

Licence mention Mathématiques - Semestre 3
Statistique
Examen de lundi 4 janvier 2016
Durée 2h00
Tout document interdit - Calculatrices autorisées

Exercice 1

1) Les dentistes associés d'un cabinet dentaire ont constaté que 37% de leurs patients ont un problème de carie dentaire. On considère un échantillon de 150 personnes prises au hasard parmi les patients de ces dentistes, suffisamment nombreux pour assimiler le choix de cet échantillon à des tirages avec remise. On note Y la variable aléatoire égale au nombre de personnes de cet échantillon ayant un problème de carie dentaire.

a) Justifier que Y suit une loi binomiale et préciser ses paramètres.

b) Déterminer la probabilité qu'exactement 50 personnes parmi les 150 aient un problème de carie dentaire.

2) Une société qui fabrique du dentifrice souhaite augmenter sa part de marché. Elle envisage pour cela de modifier le goût de son dentifrice.

a) Avant de modifier le goût, elle a demandé à ces dentistes d'interroger leurs patients. Parmi les patients de ces dentistes qui utilisent ce dentifrice, 200 personnes ont été choisies au hasard et 45% d'entre elles ont déclaré apprécier le goût du dentifrice. On désigne par p la proportion de personnes appréciant le goût du dentifrice parmi les patients qui utilisent ce dentifrice avant la modification.

Déterminer un intervalle de confiance de la proportion p au niveau de confiance de 95%.

b) Le goût a été modifié et une nouvelle étude est menée auprès de 300 personnes choisies au hasard parmi les patients des dentistes de l'association. Parmi ces personnes, 165 ont apprécié le nouveau goût du dentifrice.

Au vu des résultats, le chef de projet de la société lance la production de ce nouveau dentifrice. Que penser de la décision du chef de projet ? Justifier la réponse :

- d'abord en exploitant un deuxième intervalle de confiance au niveau de confiance de 95% ;

- puis en effectuant un test statistique adapté au risque 5% ; présenter les hypothèses de ce test et le détail des calculs mis en oeuvre pour l'effectuer.

3) En utilisant le logiciel R, on a obtenu les résultats suivants :

```
> dentifrice=matrix(c(90,165,110,135),ncol=2)
> prop.test(dentifrice,alternative="less")
```

```
2-sample test for equality of proportions with continuity correction
```

```
data: dentifrice
X-squared = 4.4101, df = 1, p-value = 0.01786
alternative hypothesis: less
95 percent confidence interval:
 -1.00000000 -0.02113271
sample estimates:
prop 1 prop 2
 0.45  0.55
```

Expliquer ce que réalise la première instruction. Préciser les hypothèses H_0 et H_1 du test statistique mis en oeuvre, puis donner, en justifiant, le résultat du test au risque 5%. Ce résultat est-il cohérent avec celui du 2) b) ?

Exercice 2

1) Dans une population donnée, on considère un caractère quantitatif, représenté par une variable aléatoire X d'espérance mathématique μ et d'écart-type σ . On considère un échantillon $(X_1, \dots, X_n)$ de taille n de X , et les estimateurs $\bar{X}$ et S_c (de μ et σ) dont la définition est rappelée dans le formulaire joint. Soit un réel $\alpha \in]0; 1[$.

a) On suppose que X suit la loi normale $\mathcal{N}(\mu; \sigma)$. Déterminer, en détaillant les calculs, un intervalle de confiance de μ au niveau $1 - \alpha$. Expliquer ce que signifie le résultat obtenu.

b) On suppose que $n \geq 30$. Quelle hypothèse sur X peut-on supprimer ? Donner alors (sans calcul) un intervalle de confiance de μ au niveau $1 - \alpha$.

Une usine fabrique en série des pompes de surface destinées à l'irrigation agricole. Le cahier de charges demande que ces pompes aient un débit de $6 \text{ m}^3 \cdot \text{h}^{-1}$ (6 m^3 par heure) avec une tolérance de $\pm 0,25 \text{ m}^3 \cdot \text{h}^{-1}$ prévue dans le cahier des charges. En sortie de chaîne de fabrication, une pompe non conforme à ce cahier des charges pour le débit présente un défaut de débit.

On désigne par X la variable aléatoire qui, à chaque pompe produite, associe son débit exprimé en $\text{m}^3 \cdot \text{h}^{-1}$.

2) On suppose que la variable aléatoire X suit une loi normale de moyenne $\mu = 6$ et d'écart type $\sigma = 0,15$.

Calculer la probabilité qu'une pompe, prélevée au hasard dans la production, présente un défaut de débit. Arrondir le résultat au millièème.

3) Une entreprise commande un nombre important de pompes. Lors de la livraison, le service qualité de l'entreprise cherche à estimer la moyenne inconnue μ , exprimée en $\text{m}^3 \cdot \text{h}^{-1}$, des débits des pompes qui lui sont livrées à partir de mesures faites sur un échantillon de cinquante pompes prises dans la livraison. On considère que cet échantillon peut être assimilé à un prélèvement au hasard et avec remise de cinquante pompes dans la livraison. Les résultats des mesures effectuées sont donnés dans le tableau ci-dessous :

Débit en $\text{m}^3 \cdot \text{h}^{-1}$	5,7	5,8	5,9	6,0	6,1	6,2	6,3
Nombre de pompes	9	8	10	9	10	3	1

a) En utilisant le logiciel R, on a obtenu les résultats suivants :

```
> x<-c(5.7,5.8,5.9,6.0,6.1,6.2,6.3)
> n<-c(9,8,10,9,10,3,1)
> debit<-rep(x,n)
> mean(debit)
[1] 5.932
> var(debit)
[1] 0.0267102
> shapiro.test(debit)
```

Shapiro-Wilk normality test

```
data: debit
W = 0.9316, p-value = 0.006335
```

Expliquer ce que réalisent les instructions saisies et indiquer ce que représentent les valeurs calculées.

Préciser les hypothèses H_0 et H_1 du test statistique mis en oeuvre, puis donner, en justifiant, le résultat du test au risque 5%. Ce résultat remet-il en cause l'hypothèse faite dans la question 2) ? Justifier la réponse.

b) Au vu de ces résultats, l'entreprise pense que le débit moyen est inférieur au $6 \text{ m}^3 \cdot \text{h}^{-1}$ annoncés par le fabricant. Effectuer un test statistique au risque 5% pour savoir si on peut considérer que l'entreprise a raison. Présenter les hypothèses de ce test et le détail des calculs mis en oeuvre pour l'effectuer.

Exercice 3

La quantité de bactéries par cm^3 de lait provenant de 8 vaches différentes a été mesurée juste après la traite, puis 24 h plus tard. Les résultats sont les suivants :

Vache n°	1	2	3	4	5	6	7	8
Juste après la traite	12000	13000	21500	17000	15000	22000	11000	21000
24h après la traite	14000	20000	31000	28000	26000	30000	16000	29000

1) Supposant les hypothèses adéquates vérifiées, effectuer un test statistique au risque 5% pour savoir si on peut considérer qu'il y a un accroissement significatif du nombre de bactéries par cm^3 de lait au cours du temps.

2) En utilisant le logiciel R, on a obtenu les résultats suivants :

```
> AT1=c(12000,13000,21500,17000,15000,22000,11000,21000)
> shapiro.test(AT1)
```

Shapiro-Wilk normality test

```
data: AT1
W = 0.8874, p-value = 0.2211
```

```
> AT2=c(14000,20000,31000,28000,26000,30000,16000,29000)
> shapiro.test(AT2)
```

Shapiro-Wilk normality test

```
data: AT2
W = 0.8694, p-value = 0.1486
```

Donner, en justifiant, le résultat des tests effectués au risque 5%. Les hypothèses mentionnées au 1) sont-elles vérifiées ?

3) a) Donner, en justifiant et au risque 5%, le résultat du test suivant, réalisé avec le logiciel R :

```
> wilcox.test(AT1,AT2,paired=TRUE)
```

Wilcoxon signed rank test with continuity correction

```
data: AT1 and AT2
V = 0, p-value = 0.01403
alternative hypothesis: true location shift is not equal to 0
```

```
Message d'avis :
In wilcox.test.default(AT1, AT2, paired = TRUE) :
impossible de calculer la p-value exacte avec des ex-aequos
```

b) Etait-il utile de le faire ? Si oui, pourquoi ? Si non, dans quel cas aurait-il été utile ?

c) Le message d'avis est-il à prendre en compte ici ? Justifier la réponse.

Exercice 4

L'objectif de cet exercice est d'identifier l'espèce microbienne présente dans une population microbienne. Pour ce faire, on réalise une préparation dont on détermine l'absorbance à intervalles de temps réguliers. L'absorbance, mesurant le trouble de la préparation, est proportionnelle à la densité de la population microbienne présente dans la préparation. Voici les résultats obtenus :

Temps x_i en minutes	30	40	50	60	70	80	90	100
Absorbance y_i (sans unité)	0,072	0,100	0,134	0,199	0,295	0,398	0,599	0,791

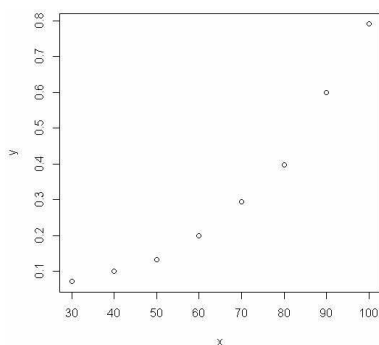
On utilise le logiciel R pour traiter ces données.

1) a) Que réalisent les instructions suivantes ?

```
x = seq(30,100,by=10)
```

```
y = c(0.072,0.100,0.134,0.199,0.295,0.398,0.599,0.791)
```

b) Quelle instruction permet d'obtenir le graphique ci-dessous ? Expliquer ce que réalisent les instructions saisies et indiquer ce que représentent les valeurs calculées ci-dessous.



```
> droitel=lm(y~x)
> droitel
```

```
Call:
lm(formula = y ~ x)
```

```
Coefficients:
(Intercept)      x
-0.32774      0.01002
```

```
> summary(droitel)
```

```
Call:
lm(formula = y ~ x)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.078595 -0.074750 -0.007095  0.045024  0.116833
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.327738   0.090896  -3.606  0.011290 *
x             0.010019   0.001319   7.597  0.000271 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

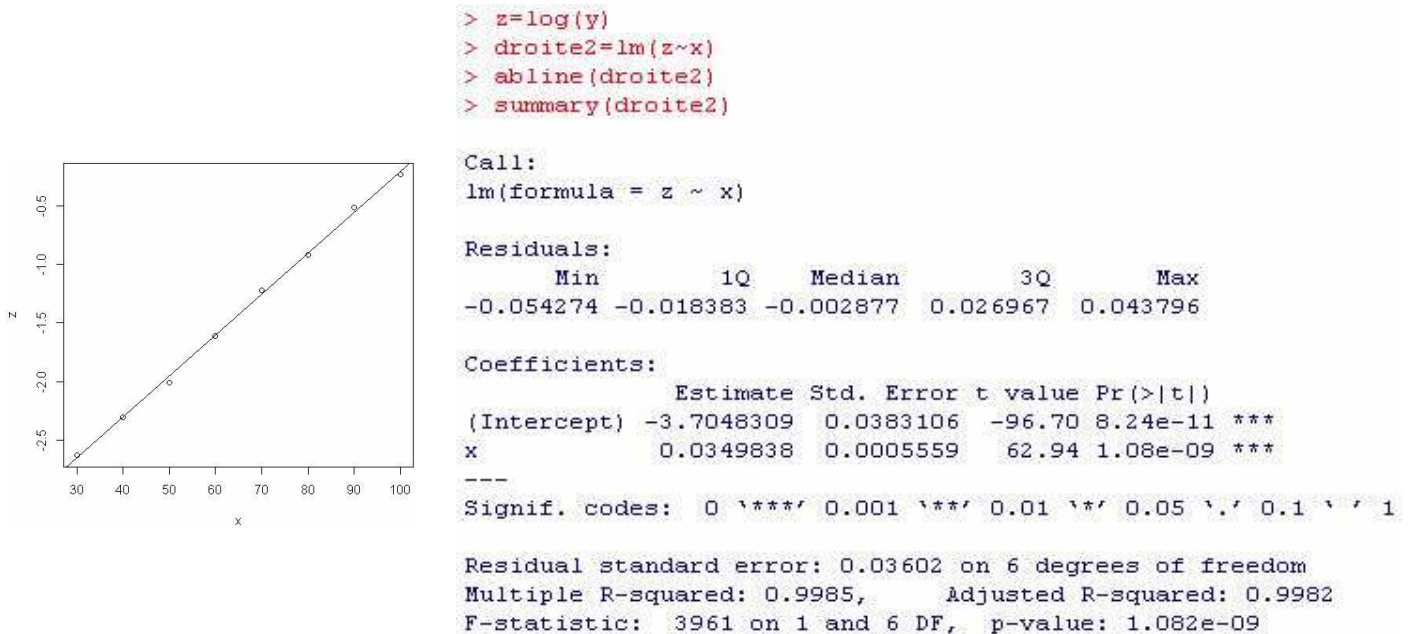
```
Residual standard error: 0.08547 on 6 degrees of freedom
Multiple R-squared:  0.9058,    Adjusted R-squared:  0.8901
F-statistic: 57.71 on 1 and 6 DF,  p-value: 0.0002708
```

c) Expliquer ce que réalisent les instructions saisies et indiquer ce que représentent les valeurs calculées ci-dessous.

```
> confint(droite1)
                2.5 %      97.5 %
(Intercept) -0.550152475 -0.10532372
x            0.006791921  0.01324617
> predict(droite1)
      1      2      3      4      5      6
-0.02716667  0.07302381  0.17321429  0.27340476  0.37359524  0.47378571
      7      8
 0.57397619  0.67416667
> residuals(droite1)[5]
      5
-0.07859524
```

d) Détailler les calculs menés pour obtenir les valeurs 0,006791921 et 0,01324617 de la 4ème ligne ci-dessus. Quelle hypothèse devrait être vérifiée pour que ces résultats soient valides ?

2) On décide d'effectuer une transformation de la variable y afin d'améliorer l'ajustement au nuage de points. On obtient les résultats suivants :



a) Exprimer la nouvelle variable z en fonction de y et donner (à partir des résultats ci-dessus) une équation de la droite de régression de z en x, ainsi que le coefficient de corrélation linéaire de z en x. Justifier que l'ajustement est meilleur que celui réalisé au 1).

b) En déduire une nouvelle expression de y en fonction de x.

c) Le coefficient directeur de la droite de régression de z en x représente ainsi la vitesse spécifique de croissance exponentielle (exprimée par minute) de la population microbienne étudiée. A partir du tableau ci-dessous, identifier l'espèce microbienne présente dans la préparation réalisée. Justifier la réponse.

Espèce microbienne	Treponema pallidum	S. cerevisiae	E. coli
Vitesse spécifique de croissance exponentielle (par heure)	0,02	0,35	2,10

Formulaire de Statistique Inférentielle

1) Estimateurs

Paramètre	Estimateur	Statistique et sa loi
μ	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$	$T = \frac{\bar{X} - \mu}{\frac{S_c}{\sqrt{n}}} : \begin{cases} \text{Student à } n - 1 \text{ d.d.l.} \\ \text{si échantillon gaussien} \end{cases}$
σ^2	$S_c^2 = \frac{n}{n-1} S^2$, avec $S^2 = \left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right) - (\bar{X})^2$	$Y^2 = \frac{n-1}{\sigma^2} S_c^2 : \begin{cases} \text{Khi deux à } n - 1 \text{ d.d.l.} \\ \text{si échantillon gaussien} \end{cases}$
p	$F = \frac{\sum_{i=1}^n X_i}{n}$	$U = \frac{F - p}{\sqrt{\frac{p(1-p)}{n}}} : \begin{cases} \text{Normale } \mathcal{N}(0; 1) \text{ (approx.)} \\ \text{si } np \geq 10 \text{ et } n(1-p) \geq 10 \end{cases}$

2) Intervalles de confiance au niveau $1 - \alpha$

Paramètre	Intervalle de confiance	Valeurs tabulées
μ	$i_\mu = \left[\bar{x} - \frac{S_c}{\sqrt{n}} t_\alpha, \bar{x} + \frac{S_c}{\sqrt{n}} t_\alpha \right]$	t_α tel que $P(-t_\alpha < T < t_\alpha) = 1 - \alpha$
σ^2	$i_{\sigma^2} = \left[\frac{n-1}{b_\alpha} S_c^2, \frac{n-1}{a_\alpha} S_c^2 \right]$	a_α et b_α tels que $\begin{aligned} P(Y^2 \geq a_\alpha) &= 1 - \frac{\alpha}{2} \\ P(Y^2 \geq b_\alpha) &= \frac{\alpha}{2} \end{aligned}$
p	$i_p = \left[f - \sqrt{\frac{f(1-f)}{n-1}} u_\alpha, f + \sqrt{\frac{f(1-f)}{n-1}} u_\alpha \right]$	u_α tel que $P(-u_\alpha < U < u_\alpha) = 1 - \alpha$

3) Tests de conformité au risque α

H_0	H_1	Statistique de test	Valeur(s) test(s)
$\mu = \mu_0$	$\mu \neq \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$T = \frac{\bar{X} - \mu_0}{\frac{S_c}{\sqrt{n}}}$	t_α tel que $P(-t_\alpha < T < t_\alpha) = 1 - \alpha$ t'_α tel que $P(T < t'_\alpha) = 1 - \alpha$, i.e. $t'_\alpha = t_{2\alpha}$ t''_α tel que $P(T \geq t''_\alpha) = 1 - \alpha$, i.e. $t''_\alpha = t_{2-2\alpha} = -t_{2\alpha}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$ $\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$	$Y^2 = \frac{n-1}{\sigma_0^2} S_c^2$	a_α et b_α tels que $P(a_\alpha < Y^2 < b_\alpha) = 1 - \alpha$ b'_α tel que $P(Y^2 \geq b'_\alpha) = \alpha$, i.e. $b'_\alpha = b_{2\alpha}$ a''_α tel que $P(Y^2 \geq a''_\alpha) = 1 - \alpha$, i.e. $a''_\alpha = a_{2\alpha}$
$p = p_0$	$p \neq p_0$ $p > p_0$ $p < p_0$	$U = \frac{F - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	u_α tel que $P(-u_\alpha < U < u_\alpha) = 1 - \alpha$ u'_α tel que $P(U < u'_\alpha) = 1 - \alpha$, i.e. $u'_\alpha = u_{2\alpha}$ u''_α tel que $P(U \geq u''_\alpha) = 1 - \alpha$, i.e. $u''_\alpha = -u_{2\alpha}$

Pour un intervalle de confiance de μ et/ou un test de conformité sur μ avec un grand échantillon (quelconque), on peut approcher la loi de Student par la loi Normale $\mathcal{N}(0; 1)$, et remplacer t_α , t'_α et t''_α par u_α , u'_α et u''_α .

4) Tests d'homogénéité au risque α

H_0	H_1	Statistique de test et sa loi sous l'hypothèse H_0	Valeur(s) test(s)
$\sigma_1 = \sigma_2$	$\sigma_1 \neq \sigma_2$	$F = \frac{S_{c,1}^2}{S_{c,2}^2}$: Snédécour à $(n_1 - 1, n_2 - 1)$ d.d.l. si échantillons indépendants gaussiens	f_α tel que $P(F \geq f_\alpha) = \frac{\alpha}{2}$ en travaillant avec $f \geq 1$
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$ $\mu_1 > \mu_2$ $\mu_1 < \mu_2$	$U = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{S_{c,1}^2}{n_1} + \frac{S_{c,2}^2}{n_2}\right)}}$: Normale $\mathcal{N}(0; 1)$ (approx.) si grands échantillons indépendants	u_α u'_α u''_α
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$ $\mu_1 > \mu_2$ $\mu_1 < \mu_2$	$T = \frac{\bar{X}_1 - \bar{X}_2}{s_{c,1,2} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$: Student à $n_1 + n_2 - 2$ d.d.l. (approx.) si petits échantillons indép. gaussiens et si $\sigma_1 = \sigma_2$ avec $s_{c,1,2}^2 = \frac{(n_1-1)s_{c,1}^2 + (n_2-1)s_{c,2}^2}{n_1+n_2-2}$	t_α t'_α t''_α
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$ $\mu_1 > \mu_2$ $\mu_1 < \mu_2$	$U = \frac{\bar{D}}{\frac{S_{c,d}}{\sqrt{n}}}$, où $D = X_1 - X_2$: Normale $\mathcal{N}(0; 1)$ (approx.) si grands échantillons appariés	u_α u'_α u''_α
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$ $\mu_1 > \mu_2$ $\mu_1 < \mu_2$	$T = \frac{\bar{D}}{\frac{S_{c,d}}{\sqrt{n}}}$, où $D = X_1 - X_2$: Student à $n - 1$ d.d.l. si petits échantillons appariés gaussiens	t_α t'_α t''_α
$p_1 = p_2$	$p_1 \neq p_2$ $p_1 > p_2$ $p_1 < p_2$	$U = \frac{F_1 - F_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)f_{1,2}(1 - f_{1,2})}}$: Normale $\mathcal{N}(0; 1)$ (approx.) si $n_1 f_1 \geq 5, n_1(1 - f_1) \geq 5,$ $n_2 f_2 \geq 5, n_2(1 - f_2) \geq 5,$ avec $f_{1,2} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$	u_α u'_α u''_α

5) Test d'ajustement à une loi théorique à r modalités au risque α

Hypothèse H_0 : le caractère suit la loi théorique définie par les probabilités p_i .

Hypothèse H_1 : $\overline{H_0}$.

Statistique de test : $D = \sum_{i=1}^r \frac{(N_i - np_i)^2}{np_i}$.

Loi de D sous l'hypothèse H_0 : khi deux à $r - 1 - k$ d.d.l.

Valeur test : b_α tel que $P(D \geq b_\alpha) = \alpha$.

6) Test d'indépendance entre deux caractères à r et s modalités au risque α

Hypothèse H_0 : les deux caractères sont indépendants.

Hypothèse H_1 : $\overline{H_0}$.

Statistique de test : $D = \sum_{i=1}^r \sum_{j=1}^s \frac{(N_{ij} - np_{ij})^2}{np_{ij}}$, avec $np_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}$, $n_{i\bullet} = \sum_{j=1}^s n_{ij}$ et $n_{\bullet j} = \sum_{i=1}^r n_{ij}$.

Loi de D sous l'hypothèse H_0 : khi deux à $(r - 1)(s - 1)$ d.d.l.

Valeur test : b_α tel que $P(D \geq b_\alpha) = \alpha$.

7) Régression linéaire simple

On considère deux variables quantitative X et Y et le modèle de régression $Y = \alpha X + \beta + \varepsilon$, où ε suit la loi normale $\mathcal{N}(0; \sigma)$. On désigne par $\rho = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$ le coefficient de corrélation linéaire entre X et Y .

Droite des moindres carrés de y en x .

Droite d'équation $y = ax + b$, avec $a = \frac{\text{cov}(x, y)}{s_x^2}$, $b = \bar{y} - a\bar{x}$,

et coefficient de corrélation linéaire r entre x et y avec : $r = \frac{\text{cov}(x, y)}{s_x s_y}$

où $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, $s_x^2 = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2$, $s_y^2 = \left(\frac{1}{n} \sum_{i=1}^n y_i^2 \right) - \bar{y}^2$ et $\text{cov}(x, y) = \left(\frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \bar{x}\bar{y}$.

Effectuant plusieurs expériences, et ainsi plusieurs échantillonnages, a , b et r apparaissent comme les valeurs observées des variables aléatoires A , B et R .

On a alors $E(A) = \alpha$, $E(B) = \beta$ et $E(R) \simeq \rho + \frac{\rho(1 - \rho^2)}{2(n - 1)}$. Des estimations ponctuelles de α , β , σ^2 et ρ sont alors respectivement a , b , $s_R^2 = \frac{n}{n - 2} (1 - r^2) s_y^2$ et r (ou plus précisément $r + \frac{r(1 - r^2)}{2(n - 3)}$).

Intervalle de confiance et tests pour un coefficient de régression linéaire α

La situation est celle décrite dans le paragraphe 1.

La variance de A est égale à $\frac{\sigma^2}{ns_x^2}$, et peut être estimée par $s_A^2 = \frac{s_R^2}{ns_x^2}$. La variable aléatoire $T = \frac{A - \alpha}{s_A}$ suit la loi de Student à $n - 2$ degrés de liberté. On détermine alors le réel t_{α_1} tel que $P(-t_{\alpha_1} < T < t_{\alpha_1}) = 1 - \alpha_1$ (table 3). On en déduit un intervalle de confiance de α au niveau $1 - \alpha_1$:

$$i_\alpha = [a - s_A t_{\alpha_1}, a + s_A t_{\alpha_1}].$$

Test (bilatéral) de $H_0 : \alpha = \alpha_0$ contre $H_1 : \alpha \neq \alpha_0$.

On calcule $t = \frac{a - \alpha_0}{s_A}$ et on décide que :

- si $t \in]-t_{\alpha_1}, t_{\alpha_1}[$, alors on ne peut rejeter H_0 ;

- si $t \notin]-t_{\alpha_1}, t_{\alpha_1}[$, alors on rejette H_0 avec une probabilité α_1 de se tromper.

Cas de β . On peut mener une étude analogue pour β en utilisant le fait que la variance de B est égale à $\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{ns_x^2} \right)$, et peut être estimée par $s_B^2 = s_R^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{ns_x^2} \right) = \frac{s_R^2}{n^2 s_x^2} \sum_{i=1}^n x_i^2$, et que la variable aléatoire $T = \frac{B - \beta}{s_B}$ suit la loi de Student à $n - 2$ degrés de liberté.

Prévision. L'ajustement affine peut servir à prévoir la valeur attendue pour Y quand l'expérimentateur fixe $X = x_0$. L'estimation ponctuelle de cette valeur est $\hat{y}_0 = ax_0 + b$, et un intervalle de confiance de cette valeur au niveau $1 - \alpha_1$: $i_{y_0} = \left[\hat{y}_0 - t_{\alpha_1} \sqrt{s_R^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{ns_x^2} \right)}, \hat{y}_0 + t_{\alpha_1} \sqrt{s_R^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{ns_x^2} \right)} \right]$.

Intervalle de confiance et tests pour un coefficient de corrélation

Soit ζ le nombre défini par $\zeta = \frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right) = \text{argth} \rho$.

Soit Z la variable aléatoire qui au cours de chaque échantillonnage prend la valeur $z = \frac{1}{2} \ln \left(\frac{1 + r}{1 - r} \right) = \text{argthr}$. Lorsque n est assez grand ($n \geq 20$ en pratique), un intervalle de confiance de ζ au niveau $1 - \alpha$:

$$i_\zeta = \left[z - \frac{u_\alpha}{\sqrt{n - 3}}, z + \frac{u_\alpha}{\sqrt{n - 3}} \right] = [z_1, z_2].$$

D'où un intervalle de confiance de ρ au niveau $1 - \alpha$: $i_\rho = [r_1, r_2] = [\text{th} z_1, \text{th} z_2]$.

Test de $H_0 : \rho = 0$ contre $H_1 : \rho \neq 0$.

Sous l'hypothèse H_0 , $T = \frac{R\sqrt{n - 2}}{\sqrt{1 - R^2}}$ suit la loi de Student à $n - 2$ degrés de liberté. On calcule $t = \frac{r\sqrt{n - 2}}{\sqrt{1 - r^2}}$:

- si $t \in]-t_\alpha, t_\alpha[$, alors on ne peut rejeter H_0 ;

- si $t \notin]-t_\alpha, t_\alpha[$, alors on rejette H_0 avec une probabilité α de se tromper.

TABLE 1

**Fonction de répartition
de la loi normale réduite**

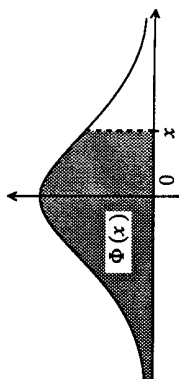
Si U suit la loi normale réduite, pour $x \geq 0$, la table donne la valeur :

$$\phi(x) = P(U \leq x).$$

La valeur x s'obtient par addition des nombres inscrits en marge.

Pour $x < 0$, on a :

$$\phi(x) = 1 - \phi(-x).$$



x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

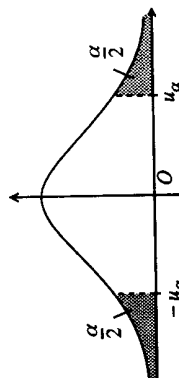
TABLE 2

**Loi normale réduite
(table de l'écart réduit)**

Si U est une variable aléatoire qui suit la loi normale réduite, la table donne, pour α choisi, la valeur u_α telle que :

$$P(|U| \geq u_\alpha) = \alpha.$$

La valeur α s'obtient par addition des nombres inscrits en marge.

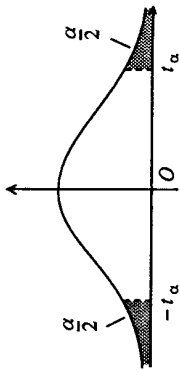


α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,1	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,2	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,3	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,4	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,5	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,6	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,7	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,8	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,9	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

TABLE 3

Lois de Student

Si T est une variable aléatoire qui suit la loi de Student à v degrés de liberté, la table donne, pour α choisi, le nombre t_α tel que $P(T \geq t_\alpha) = \alpha$.



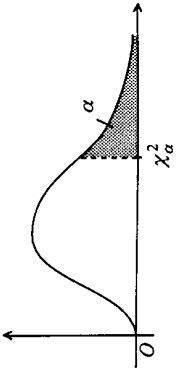
α v	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	1,000	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,816	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,765	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,134	0,741	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,727	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,718	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,711	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,706	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,703	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,700	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,697	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,695	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,694	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,692	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,691	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,690	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,689	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,688	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,688	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,687	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,686	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,686	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,685	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,685	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,684	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,684	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,684	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,683	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,683	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,683	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,681	1,050	1,303	1,684	2,021	2,423	2,704	3,551
80	0,126	0,679	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,677	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,674	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Lorsque le degré de liberté est infini, il s'agit du nombre u_α correspondant à la loi normale centrée réduite (cf. table 2).

TABLE 4

Lois de Pearson ou lois du χ^2

Si Y^2 est une variable aléatoire qui suit la loi du χ^2 à v degrés de liberté, la table donne, pour α choisi, le nombre χ_α^2 tel que $P(Y^2 \geq \chi_\alpha^2) = \alpha$.



α v	0,99	0,975	0,95	0,90	0,10	0,05	0,025	0,01	0,001
1	0,000 2	0,001	0,004	0,016	2,71	3,84	5,02	6,63	10,83
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	13,82
3	0,12	0,22	0,35	0,58	6,25	7,81	9,35	11,34	16,27
4	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	18,47
5	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	20,52
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	22,46
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,47	24,32
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	26,13
9	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	27,88
10	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	29,59
11	3,05	3,82	4,57	5,58	17,27	19,67	21,92	24,72	31,26
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	32,91
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	34,53
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	36,12
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	37,70
16	5,81	6,91	7,96	9,31	23,54	26,30	28,84	32,00	39,25
17	6,41	7,56	8,67	10,08	24,77	27,59	30,19	33,41	40,79
18	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,80	42,31
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	43,82
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	45,32
21	8,90	10,28	11,59	13,24	29,61	32,67	35,48	38,93	46,80
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	48,27
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64	49,73
24	10,86	12,40	13,85	15,66	33,20	36,41	39,37	42,98	51,18
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31	52,62
26	12,20	13,84	15,38	17,29	35,56	38,88	41,92	45,64	54,05
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96	55,48
28	13,57	15,31	16,93	18,94	37,92	41,34	44,46	48,28	56,89
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59	58,30
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89	59,70

Lorsque le degré de liberté v est tel que $v > 30$, la variable aléatoire :

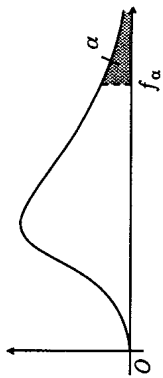
$$U = \sqrt{2Y^2} - \sqrt{2v - 1}$$

suit à peu près la loi normale réduite.

TABLE 5

Lois de Snédécór ($\alpha = 0,025$)

Si F est une variable aléatoire qui suit la loi de Snédécór à (v_1, v_2) degrés de liberté, la table donne le nombre f_α tel que $P(F \geq f_\alpha) = \alpha = 0,025$.

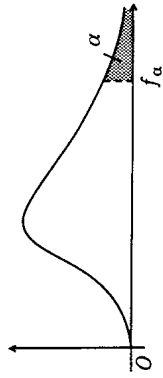


$v_1 \backslash v_2$	1	2	3	4	5	6	8	10	15	20	30	∞
1	648	800	864	900	922	937	957	969	985	993	1001	1018
2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4	39,4	39,4	39,5	39,5
3	17,4	16,0	15,4	15,1	14,9	14,7	14,5	14,4	14,3	14,2	14,1	13,9
4	12,2	10,6	9,98	9,60	9,36	9,20	8,98	8,84	8,66	8,56	8,46	8,26
5	10,0	8,43	7,76	7,39	7,15	6,98	6,76	6,62	6,43	6,33	6,23	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,60	5,46	5,27	5,17	5,07	4,85
7	8,07	6,54	5,89	5,52	5,29	5,12	4,90	4,76	4,57	4,47	4,36	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,43	4,30	4,10	4,00	3,89	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,10	3,96	3,77	3,67	3,56	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,85	3,72	3,52	3,42	3,31	3,08
11	6,72	5,26	4,63	4,28	4,04	3,88	3,66	3,53	3,33	3,23	3,12	2,88
12	6,55	5,10	4,47	4,12	3,89	3,73	3,51	3,37	3,18	3,07	2,96	2,72
13	6,41	4,97	4,35	4,00	3,77	3,60	3,39	3,25	3,05	2,95	2,84	2,60
14	6,30	4,86	4,24	3,89	3,66	3,50	3,29	3,15	2,95	2,84	2,73	2,49
15	6,20	4,76	4,15	3,80	3,58	3,41	3,20	3,06	2,86	2,76	2,64	2,40
16	6,12	4,69	4,08	3,73	3,50	3,34	3,12	2,99	2,79	2,68	2,57	2,32
17	6,04	4,62	4,01	3,66	3,44	3,28	3,06	2,92	2,72	2,62	2,50	2,25
18	5,98	4,56	3,95	3,61	3,38	3,22	3,01	2,87	2,67	2,56	2,44	2,19
19	5,92	4,51	3,90	3,56	3,33	3,17	2,96	2,82	2,62	2,51	2,39	2,13
20	5,87	4,46	3,86	3,51	3,29	3,13	2,91	2,77	2,57	2,46	2,35	2,09
22	5,79	4,38	3,78	3,44	3,22	3,05	2,84	2,70	2,50	2,39	2,27	2,00
24	5,72	4,32	3,72	3,38	3,15	2,99	2,78	2,64	2,44	2,33	2,21	1,94
26	5,66	4,27	3,67	3,33	3,10	2,94	2,73	2,59	2,39	2,28	2,16	1,88
28	5,61	4,22	3,63	3,29	3,06	2,90	2,69	2,55	2,34	2,23	2,11	1,83
30	5,57	4,18	3,59	3,25	3,03	2,87	2,65	2,51	2,31	2,20	2,07	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,53	2,39	2,18	2,07	1,94	1,64
50	5,34	3,98	3,39	3,06	2,83	2,67	2,46	2,32	2,11	1,99	1,87	1,55
60	5,29	3,93	3,34	3,01	2,79	2,63	2,41	2,27	2,06	1,94	1,82	1,48
80	5,22	3,86	3,28	2,95	2,73	2,57	2,36	2,21	2,00	1,88	1,75	1,40
100	5,18	3,83	3,25	2,92	2,70	2,54	2,32	2,18	1,97	1,85	1,71	1,35
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,19	2,05	1,83	1,71	1,57	1,00

TABLE 6

Lois de Snédécór ($\alpha = 0,05$)

Si F est une variable aléatoire qui suit la loi de Snédécór à (v_1, v_2) degrés de liberté, la table donne le nombre f_α tel que $P(F \geq f_\alpha) = \alpha = 0,05$.



$v_1 \backslash v_2$	1	2	3	4	5	6	8	10	15	20	30	∞
1	161	200	216	225	230	234	239	242	246	248	250	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,85	8,79	8,70	8,66	8,62	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96	5,86	5,80	5,75	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	4,62	4,56	4,50	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06	3,94	3,87	3,81	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,64	3,51	3,44	3,38	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,35	3,22	3,15	3,08	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,14	3,01	2,94	2,86	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,98	2,85	2,77	2,70	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,85	2,72	2,65	2,57	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,85	2,75	2,62	2,54	2,47	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,77	2,67	2,53	2,46	2,38	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,60	2,46	2,39	2,31	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,54	2,40	2,33	2,25	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,49	2,35	2,28	2,19	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,45	2,31	2,23	2,15	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,41	2,27	2,19	2,11	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,38	2,23	2,16	2,07	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,35	2,20	2,12	2,04	1,84
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,30	2,15	2,07	1,98	1,78
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,25	2,11	2,03	1,94	1,73
26	4,23	3,37	2,98	2,74	2,59	2,47	2,32	2,22	2,07	1,99	1,90	1,69
28	4,20	3,34	2,95	2,71	2,56	2,45	2,29	2,19	2,04	1,96	1,87	1,65
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,16	2,01	1,93	1,84	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,08	1,92	1,84	1,74	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	2,03	1,87	1,78	1,69	1,44
60	4,00	3,15	2,76	2,53	2,37	2,25	2,10	1,99	1,84	1,75	1,65	1,39
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,95	1,79	1,70	1,60	1,32
100	3,94	3,09	2,70	2,46	2,31	2,19	2,03	1,93	1,77	1,68	1,57	1,28
∞	3,84	3,00	2,60	2,37	2,21	2,10	1,94	1,83	1,67	1,57	1,46	1,00