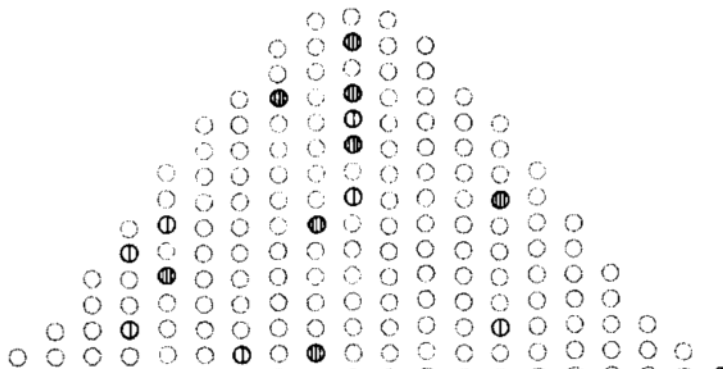
$$t = \frac{\bar{X} - \bar{Y}}{s_{\bar{X} - \bar{Y}}} \quad \underline{\underline{t \text{ di Student}}}$$

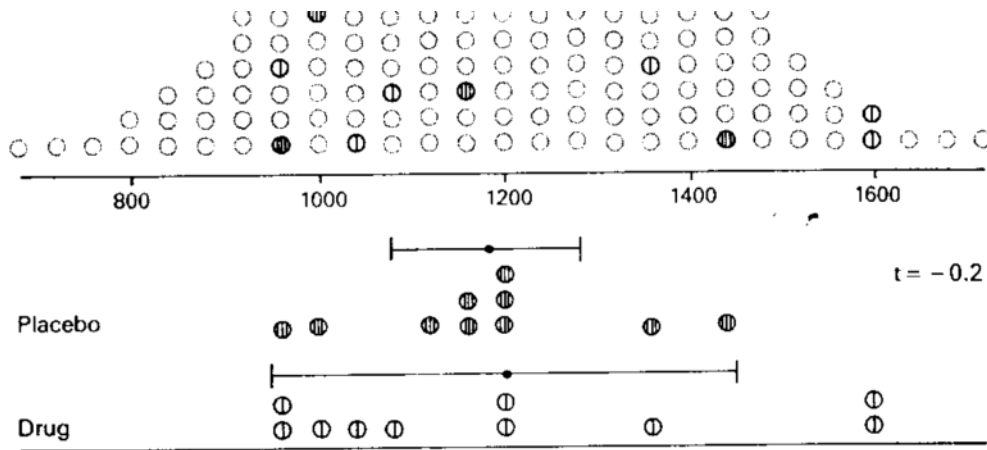
se $|t|$ è "piccola", riteniamo che $\mu_X = \mu_Y$

se $|t|$ è "grande", riteniamo che $\mu_X \neq \mu_Y$

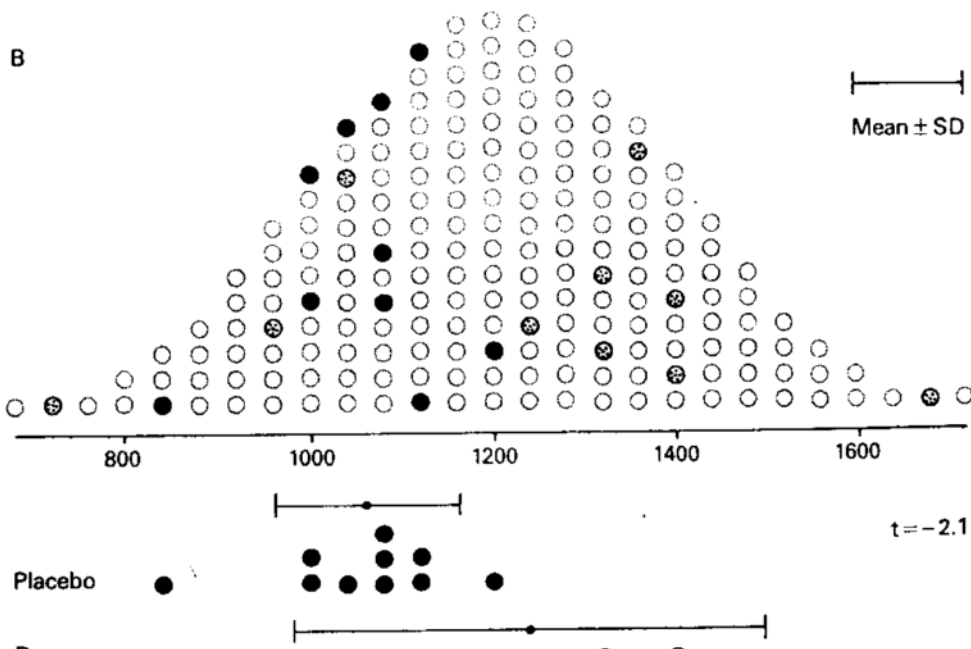
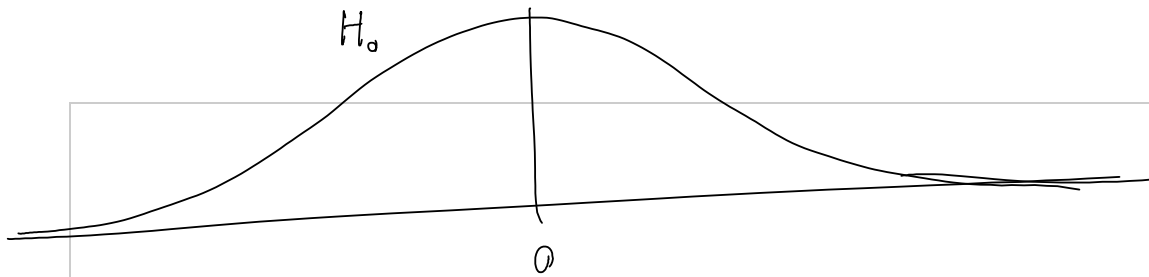
che legge ha t ?

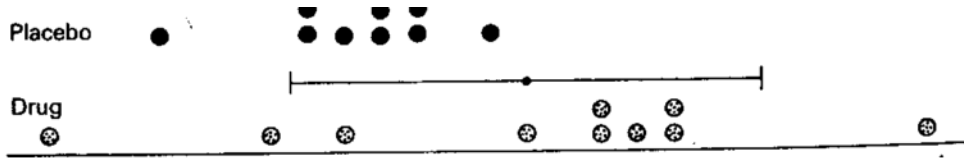
A





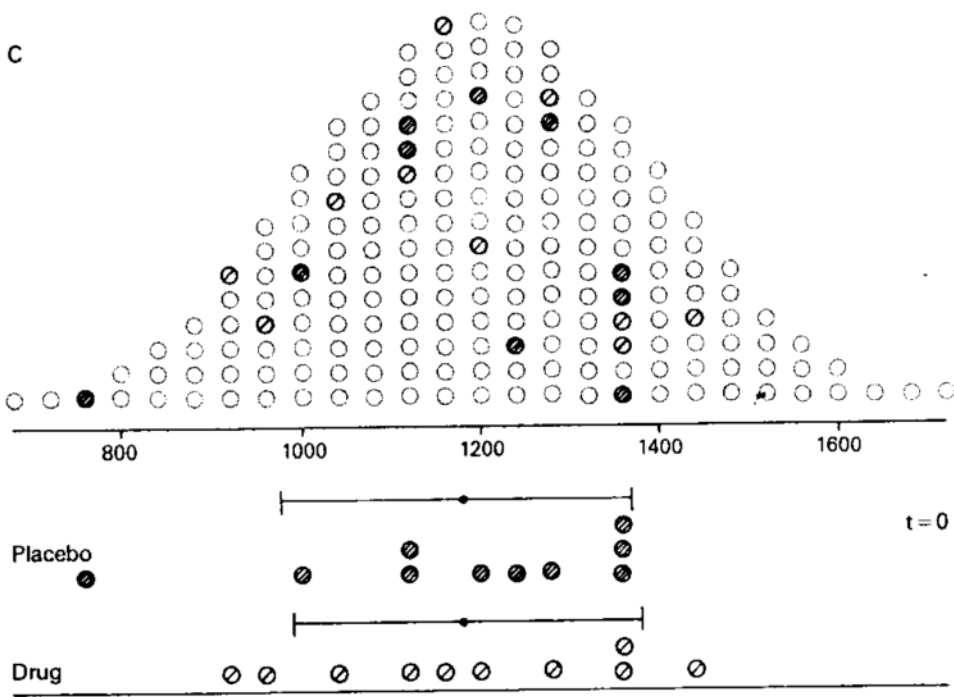
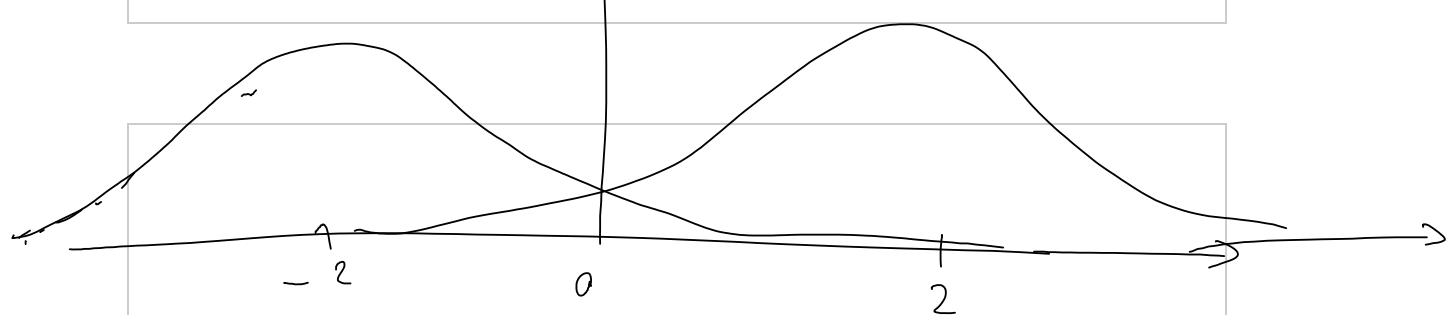
Y₀ e H₀ (cioè quando H₀: $\mu_X = \mu_Y$), $\bar{X} - \bar{Y} \sim N(0, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$
 quindi $t = \frac{\bar{X} - \bar{Y}}{s_{\bar{X} - \bar{Y}}}$ è centrata





sotto H_1 , $\mu_X > \mu_Y$ oppure $\mu_X < \mu_Y$, quindi t non è centrata (centrata: $E[t] = 0$)

t sotto H_2



regola: se $|t| \notin$ regione critica (code, prob. $\frac{\alpha}{2}$ e $\frac{\alpha}{2}$)

regola : se $|t| > t_{1-\frac{\alpha}{2}}(v)$ rigetto H_0 (altrimenti accetto H_0)

allora accettiamo $H_0: \mu_X = \mu_Y$

se $|t| \in$ regione critica, rifiuto H_0 e

accettiamo $H_1: \mu_X \neq \mu_Y$

regione critica: $\{|t| > t_{1-\frac{\alpha}{2}}(v)\}$

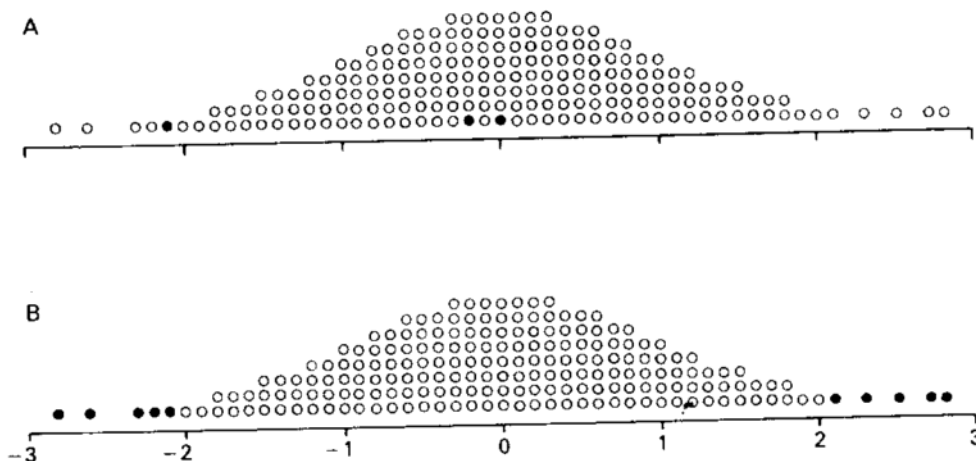
$t_{1-\frac{\alpha}{2}}(v) =$ valore critico

$\alpha =$ livello del test (piccolo, $\alpha = 5\%$, 1%)

$v =$ gradi di libertà delle t di Student

in questo caso $v = n_1 + n_2 - 2$

$\alpha = 5\%$



La regione critica è individuata dalla proprietà

$$P_0\{|t| > \alpha\} = \alpha \quad \text{con } \alpha = \text{valore critico}$$

(dove P_0 è una probabilità che rende H_0 verificata, cioè

(dove P_0 è una probabilità che rende H_0 verificata, cioè $\mu_x = \mu_y$ sotto P_0)

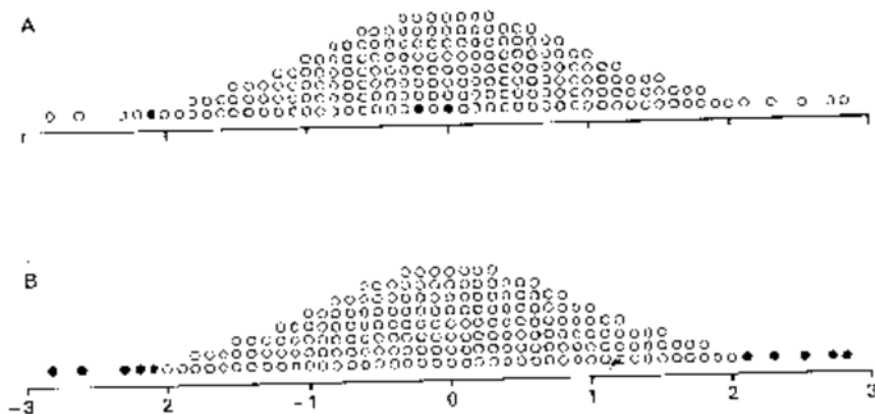
$$\text{Questo significa } 1 - \alpha = P_0 \{ |t| \leq n \} = P_0 \{ -n \leq t \leq n \} =$$

$$= F_t(n) - F_t(-n)$$

Ma t è simmetrica $\Rightarrow F_t(-n) = 1 - F_t(n)$, e

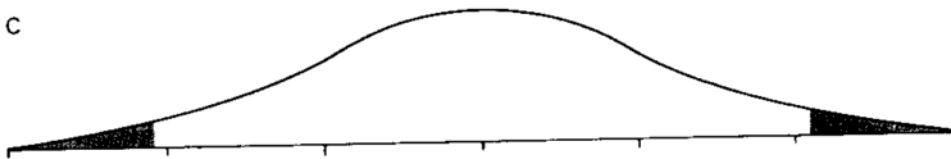
$$1 - \alpha = F_t(n) - (1 - F_t(n)) = 2F_t(n) - 1$$

$$2F_t(n) = 1 - \alpha ; \quad F_t(n) = 1 - \frac{\alpha}{2} \quad n = t_{1-\frac{\alpha}{2}}(\nu)$$



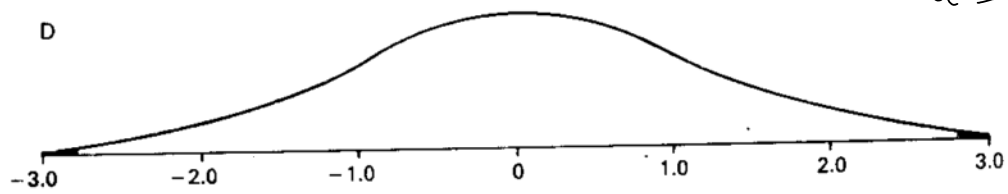
$$\alpha = 5\%$$

C



$$\alpha = 1\%$$

D



4.7 Quantili della distribuzione $t(n)$ di Student

$$\mathbb{P}\{t \leq x\} = \alpha, \quad t \sim t(\nu)$$

Nella tabella sono riportati vari ν significativi e $\alpha = 0.95, 0.975, 0.99$ e 0.995 .
Nell'intersezione tra α e ν si legge $x = t_\alpha(\nu)$.

ν	α	0.95	0.975	0.99	0.995	0.9975	0.999
1	6.313	12.706	31.820	63.657	127.321	318.309	
2	2.919	4.302	6.964	9.924	14.089	22.327	
3	2.353	3.182	4.540	5.840	7.453	10.215	
4	2.131	2.776	3.747	4.604	5.598	7.173	
5	2.015	2.570	3.364	4.032	4.773	5.893	
6	1.943	2.446	3.142	3.707	4.317	5.208	
7	1.894	2.364	2.998	3.499	4.029	4.785	
8	1.859	2.306	2.896	3.355	3.833	4.501	
9	1.833	2.262	2.821	3.249	3.690	4.297	
10	1.812	2.228	2.763	3.169	3.581	4.144	
11	1.795	2.201	2.718	3.105	3.497	4.025	
12	1.782	2.178	2.681	3.054	3.428	3.930	
13	1.770	2.160	2.650	3.012	3.372	3.852	
14	1.761	2.144	2.624	2.976	3.326	3.787	
15	1.753	2.131	2.602	2.946	3.286	3.733	
16	1.745	2.119	2.583	2.920	3.252	3.686	
17	1.739	2.109	2.566	2.898	3.222	3.646	
18	1.734	2.100	2.552	2.878	3.197	3.610	
19	1.729	2.093	2.539	2.861	3.174	3.579	
20	1.724	2.086	2.528	2.845	3.153	3.552	
21	1.720	2.079	2.517	2.831	3.135	3.527	
22	1.717	2.073	2.508	2.818	3.119	3.505	
23	1.713	2.068	2.499	2.807	3.104	3.485	
24	1.710	2.063	2.492	2.796	3.091	3.467	
25	1.708	2.059	2.485	2.787	3.078	3.450	
26	1.705	2.055	2.478	2.778	3.067	3.435	
27	1.703	2.051	2.472	2.770	3.057	3.421	
28	1.701	2.048	2.467	2.763	3.047	3.408	
29	1.699	2.045	2.462	2.756	3.038	3.396	
30	1.697	2.042	2.457	2.750	3.030	3.385	
40	1.683	2.021	2.423	2.704	2.971	3.307	
50	1.676	2.009	2.403	2.678	2.937	3.261	
60	1.670	2.000	2.390	2.660	2.915	3.232	
70	1.667	1.994	2.381	2.648	2.899	3.211	
80	1.664	1.990	2.373	2.638	2.887	3.195	
90	1.662	1.987	2.368	2.632	2.878	3.183	
100	1.660	1.984	2.364	2.626	2.871	3.174	
120	1.657	1.979	2.357	2.617	2.860	3.160	
200	1.653	1.972	2.345	2.601	2.839	3.131	
∞	1.644	1.960	2.326	2.575	2.807	3.090	