

Thème : Séries statistiques à une variable

Cet énoncé est celui de la deuxième épreuve orale (épreuve sur dossier) du Capes Externe de mathématiques, proposé aux candidat(e)s le 30 Juin 2006.

Pour consulter les archives de cette épreuve orale, depuis la session 2005, on se reportera au site officiel du jury, à l'adresse <http://capes-math.org/>

1. L'exercice proposé au candidat

Le tableau ci-contre indique, pour chaque mois de l'année 2004, trois données concernant le site web du CAPES (la « bande passante » représente le volume d'information qui a été chargé).

Mois	Visiteurs différents	Visites	Bande passante
Janvier 2004	353	425	62 Mo
Février 2004	577	744	144 Mo
Mars 2004	834	1 151	169 Mo
Avril 2004	650	803	132 Mo
Mai 2004	2 498	3 404	1 021 Mo
Juin 2004	2 324	3 254	907 Mo
Juillet 2004	2 636	3 482	589 Mo
Août 2004	1 410	1 916	274 Mo
Septembre 2004	2 525	3 553	681 Mo
Octobre 2004	2 897	4 135	2 600 Mo
Novembre 2004	3 861	5 232	4 372 Mo
Décembre 2004	2 452	3 157	2 499 Mo

- Donner pour ces trois séries de données le tableau des effectifs cumulés croissants. À quels types de questions ces tableaux permettent-ils de répondre ?
- Calculer la moyenne du nombre des visiteurs et la moyenne du nombre des visites. On s'intéresse au nombre moyen de visites par visiteurs : un élève propose de le calculer chaque mois et de faire la moyenne des résultats obtenus. Un autre propose de faire le quotient (moyenne des visites)/(moyenne des visiteurs). Obtient-on le même résultat ? Pourquoi ? En moyenne quelle est la bande passante utilisée par un visiteur ?
- Proposer une ou deux représentations graphiques permettant de visualiser les données du tableau.

2. Le travail demandé au candidat

En aucun cas, le candidat ne doit rédiger sur sa fiche sa solution de l'exercice. Celle-ci pourra néanmoins lui être demandée partiellement ou en totalité lors de l'entretien avec le jury

Pendant sa préparation, le candidat traitera les questions suivantes:

- Q.1) Préciser à quel niveau d'enseignement une telle activité peut trouver sa place. Indiquer comment vous la mettriez en oeuvre dans une classe.


- Q.2) Quelles représentations graphiques peut-on obtenir en réponse à la question 3) ? Montrer sur un écran de calculatrice une de ces représentations.
- Q.3) Donner au moins un autre exemple permettant d'illustrer l'intérêt et les limites de la notion de moyenne, vous pourrez (sans que ce soit une obligation) utiliser le graphique ci-contre. Énoncer les théorèmes mis en jeu dans l'exercice.

Sur ses fiches, le candidat rédigera et présentera :

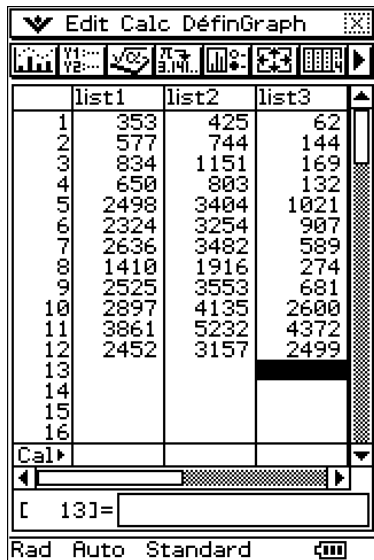
- sa réponse à la question Q.3.
- un ou plusieurs exercices sur le thème : « Séries statistiques à une variable. »

Proposition de corrigé avec le Classpad 300

L'exercice proposé au candidat

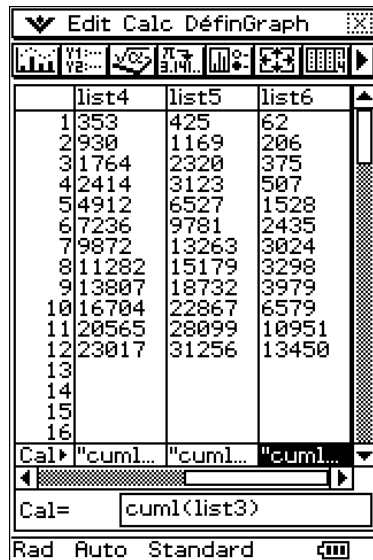
1. On passe dans , et on choisit « Edit/Tout effacer » pour faire place nette.

On entre les données de la colonne « Visiteurs » dans `list1`, celles de la colonne « Visites » dans `liste2`, et enfin celles de la colonne « Bande passante » dans `liste3` (cf fig1)



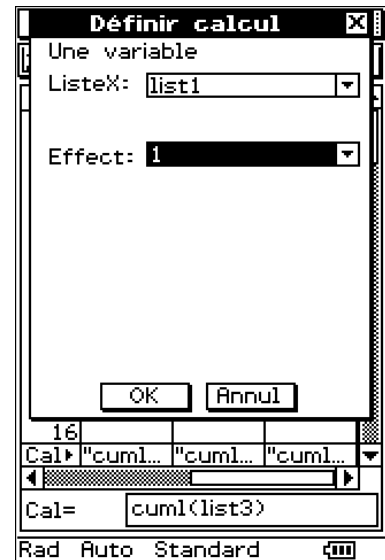
	list1	list2	list3
1	353	425	62
2	577	744	144
3	834	1151	169
4	650	803	132
5	2498	3404	1021
6	2324	3254	907
7	2636	3482	589
8	1410	1916	274
9	2525	3553	681
10	2897	4135	2600
11	3861	5232	4372
12	2452	3157	2499
13			
14			
15			
16			

fig1 : tableau des données



	list4	list5	list6
1	353	425	62
2	2930	1169	206
3	31764	2320	375
4	42414	3123	507
5	54912	6527	1528
6	67236	9781	2435
7	79872	13263	3024
8	811282	15179	3298
9	913807	18732	3979
10	1016704	22867	6579
11	1120565	28099	10951
12	1223017	31256	13450
13			
14			
15			
16			

fig2 : effectifs cumulés



Définir calcul

Une variable

ListeX: list1

Effect: 1

OK Annul

16

Cal▶ "cuml... "cuml... "cuml..."

Cal= cuml(list3)

fig3 : stats sur list1

On positionne ensuite le curseur sur la cellule `Cal▶` de la colonne `list4`.

Dans le champ `Cal=`, on tape `cuml(list1)` et on valide.


La colonne `list4` se remplit alors avec les effectifs cumulés de la première colonne.



On procède de même pour placer dans les colonnes `list5` et `list6` les effectifs cumulés des colonnes `list2` et `list3` (cf fig2).

2. Pour calculer la moyenne du nombre de visiteurs, on sélectionne « Calc/Une variable ».

Une fenêtre s'ouvre (fig3) demandant la liste sur laquelle doit porter le calcul (le choix par défaut `list1` nous convient très bien) et la liste des effectifs (par défaut le choix est 1 ce qui nous convient aussi).

Après avoir validé, le Classpad affiche les résultats pour `list1`, et on y lit en particulier la moyenne légèrement supérieure à 1918 (fig4). On peut procéder de même pour `list2` et on trouve une moyenne légèrement inférieure à 2605 visites par mois (fig5).

Une autre méthode consiste à ouvrir une fenêtre de calcul formel (utiliser ) depuis l'écran de statistiques et à évaluer les expressions `mean(list1)` et `mean(list2)`. Mais puisqu'on a formé les effectifs cumulés, les deux moyennes peuvent également être obtenues en divisant par 12 le dernier élément de `list3` et de `list4` (fig6).

NB : on utilisera  (ou on passera en mode `Decimal`) pour obtenir les résultats sous forme réelle plutôt que rationnelle. On réglera également le format d'affichage des nombres réels (menu «  Format de base »). Ici nous sommes en « `Fixe 3` ».

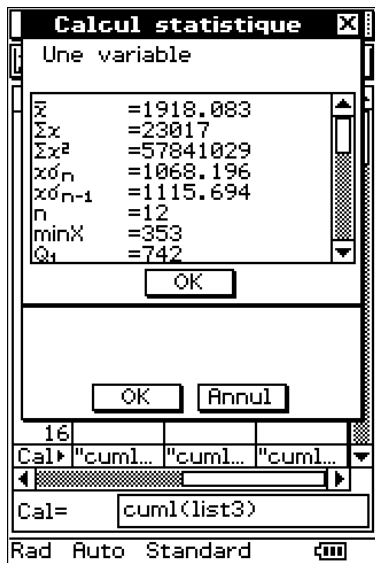


fig4 : résultats pour list1

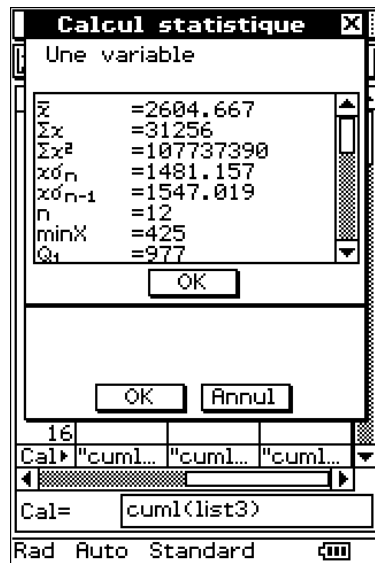


fig5 : résultats pour list2

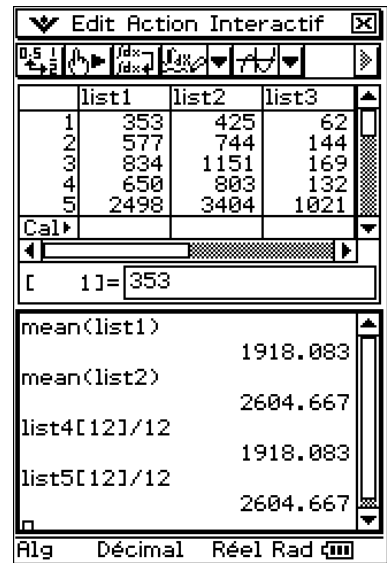


fig6 : les deux moyennes

L'énoncé demande comment calculer au mieux le nombre moyen de visites par visiteur :

- Première idée : on calcule le quotient (moyenne des visites)/(moyenne des visiteurs). Dans l'environnement de calcul formel, il suffit d'évaluer $\text{mean}(\text{list2})/\text{mean}(\text{list1})$. On trouve alors environ 1.358 visites par visiteur (fig7).
- Deuxième idée : la moyenne des quotients (visites d'un mois)/(visiteurs de ce mois). Dans l'environnement de calcul formel, il suffit d'évaluer $\text{mean}(\text{list2}/\text{list1})$. En effet l'expression $\text{list2}/\text{list1}$ renvoie la liste des quotients deux à deux des éléments de list2 (les visites mensuelles) par ceux de list1 (les visiteurs mensuels). Avec cette deuxième idée, on trouve environ 1.336 visites par visiteur (fig7)

Même si les deux résultats obtenus sont assez proches, ils sont différents.

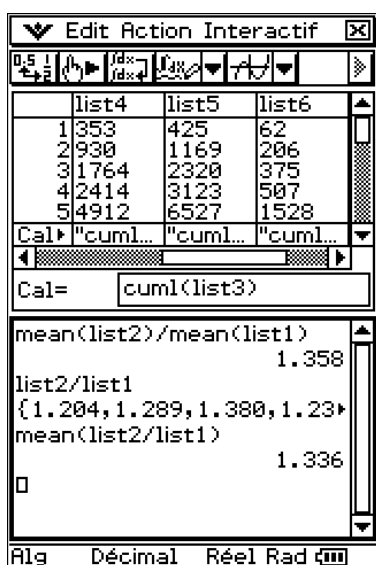


fig7 : quotient des moyennes et moyenne des quotients

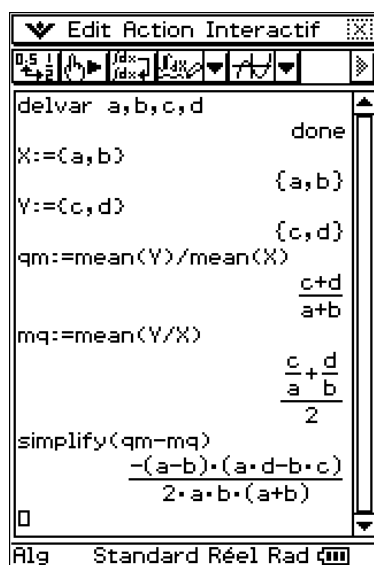


fig8 : une vérification par le calcul formel

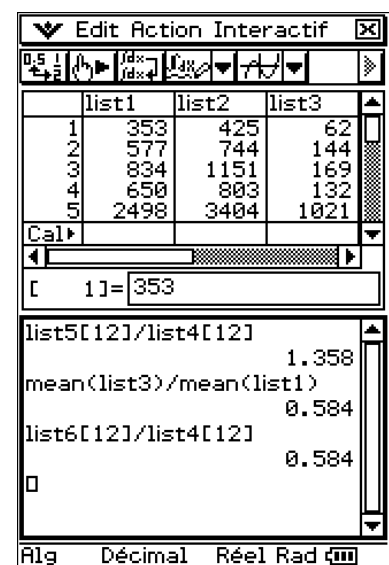


fig9 : bande passante moyenne par visiteur

La question posée par l'énoncé est "pourquoi n'obtient-on pas le même résultat?".

Il serait sans doute plus judicieux de demander pourquoi on devrait obtenir deux résultats identiques. En effet, dans un cas, on calcule le quotient des moyennes de deux listes, et dans l'autre on effectue la moyenne des quotients deux à deux des éléments de cette liste.

Pour se convaincre qu'il n'y a aucune que ces deux valeurs coïncident, il suffit de prendre l'exemple de deux listes réduites à deux éléments.

Pour le détail des calculs, voir (fig8).

- On définit les deux listes $X = \{a, b\}$ et $Y = \{c, d\}$.
- On place dans qm le quotient des moyennes $\frac{c+d}{2}$ et $\frac{a+b}{2}$, c'est-à-dire $\frac{c+d}{a+b}$.
- On place dans mq la moyenne des quotients $\frac{c}{a}$ et $\frac{d}{b}$, donc $\frac{1}{2}\left(\frac{c}{a} + \frac{d}{b}\right)$.
- On vérifie que $qm - mq = \frac{(b-a)(ad-bc)}{2ab(a+b)}$.
- Il n'y a donc aucune raison d'avoir $qm = mq$, sauf dans deux cas évidents :
 - ◊ Si $a = b$: alors $qm = mq = \frac{c+d}{2a}$ (moy. de Y divisée par valeur constante de X).

Il en aurait été de même, dans l'exemple de l'exercice-jury, si le nombre des visiteurs avait été constant chaque mois de l'année.

- ◊ Si $ad = bc$ c'est-à-dire si $\frac{c}{a} = \frac{d}{b}$, donc s'il y a un coefficient de proportionnalité λ de la liste Y relativement à la liste X . Dans ce cas $qm = mq = \lambda$.

Il en aurait été de même, dans l'exemple de l'exercice-jury, si le nombre de visites avait été proportionnel, tout au long de l'année, au nombre de visiteurs.


Il semble plus judicieux de calculer le « quotient des moyennes » (ce qui revient à faire un bilan de l'assiduité des visiteurs en fin d'année) plutôt que la moyenne des quotients : cette dernière façon de procéder donnerait en effet la même importance à chaque mois indépendamment du nombre de visiteurs, ce qui n'est pas très prudent (dans notre exemple, novembre est beaucoup plus représentatif que janvier).

Remarque : calculer le « quotient des moyennes » revient à faire le quotient des valeurs finales dans les listes d'effectifs cumulés. Voir par exemple (fig9) où on retrouve le quotient des moyennes de `list1` et `list2` en calculant `list5[12]/list4[12]`.

Pour calculer la bande passante utilisée en moyenne par un visiteur, on calcule le quotient de la moyenne de `list3` par la moyenne de `list1` ou (ce qui revient au même et semble ici opportun puisqu'on nous a demandé de calculer les effectifs cumulés) le quotient de `list6[12]` (bande passante totale consommée sur l'année) par `list4[12]` (nombre total de visiteurs sur l'année).

On voit (fig9) que la bande passante moyenne par visiteur est environ égale à 0.584Mo.

3. On va maintenant illustrer les données de l'exercice par des graphiques.


On revient à l'application , dans laquelle on a entré les données du tableau initial puis calculé les trois colonnes d'effectifs cumulés.

Le menu « DéfinGraph » ouvre une fenêtre (fig10) permettant de choisir et de paramétrer les tracés statistiques. Il est possible de définir simultanément neuf tracés, mais seuls seront effectivement dessinés ceux qui sont sélectionnés (case).

Le choix « Réglage... » ouvre une fenêtre à onglets dans laquelle on peut entrer et mémoriser les paramètres de chacun des neuf tracés statistiques possibles.

L'essentiel d'un tracé statistique tient en le choix de deux colonnes (une pour les abscisses, l'autre pour les ordonnées), puis l'affichage du nuage de points correspondant. Il est recommandé d'avoir coché la case « Oui » dans l'option « Fenêtre Stat Auto », pour que le Classpad ajuste automatiquement l'échelle en x et l'échelle en y aux données.

On peut par exemple tracer le nuage des (x_k, y_k) où x_k et y_k représentent respectivement le nombre de visiteurs et la bande passante pour le k -ème mois de l'année.

Pour cela, on choisit (dans GraphStat1 par exemple) les listes list1 et list3 comme indiqué (fig11). On valide le choix et on trace en choisissant .

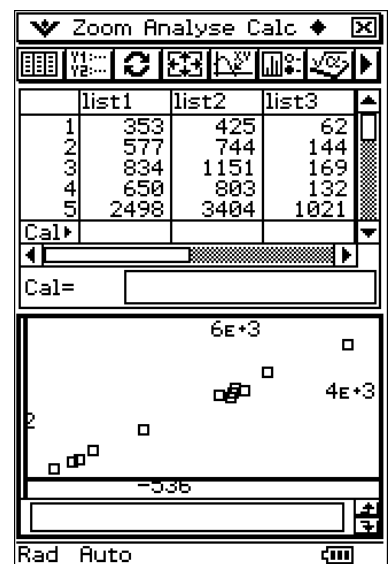
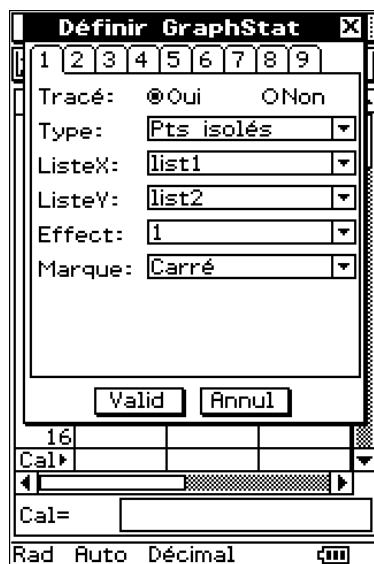
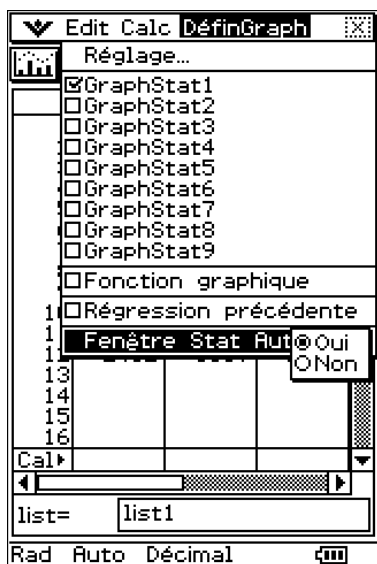


fig10 : le menu DéfinGraph fig11 : choix de GraphStat1 fig12 : (visiteurs,visites)

On constate que les points obtenus sont bien alignés : tout au long de l'année, le nombre de visite augmente donc linéairement avec le nombre de visiteurs (en principe, un même visiteur, même s'il charge frénétiquement la page d'accueil du site, ne sera compté qu'une fois tant qu'il conserve la même adresse IP).

Pour visualiser et quantifier l'alignement des points, on peut choisir la fonction « Cal/Rég linéaire ». On renseigne alors la fenêtre de saisie comme indiqué (fig13).

La réponse du Classpad est d'abord numérique (fig14) : on voit que les points (x_k, y_k) sont sensiblement alignés sur la droite d'équation $y = 1.384x - 49.526$ (l'alignement est excellent, car le coefficient de corrélation r est très proche de 1).

La réponse est ensuite graphique (fig15), la droite de régression venant se superposer au nuage de points (on a ici maximisé la fenêtre graphique, et effectué un zoom boîte pour que le tracé occupe la plus grande partie possible de l'écran).

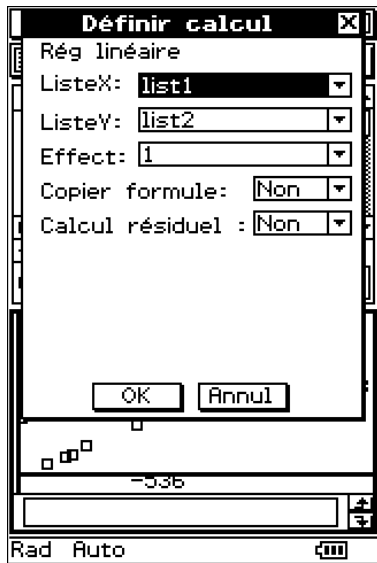


fig13 : Définir calcul stat

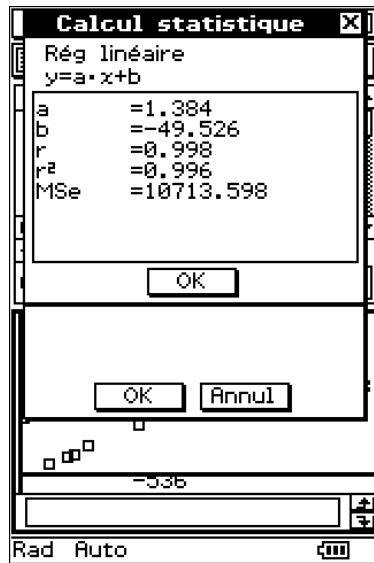


fig14 : régression linéaire

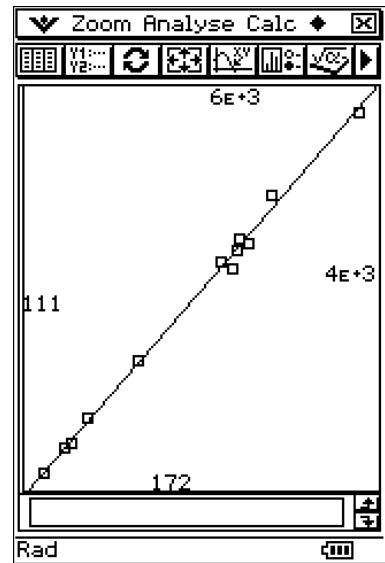



fig15 : nuage + droite

On peut aussi tracer de nombreux graphiques liés au tableau des données de l'exercice, mais à condition de passer dans l'environnement  (nous aurions d'ailleurs très bien pu effectuer tous les calculs précédents dans cet environnement).

On ouvre donc , et on fait éventuellement place nette avec « Fich/Nouveau ».

Pour l'instant, les données ont été placées dans les variables list1 à list6.

On commence par importer le contenu de list1 dans la première colonne du tableau, c'est-à-dire à partir de la colonne A1 (fig16). On procède de même, si on le souhaite, pour placer le contenu des listes list2 à list6 dans les colonnes B à F (fig17).

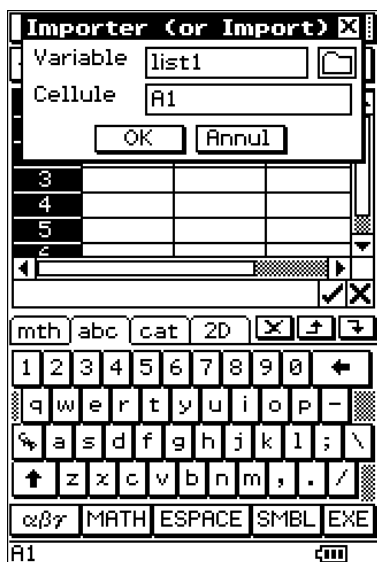


fig16 : importer list1

	A	B	C
1	353	425	62
2	577	744	144
3	834	1151	169
4	650	803	132
5	2498	3404	1021
6	2324	3254	907
7	2636	3482	589
8	1410	1916	274
9	2525	3553	681
10	2897	4135	2600
11	3861	5232	4372
12	2452	3157	2499
13			
14			
15			

fig17 : import terminé

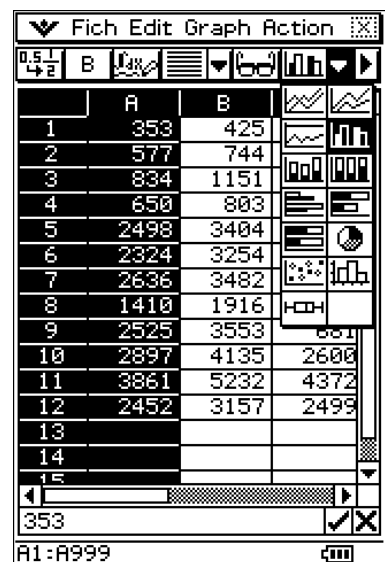


fig18 : styles de graphique

On peut ensuite sélectionner par exemple la colonne A (qui donne le nombre de visiteurs par mois) puis choisir le type de graphique parmi ceux, et ils sont nombreux, que propose le classpad (fig18).

On peut par exemple choisir de représenter les données de la colonne A sous la forme d'un camembert (fig19). Si on pointe le stylet sur l'une des parts, on constate à quelle cellule (c'est-à-dire quelle ligne, donc pour nous quel mois) cette part correspond (ici nous avons marqué la part relative au cinquième mois : 2498 visiteurs, soit 11% du total).

On peut aussi, par exemple, représenter les données de C (la bande passante utilisée) sous la forme d'un diagramme en bâtons verticaux (fig 20). Si on pose le stylet sur l'un de ces bâtons, on voit à quel mois et à quel effectif il correspond (ici nous avons sélectionné le onzième mois, celui où la bande passante utilisée est la plus importante).

On peut aussi sélectionner plusieurs colonnes (ici A, B et C) et afficher les lignes polygonales traduisant, mois après mois, les variations des données du tableau (poser le stylet sur un des points permet d'identifier à quel donnée il correspond). Ici, par exemple (fig21), nous avons marqué le point correspondant à la bande passante du mois de décembre.

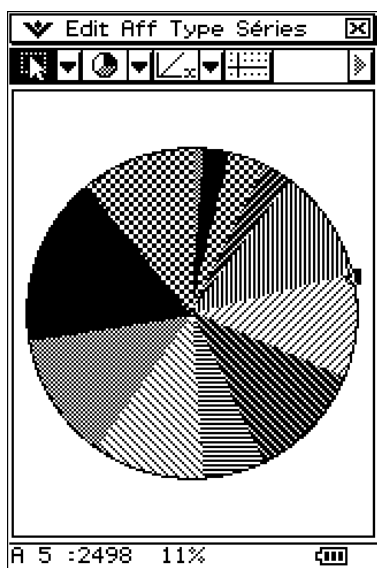


fig19 : camemberts visiteurs

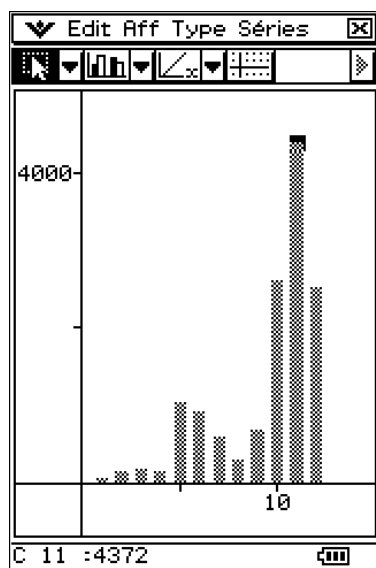


fig20 : bande passante

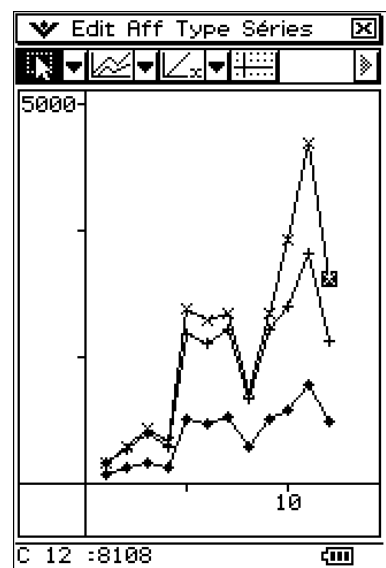


fig21 : variations comparées

On peut aussi (et ce sera là notre dernier exemple, car il y a bien d'autres possibilités) dessiner un histogramme de données (par exemple pour la colonne A).

On peut définir l'origine et l'amplitude des classes avec la fonction « Ampl. Classe » du menu « Séries ».

Ici, par exemple, l'origine est en 0 et l'amplitude des classes est de 500. Nous formons donc les classes $[0, 500[$, $[500, 1000[$, etc (fig22).

Poser le stylet sur l'écran permet de savoir précisément à quelle classe on a affaire. On voit ici que la classe $[2500, 3000[$ contient trois éléments (autrement dit, il y a eu trois mois pendant lesquels le nombre de visiteurs s'est situé dans cet intervalle).

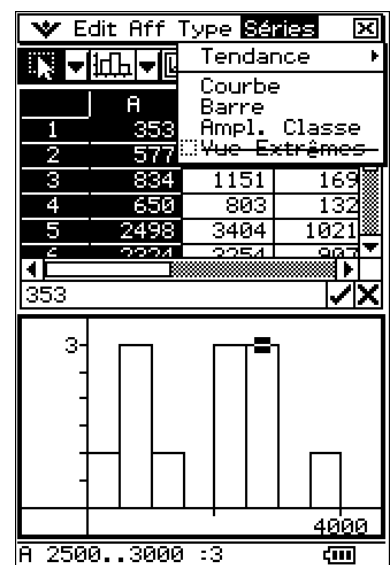


fig22 : histogramme visiteurs