PASS/LAS

Correction

UE5 – ED Tut'rentrée 1

23 / 24 / 25 Août 2021

Fait par les bbs D1 de l'UE gangstat

Relu par ces mêmes babies

QCM 1 : AB

A. VRAI. La population est un ensemble très grand, parfois même innombrable donc les **études exhaustives y sont impossibles**.
B. VRAI.
C. FAUX. Elles **ne sont pas mesurables** car ce sont des **variables QUALITATIVES**. En revanche, elles peuvent être ordonnées et classées.
D. FAUX. Le **mode** est une mesure de la tendance centrale qui ne peut se déterminer **QUE** pour les **variables quantitatives**. De plus, les variables qualitatives ne se mesurent pas.
E. FAUX. Le groupe sanguin est une variable qualitative **pure**. En effet, on ne peut ni classer ni ordonner les groupes sanguins comme on pourrait le faire pour l'intensité d'une douleur par exemple.

QCM 2 : AE

En considérant X comme la variable "nombre de vertèbres" et x_i la valeur de X mesurée pour chaque girafon. Ici, on a :

- *N = 12 girafons*
- *Valeurs prises par X: 2, 2, 2, 2, 2, 3, 7, 12, 12, 13, 13, 14*

A. VRAI. La **moyenne** est égale à la somme des valeurs prises par X multipliées par leur effectifs respectifs (n_i) divisée par l'effectif total (N) soit :

$$\overline{X} = \frac{\Sigma(x_i \times n_i)}{N} = \frac{2\times5 + 3\times1 + 7\times1 + 12\times2 + 13\times2 + 14\times1}{12} = \frac{10+3+7+24+26+14}{12} = \frac{84}{12} = \textbf{7 vertèbres.}$$

B. FAUX. La **médiane** est la **valeur de X qui divise l'échantillon en 2 parties égales**, c'est-à-dire autant de girafons qui ont un nombre de vertèbres au-dessus de cette médiane (50%) que de girafons qui ont un nombre de vertèbres en-dessous (50%).

⚠️ Ici nous sommes devant une série paire. La médiane se situe donc entre la $n^{ième}$ et $(n+1)^{ième}$ avec $n = \frac{N}{2}$.

Donc :

2, 2, 2, 2, 2, 3, 7, 12, 12, 13, 13, 14
6ième valeur ↑ 7ième valeur
MÉDIANE

Médiane = $\frac{n^{ième} + (n+1)^{ième}}{2} = \frac{3+7}{2} =$ **5 vertèbres.**

C. FAUX. Le **mode** est la **valeur de X la plus souvent rencontrée** dans l'échantillon. Ici, le mode vaut **2 vertèbres** car cette valeur de X est répétée le plus de fois (5 fois).
Tandis que la différence entre les valeurs extrêmes de l'échantillon correspond à l'**étendue** et vaut **12**.
D. FAUX. La **mesure de la dispersion** est permise par la variance, l'écart-type et l'**étendue**. Alors que la moyenne, le mode et la médiane permettent de **mesurer la tendance centrale**.

E. VRAI. L'**écart-type** $\sigma(X)$est égale à la racine carré de la variance $s^2(X)$. Donc : $\sigma(X) = \sqrt{s^2(X)}$

$$\sqrt{s^2(X)} = \sqrt{\frac{\sum n_i (x_i - \bar{X})^2}{N}} = \sqrt{\frac{5\times(2-7)^2 + 1\times(3-7)^2 + 1\times(7-7)^2 + 2\times(12-7)^2 + 2\times(13-7)^2 + 1\times(14-7)^2}{12}}$$

$$= \sqrt{\frac{5\times(-5)^2 + (-4)^2 + 2\times(5)^2 + 2\times(6)^2 + (7)^2}{12}} = \sqrt{\frac{5\times25 + 16 + 2\times25 + 2\times36 + 49}{12}}$$

$$= \sqrt{\frac{312}{12}} = \sqrt{26} \approx \sqrt{25} = \textbf{5 vertèbres}$$

QCM 3 : ABD

A. VRAI. Le **mode** est la valeur la plus souvent rencontrée, ici le mode est donc bien égal à **7**.

Pour la suite de l'exercice on va effectuer la méthode du changement de variable grâce au tableau suivant :
Nous utilisons cette méthode car elle facilite les calculs.

N_i	X_i	$Y_i = X_i$ - mode	Y_i^2	Y_iN_i	$Y_i^2N_i$
1	10	10-7 = 3	3^2 = 9	3x1 = 3	9x1 = 9
2	9	9-7 = 2	2^2 = 4	2x2 = 4	4x2 = 8
3	7	7-7 = 0	0^2 = 0	0x3 = 0	0x3 = 0
1	4	4-7 = -3	$(-3)^2$ = 9	-3x1 = -3	9x1 = 9
2	2	2-7 = -5	$(-5)^2$ = 25	-5x2 = -10	25x2 = 50
1	0	0-7 = -7	$(-7)^2$ = 49	-7x1 = -7	49x1 = 49
N = 10				**T1 = -13**	**T2 = 125**

B. VRAI. La moyenne de Y_i : $m(Y_i) = \frac{T1}{N} = \frac{-13}{10} =$ **-1,3**

⚠️ *Attention à ne pas oublier de repasser ensuite à X_i en ajoutant le mode* ⚠️

La **moyenne** de X_i : $m(X_i) = m(Y_i) + mode = -1,3 + 7 = 5,7$

Il est également possible de calculer la moyenne sans passer par le tableau. Pour cela, nous utilisons la formule suivante :

$$m(X_i) = \frac{\Sigma(x_i \times n_i)}{N} = \frac{1\times10 + 2\times9 + 3\times7 + 1\times4 + 2\times2 + 1\times0}{10} = \frac{10 + 18 + 21 + 4 + 4 + 0}{10} = 5,7$$

C. FAUX. $Var(Y_i) = \frac{T2 - \frac{T1^2}{N}}{N} = \frac{125 - \frac{(-13)^2}{10}}{10} = \frac{125-16,9}{10} = \frac{108,1}{10} =$ **10,81**

Dans le cas d'un changement de variable de la forme $Y_i = X_i$ - mode, la **variance de Y est la même que celle de X**. Donc $Var(X_i) = 10,81\ gaufres^2$

D. VRAI. *cf. correction item C.*

E. FAUX. L'**écart type** est égal à la racine carré de la variance donc $\sigma(X_i) = \sqrt{\sigma^2(X_i)} = \sqrt{10,81} \approx \sqrt{10} = 3,16$

L'écart type est de **3,16 gaufres**.

→ Remarque : *Ici, pas besoin de connaître la valeur de $\sqrt{10}$, on sait que l'écart-type correspond à la racine de la variance, ils ne peuvent donc pas avoir la même valeur.*

QCM 4 : AC

A. VRAI. Ici, on souhaite déterminer le **paramètre dans la population** à partir de celui d'un é**C**hantillon. On réalise ainsi un **intervalle de Confiance.**
B. FAUX. L'**intervalle de Pari** détermine le **paramètre dans l'échantillon** à partir de celui de la **P**opulation.
C. VRAI. Nous sommes ici dans un **intervalle de confiance** avec une **variable qualitative** (proportion de personnes qui disent "chocolatine" = fréquence). Ainsi, les conditions à vérifier sont **NP et NQ ≥ 5**.
   → Seulement, on ne connaît pas encore ni P ni Q, on doit donc vérifier les conditions ***À POSTÉRIORI***.

À la fin on trouve : P ∈ [0,12 ; 0,28] (cf. correction QCM 5) et on sait que N = 100 et que Q = 1 - P

- *Pour P = 0,12*
  - *NP = 100 x 0,12 = 12*
  - *NQ = 100 x 0,88 = 88*
- *Pour P = 0,28*
  - *NP = 100 x 0,28 = 28*
  - *NQ = 100 x 0,72 = 72*

NP et NQ ≥ 5 sur tout l'intervalle. Les conditions sont respectées.

D. FAUX. *cf. correction item C.*
E. FAUX. Nous sommes ici dans un **intervalle de confiance** avec une **variable qualitative** (proportion de personnes qui disent "chocolatine" = fréquence). Ainsi, le paramètre calculé appartient à l'intervalle :

$$\left[f \pm U_\alpha \times \sqrt{\frac{f\times(1-f)}{N}}\right]$$

QCM 5 : CE

Dans tous les QCMs de "calculs" en UE5, il faut apprendre à résumer par étapes :

1) Trouver la bonne variable et son type
   → *Ici, variable qualitative : "proportion de gens qui disent chocolatine"*
2) Comprendre ce que l'on cherche à partir de ce que l'on a
   → *On cherche la fréquence dans la population, en connaissant la fréquence d'un échantillon ⇒* ***intervalle de confiance*** *(A et D faux d'office)*
3) Vérifier les conditions d'applications → *NP et NQ ≥ 5 (⚠️ Recherchées A POSTÉRIORI ⚠️)*
4) Trouver le U_α correspondant au α → *α = 0,05 ⇒ U_α = 1,96 (≈ 2)*
5) Faire le calcul de l'intervalle : $P \in \left[f \pm U_\alpha \times \sqrt{\frac{(1-f)}{N}}\right] = \left[0,2 \pm 1,96 \times \sqrt{\frac{0,2\times(1-0,2)}{100}}\right]$

$$P \in \left[0,2 \pm 2 \times \sqrt{\frac{(0,2\times0,8)}{100}}\right]$$

$$P \in \left[0,2 \pm 2 \times \sqrt{\frac{0,16}{100}}\right]$$

$$P \in [0,2 \pm 2 \times 0,04] = [0,12 ; 0,28]$$

Ainsi, la proportion dans la population de gens qui disent chocolatine est d'environ **20 % ± 8%**.

A. FAUX. *(On cherche à encadrer le paramètre "P" et non "m")*
B. FAUX. *cf. correction ci-dessus.*
C. VRAI. *cf. correction ci-dessus.*

D. FAUX. ⚠️ On ne pourra jamais avoir comme résultat une seule valeur, on établit toujours un intervalle.

E. VRAI. En effet, plus le **risque α est grand**, plus **l'intervalle** qui contient la variable **est petit**.

Si on raisonne avec la formule : $P \in \left[f \pm U_{\alpha} \times \sqrt{\frac{(1-f)}{N}}\right]$. Quand **α augmente**, **U_{α} diminue** et **l'intervalle devient plus précis**.

QCM 6 : BE

A. FAUX. On connaît ici la valeur de la **fréquence dans la population P** et on réalise **des échantillons dont on cherche la valeur de la fréquence** pour chacun : c'est le principe de la **fluctuation d'échantillonnage dans le cadre d'un intervalle de Pari**. En d'autres termes, c'est le paramètre de l'échantillon qu'on cherche à encadrer dans l'intervalle et non celui de la population.

*Moyen mnémo : Pour un intervalle de **P**ari, on **P**art de la **P**opulation.*

B. VRAI. Nous connaissons les différents paramètres nous permettant d'appliquer la formule suivante pour réaliser notre intervalle de pari :

$$f \in \left[P \pm U_{\alpha} \times \sqrt{\frac{PQ}{N}}\right]$$

⚠️ Avant de commencer nos calculs, nous devons impérativement vérifier que les conditions sont bien respectées : **NP≥5 et NQ≥5**.

On a donc :

- NP = 64×0,2 = 12,8
- NQ = 64×0,8 = 51,2 > 5

Nous pouvons ainsi passer au calcul.

$$f \in \left[P \pm U_{\alpha} \times \sqrt{\frac{PQ}{N}}\right]$$ *Avec $U_{\alpha} \simeq 2$ pour un risque de 5%.*

$$f \in \left[0,2 \pm 2 \times \sqrt{\frac{0,2 \times 0,8}{64}}\right]$$

$$f \in \left[0,2 \pm 2 \times \sqrt{\frac{0,16}{64}}\right]$$

$$f \in \left[0,2 \pm 2 \times \frac{\sqrt{0,16}}{\sqrt{64}}\right]$$

$$f \in \left[0,2 \pm 2 \times \frac{0,4}{8}\right]$$

$$f \in \left[0,2 \pm \frac{0,8}{8}\right]$$

$$f \in [0,2 \pm 0,1]$$

$$f \in [10\% ; 30\%]$$

C. FAUX. Le risque α n'étant pas le même, l'intervalle est forcément différent : **plus le risque α de se tromper est faible, plus l'intervalle sera grand et peu précis**.

D. FAUX. Comme pour l'item B, nous avons tous les paramètres pour appliquer la formule, il ne nous reste plus qu'à **vérifier les conditions** : NP≥5 et NQ≥5.

On a donc NP = 16×0,2 = **3,2 < 5**. **Les conditions ne sont pas respectées, l'intervalle proposé est rejeté** : il faut un plus grand échantillon pour une meilleure représentativité.

E. VRAI. En effet, on peut remarquer mathématiquement que l'échantillon N se trouve au dénominateur de la formule : **plus N croît, plus l'écart se réduit et l'intervalle devient précis**.

QCM 7 : BD

A. FAUX. Cette étude s'intéresse à une **variable quantitative** : le nombre de lapins adoptés.

B. VRAI. Ici, on part de la Population pour encadrer le paramètre dans un échantillon : il s'agit donc d'un **intervalle de Pari**.

C. FAUX. <u>ATTENTION</u> : il faut utiliser l'**<u>écart type</u>** et non la variance. L'écart type est la **racine carrée de la variance** ⇒ $\sigma = \sqrt{Var} = \sqrt{9} = 3$.

D. VRAI. On réalise ici une fluctuation d'échantillonnage. On cherche à encadrer la moyenne m de l'échantillon dans un intervalle de pari.

On utilise la formule : $m \in \left[\mu \pm U_{\alpha} \frac{\sigma}{\sqrt{N}}\right] = \left[3 \pm 2\frac{3}{\sqrt{36}}\right]$ *(pour α = 5% → U_α ≈ 2)*

$$m \in \left[3 \pm \frac{2 \times 3}{6}\right]$$
$$m \in [3 \pm 1]$$
$$m \in [2 ; 4]$$

E. FAUX. Les <u>conditions ne sont pas respectées</u> car la moitié du groupe est égale à 18.

→ **N doit être ≥ 30** pour effectuer un intervalle de pari sur une variable quantitative.

<u>QCM 8 : B</u>

A. FAUX. C'est l'inverse, H0 serait plutôt "Il n'y a **<u>pas</u>** de différence significative entre l'échantillon et la population".

B. VRAI. Autre formulation possible.

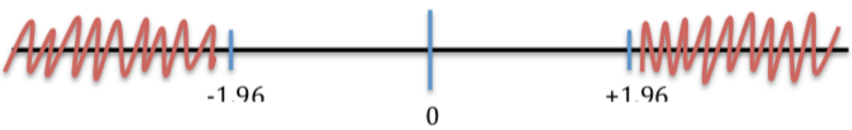
C. FAUX. Pour des variables qui sont **<u>quantitatives</u>**, comme c'est le cas ici (car nous pouvons mesurer la variable qui sont « les dépenses »), la condition d'utilisation du test de la loi Normale est **N ≥ 30**.

D. FAUX. Pour des variables qui sont **<u>qualitatives</u>**, la condition d'utilisation du test de la loi Normale est **NP et NQ ≥ 5**.

E. FAUX. Plus la région critique est grande, plus notre paramètre a de chance de se trouver dans cette région critique donc plus on a de chance de rejeter H_0 (et <u>moins on a de chance</u> d'accepter H_0).

<u>QCM 9 : ACD</u>

A. VRAI. Pour α = 0,05, U_α = 1,96 (â connaître par <3). Pour le seuil α = 0,05, la région critique est la région hachurée :



B. FAUX. En présence d'une variable **<u>quantitative</u>**, on calcule le paramètre U avec la formule suivante :

$$U = \frac{m - \mu}{\frac{\sigma}{\sqrt{N}}}$$ → (⚠️ *aux détails)*

C. VRAI. Comme les conditions sont respectées *(cf. QCM 8)*, nous pouvons calculer le paramètre :

$$U = \frac{m-\mu}{\frac{\sigma}{\sqrt{N}}} = \frac{50-25}{\frac{10}{\sqrt{36}}} = \frac{25}{\frac{10}{6}} = \frac{25}{\frac{5}{3}} = \frac{25 \times 3}{5} = \frac{5 \times 5 \times 3}{5} = 5 \times 3 = 15$$

D. VRAI. **U ∈ RC**, on **rejette H_0** au risque **α de 1ère espèce** : l'échantillon n'est pas représentatif de la population, il existe une différence significative entre m et μ.

E. FAUX. *cf. correction item D.*

<u>QCM 10 : AC</u>

A. VRAI. *<u>Remarque :</u> L'hypothèse H1 serait l'inverse « Il y a une différence significative entre les deux fréquences ».*

B. FAUX. Les conditions à vérifier pour le test basé sur la loi Normale avec des **variables <u>qualitatives</u>** sont **NP ≥ 5 et NQ ≥ 5**.

Les conditions sont donc vérifiées puisqu'on a :

- NP = $70 \times 0,1 = 7$
- NQ = $70 \times 0,9 = 63$

C. VRAI. Le paramètre U pour le test de la loi normale avec des valeurs qualitatives se calcule avec la formule suivante : $U = \frac{f - P}{\sqrt{\frac{PQ}{N}}}$.

On a donc : $U = \frac{0,15 - 0,1}{\sqrt{\frac{0,1 \times 0,9}{70}}} = \frac{0,05}{\sqrt{\frac{0,09}{70}}} \approx \frac{0,05}{0,03} \approx \frac{5}{3} \simeq$ **1,67**

D. FAUX. Pour un risque $\alpha \leq 5\%$, on a $U\alpha = 1,96$ (voir TER). Ici, $U \simeq 1,67$ et **n'appartient donc pas à la RC** : on **accepte H_0** au **risque β de seconde espèce**.

E. FAUX. Peu importe le résultat de notre test, on ne peut **jamais établir de causalité** dans les conclusions d'un test statistique de la loi Normale. On ne peut parler QUE de **lien statistique**.