

Module 7 - Estimation ponctuelle et lois échantillonales

STT-1000 Probabilités et statistique

Département de mathématiques et de statistique



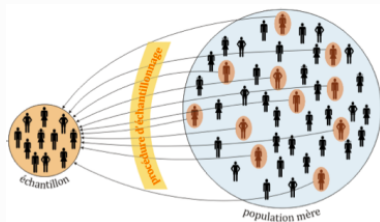
UNIVERSITÉ
LAVAL

Concepts d'inférence statistique

L'inférence statistique

Tirer des conclusions au sujet d'une population à l'aide d'un échantillon.

- Faire rouler sur un circuit 20 voitures neuves prises au hasard afin d'étudier la fiabilité de la suspension.
- Sonder une population d'acheteurs potentiels d'un nouveau produit sur l'attrait d'une option.



Deux types d'inférence

Estimation de paramètres	Vérification d'hypothèses
Estimation ponctuelle : • μ estimé par... • σ^2 estimé par...	Tests d'hypothèses : • A-t-on $\mu \leq 60$? • A-t-on $\sigma^2 < 10$? • A-t-on $\mu_1 = \mu_2 = \mu_3$?
Intervalles de confiance : associer une marge d'erreur aux estimations ponctuelles	

Modèle, loi, paramètres



On suppose un **modèle probabiliste** pour la population (une loi, avec des **paramètres inconnus**, ex. μ et σ^2). À partir d'un **échantillon aléatoire**, on calcule des **paramètres estimés** (ex. $\bar{x}$ et s^2).

Exemple 1 : masse des camions

Population :	l'ensemble des camions qui passent sur un viaduc
Caractéristique d'intérêt :	X : masse d'un camion
Modèle probabiliste :	$X \sim \mathcal{N}(\mu, \sigma^2)$
Paramètre d'intérêt :	μ : espérance de la masse des camions sur ce viaduc
Taille d'échantillon :	$n = 25$
Échantillon aléatoire :	$X_1, X_2, \dots, X_{25}$
Échantillon observé :	$x_1, x_2, \dots, x_{25}$
Estimation ponctuelle :	$\bar{x} = 16,5$ tonnes

Estimation ponctuelle

Paramètres classiques : μ et σ^2

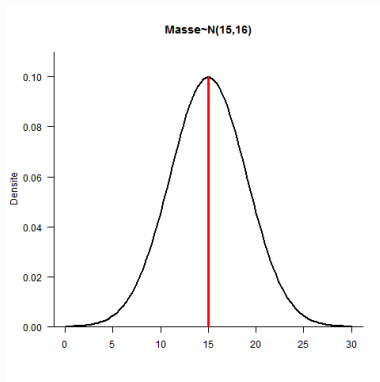
- Par quelles quantités pourrait-on estimer l'espérance et la variance de la population ?

Par $\bar{X}$ et S^2

- Ces mesures échantillonnelles sont-elles des variables aléatoires ?
- Elles ont donc une loi de probabilité...

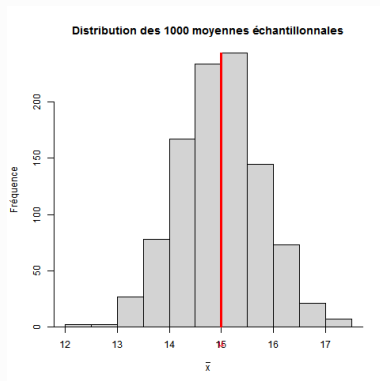
Exemple 2 : Simulation de données

Prenons un échantillon de 25 camions et mesurons leur masse (distribuée selon le modèle normal). On s'attend à ce que $\bar{x}$ diffère un peu de μ .



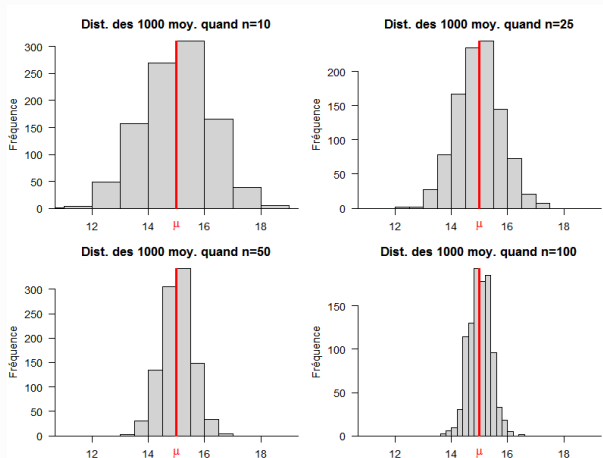
Exemple 2 : Et si on recommençait 1000 fois ?

Comment seraient distribuées nos 1000 moyennes (chaque fois avec $n = 25$) ?



Exemple 2 : Si on changeait la taille d'échantillon ?

Qu'advierait-il de la distribution des $\bar{x}$ avec un n différent ?



Qualités d'un estimateur

Le choix de $\bar{X}$ et S^2 est-il le meilleur pour estimer μ et σ^2 ?

Quelles qualités devrait avoir un bon estimateur ?

1. Être **sans biais**, i.e. avoir une espérance égale au paramètre estimé.
2. Avoir une **variance faible**, idéalement qui diminue lorsque n grandit.

Estimation de l'espérance μ par $\bar{X}$

Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes d'espérance μ et de variance σ^2 .

Alors $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ est-il un bon estimateur de μ ?

- $\bar{X}$ est-il sans biais ?
- $V(\bar{X})$ diminue-t-elle quand n augmente ?

Estimation d'une variance σ^2 par S^2

Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes d'espérance μ et de variance σ^2 .

$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ est-il un bon estimateur de σ^2 ?

- S^2 est-il sans biais ?
- $V(S^2)$ diminue-t-elle quand n augmente ?

On montrera plus tard que si les X_i proviennent d'une loi normale, la variance de S^2 est

$$V(S^2) = \frac{2\sigma^4}{n-1}.$$

- L'espérance de la taille des hommes canadiens est de 1,78 m. Supposons $X \sim \mathcal{N}(1,78; \sigma^2)$.
- Jacques mesure 250 Canadiens choisis au hasard.
- Marie mesure 1000 Canadiens choisis au hasard.

Vrai ou Faux ?

Jacques a plus de chances que Marie d'obtenir...

- une taille moyenne supérieure à 1,78 m.
- une taille moyenne supérieure à 1,82 m.

Faux pour le premier énoncé : $P(\bar{X} > 1,78) = 0,5$ pour les deux, peu importe n (1,78 est le centre de la distribution de $\bar{X}$ dans les deux cas). **Vrai** pour le second : la distribution de $\bar{X}$ de Jacques ($n = 250$) est plus dispersée que celle de Marie ($n = 1000$), donc $P(\bar{X} > 1,82)$ est plus grande pour Jacques.

Lois échantillonnales

Distribution de certaines statistiques

- Les estimateurs sont des variables aléatoires.
- Leur distribution dépend souvent de la loi de X dans la population.
- Pour associer une marge d'erreur à un estimateur, on doit trouver sa loi.
- Les lois qui impliquent des estimateurs sont appelées des *lois échantillonnales*. Nous en verrons trois :
 - Normale $\mathcal{N}(\mu, \sigma^2/n)$
 - Khi-carré χ_v^2
 - Student t_v

Distribution de certaines statistiques

Dans le cas où $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ indépendantes (i.i.d.), on connaît la loi de plusieurs statistiques.

- $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
- $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$
- $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$

Loi Normale

Propriétés de la loi normale

- Si $X_1, \dots, X_n$ sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, alors

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

- En standardisant,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

Table de la loi normale

Table II Les valeurs de la fonction de répartition de la loi normale centrée réduite*

		$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$									
		Direction des écarts de z									
z	0,00	0,04	0,08	0,12	0,16	0,20	0,24	0,28	0,32	0,36	0,40
0,0	0,50000	0,50399	0,50798	0,51197	0,51596	0,51994	0,52391	0,52789	0,53186	0,53583	0,53979
0,1	0,53983	0,54379	0,54776	0,55172	0,55569	0,55966	0,56362	0,56759	0,57156	0,57553	0,57949
0,2	0,57930	0,58327	0,58724	0,59120	0,59517	0,59914	0,60311	0,60708	0,61105	0,61502	0,61899
0,3	0,61878	0,62275	0,62672	0,63069	0,63466	0,63863	0,64260	0,64657	0,65054	0,65451	0,65848
0,4	0,65849	0,66246	0,66643	0,67040	0,67437	0,67834	0,68231	0,68628	0,69025	0,69422	0,69819
0,5	0,69814	0,70211	0,70608	0,71005	0,71402	0,71799	0,72196	0,72593	0,72990	0,73387	0,73784
0,6	0,73786	0,74183	0,74580	0,74977	0,75374	0,75771	0,76168	0,76565	0,76962	0,77359	0,77756
0,7	0,77759	0,78156	0,78553	0,78950	0,79347	0,79744	0,80141	0,80538	0,80935	0,81332	0,81729
0,8	0,81734	0,82131	0,82528	0,82925	0,83322	0,83719	0,84116	0,84513	0,84910	0,85307	0,85704
0,9	0,85709	0,86106	0,86503	0,86899	0,87296	0,87693	0,88090	0,88487	0,88884	0,89281	0,89678
1,0	0,89685	0,90082	0,90479	0,90876	0,91273	0,91670	0,92067	0,92464	0,92861	0,93258	0,93655
1,1	0,93659	0,94056	0,94453	0,94850	0,95247	0,95644	0,96041	0,96438	0,96835	0,97232	0,97629
1,2	0,97634	0,98031	0,98428	0,98825	0,99222	0,99619	0,99999	0,99999	0,99999	0,99999	0,99999
1,3	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
1,4	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
1,5	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
1,6	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
1,7	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
1,8	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
1,9	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
2,0	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
2,1	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
2,2	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
2,3	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
2,4	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
2,5	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
2,6	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
2,7	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
2,8	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
2,9	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
3,0	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
3,1	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
3,2	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
3,3	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
3,4	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
3,5	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
3,6	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
3,7	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
3,8	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999
3,9	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999	0,99999

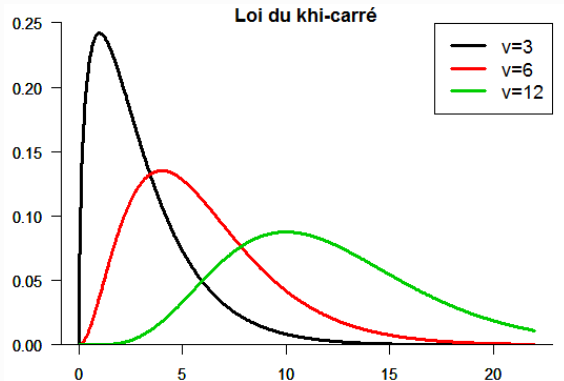


*Les valeurs figurant dans cette table sont celles de $\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$, où $Z = -\infty$ à $Z = +\infty$. La première décimale de z est indiquée dans la 1^{re} colonne, et sa deuxième décimale est indiquée sur la 1^{re} ligne. (Par exemple, $\Phi(1,23) = 0,89065$.)

Loi du khi-carré

Loi du khi-deux (ou loi du khi-carré)

- valeurs possibles d'une variable suivant une $\chi_v^2 : [0, \infty)$
- un seul paramètre qu'on appelle "degrés de liberté"



Propriétés de la loi du khi-carré

- Si $K \sim \chi_v^2$, alors $\begin{cases} E(K) = v \\ V(K) = 2v \end{cases}$
- Si $Z \sim \mathcal{N}(0, 1)$, alors $Z^2 \sim \chi_1^2$
- Si $K_1, \dots, K_n$ sont des v.a. indépendantes et que $K_i \sim \chi_{v_i}^2$, alors

$$K_1 + \dots + K_n \sim \chi_{v_1 + \dots + v_n}^2$$

- Si $X_1, \dots, X_n$ sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, alors

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi_n^2 \quad \text{et} \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Table de la loi du khi-deux

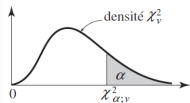


Table III Les quantiles de la loi du khi-carré*

α v	Probabilité										
	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
1	0.00+	0.00+	0.00+	0.00+	0.02	0.45	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	1.39	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99

Degrés de liberté

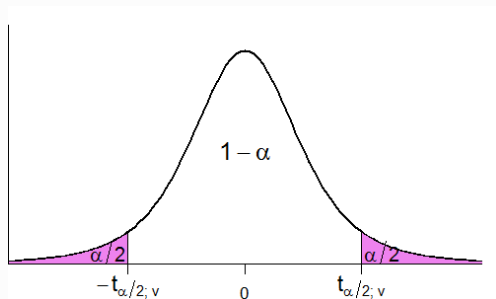
Exemple 3

- Supposons que la température d'un four à peinture 30 minutes après l'allumage à 200°C suit une loi $\mathcal{N}(197, 38)$.
- Lors de 20 jours différents, on mesure la température après 30 minutes.
- Quelle est la probabilité que l'écart-type échantillonnal des mesures excède $7,38^{\circ}\text{C}$?

Loi de Student

Loi de Student (loi t_v)

- ressemble à la $\mathcal{N}(0, 1)$, sauf qu'elle a des queues plus lourdes (des valeurs éloignées de 0 sont plus probables que sous la loi $\mathcal{N}(0, 1)$).
- un seul paramètre appelé "degrés de liberté" (plus le nombre de degrés de liberté est faible, plus les queues sont lourdes).



Propriétés de la loi t

- Si $T \sim t_\nu$ (et $\nu > 1$), alors $E(T) = 0$.
- Si $T \sim t_\nu$ (et $\nu > 2$), alors $V(T) = \frac{\nu}{\nu - 2}$.
- Si $Z \sim \mathcal{N}(0, 1)$, $K \sim \chi_\nu^2$, Z et K sont indépendantes,

$$\text{alors } \frac{Z}{\sqrt{K/\nu}} \sim t_\nu$$

- Si $X_1, \dots, X_n$ sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$, alors

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Table de la loi de Student

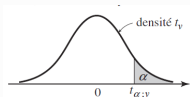


Table IV Les quantiles de la loi t de Student*

α v	Probabilité									
	0,40	0,25	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	0,325	1,000	3,078	6,314	12,706	31,821	63,657	127,32	318,31	636,62
2	0,289	0,816	1,886	2,920	4,303	6,965	9,925	14,089	23,326	31,598
3	0,277	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,213	12,924
4	0,271	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,267	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869
6	0,265	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,263	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,262	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,261	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,260	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,260	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,259	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,259	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,258	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,258	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,258	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,257	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	0,257	0,688	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	0,257	0,688	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	0,257	0,687	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21	0,257	0,686	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22	0,256	0,686	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	0,256	0,685	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,767
24	0,256	0,685	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	0,256	0,684	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26	0,256	0,684	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	0,256	0,684	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,690
28	0,256	0,683	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674

Degrés de liberté

Exemple 4

On effectue 9 forages sur un terrain afin de déterminer la profondeur à laquelle un certain minerais peut être extrait.

- X_i = profondeur du minerais lors du forage i .
- On suppose que les X_i sont i.i.d. $\mathcal{N}(\mu, \sigma^2)$.

Quelle est la probabilité que la moyenne échantillonnale soit supérieure à la vraie moyenne μ par plus d'un écart-type échantillonnal ?

Réponses des exemples

Réponses des exemples

- **Exemple 3 (four à peinture, $X \sim \mathcal{N}(197, 38)$, $n = 20$) :** $\frac{19S^2}{38} \sim \chi_{19}^2$;
 $S > 7,38 \Leftrightarrow \chi_{19}^2 > 19 \times 7,38^2 / 38 \approx 27,23$, **donc** $P(S > 7,38) \approx 0,10$ (**table du khi-carré, 19 d.l., puisque** $\chi_{0,10;19}^2 = 27,204$).
- **Exemple 4 (9 forages, X_i i.i.d. $\mathcal{N}(\mu, \sigma^2)$) :** $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_8$; l'événement $\bar{X} - \mu > S$ équivaut à $T > \sqrt{9} = 3$, **donc** $P(T_8 > 3)$ **est entre 0,005 et 0,01 (table de Student, 8 d.l.), soit environ 0,01.**

Bibliographie

Ce document est rédigé avec Beamer ainsi qu'une classe .cls fournie par Jérôme Soucy. Les diapositives sont adaptées des diapositives créées par Thierry Duchesne et Emmanuelle Reny-Nolin pour le cours STT-1900. Les graphiques proviennent de ces mêmes diapositives.

Pour signaler une erreur dans ce document, veuillez écrire à l'adresse : julien.miron@mat.ulaval.ca.