Solutions aux problèmes supplémentaires pour l'examen final

STT1700 (Automne 2009)

1a) Soit Y le nombre de boîtes contenant moins de 200g parmi les 20 boîtes.
Alors, $Y \sim B(20, p)$, où p = probabilité de contenir moins de 200g si $\mu = 200$g.
Soit X le poids d'une boîte.
Sous H_o, on a $X \sim N(200{,}4) \Rightarrow p(X \leq 200|\mu = 200) = P(z \leq 0) = \frac{1}{2}$.
Maintenant, on cherche C tel que

$$P\left(Y \geq C \middle| \underbrace{H_o est\ vraie}_{\text{p}=1/2}\right) \leq 0.05$$

C	$P(Y \geq C \vert \text{p} = 1/2)$
11	0.411
12	0.251
13	0.131
14	0.057
15	0.02

⇒On rejette H_o si $Y \geq 15$ et la taille de la région critique est 0.02.

b) Fonction de puissance : $P(\text{rejeter } H_o \,|H_o \text{ est fausse}) = P(\text{Y}\geq 15|\mu < 200\text{g})$
Il faut calculer : $\emptyset(\mu) = P(\text{X}\leq 200|\mu)$ pour μ, $\mu = 196, 197, \dots, 199.8$

μ	P	Y≥15	μ	P	P(Y≥15)
196	0.977250	0.999959	199.5	0.598706	0.123201
197	0.933193	0.9984796	199.75	0.549736	0.055072
198	0.841345	0.5158312	199.8	0.533826	0.042605
199	0.691462	0.38414	200	0.50	0.02
199.25	0.646170	0.234336			

c) On cherche D telque $P(\bar{X} \leq D|\mu = 200) = 0.02$ (0.02 qui est la taille de la région critique en (a)).
Or $\bar{X} \sim N(200{,}0.2)$ sous H_o.

$$\Rightarrow P\left(Z \leq \frac{D-200}{\sqrt{0.2}} \middle| \mu = 200\right) = 0.02 \Leftrightarrow \frac{D-200}{\sqrt{0.2}} = -2.04$$

$\Leftrightarrow D = 199.08$

$\Rightarrow$ On rejette H_o si $\bar{X} \leq 199.08$

d) Fonction de puissance : $\emptyset(\mu) = P(\bar{X} \leq 199.08 | \mu < 200)$

μ	$\emptyset(\mu)$
196	1.0000
197	0.9999
198	0.9873
199	0.5
199.25	0.2881
199.5	0.1318
199.75	0.0468
199.8	0.0368

3) D'après le TLC, on a $\bar{X} \sim N(\mu, 9)$

a) $P(\bar{X} > 186 | \mu = 175) = P\left(Z > \frac{186-175}{3}\right) = P(Z > 3.66)$

b) $\alpha = P($rejeter H_o $|H_o$ est vraie$) = P(\bar{X} > 186 | \mu \leq 180)$.

La valeur maximum de α, α_{max}, survient lorsque $\mu = 180$ (vérifiez le en essayant plusieurs valeurs de μ)

$$\Rightarrow \alpha_{max} = P(\bar{X} > 186 | \mu = 180) = P\left(Z > \frac{186 - 180}{3}\right) = P(Z > 2) = 0.0228$$

c) $\beta = P($accepter H_o $|H_o$ est fausse$) = P(\bar{X} < 186 | \mu > 180)$

La valeur maximum de β, suivant lorsque $\mu = 180$.

$\Rightarrow \beta_{max} = P(\bar{X} < 186 | \mu = 180)$

4a) $H_0: \mu = 61.8$

$H_1: \mu \neq 61.8$

$n = 20, \bar{X} = 66.05, S = 9.25, t_{19;0.25} = 2.09$

$$t = \frac{X - 0.618}{9.25/\sqrt{20}} = 2.05 < 2.09$$

$\Rightarrow$ On ne rejette pas H_0. On ne peut conclure que la moyenne de la population et differente de l'idéal grec.

5a) $H_0: \mu_1 - \mu_2 = 2$

$H_1: \mu_1 - \mu_2 \neq 2$

$n_1 = 50, \bar{X}_1 = 26.6, \sigma_1 = 1.2$

$n_2 = 40, \bar{X}_2 = 23.8, \sigma_2 = 1.4$

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - 2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = 2.868 > 1.96 \Rightarrow On\ rejette\ H_0$$

b) $H_o: \mu_1 - \mu_2$ vs $H_1: \mu_1 - \mu_2 > 2$.

$Z = 2.868 > 1.645 \Rightarrow$ On rejette H_0.

c) $P\left(\underbrace{H_0}_{|Z|>1.96} \middle| \underbrace{H_0 \text{ est fausse}}_{\mu_1-\mu_2=2.5}\right)$

En a) On rejette H_o si $|Z| > 1.96$, i.e. si $\bar{X}_1 - \bar{X}_2 > 2.79$ ou si $\bar{X}_1 - \bar{X}_2 < 1.20$

$\Rightarrow P(\text{ rejeter } H_0 \,|\mu_1 - \mu_2 = 2.5) = P(\bar{X}_1 - \bar{X}_2 > 2.79 \,|\mu_1 - \mu_2 = 2.5)$

$+ P(\bar{X}_1 - \bar{X}_2 < 1.20 \,|\mu_1 - \mu_2 = 2.5)$

Mais $\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \Rightarrow$ À vous de continuer.

Même idée pour le test en b sauf que la région critique n'est que d'un seul côté.

6) $H_0: \mu_{1975} - \mu_{1950} = 1$

$H_1: \mu_{1975} - \mu_{1950} < 1$

$Z = \frac{\bar{Y} - \bar{X} - 1}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = -0.07$

$\Rightarrow$ On ne rejette pas H_0.

7) On a $t = 1.007$ avec $S = 2.667 = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$

Comme $t < 1.94 = t_{6;0.05}$, on ne rejette pas H_0.

8) On a $t = 3.26$ avec $S = 2.766$

Comme $t = 3.26 > t_{10;0.05} = 3.17$, on rejette H_0 et on conclut que $\mu_1 \neq \mu_2$.

9) Soit μ_1 le nombre moyen de bactéries dans les chambres avec tapis

Soit μ_2 le nombre moyen de bactéries dans les chambres sans tapis

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

On a $\bar{X}_1 = 1.04125$, $S_1 = 0.2498$

$\bar{X}_2$ = 0.9075, $S_2 = 02975$

$\Rightarrow S = 0.2747$

$t = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = 0.9737 < 2.14 = t_{14;0.025}$ ⇒ On ne rejette pas H_0.

10) On a t=3.3$> 2.12 = t_{16;0.025}$ ⇒ Les lettres ont fort probablement été écrites par des auteurs différents.

11) $H_0: p_1 = p_2$

$H_1: p_1 > p_2$

$p_1 =$ prop. dans la population des individus qui se sentent tendus et nerveux parmi les consommateurs de cola.

$p_2 =$ prop. dans la population parmi les non-consommateurs de cola.

$$\widehat{p_1} = 58\%, \widehat{p_2} = 40\% \Rightarrow Z = \frac{0.58-0.40}{\sqrt{\frac{0.58 \times 0.42}{57}+\frac{0.40 \times 0.60}{112}}} = 2.247 < -1.645$$

$\Rightarrow$ On rejette H_0.

12) $H_0: p = 0.46$

$H_1: p < 0.46$

$\hat{p} = \frac{6}{40} \Rightarrow Z = \frac{6/40 - 0.46}{\sqrt{\frac{0.46 \times 0.54}{40}}} = -3.93 < -1.645$

$\Rightarrow$ On rejette H_0.

13) $H_0: p_1 = p_2$

$H_1: p_1 \neq p_2$

$n_1 = 53, \widehat{p_1} = {}^{3}/_{53}, n_2 = 39, \widehat{p_2} = {}^{11}/_{39}$

$$\Rightarrow Z = \frac{\widehat{p_1} - \widehat{p_2}}{\sqrt{\frac{\widehat{p_1}\,\widehat{q_1}}{n_1}+\frac{\widehat{p_2}\widehat{q_2}}{n_2}}} = -2.863 < -1.645$$

$\Rightarrow$ On rejette H_0.

14) $H_0: p_1 = p_2$

$H_1: p_2 > p_1$

$\widehat{p_1} = {}^{12}/_{28} = 0.4286$, $\widehat{p_2} = {}^{23}/_{36}$ = 0.6389

$\Rightarrow Z = 1.7084 > 1.645$

$\Rightarrow$ On rejette H_0.

15) C'est un test avec données appariées.

Soit X_i le résultat du test pour le i^eenfant sous placébo

Soit Y_i le résultat du test pour le i^eenfant sous médication.

Soit $D_i = X_i - Y_i$

On a : D_i :-16, -7, +5, -24, -9, -11, -6, -2, -10, -30

$H_0: \mu_D = 0 \;\; vs \;\; H_1: \mu_D < 0$

$t = \frac{\bar{D}-0}{S_D/\sqrt{n}} = -11/10.21 = -1.07 > -1.83$ ⇒ On rejette pas H_0.

16a) On a $E(S_1^2) = \sigma^2$ et $E(S_2^2) = \sigma^2$

$\Rightarrow E(aS_1^2 + bS_2^2) = aE(S_1^2) + bE(S_2^2) = a\sigma^2 + b\sigma^2$

$= (a+b)\sigma^2 = \sigma^2$ ssi $a + b = 1$ ou $b = 1 - a$

b) Considérons les estimateurs de la forme : $aS_1^2 + (1-a){S_2^2}^2$

D'abord, rappelons nous que $(n_1 - 1)\frac{S_1^2}{\sigma^2} \sim X^2_{n_1-1}$ et $(n_2 - 1)\frac{S_2^2}{\sigma^2} \sim X^2_{n_2-1}$.

Donc $V\left(\frac{(n_1-1)S_1^2}{\sigma^2}\right) = 2(n_1 - 1) \Rightarrow V(S_1^2\,) = \frac{2\sigma^4}{(n_1-1)}$. Idem pour S_2^2.

$\Rightarrow V(aS_1^2 + (1-a)S_2^2) \;=\; a^2V(S_1^2\,) + (1-a)^2V(S_2^2\,)$

$$= a^2\frac{2\sigma^4}{(n_1-1)} + (1-a)^2\frac{2\sigma^4}{(n_2-1)} \equiv V$$

On cherche a qui minimize V.

$$\frac{\partial V}{\partial a} = 4a\frac{\sigma^4}{(n_1-1)} - 4(1-a)\frac{\sigma^4}{(n_2-1)} = 0 \Leftrightarrow \frac{a}{(n_1-1)} - (1-a)\frac{1}{(n_2-1)} = 0$$

$\Leftrightarrow a = \frac{(n_1-1)}{n_1+n_2-2}$

$\Rightarrow S_{opt}^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1+n_2-2)}$ (formule vue en cours)

17) $H_0: \mu = 11.3$

$H_1: \mu > 11.3$

$t = \frac{\bar{X}-11.3}{S/\sqrt{n}} = 1.28 < t_{11;005} = 2.201$

$\Rightarrow$ On ne rejette pas H_0.

18) H_0: Le mois de l'année n'a pas d'effet sur la mortalité

i.e., $X = (X_1, \dots, X_{12}) \sim MN\left(1855, \frac{1}{12}, \frac{1}{12}, \dots, \frac{1}{12}\right)$

H_1: non H_0.

a) $\chi^2 = \sum \frac{(O_i-T_i)^2}{T_i} = \frac{(157-154.583)^2}{154.583} + \cdots - \frac{(158-154.583)}{154.583} = 1.58437$
or $1.58437 < X_{11;005}^2 = 19.675$ On ne rejette pas H_0.

b) On calcule $P(X^2 < 1.58437) = 0.0004562$
Il est très improbable d'observer un χ^2 aussi petit que celui que l'on a observé $\Rightarrow$ les résultats sont trop beaux pour être vrais.

19a) On donne la distribution conditionnelle de la région de l'époux étant donné la région de l'époux.

	A	B	C	EU	Total
A	0.425	0.039	0.268	0.265	1.00
B	0.149	0.446	0.199	0.211	1.00
C	0.119	0.023	0.699	0.157	1.00
EU	0.159	0.035	0.231	0.573	1.00

Le tableau précédent montre clairement qu'il y a dépendance entre la région de l'époux et celle de l'épouse.

b) Soit p_B la probabilité qu'une femme Baptiste épouse un coréligionnaire
$H_0: p_B = \frac{1}{2}$ vs $H_1: p_B \neq \frac{1}{2}$

$$z = \frac{(2222/4769) - 0.5}{\sqrt{\frac{0.5 \times 0.5}{4.769}}} = -4.69 < -1.96$$

$\Rightarrow$ On rejette H_0.

c) Soit p_B la proportion de mariages impliquant baptiste et catholiques dans lesquels l'homme est catholique et la femme baptiste.

$H_0: p_{cB} = {}^{1}\!/_{2}$ vs $H_1: p_{cB} \neq {}^{1}\!/_{2}$

$n = 1906, \hat{p}_{cB} = {}^{914}\!/_{1906}$

$Z = \frac{(914/1906) - 0.5}{\sqrt{0.5 \times 0.5/1306}} = -1.79 < -1.96$ $\Rightarrow$ On ne rejette pas H_0.

d) H_0: La région de l'époux est indépendante de la région de l'épouse.
Trop de calculs pour moi ! C'est un test classique d'indépendance.
Il n'y a donc aucune difficulté.