Tests paramétriques - Variables aléatoires continues

Grands échantillons

Yohann.Foucher@univ-nantes.fr

Equipe d'Accueil 4275 "Biostatistique, recherche clinique et mesures subjectives en santé", Université de Nantes

Ecole Doctorale - Module Biostatistique I, Mardi 27 Mars 2012

UNIVERSITÉ DE NANTES

CENTRE HOSPITALIER UNIVERSITAIRE DE NANTES

itun institut transplantation urologie néphrologie INSERM - UMR 643

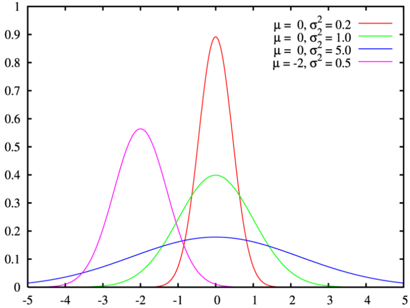
Plan

1. Rappels
   - La loi normale
   - Les tests d'hypothèses
2. Comparaison avec une moyenne de référence
   - Bilatéral
   - Unilatéral
3. Comparaison de deux moyennes (échantillons indépendants)
   - Bilatéral
   - Unilatéral
4. Comparaison de deux moyennes (échantillons appariés)

Plan

Loi normale

- Caractérisée par sa moyenne (μ) et son écart-type (σ) : $X \sim N(\mu, \sigma)$
- Loi normale ou loi de Laplace-Gauss
- La distribution est symétrique : Mode $=$ Moyenne $=$ Médiane

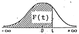


Loi normale

- Si $X \sim N(\mu, \sigma)$ alors $T = (X - \mu)/\sigma \sim N(0, 1)$.
- $N(0, 1)$ est la loi normale centrée et réduite.
- La probabilité $P(T < t)$ est donnée dans la table :

FONCTION DE REPARTITION F(t)

DE LA LOI NORMALE CENTREE, REDUITE



$$F(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-t^2/2}\, dt.$$

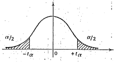
t	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	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	0,7290	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	0,9779	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 pour les grandes valeurs de t

t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
F(t)	0,99865	0,99904	0,99931	0,99952	0,99966	0,99976	0,999841	0,999928	0,999968	0,999997

TABLE DE STUDENT

BILATERALE



ν \ α	0,10	0,05	0,02	0,01
1	6,314	12,706	31,821	63,657
2	2,920	4,303	6,965	9,925
3	2,353	3,182	4,541	5,841
4	2,132	2,776	3,747	4,604
5	2,015	2,571	3,365	4,032
6	1,943	2,447	3,143	3,707
7	1,895	2,365	2,998	3,499
8	1,860	2,306	2,896	3,355
9	1,833	2,262	2,821	3,250
10	1,812	2,228	2,764	3,169
11	1,796	2,201	2,718	3,106
12	1,782	2,179	2,681	3,055
13	1,771	2,160	2,650	3,012
14	1,761	2,145	2,624	2,977
15	1,753	2,131	2,602	2,947
16	1,746	2,120	2,583	2,921
17	1,740	2,110	2,567	2,898
18	1,734	2,101	2,552	2,878
19	1,729	2,093	2,539	2,861
20	1,725	2,086	2,528	2,845
21	1,721	2,080	2,518	2,831
22	1,717	2,074	2,508	2,819
23	1,714	2,069	2,500	2,807
24	1,711	2,064	2,492	2,797
25	1,708	2,060	2,485	2,787
26	1,706	2,056	2,479	2,779
27	1,703	2,052	2,473	2,771
28	1,701	2,048	2,467	2,763
29	1,699	2,045	2,462	2,756
30	1,697	2,042	2,457	2,750
∞	1,645	1,960	2,326	2,576

TABLE DU χ^2

ν \ α	.05	.02	.01
1	3,84	5,41	6,64
2	5,99	7,82	9,21
3	7,82	9,84	11,34
4	9,49	11,67	13,28
5	11,07	13,39	15,09
6	12,59	15,03	16,81
7	14,07	16,62	18,48
8	15,51	18,17	20,09
9	16,92	19,68	21,67
10	18,31	21,16	23,21
11	19,68	22,62	24,72
12	21,03	24,05	26,22
13	22,36	25,47	27,69
14	23,68	26,37	29,14
15	25,00	28,26	30,58
16	26,30	29,63	32,00
17	27,59	31,00	33,41
18	28,87	32,35	34,80
19	30,14	33,69	36,19
20	31,41	35,02	37,57
21	32,67	36,34	38,93
22	33,92	37,66	40,29
23	35,17	38,97	41,64
24	36,42	40,27	42,98
25	37,65	41,57	44,31
26	38,88	42,86	45,64
27	40,11	44,14	46,96
28	41,34	45,42	48,28
29	42,56	46,69	49,59
30	43,77	47,96	50,89

Loi normale

- Symétrie $\rightarrow P(T > t) = 1 - P(T \leq t) = P(T < -t)$.
- Valeurs importantes à retenir :
  - $P(-1,96 < T < 1,96) = 0,95$
  - $P(T < -1,64) = P(T > 1,64) = 0,95$
- Si $X \sim N(\mu, \sigma)$ alors

$$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$$

- A partir du théorème Central Limite ($n > 30$), quelque soit la distribution de X, on admet que :

$$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$$

Les tests d'hypothèses

- Principe de la statistique : Est ce que la différence que j'observe est due à la fluctuation d'échantillonnage ?
- Exemple : J'observe deux moyennes (et leur variance) sur deux groupes différents. Deux hypothèses :
  - La différence observée est minime. Elle est due au fait que tous les sujets des deux groupes n'ont pas été inclus. Si on avait inclut tout le monde, les moyennes seraient égales (Hypothèse nulle : H_0).
  - La différence observée est importante. Elle ne peut pas être dûe au fait que tous les sujets des deux groupes n'ont pas été inclus. Il est évident que si on avait inclut tout le monde, les moyennes seraient différentes (Hypothèse alternative : H_1).
- Probabilité de me tromper si je conclu à H_1 (α, risque de 1ère espèce)
- Probabilité de me tromper si je conclu à H_0 (β, risque de 2nd espèce)

Les tests d'hypothèses

- Calculer la probabilité de me tromper si je conclu à H_1 (risque de 1ère espèce).
- Principe : Calculer la probabilité d'observer la différence en supposant H_0 vraie.
- Si cette probabilité est très faible, on rejettera H_0 avec une assez grande certitude.
- Les objectifs de ce cours sont de comparer :
  1. une moyenne observée par rapport à une moyenne de référence.
  2. deux variances observées.
  3. deux moyennes observées (échantillons indépendants).
  4. deux moyennes observées (échantillons appariés).

Plan

Test bilatéral - Grands échantillons (n≥30)

- Définition de la variable aléatoire X
- Choix des hypothèses
  $H_0 : \mu = \mu_0$
  $H_1 : \mu \neq \mu_0$
- Sous H_0, comme $n > 30$:

$$\bar{X} \sim N(\mu_0, \sigma/\sqrt{n}))$$

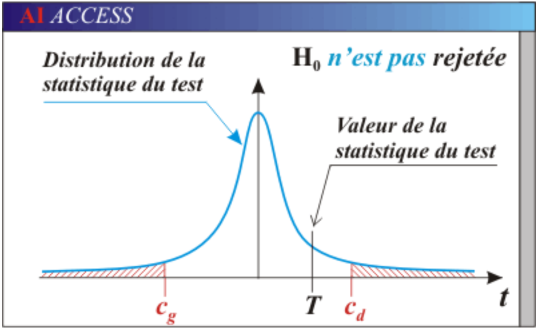
donc

$$U = (\bar{X} - \mu_0)/(\sigma/\sqrt{n}) \sim N(0,1)$$

- AN : $u = (\bar{x} - \mu_0)/(s/\sqrt{n})$
- La probabilité de se tromper en rejetant H_0 est la probabilité d'observer un échantillon plus éloigné de H_0 (en se plaçant toujours sous cette hypothèse).

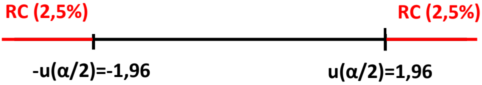
Test bilatéral - Grands échantillons (n$\geq$30)

- Définition du risque de 1ère espèce maximum :α
- Souvent $\alpha = 5\%$



Test bilatéral - Grands échantillons ($n \geq 30$)

- $\alpha = 5\%$, test bilatéral, loi normale



- Si $u \in RC \rightarrow p < \alpha$.
- p est la probabilité de se tromper en rejetant H_0 (p-value).
  - Rejet de H_0 car moins de 5% de chance qu'elle soit vraie.
  - Il semble que l'échantillon observé ne soit pas issu d'une population de moyenne μ_0.
  - Il semble que la moyenne observée soit $\neq$ de μ_0.
- Si $u \notin RC \rightarrow p > \alpha$.
  - Non rejet de H_0 car plus de 5% de chance qu'elle soit vraie.
  - Cela ne nous renseigne en rien sur la véracité de H_0.
  - On ne peut pas montrer que l'échantillon observé n'est pas issu d'une population de moyenne μ_0.
  - On ne peut pas montrer que la moyenne observée est $\neq$ de μ_0.

Test unilatéral - Grands échantillons (n≥30)

- Définition de la variable aléatoire X
- Choix des hypothèses

  $H_0 : \mu = \mu_0$
  $H_1 : \mu > \mu_0$
- Sous H_0, comme $n > 30$:

$$\bar{X} \sim N(\mu_0, \sigma/\sqrt{n}))$$

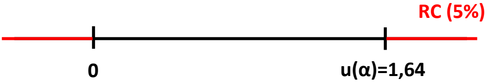
  donc

$$U = (\bar{X} - \mu_0)/(\sigma/\sqrt{n}) \sim N(0,1)$$

- AN : $u = (\bar{x} - \mu_0)/(s/\sqrt{n})$

Test unilatéral - Grands échantillons (n≥30)

- $\alpha = 5\%$, test unilatéral, loi normale



- Si $u \in RC \rightarrow p < \alpha$.
  - Rejet de H_0 car moins de 5% de chance qu'elle soit vraie.
  - Il semble que l'échantillon observé ne soit pas issu d'une population de moyenne μ_0.
  - Il semble que la moyenne observée soit $>$ à μ_0.
- Si $u \notin RC \rightarrow p > \alpha$.
  - Non rejet de H_0 car plus de 5% de chance qu'elle soit vraie.
  - Cela ne nous renseigne en rien sur la véracité de H_0.
  - On ne peut pas montrer que la moyenne observée est $>$ à μ_0.

Plan

Test bilatéral - Grands échantillons (n_A et n_B $\geq$30)

- Définition des variables aléatoires :
  - X_A dans le groupe A de taille n_A
  - X_B dans le groupe B de taille n_B
- Choix des hypothèses

  $H_0 : \mu_A - \mu_B = k$

  $H_1 : \mu_A - \mu_B \neq k$
- Sous H_0, comme n_A et $n_B > 30$:

$$\bar{X}_A \sim N(\mu_A, \sigma_A/\sqrt{n_A}) \text{ et } \bar{X}_B \sim N(\mu_B, \sigma_B/\sqrt{n_B})$$

donc

$$(\bar{X}_A - \bar{X}_A) \sim N(k, \sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B})$$

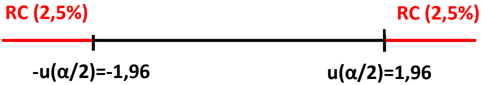
et

$$U = (\bar{X}_A - \bar{X}_B - k)\Big/\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B} \sim N(0, 1)$$

- AN : $u = (\bar{x}_A - \bar{x}_B - k)\Big/\sqrt{s_A^2/n_A + s_B^2/n_B}$

Test bilatéral - Grands échantillons (n_A et $n_B \geq 30$)

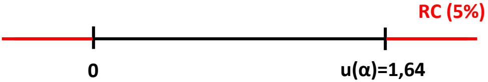
- $\alpha = 5\%$, test bilatéral, loi normale



- Si $u \in RC \rightarrow p < \alpha$.
  - Rejet de H_0 car moins de 5% de chance qu'elle soit vraie.
  - Il semble que l'écart entre les moyennes des deux populations soit $\neq$ de k.
  - Souvent $k = 0$, les moyennes des deux populations semblent $\neq$.
- Si $u \notin RC \rightarrow p > \alpha$.
  - Non rejet de H_0 car plus de 5% de chance qu'elle soit vraie.
  - On ne peut pas montrer que l'écart entre les moyennes des deux populations est $\neq$ de k.
  - Si $k = 0$, on ne peut pas montrer que les moyennes des deux populations sont $\neq$.

Test unilatéral - Grands échantillons (n_A et n_B $\geq$30)

- Identique au cas bilatéral, mais....
- $H_1 : \mu_A - \mu_B > k$
- $\alpha = 5\%$, test unilatéral, loi normale



- Si $u \in RC \rightarrow p < \alpha$.
  - Il semble que l'écart entre les moyennes de A et B soit $>$ à k.
  - Souvent $k = 0$, la moyenne de A semble supérieure à celle de B.
- Si $u \notin RC \rightarrow p > \alpha$.
- On ne peut pas montrer que la différence entre la moyenne de A et celle de B est $>$ à k.
- Si $k = 0$, on ne peut pas montrer que la moyenne de A est $>$ à la celle de B.

Plan

www.divat.fr

Rappels
La loi normale
Les tests d'hypothèses
Comparaison avec une moyenne de référence
Bilatéral
Unilatéral
Comparaison de deux moyennes (échantillons indépendants)
Bilatéral
Unilatéral
Comparaison de deux moyennes (échantillons appariés)

Comparaison de deux moyennes (échantillons appariés)

- On observe le couple (X_A, X_B) pour chaque sujet.
- Définition de la variable aléatoire $X = X_A - X_B$.
- μ : moyenne de la différence dans la population.
- Bilatéral :

  $H_0 : \mu = 0$
  $H_1 : \mu \neq 0$
- Unilatéral :

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

$\Longrightarrow$ Voir les tests de comparaison d'un moyenne observée à une valeur de référence (avec $\mu_0 = 0$).