STATISTICA

VERIFICA D'IPOTESI - 3

Esempio 2

Il peso medio di un campione di 50 studenti che dichiarano di svolgere **molta attività fisica** (almeno 3 ore di palestra/allenamento sportivo a settimana) è di 71.2 kg con dev. standard di 2.5 kg.

Il peso medio di un campione di 35 studenti che dichiarano **vita sendentaria** (meno di 1 ora…) hanno un peso medio di 70 kg con devi. standard di 3.2 kg.

Sottoporre a verifica l'ipotesi nulla che lo sport non abbia effetto sul peso.

Esempio 2

Il peso medio di un campione di 50 studenti che dichiarano di svolgere **molta attività fisica** (almeno 3 ore di palestra/allenamento sportivo a settimana) è di 71.2 kg con dev. standard di 2.5 kg.

Il peso medio di un campione di 35 studenti che dichiarano **vita sendentaria** (meno di 1 ora…) hanno un peso medio di 70 kg con devi. standard di 3.2 kg.

Sottoporre a verifica l'ipotesi nulla che lo sport non abbia effetto sul peso.

$\left(X_1, \dots, X_{n_1}\right)$ i.i.d $\sim N(\mu_1, {\sigma_1}^2)$ **camp. indip.** $(Y_1, \dots, Y_{n_2})$ i.i.d $\sim N(\mu_2, {\sigma_2}^2)$

"sportivi" "sedentari"

Esempio 2

Il peso medio di un campione di 50 studenti che dichiarano di svolgere **molta attività fisica** (almeno 3 ore di palestra/allenamento sportivo a settimana) è di 71.2 kg con dev. standard di 2.5 kg.

Il peso medio di un campione di 35 studenti che dichiarano **vita sendentaria** (meno di 1 ora…) hanno un peso medio di 70 kg con devi. standard di 3.2 kg.

Sottoporre a verifica l'ipotesi nulla che lo sport non abbia effetto sul peso.

$\left(X_1, \dots, X_{n_1}\right)$ i.i.d $\sim N(\mu_1, {\sigma_1}^2)$ **camp. indip.** $(Y_1, \dots, Y_{n_2})$ i.i.d $\sim N(\mu_2, {\sigma_2}^2)$

"sportivi" "sedentari"

$\boldsymbol{H_0 : \mu_1 = \mu_2}$ $\boldsymbol{H_1 : \mu_1 \neq \mu_2}$

Verifica d'ipotesi: μ_1 e μ_2

$(X_1, \dots, X_{n_1})$ campione aleatorio $N(\mu_1, {\sigma_1}^2)$

$(Y_1, \dots, Y_{n_2})$ campione aleatorio $N(\mu_2, {\sigma_2}^2)$

$\boldsymbol{{\sigma_1}^2 = {\sigma_2}^2}$ **ma non note**

$\boldsymbol{H_0}$	$\boldsymbol{H_1}$	Rifiutiamo $\boldsymbol{H_0}$ se:
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\left\| \dfrac{\boldsymbol{\bar{x}_1 - \bar{x}_2}}{\sqrt{una\ varianza\ opportuna}} \right\| > t(n_1 + n_2 - 2)_{\alpha/2}$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$\dfrac{\boldsymbol{\bar{x}_1 - \bar{x}_2}}{\sqrt{una\ varianza\ opportuna}} > t(n_1 + n_2 - 2)_{\alpha}$
…	…	…

Verifica d'ipotesi: μ_1 e μ_2

$(X_1, \ldots, X_{n_1})$ campione aleatorio $N(\mu_1, {\sigma_1}^2)$

$(Y_1, \ldots, Y_{n_2})$ campione aleatorio $N(\mu_2, {\sigma_2}^2)$

${\sigma_1}^2 = {\sigma_2}^2$ ma non note

$$\text{"}una\ varianza\ opportuna\text{"} \ : \ {s_p}^2 \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

(*pooled*)

con

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

Verifica d'ipotesi: μ_1 e μ_2

$(X_1, \dots, X_{n_1})$ campione aleatorio $N(\mu_1, \sigma^2)$

$(Y_1, \dots, Y_{n_2})$ campione aleatorio $N(\mu_2, \sigma^2)$

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

H_0	H_1	Rifiutiamo H_0 se:
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\left\lvert \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^{\,2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right\rvert > t(n_1 + n_2 - 2)_{\alpha/2}$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^{\,2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > t(n_1 + n_2 - 2)_{\alpha}$
…	…	…

Verifica d'ipotesi: μ_1 e μ_2

$(X_1, \dots, X_{n_1})$ campione aleatorio $E(X_i) = \mu_1, Var(X_i) = \sigma^2$ $n_1 > 30$

$(Y_1, \dots, Y_{n_2})$ campione aleatorio $E(Y_i) = \mu_2, Var(Y_i) = \sigma^2$ $n_2 > 30$

$\boldsymbol{H_0}$	$\boldsymbol{H_1}$	Rifiutiamo $\boldsymbol{H_0}$ se:
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\left\lvert \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{{s_p}^2 \times \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \right\rvert > t(n_1 + n_2 - 2)_{\alpha/2}$
$\mu_1 = \mu_2$	$\mu_1 > \mu_2$	$\dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{{s_p}^2 \times \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} > t(n_1 + n_2 - 2)_{\alpha}$
…	…	…

Esempio 2

Il peso medio di un campione di 50 studenti che dichiarano di svolgere **molta attività fisica** (almeno 3 ore di palestra/allenamento sportivo a settimana) è di 71.2 kg con dev. standard di 2.5 kg.

Il peso medio di un campione di 35 studenti che dichiarano **vita sendentaria** (meno di 1 ora…) hanno un peso medio di 70 kg con devi. standard di 3.2 kg.

Sottoporre a verifica l'ipotesi nulla che lo sport non abbia effetto sul peso.

"sportivi" $\left(X_1, \dots, X_{n_1}\right)$ i.i.d $\sim N(\mu_1, \sigma^2)$

camp. indip.

"sedentari" $(Y_1, \dots, Y_{n_2})$ i.i.d $\sim N(\mu_1, \sigma^2)$

$$s_p{}^2 = \frac{49 \times 2.5^2 + 34 \times 3.2^2}{50 + 35 - 2} = 7.88 \Longrightarrow s = 2.81$$

Esempio 2

Il peso medio di un campione di 50 studenti che dichiarano di svolgere **molta attività fisica** (almeno 3 ore di palestra/allenamento sportivo a settimana) è di 71.2 kg con dev. standard di 2.5 kg.

Il peso medio di un campione di 35 studenti che dichiarano **vita sendentaria** (meno di 1 ora…) hanno un peso medio di 70 kg con devi. standard di 3.2 kg.

Sottoporre a verifica l'ipotesi nulla che lo sport non abbia effetto sul peso.

"sportivi" $(X_1, \dots, X_{n_1})$ i.i.d $\sim N(\mu_1, \sigma^2)$

camp. indip.

"sedentari" $(Y_1, \dots, Y_{n_2})$ i.i.d $\sim N(\mu_1, \sigma^2)$

$$\left|\frac{\bar{x}_1 - \bar{x}_2}{s_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right| = \frac{|71.2 - 70|}{2.81 \times \sqrt{\frac{1}{50} + \frac{1}{35}}} = 1.94$$

$H_0 : \mu_1 = \mu_2$ $\quad$ $H_1 : \mu_1 \neq \mu_2$

cfr. $t(83)_{0.025} \approx 1.9901$

Esempio 2

Il peso medio di un campione di 50 studenti che dichiarano di svolgere **molta attività fisica** (almeno 3 ore di palestra/allenamento sportivo a settimana) è di 71.2 kg con dev. standard di 2.5 kg.

Il peso medio di un campione di 35 studenti che dichiarano **vita sendentaria** (meno di 1 ora…) hanno un peso medio di 70 kg con devi. standard di 3.2 kg.

Sottoporre a verifica l'ipotesi nulla che lo sport non abbia effetto sul peso.

"sportivi" | "sedentari"

$(X_1, \dots, X_{n_1})$ i.i.d $\sim N(\mu_1, \boldsymbol{\sigma_1}^2)$ $\boxed{\textbf{camp. indip.}}$ $(Y_1, \dots, Y_{n_2})$ i.i.d $\sim N(\mu_1, \boldsymbol{\sigma_2}^2)$

Se non è possibile assumerle uguali, c'è una soluzione, che agisce sui gradi di libertà della distribuzione t, usualmente implementata nei software statistici.

Ese

Il peso me
svolger
pales
stand

Il peso ta
sende 70 kg con
devi. sta

Sottoporre a ver port non abbia effetto
sul peso.

Si potrebbe fare un test

$$H_0 : \sigma_1^2 = \sigma_2^2$$

contro

$$H_0 : \sigma_1^2 \neq \sigma_2^2?$$

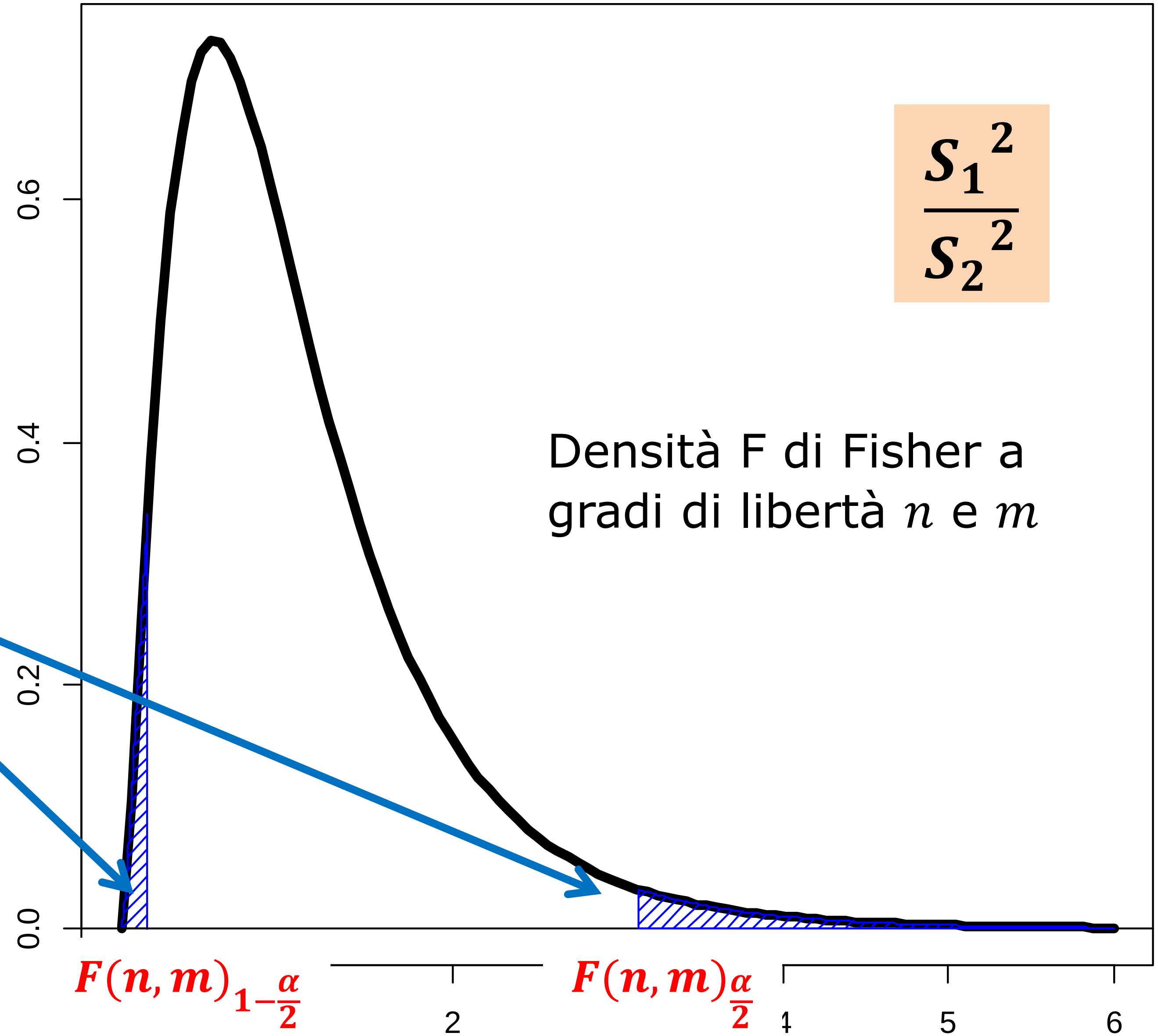
"sportivi"		"sedentari"
$(X_1, \dots, X_{n_1})$ i.i.d $\sim N(\mu_1, \sigma_1^2)$	**camp. indip.**	$(Y_1, \dots, Y_{n_2})$ i.i.d $\sim N(\mu_1, \sigma_2^2)$

Se non è possibile assumerle uguali, c'è una soluzione, che agisce sui gradi di libertà della distribuzione t, usualmente implementata nei software statistici.

In risposta a una domanda

$$H_0 : \frac{\sigma_1^{\,2}}{\sigma_2^{\,2}} = 1$$

$$H_1 : \frac{\sigma_1^{\,2}}{\sigma_2^{\,2}} \neq 1$$



Esercizio 2

Durante uno studio sul quoziente intellettivo un gruppo di 20 uomini scelti a caso ed uno di 20 donne scelte a caso sono stati sottoposti ad un test per la misura del QI ottenendo i seguenti punteggi medi: $\bar{x}_U = 115$ e $\bar{x}_D = 111.9$, con le rispettive varianze: ${s_U}^2 = 624.31$ e ${s_D}^2 = 561.04$.

a) in quale dei due gruppo la variabilità campionaria del QI è maggiore?

b) C'è abbastanza evidenza nei dati per poter affermare che gli uomini hanno un QI medio superiore a quello delle donne?

Esercizio 2

Durante uno studio sul quoziente intellettivo un gruppo di 20 uomini scelti a caso ed uno di 20 donne scelte a caso sono stati sottoposti ad un test per la misura del QI ottenendo i seguenti punteggi medi: $\bar{x}_U = 115$ e $\bar{x}_D = 111.9$, con le rispettive varianze: ${s_U}^2 = 624.31$ e ${s_D}^2 = 561.04$.

a) in quale dei due gruppo la variabilità campionaria del QI è maggiore?

b) C'è abbastanza evidenza nei dati per poter affermare che gli uomini hanno un QI medio superiore a quello delle donne?

a) $\mathrm{CV-U}\ : \quad \frac{\sqrt{624.31}}{115} = 0.217, \quad \mathrm{CV-D}: \quad \frac{\sqrt{561.04}}{111.9} = 0.211$

Esercizio 2

Durante uno studio sul quoziente intellettivo un gruppo di 20 uomini scelti a caso ed uno di 20 donne scelte a caso sono stati sottoposti ad un test per la misura del QI ottenendo i seguenti punteggi medi: $\bar{x}_U = 115$ e $\bar{x}_D = 111.9$, con le rispettive varianze: ${s_U}^2 = 624.31$ e ${s_D}^2 = 561.04$.

b) C'è abbastanza evidenza nei dati per poter affermare che gli uomini hanno un QI medio superiore a quello delle donne?

due campioni indipendenti

$X_i \sim N(\mu_U, \sigma^2)$ QI uomini

$H_0 : \mu_U = \mu_D \quad H_1 : \mu_U > \mu_D$

$Y_i \sim N(\mu_D, \sigma^2)$ QI donne

Esercizio 2

Durante uno studio sul quoziente intellettivo un gruppo di 20 uomini scelti a caso ed uno di 20 donne scelte a caso sono stati sottoposti ad un test per la misura del QI ottenendo i seguenti punteggi medi: $\bar{x}_U = 115$ e $\bar{x}_D = 111.9$, con le rispettive varianze: ${s_U}^2 = 624.31$ e ${s_D}^2 = 561.04$.

b) C'è abbastanza evidenza nei dati per poter affermare che gli uomini hanno un QI medio superiore a quello delle donne?

due campioni indipendenti

$X_i \sim N(\mu_U, \sigma^2)$ QI uomini

$Y_i \sim N(\mu_D, \sigma^2)$ QI donne

$H_0 : \mu_U = \mu_D \qquad H_1 : \mu_U > \mu_D$

$$s_p = \sqrt{\frac{19 \times 624.31 + 19 \times 561.04}{20 + 20 - 2}} = \sqrt{592.675} = 24.34$$

Esercizio 2

Durante uno studio sul quoziente intellettivo un gruppo di 20 uomini scelti a caso ed uno di 20 donne scelte a caso sono stati sottoposti ad un test per la misura del QI ottenendo i seguenti punteggi medi: $\bar{x}_U = 115$ e $\bar{x}_D = 111.9$, con le rispettive varianze: ${s_U}^2 = 624.31$ e ${s_D}^2 = 561.04$.

b) C'è abbastanza evidenza nei dati per poter affermare che gli uomini hanno un QI medio superiore a quello delle donne?

due campioni indipendenti

$X_i \sim N(\mu_U, \sigma^2)$ QI uomini

$H_0 : \mu_U = \mu_D \qquad H_1 : \mu_U > \mu_D$

$Y_i \sim N(\mu_D, \sigma^2)$ QI donne

$s_p = 24.34$

$$\frac{\bar{x}_U - \bar{x}_D}{s_p \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{115 - 111.9}{24.34 \times \sqrt{\frac{2}{20}}} = 0.40$$

Esercizio 2

Durante uno studio sul quoziente intellettivo un gruppo di 20 uomini scelti a caso ed uno di 20 donne scelte a caso sono stati sottoposti ad un test per la misura del QI ottenendo i seguenti punteggi medi: $\bar{x}_U = 115$ e $\bar{x}_D = 111.9$, con le rispettive varianze: ${s_U}^2 = 624.31$ e ${s_D}^2 = 561.04$.

b) C'è abbastanza evidenza nei dati per poter affermare che gli uomini hanno un QI medio superiore a quello delle donne? no!

due campioni indipendenti

$X_i \sim N(\mu_U, \sigma^2)$ QI uomini

$H_0 : \mu_U = \mu_D \quad H_1 : \mu_U > \mu_D$

$Y_i \sim N(\mu_D, \sigma^2)$ QI donne

cfr. $t(38)_{0.05} \approx t(40)_{0.05} = 1.6839$

Un ***contro***esempio

Effetto di una sostanza omeopatica contro l'ipertensione:

Un campione casuale di individui di cui misuriamo la pressione **prima** e **dopo** la somministrazione del farmaco.

Individuo	Pres. min. PRIMA	Pres. min. DOPO
A. Bianchini	83	75
B. Rossini	90	87
C. Verdini	66	62
E. Giallini	85	86
F. Turchini	72	70
…	…	…

Un ***contro***esempio

Effetto di una sostanza omeopatica contro l'ipertensione:

Un campione casuale di individui di cui misuriamo la pressione **prima** e **dopo** la somministrazione del farmaco.

Individuo	Pres. min. PRIMA	Pres. min. DOPO
A. Bianchini	83	75
B. Rossini	90	87
C. Verdini	66	62
E. Giallini	85	86
F. Turchini	72	70
…	…	…

I due campioni **non** sono **indipendenti** perchè a coppie sono "collegati"

Esempio

Altra situazione tipica: misura di una certa variabile effettuata su due unità statistiche legate tra loro: *campioni accoppiati*.

Coppie sposate	Marito	Moglie	Differenza (d, in anni)
1	Titolo di studio della moglie	Titolo di studio	Per verificare se il marito tende ad aumentare il grado di educazione della moglie
2			
…			
153			

Social and Psychological Factors Affecting Fertility
P. Kiser & C. Whelpton. Milbank Memorial Fund (1950)
(Anderson & Finn, p.433)

Esempio

Altra situazione tipica: misura di una certa variabile effettuata su due unità statistiche legate tra loro: *campioni accoppiati*.

Coppie sposate	Marito X_i	Moglie Y_i	Differenza (d, in anni)
1	Titolo di studio della moglie	Titolo di studio	d_1
2			d_2
…			…
153			d_{153}

$$D_i = X_i - Y_i$$

$$D_i \sim N(\mu, \sigma^2)$$

Social and Psychological Factors Affecting Fertility
P. Kiser & C. Whelpton. Milbank Memorial Fund (1950)
(Anderson & Finn, p.433)

$$H_0: \mu = 0 \quad H_1: \mu > 0$$

Esempio

Altra situazione tipica: misura di una certa variabile effettuata su due unità statistiche legate tra loro: *campioni accoppiati*.

Coppie sposate	Marito X_i	Moglie Y_i	Differenza (d, in anni)
1	Titolo di studio della moglie	Titolo di studio	$\boldsymbol{d_1}$
2			$\boldsymbol{d_2}$
…			…
153			$\boldsymbol{d_{153}}$

$D_i = X_i - Y_i$

$D_i \sim N(\mu, \sigma^2)$

$\bar{d} = 0.32,$
$s_d = 1.07\ anni$

Social and Psychological Factors Affecting Fertility
P. Kiser & C. Whelpton. Milbank Memorial Fund (1950)
(Anderson & Finn, p.433)

$H_0: \mu = 0 \quad H_1: \mu > 0$

$$\frac{\bar{d} - 0}{s_d/\sqrt{153}} = \frac{0.32}{1.07/\sqrt{153}} = 3.699 \geq t(152)_{0.01} \approx z_{0.01} = 2.3263$$

STATISTICA

Relazioni tra fenomeni

"qualitativi"

Le tabelle di contingenza

	Infarto miocardico		
Gruppo	**Sì**	**No**	
Placebo	189	10845	11034
Aspirina	104	10933	11037

Tipo di farmaco, Infarto miocardico

Le tabelle di contingenza

	Infarto miocardico		
Gruppo	**Sì**	**No**	
Placebo	189	10845	11034
Aspirina	104	10933	11037

Tipo di farmaco, Infarto miocardico

Appartenenza etnica e preferenze politiche

Esposizione ad una data sostanza e insorgenza di malattie

Livello sociale della famiglia e riuscita scolastica dei figli

Metodo colturale e produzione

Dimensione del cervello e QI

Genere e percorso di studi

Le tabelle di contingenza

$$(X, Y)$$

Tipo di farmaco, Infarto miocardico

Appartenenza etnica e preferenze politiche

Esposizione ad una data sostanza e insorgenza di malattie

Livello sociale della famiglia e riuscita scolastica dei figli

Metodo colturale e produzione

Dimensione del cervello e QI

Genere e percorso di studi

Le tabelle di contingenza

(X, Y)

tabella a
doppia entrata

		Y				
		y_1	y_2	…	y_k	TOT
X	x_1					$n_{1\cdot}$
	x_2		n_{ij}			$n_{2\cdot}$
	…					…
	x_h					$n_{h\cdot}$
	TOT	$n_{\cdot 1}$	$n_{\cdot 2}$	…	$n_{\cdot k}$	n

Le tabelle di contingenza

(X, Y)

tabella a
doppia entrata

		Y				
		y_1	y_2	…	y_k	TOT
X	x_1					$n_{1\cdot}$
	x_2					$n_{2\cdot}$
	…					…
	x_h					$n_{h\cdot}$
	TOT	$n_{\cdot 1}$	$n_{\cdot 2}$	…	$n_{\cdot k}$	n

n_{ij} distribuzione
di
frequenza
congiunta

Le tabelle di contingenza

Genere, Sopravvivenza

Classe di viaggio, Sopravvivenza

Tipo di farmaco, Infarto miocardico

(X, Y)

tabella a doppia entrata

		Y				
		y_1	y_2	...	y_k	TOT
X	x_1					$n_{1\cdot}$
	x_2					$n_{2\cdot}$
	...					
	x_h					$n_{h\cdot}$
	TOT	$n_{\cdot 1}$	$n_{\cdot 2}$		$n_{\cdot k}$	n

distribuzione **congiunta**

distribuz. **marginale** di X (somma nelle righe)

distribuz. **marginale** di Y (somma nelle colonne)

Le tabelle di contingenza

	Soddisfazione		
Gruppo	**Alta**	**Bassa**	
I classe	203	122	**325**
II classe	118	167	**285**
III classe	178	528	**706**
	499	**817**	$\boldsymbol{n = 1316}$

Le tabelle di contingenza

	Soddisfazione		
Gruppo	**Alta**	**Bassa**	
I classe	203	122	**325**
II classe	118	167	**285**
III classe	178	528	**706**
	499	**817**	$\boldsymbol{n = 1316}$

$freq.ass./_{1316}$

Le tabelle di contingenza

	Soddisfazione		
Gruppo	**Alta**	**Bassa**	
I classe	0.154	0.093	**0.247**
II classe	0.090	0.127	**0.217**
III classe	0.135	0.401	**0.536**
	0.379	**0.621**	**1**

$\frac{n_{ij}}{n}$

$\frac{n_{i\cdot}}{n}$

$\frac{n_{\cdot j}}{n}$

0.090=
prob. che una
passeggero scelto
a caso abbia
viaggiato in II cl.
e sia molto
soddisfatto

Le tabelle di contingenza

	Soddisfazione		
Gruppo	**Alta**	**Bassa**	
I classe	0.154	0.093	**0.247**
II classe	0.090	0.127	**0.217**
III classe	0.135	0.401	**0.536**
	0.379	**0.621**	**1**

$\frac{n_{ij}}{n}$

$\frac{n_{i\cdot}}{n}$

$\frac{n_{\cdot j}}{n}$

0.247=
prob. che una passeggero scelto a caso abbia viaggiato in I classe

0.621=
prob. che una passeggero scelto a caso sia stato poco soddisfatto

Le tabelle di contingenza

	Soddisfazione		
Gruppo	**Alta**	**Bassa**	
I classe	203	122	**325**
II classe	118	167	**285**
III classe	178	528	**706**
	499	**817**	$n = 1316$

$$\frac{203}{325} = 0.625$$

Prob. che un passeggero sia stato molto soddisfatto **sapendo** che era in I classe

Le tabelle di contingenza

	Soddisfazione		
Gruppo	**Alt**	**ssa**	
I classe			**325**
II classe			**285**
III class			**706**

$n = 1316$

Il grado di soddisfazione dipende dalla classe di viaggio?

$$\frac{203}{325} = 0.625$$

Prob. che un passeggero sia stato molto soddisfatto **sapendo** che era in I classe

Le tabelle di contingenza

	Soddisfazione			Prob. di A. sodd. | classe	
Gruppo	**Alta**	**Bassa**			
I classe	203	122	325	203/325	0.625
II classe	118	167	285	118/285	0.706
III classe	178	528	706	178/706	0.252
	499	**817**	$\boldsymbol{n = 1316}$		

Prob. di alta soddisfazione **sapendo la classe di appartenenza**

Le tabelle di contingenza

	Soddisfazione			Prob. di A. sodd. | classe	
Gruppo	**Alta**	**Bassa**			
I classe	203	122	325	203/325	0.625
II classe	118	167	285	118/285	0.706
III classe	178	528	706	178/706	0.252
	499	**817**	$n = 1316$		

Prob. di alta soddisfazione **sapendo la classe di appartenenza**

fossero state tutte uguali avremmo concluso che la soddisfazione non dipende dalla classe di viaggio

Le tabelle di contingenza

	Soddisfazione		
Gruppo	**Alta**	**Bassa**	
I classe			**325**
II classe			**285**
III classe			**706**
	499	**817**	***n*=1316**

ferme restando
le distribuzioni
marginali

l'unico modo per assicurarsi prob. condizionate tutte uguali è quello corrispondente all'**indipendenza delle variabili**

Le tabelle di contingenza

	Soddisfazione		
Gruppo	**Alta**	**Bassa**	
I classe	123.2	201.8	**325**
II classe	108.1	176.9	**285**
III classe	267.7	438.3	**706**
	499	**817**	***n*=1316**

ferme restando le distribuzioni marginali

n^*_{ij} frequenze *attese* o *teoriche* nell'ipotesi di **indipendenza**

$$\frac{n^*_{ij}}{n} = \frac{n_{i.}}{n} \times \frac{n_{.j}}{n} \Leftrightarrow n^*_{ij} = \frac{n_{i.} \times n_{.j}}{n}$$

$$p.es., P(\mathrm{Icl}\ \cap Sodd = A) = P(\mathrm{Icl}) \times P(Sodd = A)$$

Le tabelle di contingenza

	Soddisfazione		
Gruppo	**Alta**	**Bassa**	
I classe	123.2	201.8	**325**
II classe			**285**
III classe			**706**
	499	**817**	***n*=1316**

l'unico modo per assicurarsi prob. condizionate tutte uguali è quello nella tabella, corrispondente all'**indipendenza delle variabili**

$$n^*_{11} = \frac{n_{1.} \times n_{.1}}{n}$$

Le tabelle di contingenza

	Soddisfazione		
Gruppo	**Alta**	**Bassa**	
I classe	123.2	201.8	**325**
II classe	108.1	176.9	**285**
III classe			**706**
	499	**817**	***n*=1316**

l'unico modo per assicurarsi prob. condizionate tutte uguali è quello nella tabella, corrispondente all'**indipendenza delle variabili**

$$n^*_{21} = \frac{n_{2\cdot} \times n_{\cdot 1}}{n}$$

Le tabelle di contingenza

	Soddisfazione		
Gruppo	**Alta**	**Bassa**	
I classe	123.2	201.8	**325**
II classe	108.1	176.9	**285**
III classe	267.7	438.3	**706**
	499	**817**	***n*=1316**

l'unico modo per assicurarsi prob. condizionate tutte uguali è quello nella tabella, corrispondente all'**indipendenza delle variabili**

$$n^*_{31} = \frac{n_{3.} \times n_{.1}}{n}$$

Le tabelle di contingenza

	Soddisfazione		
Gruppo	**Alta**	**Bassa**	
I classe	123.2	201.8	**325**
II classe	108.1	176.9	**285**
III classe	267.7	438.3	**706**
	499	**817**	***n*=1316**

l'unico modo per assicurarsi prob. condizionate tutte uguali è quello nella tabella, corrispondente all'**indipendenza delle variabili**

$$n^*_{32} = \frac{n_{3\cdot} \times n_{\cdot 2}}{n}$$

Le tabelle di contingenza

frequenze **attese** nel caso dell'**indipendenza**

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	123.2	201.8
II classe	108.1	176.9
III classe	267.7	438.3

frequenze **osservate**

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	203	122
II classe	118	167
III classe	178	528

Le tabelle di contingenza

frequenze **attese**

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	123.2	201.8
II classe	108.1	176.9
III classe	267.7	438.3

frequenze **osservate**

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	203	122
II classe	118	167
III classe	178	528

Le tabelle di contingenza

frequenze **attese** $n_{ij}{}^{*}$

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	123.2	201.8
II classe	108.1	176.9
III classe	267.7	438.3

$$\chi^2 = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{\left(n_{ij} - n^{*}{}_{ij}\right)^2}{n^{*}{}_{ij}}$$

frequenze **osservate** n_{ij}

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	203	122
II classe	118	167
III classe	178	528

Le tabelle di contingenza

frequenze **attese** n_{ij}^{*}

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	123.2	201.8
II classe	108.1	176.9
III classe	267.7	438.3

$\chi^2 = 0$ indica assenza totale di associazione, o **indipendenza**

$$\chi^2 = \sum_{i=1}^{h} \sum_{j=1}^{k} \frac{\left(n_{ij} - n^{*}_{ij}\right)^2}{n^{*}_{ij}}$$

frequenze **osservate** n_{ij}

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	203	122
II classe	118	167
III classe	178	528

L’indice del chi-quadrato

Perfetta interdipendenza tra le due variabili

	Soddisfazione	
Gruppo	Alta	Bassa
I&II classe	500	0
III classe	0	816

Perfetta dipendenza della soddisfazione dalla classe di viaggio.

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	280	0
II classe	220	0
III classe	0	816

L’indice del chi-quadrato

Indicazione di dipendenza
tra le due variabili

	Tipo di vacanza		
N. figli	Estero	Mare	Montagna
0	120	0	30
1	20	0	80
≥ 2	0	100	10

Perfetta dipendenza
della soddisfazione dalla classe di viaggio.

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	280	0
II classe	220	0
III classe	0	816

L’indice del chi-quadrato

frequenze **attese** $n_{ij}{}^{*}$

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	123.2	201.8
II classe	108.1	176.9
III classe	267.7	438.3

frequenze **osservate** n_{ij}

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	203	122
II classe	118	167
III classe	178	528

$$\chi^2 = \frac{(123. - 203)^2}{123.2} + \frac{(201.8 - 122)^2}{201.8} + \frac{(108.1 - 118)^2}{108.1} + \cdots + \frac{(438.3 - 528)^2}{438.3} = \mathbf{= 133.05}$$

Promemoria: $\chi^2 = 0$ corrisponde all’**indipendenza**

Il **test** del chi-quadrato

Campione casuale : $(X_1, Y_1), \ldots, (X_n, Y_n)$ (X e Y variabili categoriche)

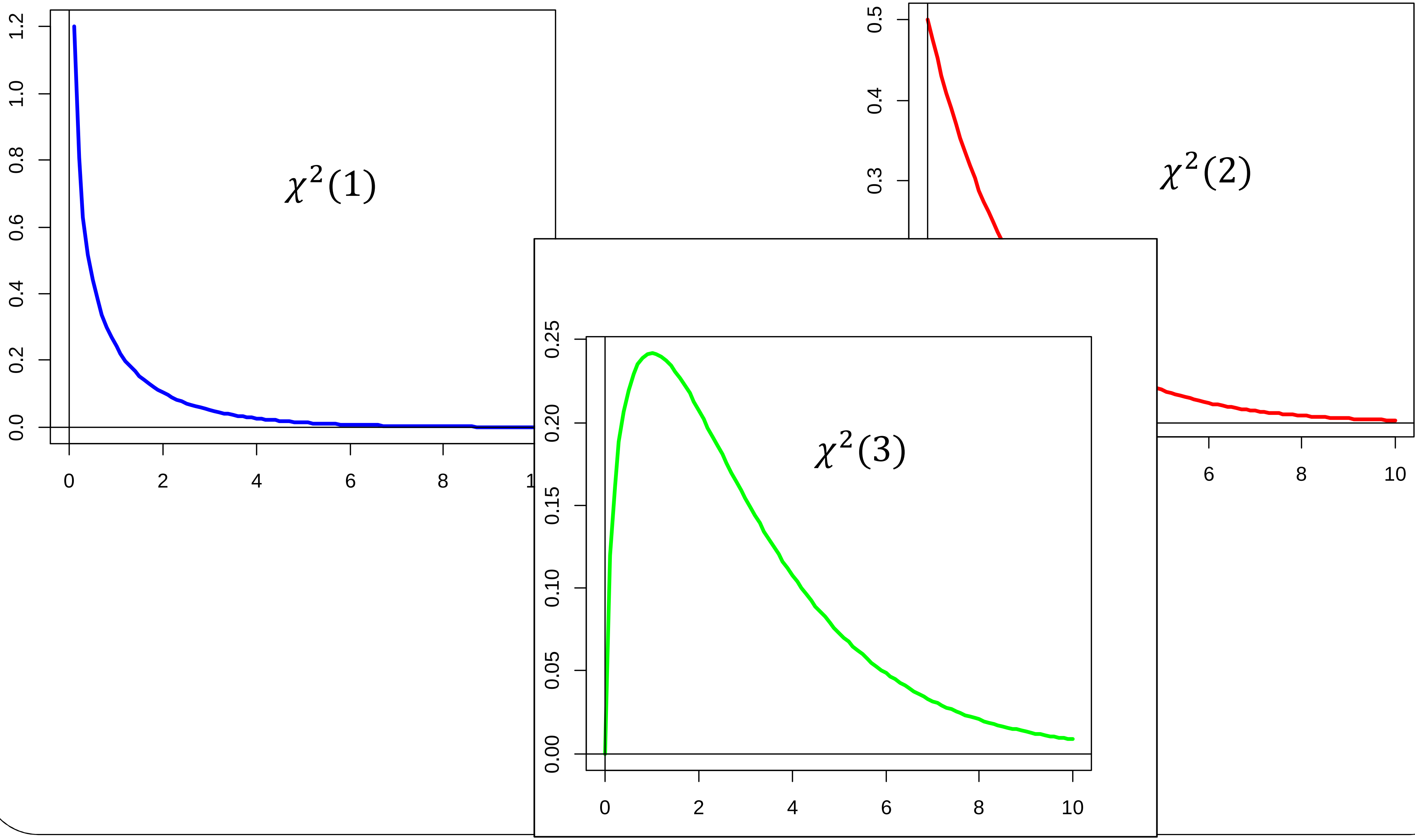
H_0 : indipendenza $(\chi^2 = 0)$ H_1 : associazione $(\chi^2 \neq 0)$

Se $n^*_{ij} \geq 5$ per tutti i valori, **si rifiuta** l'ipotesi di indipendenza se

$$\underbrace{\chi^2 = \sum_{i=1}^{h}\sum_{j=1}^{k}\frac{(n_{ij} - n^*_{ij})^2}{n^*_{ij}}}_{\text{indice}} > \underbrace{\chi^2(\underbrace{(r-1)\times(c-1)}_{\text{gradi di libertà}})_\alpha}_{\text{distribuzione (densità)}}$$

v. 24/11

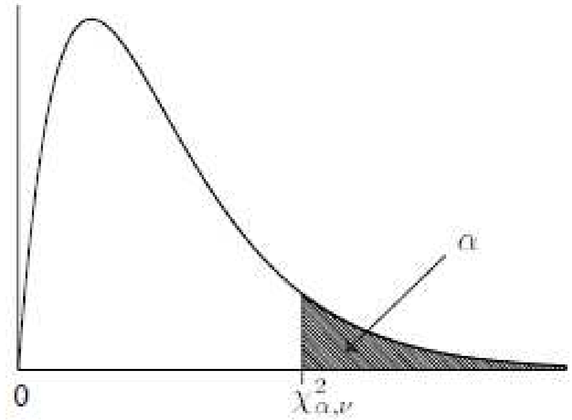
La densità del chi-quadrato



Simile alla tavola della *t*-Student

Tavola 3: Valori critici della Distribuzione Chi-Quadrato

$P(\chi^2 \geq \chi^2_{\alpha,\nu}) = \alpha$



$$\boxed{\chi^2(2 \times 1)_{0.05} = 5.9915}$$

g.d.l.

Tavola 3 *(segue)*: Valori critici della Distribuzione Chi-Quadrato

α piccolo, per i test

ν	0.20	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.0001
1	1.6424	2.7055	3.8415	5.0239	6.6349	7.8794	10.8274	12.1153	15.1343
2	3.2189	4.6052	$\boxed{5.9915}$	7.3778	9.2104	10.5965	13.8150	15.2014	18.4247
3	4.6416	6.2514	7.8147	9.3484	11.3449	12.8381	16.2660	17.7311	21.1040
4	5.9886	7.7794	9.4877	11.1433	13.2767	14.8602	18.4662	19.9977	23.5064
5	7.2893	9.2363	11.0705	12.8325	15.0863	16.7496	20.5147	22.1057	25.7507
6	8.5581	10.6446	12.5916	14.4494	16.8119	18.5475	22.4575	24.1016	27.8527
7	9.8032	12.0170	14.0671	16.0128	18.4753	20.2777	24.3213	26.0179	29.8814
8	11.0301	13.3616	15.5073	17.5345	20.0902	21.9549	26.1239	27.8674	31.8268

L'indice del chi-quadrato

frequenze **attese** n_{ij}^*

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	123.2	201.8
II classe	108.1	176.9
III classe	267.7	438.3

frequenze **osservate** n_{ij}

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	203	122
II classe	118	167
III classe	178	528

$$\chi^2 = \frac{(123.-203)^2}{123.2} + \frac{(201.8-122)^2}{201.8} + \frac{(108.1-118)^2}{108.1} + \cdots + \frac{(438.3-528)^2}{438.3} = \mathbf{= 133.05}$$

C'è una forte evidenza **contro** l'ipotesi nulla di indipendenza.

$$\chi^2(2 \times 1)_{0.05} = 5.9915$$

L'indice del chi-quadrato

frequenze **attese** $n_{ij}{}^*$

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	123.2	201.8
II classe	108.1	176.9
III classe	267.7	438.3

frequenze **osservate** n_{ij}

	Soddisfazione	
Gruppo	**Alta**	**Bassa**
I classe	203	122
II classe	118	167
III classe	178	528

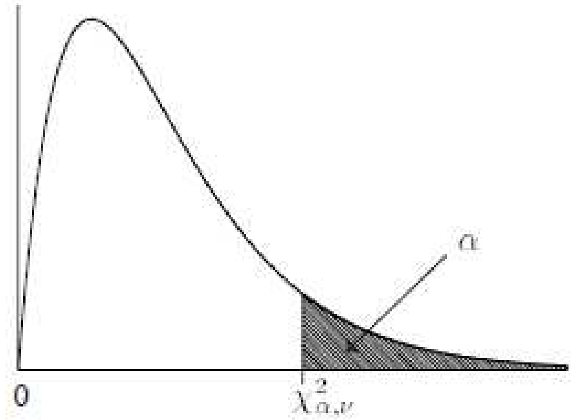
$$\chi^2 = \frac{(123.-203)^2}{123.2} + \frac{(201.8-122)^2}{201.8} + \frac{(108.1-118)^2}{108.1} + \cdots + \frac{(438.3-528)^2}{438.3} = \mathbf{= 133.05}$$

Quanto vale il *p-value* del test?

Simile alla tavola della *t*-Student

Tavola 3: Valori critici della Distribuzione Chi-Quadrato

$P(\chi^2 \geq \chi^2_{\alpha,\nu}) = \alpha$



$p - value < 0.0001$

g.d.l.

Tavola 3 *(segue)*: Valori critici della Distribuzione Chi-Quadrato

α piccolo, per i test

ν	0.20	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.0001
1	1.6424	2.7055	3.8415	5.0239	6.6349	7.8794	10.8274	12.1153	15.1343
2	3.2189	4.6052	5.9915	7.3778	9.2104	10.5965	13.8150	15.2014	$\boxed{18.4247}$
3	4.6416	6.2514	7.8147	9.3484	11.3449	12.8381	16.2660	17.7311	21.1040
4	5.9886	7.7794	9.4877	11.1433	13.2767	14.8602	18.4662	19.9977	23.5064
5	7.2893	9.2363	11.0705	12.8325	15.0863	16.7496	20.5147	22.1057	25.7507
6	8.5581	10.6446	12.5916	14.4494	16.8119	18.5475	22.4575	24.1016	27.8527
7	9.8032	12.0170	14.0671	16.0128	18.4753	20.2777	24.3213	26.0179	29.8814
8	11.0301	13.3616	15.5073	17.5345	20.0902	21.9549	26.1239	27.8674	31.8268