Université Paris Paris Ouest Nanterre La Défense — Année 2010/2011

Licence de gestion apprentissage — **Statistiques**

Corrigé de l'examen du 27 janvier 2011

Exercice 1

1) **A-** Soit $Z = \dfrac{X - 1,2}{0,1}$ suit une loi $N(0,1)$.

D'où $P(X > 1,3) = P(Z > 1) = 1 - 0,8413 = 0,1587$.

B- Soit $Z = \sqrt{25}\,\dfrac{\bar{X}_n - 1,2}{0,1}$ suit une loi $N(0,1)$.

Posons $0,95 = P(a < \bar{X}_n < b) = P(-t < Z < t)$ où le quantile $t = 1,96$ est lu dans la table de $N(0,1)$.

Donc $a = 1,2 - \dfrac{1,96 * 0,1}{5} = 1,1608$ et $b = 1,2 + \dfrac{1,96 * 0,1}{5} = 1,2392$

2) **A-** $\bar{X}_n$ est un estimateur sans biais de m car $E(\bar{X}_n) = m$.

S^{*2} est un estimateur sans biais de σ^2 car $E(S^{*2}) = \sigma^2$.

B- $\bar{x}_n = \dfrac{31,5}{25} = 1,26$ et $\hat{s}^2 = \dfrac{25}{24}\left(\dfrac{40,22}{25} - (1,26)^2\right) = 0,0221$

C-

(a) α = risque de dire que le taux limite est atteint alors qu'il ne l'est pas.
β = risque de dire que le taux limite n'est pas atteint alors qu'il l'est.

(b) Le distributeur car il veut minimiser le premier risque.

(c) La statistique de test est $\bar{X}_n$ car $\bar{X}_n$ est un estimateur sans biais de m.

Soit $Z = \sqrt{25}\,\dfrac{\bar{X}_n - 1,2}{0,1}$ suit une loi $N(0,1)$.

La région de rejet est $W = (\bar{X}_n > K)$ et $\alpha = 0.05 = P_{H_0}(W) = P_{H_0}\left(Z > \sqrt{25}\,\dfrac{K - 1,2}{0,1}\right)$.

D'où $K = 1,2 + \dfrac{1,645 * 0,1}{5} = 1,2329$.

On rejette H_0 au risque 5% si $\bar{X}_n > 1,2329$.

$\bar{x}_n = 1,26 > 1,2329$, on rejette donc H_0, le taux limite est atteint.

(d) $\beta = P_{H_1}(\bar{W}) = P_{H_0}(Z \leq \sqrt{25}\ \frac{1,2329-1,3}{0,1}) = P(Z \leq -3,35) \approx 0$. La puissance du test est $\eta = P_{H_1}(W) \approx 1$.

D- La statistique de test est $\bar{X}_n$ car $\bar{X}_n$ est un estimateur sans biais de m.

Soit $Z = \sqrt{25}\ \frac{\bar{X}_n - 1,2}{S^*}$ suit une loi $T(24)$.
La région de rejet est $W = (\bar{X}_n > K)$ et $\alpha = 0.05 = P_{H_0}(W) = P_{H_0}(Z > \sqrt{25}\ \frac{K-1,2}{S^*})$.

D'où $K = 1,2 + \frac{1,711 * 0,1486}{5} = 1,2508$.
On rejette H_0 au risque 5% si $\bar{X}_n > 1,2508$.
$\bar{x}_n = 1,26 > 1,2508$, on rejette donc H_0, le taux limite est atteint.

E- La statistique de test est S^{*2} car S^{*2} est un estimateur sans biais de σ^2.

Soit $Z = \frac{24 * S^{*2}}{0,1^2}$ suit une loi $khi-deux(14)$.
La région de rejet est $W = (S^{*2} > K)$ et $\alpha = 0.05 = P_{H_0}(W) = P_{H_0}(Z > \frac{24*K}{0,1^2})$.
36,415 * $\frac{0,1^2}{24} = 0,0152$.
On rejette H_0 au risque 5% si $S^{*2} > 0,0152$.
$s^{*2} = 0,0221 > 0,0152$, on rejette donc H_0.

Exercice 2

Soit X_i la variable aléatoire qui vaut 1 si le ième individu est décidé à modifier ses habitudes de transport avec la probabilité p et qui vaut 0 si l'individu n'est pas décidé avec la probabilité 1-p. La variable X_i suit une loi de $Be(p)$.

D'après le théorème central limite appliqué à l'échantillon $(X_1, ..., X_n)$, la variable aléatoire $Z = \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}$ converge vers une v.a. de loi normale centrée, réduite $N(0,1)$ dès que $n > 30$, $np > 5$ et $n(1-p) > 5$.

Donc, pour $n = 25$, la condition n'est pas vérifiée.

Pour $n = 100$, posons $0.9 = P(-t < Z < t)$ où le quantile $t = 1.645$ est lu dans la table de la loi normale et, ainsi, l'intervalle de confiance pour p, est donné par

$$P\left(\bar{X}_n - 1.645\sqrt{\frac{p(1-p)}{n}} < p < \bar{X}_n + 1.645\sqrt{\frac{p(1-p)}{n}}\right) = 0.90$$

En substituant, de part et d'autre de l'inégalité, le paramètre p par son estimation $\bar{x}_n = 0,2$ on obtient finalement une fourchette d'estimation asymptotique pour p, à partir de l'échantillon des observations $(x_1, x_2, ..., x_n)$:

$$\bar{x}_n - 1.645\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} < p < \bar{x}_n + 1.645\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}.$$

On obtient donc

$0.1342 \ < \ p \ < \ 0.2658.$

Vérifions à postériori les conditions d'application du TCL: $n * 0.1342 > 5$ et $n * (1 - 0.2658) > 5$, l'intervalle de confiance est bien validé.

Exercice 3

(a) Test du khi-deux d'indépendance.
Notons X = nombre d'années de scolarité et Y = attitude face à l'avortement.

(b) On veut tester au niveau 10%:

$$H_0 : X \text{et } Y \text{sont indépendantes} \quad \textit{contre}$$
$$H_1 : X \text{et } Y \text{sont liées}$$

On construit le tableau des effectifs théoriques sous H_0:

attitude face à l'avortement / nombre d'années de scolarité	pour	indifférent	contre
inférieur à 8 ans	45.1	21.4	43.5
entre 8 et 12 ans	179.1	85	172.9
supérieur à 12 ans	93.8	44.6	90.6

Sous H_0, comme $n \geq 30$ et que les effectifs théoriques sont supérieurs à 5, T_n =suit approximativement une loi $\chi^2(4)$.

La région de rejet, au niveau 10% de la forme

$$W = (T_n \geq 7.779)$$

On calcule la valeur de T_n sur l'échantillon et on obtient $t_n = 17,71 \geq 7.779$ qui est dans la région de rejet, on accepte au niveau 10% l'hypothèse d'une liaison entre le nombre d'années de scolarité et l'attitude face à l'avortement.

(c) Soit $X = 1$ si l'individu est contre l'avortement avec la probabilité p et 0 sinon avec la probabilité $1 - p$.
En se limitant aux $n = 110$ individus dont la durée de scolarité est inférieure à 8 ans, on veut tester au niveau 10%:

$$H_0 : p = 0,5 \textit{ contre}$$

$$H_1 : p > 0,5$$

La statistique de test est $\bar{X}_n$

Comme $n > 30$, $np_0 = 110 * 0.5 \geq 5$ et $n(1 - p_0) \geq 5$ on peut appliquer le théorème central limite à l'échantillon $(X_1, ..., X_n)$ de loi de bernoulli(p) :

ainsi $Z = \dfrac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} = \sqrt{110}\ \dfrac{\bar{X}_n - 0.5}{\sqrt{0.5 * 0.5}}$ suit approximativement une loi $N(0, 1)$.

$W = (\bar{X}_n > K)$ et $\alpha = 0.1 = P_{H_0}(W) = P_{H_0}(Z > \sqrt{110}\ \dfrac{K - 0.5}{\sqrt{0.5 * 0.5}})$.

D'où $K = 0.5 + 1.28 * \dfrac{\sqrt{0.5 * 0.5}}{\sqrt{110}} = 0.561$. On rejette H_0 au risque 10% si $\bar{X}_n > 0.561$.
$\bar{x}_n = \dfrac{56}{110} = 0,5091 < 0,561$, on ne rejette donc pas H_0 et on ne peut pas accepter l'hypothèse que le rejet de l'avortement est majoritaire chez ces individus.