1. Objet du TP. On va tester une hypothèse d'équirépartition. Il peut s'agir du nombre de clients par jour ouvrable dans une banque (y a-t-il le même nombre de clients chaque jour ?), l'équirépartition des chiffres dans les décimales du nombre pi, ou l'exemple suivant, sur lequel on s'appuiera : on lance 300 fois un dé, et on obtient les résultats suivants :

Face	1	2	3	4	5	6
Effectif	41	54	46	61	39	59

Au vu des résultats, peut-on conclure que le dé est pipé ou honnête ?

1ère méthode (terminale S et ES depuis le bac 2003)
On va lancer un grand nombre n de fois un dé non pipé, et noter les fréquences f_i d'apparition de chaque face i. Pratiquement, on fait une simulation sur ordinateur, et on note
$d^2 = \sum_{i=1}^{6}(f_i - 1/6)^2$
Plus le nombre n de simulations sera grand, et plus les f_i vont se rapprocher de 1/6, et donc plus d^2 va diminuer. Pour comparer deux simulations d'effectifs différents, on va en fait calculer nd^2. On va donc considérer la variable aléatoire ND^2 dont les réalisations sont les nd^2 obtenus par simulation.
Pour chaque simulation de n lancers, on obtient une nouvelle valeur de nd^2, et c'est cette série de nombres qui va nous permettre de prendre une décision.

On choisit l'hypothèse nulle H_0 selon laquelle le dé est honnête que l'on teste contre l'hypothèse H_1 selon laquelle le dé est pipé.

On choisit un seuil α , risque de rejet à tort de l'hypothèse H_0, et on en déduit l'intervalle d'acceptation $J(\alpha) := [0;y(\alpha)]$ tel que $P(ND^2 \notin J(\alpha)) = \alpha$. Pour ce premier TP, on choisira $\alpha=10\%$, de sorte que $J(10\%)=[0;D_9]$, en appelant D_9 le 9ème décile de la série des nd^2 obtenus par simulation.

On conclut : si la valeur de nd^2 observée pour le dé à tester n'appartient pas à $J(\alpha)$, donc pour ce premier TP si nd^2 observé $>D_9$, on rejette l'hypothèse H_0 et le dé est déclaré pipé, avec un risque d'erreur de 10%. Sinon, l'hypothèse H_0 n'est pas rejetée, et on en déduit que les écarts de fréquences observés ne permettent pas de mettre en doute le fait que le dé soit honnête.

2ème méthode
On peut éviter de faire cette série de simulations de nd^2 avec un dé honnête, car il existe un outil mathématique qui donne plus simplement les résultats. En appelant O_i l'effectif observé et T_i l'effectif théorique attendu, (ici $T_i = 300*1/6 = 50$), on fabrique $\sum_{i=1}^{6}\frac{(O_i - T_i)^2}{T_i}$ appelé khi-deux de l'échantillon.
Comme on connaît l'effectif total, l'un des 6 effectifs observés peut se déduire des 5 autres : on dit alors que c'est un khi-deux à 5 degrés de liberté. Cette quantité est aléatoire comme on l'a vu précédemment. On peut montrer que lorsque la taille de l'échantillon est suffisamment grande (ici 300) , la loi de probabilité du khi-deux a une densité f(x) qu'on peut assimiler à celle d'une variable aléatoire notée $\chi^2(5)$.
Comme précédemment, on choisit un seuil α, risque de rejet à tort de l'hypothèse H_0, et on en déduit l'intervalle d'acceptation $I(\alpha) := [0;x(\alpha)]$ tel que $P(\chi^2(5)>x(\alpha)) = \alpha$.

Lien entre les deux méthodes

Dans notre exemple, $O_i = nf_i$ et $T_i = n/6$, de sorte que $\frac{(O_i - T_i)^2}{T_i} = n\frac{(f_i - 1/6)^2}{1/6} = 6n(f_i - 1/6)^2$ d'où
$\chi^2(5) = 6.(ND^2)$.

2. Simulation
Ouvrir un nouveau classeur qu'on enregistre sous le nom **Simul_ Khi2**.
Faire un clic droit sur l'onglet **Feuil1** et le renommer **Simul**.
En **A2**, simuler le résultat d'un dé équilibré. Vérifier avec F9
De **B1** à **G1**, écrire les entiers de 1 à 6.

En **B2**, écrire une formule indiquant 1 si A2=B1, et 0 sinon. Cette même formule sera recopiée jusqu'en **G2**, puis la plage **A2:G2** sera recopiée jusqu'à la ligne 1001. Il faut donc en **B2** placer les signes $ là où ils doivent être, puis recopier comme indiqué.
En **B1002**, écrire la fréquence f_1 d'apparition du 1. Recopier la formule jusqu'en **G1002**.
En **B1003**, calculer $\mathbf{(f_1 - 1/6)^2}$ et recopier la formule jusqu'en **G1003**.
Pour nommer **écart** la cellule **I1**, cliquer sur **I1**, puis :
Insertion – Nom – Définir – Nom dans le classeur : écart -OK
et y introduire la formule **=1000*SOMME(B1003:G1003)**.
Aligner à gauche. En **H1** écrire **1000d^2=** et aligner à droite. Enfin mettre en gras la plage **H1:I1**
Appuyer plusieurs fois sur F9 pour observer les fluctuations de la cellule **I1**.
Ces fluctuations sont très importantes : nous allons les étudier.

3. Création d'une série de 500 simulations

Pour étudier la répartition des différentes valeurs apparaissant dans **I1**, on va fabriquer une macro.
Renommer **Série** la **Feuil2**.
Nommer **début** la cellule **A1**. Méthode rapide : cliquer dans **A1**, puis dans la **Zone Nom** (au dessus de la colonne **A**). **A1** devient blanc sur fond bleu. Écrire par-dessus le nom qu'on attribue à la cellule. Valider.
De même, nommer **plage** la plage **A1:A500**
Faire apparaître la barre d'outils **Commandes** (si elle n'est pas présente) par **Affichage – Barre d'outils – Commandes**
On crée maintenant la macro. Cliquer sur le rectangle gris **Bouton de commande** de la nouvelle barre d'outils. Déplacer la souris vers le milieu de la feuille : la flèche se transforme en croix noire fine. Cliquer et déplacer la souris en laissant appuyé, ce qui crée un bouton. On peut redimensionner, déplacer et renommer ce bouton.
Pour écrire le code : clic droit sur le bouton, Visualiser le code.

Écrire après la ligne **Private Sub CommandButton1_Click()** le code suivant : ˽ signifie 'espace'.

```
For˽i˽=˽0˽To˽499
a˽=˽Sheets("Simul").Range("écart")
Sheets("Série").Range("début").Offset(i,0)˽=˽a
Next˽i
Range ("plage").sort˽Key1:=Range("début")
```

La première et la quatrième ligne génèrent une boucle ; la deuxième place dans la variable **a** ce que contient la cellule "écart" de la feuille"Simul" ; et la troisième le restitue à partir de la cellule "début" de la feuille "Série" avec un décalage de i lignes et 0 colonne. Enfin la dernière ligne trie les résultats dans l'ordre croissant.
Fermer Visual Basic en cliquant en haut à droite sur la croix blanche sur fond rouge.
Pour renommer le bouton, clic droit sur le bouton, **Objet Bouton de commande – Edition**. Écrire par exemple **Action !** Puis déplacer le bouton sur la plage **B1:B2**
Cliquer sur le bouton **Désactiver le mode création** (équerre bleue). Appuyer sur **Action !**
En C1, écrire la formule **=A450** qui donne le neuvième décile de la série. Renommer **Résultats** la **Feuil3.**
Sur la feuille **Résultats**, écrivez en **A1** votre nom. En **A3**, écrire **D9 observé** et écrire de A4 à A8 le neuvième décile de 5 séries (arrondi au millième). En **D8** écrire **D9 =** et en **E8** calculer la moyenne des 5 valeurs observées pour **D9**. Arrondir au millième. Aligner **D8** à droite, **E8** à gauche et mettre les deux cellules en gras.

4. Utilisation du neuvième décile

On lance 300 fois un dé et on veut tester s'il est équilibré. Les résultats sont ceux indiqués au paragraphe 1.
Reproduire ce tableau dans la plage **C3:I4**. En **C5** écrire $\mathbf{f_i}$ et en **C6** $\mathbf{(f_i\text{-}1/6)\hat{}2}$. Compléter la plage **D5:I6** et mettre des bordures au tableau. Corriger les largeurs de colonnes pour que tout rentre dans une page. Centrer toutes les cellules du tableau. En **G8** écrire **nd^2 =** et en **H8** le calculer pour le dé à tester. Même présentation que pour la plage **D8:E8**.

5. Utilisation du khi-deux

En **D19** écrire **khi2_th** (pour théorique) et en **E19** écrire la formule **=KHIDEUX.INVERSE(0,1;5)**. Même présentation que pour la plage **D8:E8**.
En **G19** écrire **khi_2obs** (pour observé) et en **H19** le calculer (voir lien entre les deux méthodes de la première page). Même présentation que pour la plage **D8:E8**.

6. Courbe du khi-deux et histogramme de la simulation.

On a vu qu'il fallait multiplier par 6 les valeurs de **nd^2** pour les comparer à celles du khi-deux. On va donc, dans la feuille **Simul**, écrire **6000d^2** dans **H1** et corriger la formule dans **I1**. Dans la feuille **Série**, on va insérer deux lignes au dessus de la ligne 1 de la manière suivante : sélectionner la ligne 1 par clic droit, insertion. Puis recommencer. De **C1** à **Z1**, écrire les entiers de 0 à 23.
Appelons momentanément F la fonction de répartition de $\chi^2(5)$. Pour approcher sa densité en x, on va calculer $\dfrac{F(x+0{,}001)-F(x)}{0{,}001} = 1000\,(F(x+0{,}001)-F(x))$.
Sous Excel, **LOI.KHIDEUX(x ;5)** = $P(\chi^2(5)>x) = 1- P(\chi^2(5)\leq x) = 1-F(x)$.
En **C2** écrire la densité en **C1** et recopier la formule jusqu'en **Z2**.

Pour construire l'histogramme, il faut compter le nombre de nd^2 autour de chaque nombre de la première ligne. Ainsi, dans la colonne **K**, on va compter le nombre de nd^2 dans l'intervalle [7,5 ; 8,5[. Donc sur chaque ligne, il faudra obtenir un 1 et des 0 ailleurs.

En **C3**, taper la formule **=si(et(C$1–0,5< = $A3;$A3<C$1+0,5);1;0)**. La recopier jusqu'en **Z3**, puis recopier la plage **C3:Z3** jusqu'à la ligne **502**. Il faut donc bien placer les $ dans **C3**.

En **C503**, écrire **=MOYENNE(C3:C502)** et recopier cette formule jusqu'en **Z503**.
À l'aide de l'assistant graphique, faire sur un même graphique un histogramme de la simulation (plage **C503:Z503** et une courbe du khi-deux (plage **C2:Z2**) avec comme étiquette des abscisses la plage **C1:Z1**. Mettre sur fond blanc et effacer les étiquettes **Série1** et **Série2**. Recopier ce graphique dans la feuille **Résultats** à partir de la ligne 30.
Imprimer la feuille **Résultats** puis écrire à la main dans les deux plages blanches les conclusions que l'on peut tirer par chacune des méthodes sur la qualité du dé testé.