

NB: pour gagner du temps, traitez les parties A, B et C directement sur l'énoncé et placez-le à l'intérieur de la copie qui vous a été remise

A. Structures de la langue (4 points – niveau B1)

1°) Usage des articles: corrigez 5 erreurs d'article dans ce petit texte:

~~The~~ mathematical probabilities are ~~at~~ the root of ~~the~~ statistics. An universal principle is the law of ~~the~~ large numbers, which supports an empirical approach of probabilities by means of ~~a~~ repeated experiments.

2°) Questions et réponses compléter à gauche avec des questions ou à droite avec des réponses:

– ? The case happened in Holland

– Was the nurse guilty ? (négative courte)

– ? She spent six years in prison

– ? Yes, statistics can be quite counter-intuitive.

3°) Données quantitatives écrivez dates, valeurs et symboles en toutes lettres, comme lorsque vous les prononcez à voix haute:

Pascal and Fermat invented the theory of probabilities in 1654 (.....)

but it was developed only in the 18th century by De Moivre, Bernoulli and Bayes

1 / 342,000 (.....) is approximately equal to 2.9

(.....)

The golden ratio is : $\phi = (1 + \sqrt{5}) / 2$ (.....)

4°) Comparatifs et superlatifs corrigez les erreurs de forme dans le texte qui suit:

Statistics are certainly the more difficult field of applied maths. Even specialists who are

have a more clear view of the problems than ordinary people may make gross mistakes

time. The, seriousest error of method is to derive your hypotheses from the data. However

frequent than some people think.

B. Décomposition et prononciation des mots (3 points – niveau B1)

Indiquer la catégorie (N=nom, V=verbe, A=adjectif, AV=adverbe) et la prononciation en notation API des mots suivants du texte (après les avoir éventuellement décomposés en écrivant les préfixes et suffixes EN MAJUSCULES et en parenthésant au besoin les étapes de la décomposition) catégorie finale : mot = décomposition des préfixes et suffixes => /prononciation déduite/ notez les catégories avec N pour nom, V pour verbe, A pour adjectif et AV pour adverbe par exemple: N : synthesizer = V: (N: (SYN- + N: thesis) + -IZE) + -ER => /'sɪnθəsaɪzər/

.... : unlikely = => /...../

.... : continued = => /...../

.... : prosecution = => /...../

.... : investigators = => /...../

Dans l'extrait qui suit, soulignez toutes les occurrences de la voyelle /ɪ:/ (comme dans « please ») et de la diphthongue /eɪ/ (comme dans « play ») :

~~It was barely based on a figure of one in 342 million against. Even if we found errors in~~
~~and believe me, we will — as in our previous story, the right answer is still being~~
~~rence, as we have already seen repeatedly, the interesting thing about statistics is not the~~
~~icky maths, but what the numbers mean.~~

this figure — a
 important Don
 tr

C. Compréhension écrite élémentaire (4 points – niveau B1)

Les justifications doivent être de très courts extraits du texte judicieusement choisis

a) Ce texte est un extrait d'un article de presse sur les erreurs judiciaires VRAI – FAUX

Justification:

b) proposez un synonyme ou une définition en anglais de “irrelevant”

.....

c) l'accusation avait négligé certaines données statistiques VRAI – FAUX

Justification:

d) proposez un synonyme en anglais de “startling”

.....

e) proposez trois synonymes en anglais de “odd”

.....

f) l'expert a utilisé des données brutes erronées VRAI – FAUX

Justification:

.....

D. Reformulation et synthèse (3 points – niveau B2)

Answer each question in English in 5 to 10 lines ; carefully summarise and rephrase, and don't just copy-paste whole sentences from the original text: any select citation should be short and explicit (enclosed within quotation marks)

- 1°) Explain the title: “Losing the lottery”
- 2°) How many different mistakes were made in this case?
- 3°) Propose a little story making the same point as the example of the machine gun.

D. Production écrite élémentaire (3 points – niveau B1 - 150 mots)

NB : les étudiants des niveaux 2 & 3 doivent traiter ce sujet en priorité car l'évaluation de cette production portera davantage sur le contenu que sur la correction de la langue

Describe how computer specialists can help the police in certain criminal cases

F. Production écrite avancée (3 points – niveau B2 – 250 mots)

NB : l'évaluation de cette production portera à la fois sur le contenu, le lexique et la syntaxe

How is it possible to avoid that biased or incompetent “experts” cause a miscarriage of justice ? Identify problems and propose solutions.

Losing the lottery

It is possible to be very unlucky indeed. A nurse called Lucia de Berk has been in prison for six years in Holland, convicted of seven counts of murder and three of attempted murder. An unusually large number of people died when she was on shift, and that, essentially, along with some very weak circumstantial evidence, is the substance of the case against her. She has never confessed, she has continued to protest her innocence, and her trial has generated a small collection of theoretical papers in the statistics literature.

The judgement was largely based on a figure of 'one in 342 million against'. Even if we found errors in this figure — and believe me, we will — as in our previous story, the figure itself would still be largely irrelevant. Because, as we have already seen repeatedly, the interesting thing about statistics is not the tricky maths, but what the numbers mean.

There is also an important lesson here from which we could all benefit: unlikely things do happen. Somebody wins the lottery every week; children are struck by lightning. It's only weird and startling when something very, very specific and unlikely happens if you have specifically predicted it beforehand. [...] Here is an analogy. Imagine I am standing near a large wooden barn with an enormous machine gun. I place a blindfold over my eyes and laughing maniacally — I fire off many thousands and thousands of bullets into the side of the barn. I then drop the gun, walk over to the wall, examine it closely for some time, all over, pacing up and down. I find one spot where there are three bullet holes close to each other, then draw a target around them, announcing proudly that I am an excellent marksman.

You would, I think, disagree with both my methods and my conclusions for that deduction. But this is exactly what has happened in Lucia's case: the prosecutors found seven deaths, on one nurse's shifts, in one hospital, in one city, in one country, in the world, and then drew a target around them.

This breaks a cardinal rule of any research involving statistics: you cannot find your hypothesis in your results. Before you go to your data with your statistical tool, you have to have a specific hypothesis to test. If your hypothesis comes from analysing the data, then there is no sense in analysing the same data again to confirm it.

This is a rather complex, philosophical, mathematical form of circularity: but there were also very concrete

~~refined statistics in the case. To collect more data, the investigators went back to the wards to see if~~
they could find more suspicious deaths. But all the people who were asked to remember 'suspicious incidents' knew that they were being asked because Lucia might be a serial killer. There was a high risk that 'an incident anonymous with 'Lucia was present'. Some sudden deaths when Lucia was not present calculations, by definition: they are in no way suspicious, because Lucia was not

asked to make a list of incidents that happened during or shortly after Lucia's shifts said in this manner more patterns were unearthed, and so it became even more likely that more suspicious deaths on Lucia's shifts. Meanwhile, Lucia waited in prison for her nightmares.

amount of corollary statistical information was almost completely ignored. In the three ed on the ward in question, there were seven deaths. In the three years that she did work six deaths.

~~Here's a thought: a killer should go down on a ward at the precise moment that a serial~~
killer — on a killing spree — arrives. If Lucia killed them all, then there must have been no natural deaths on that ward at all in the whole of the three years that she worked there.

[...]

But the strangest thing of all is this. In generating his obligatory, spurious figure — which this time was 'one in 342 million' — the prosecution's statistician made a simple, rudimentary mathematical error. He combined individual statistical tests by multiplying p-values, the mathematical description of chance, or statistical significance. If you multiply p-values together, then harmless and probable incidents rapidly appear vanishingly unlikely. Let's say you worked in twenty hospitals, each with a harmless incident pattern: say $p=0.5$. If you multiply those harmless p-values, of entirely chance findings, you end up with a final p-value of 0.5 to the power of twenty, which is $p < 0.000001$, which is extremely, very highly statistically significant. With this mathematical error, by his reasoning, if you change hospitals a lot, you automatically become a suspect. Have you worked in twenty hospitals? For God's sake don't tell the Dutch police if you have.

was suspicious' became s would not be listed in the present.

It gets worse. 'We were a one hospital employee. If investigators would find trial. This is the stuff of r At the same time, a huge years' before Lucia work on the ward, there were