

## **Bonus : le mystère des temps de course des 20km de paris 2004**

Ce 17 octobre 2004, aux abords de la Tour Eiffel, s'élancent les 20km de Paris avec un peu moins de 17000 coureurs et coureuses pour la plupart amateurs.

Pas de records cette année-là, mais cet événement reste dans les mémoires des amoureux de data pour une autre raison : la distribution des temps d'arrivée. En effet, si l'on compare les distributions des temps d'arrivée de la course, on remarque une chose étrange : la distribution habituellement Gaussienne ne l'est pas. Plus troublant encore, un pic d'arrivées vers 1h50, un gouffre juste avant les 2h puis un autre pic vers 2h10, des anomalies que l'on ne retrouve que sur cette course en particulier. Il est en effet courant de trouver des dérivations dans la courbe Gaussienne aux alentours des "temps symboliques" 1h45 ou 2h dans ce cas.

Mais alors que s'est-il passé au pont d'Iéna ce jour-là?

On pourrait se dire qu'il y a eu un incident sur le parcours uniquement pour cette tranche de coureurs. Malheureusement, aucun incident de ce type n'a été à déplorer dans Paris ce jour-là.

On pourrait aussi se dire que la météo était mauvaise, ou que la pluie s'est invitée sur le parcours que les premiers concurrents auraient éviter. Là encore, rien d'anormal : la station de Paris-Montsouris a relevé une température minimale de 6,5°C et une maximale de 13,9°C, un vent modéré et aucune précipitation, rien d'anormal pour une journée d'octobre donc, ou de spécial à cette tranche de coureurs.

Penchons-nous à présent plus en profondeur vers les données de la course.

Si l'on reprend les temps de course, disponible sur le site des 20km de paris pour toutes les éditions, on voit que des temps intermédiaires sont pris aux 10km, jusque là, rien d'étonnant sur une course de 20km.

On remarque aussi qu'il y est indiqué un temps réel et un temps officiel.

Si le temps réel est le temps brut entre l'heure à laquelle la puce a passé la ligne de départ et l'heure à laquelle elle a passé la ligne d'arrivée, comment est calculé le temps officiel ? Est-ce la prise en compte des différentes vagues de départ de la course?

Les élites, forcément partis devant n'ont pas été impactés par ce calcul officiel. Les suivants beaucoup plus, ce qui corrobore un calcul sur un départ différé.

Si l'on trace la distribution basée sur les temps réels, on retrouve une forme beaucoup plus normale, avec des pics juste avant les barrières de 1h45 et 2h (graphique ci-dessous).

Selon ces données, l'hypothèse la plus probable serait de dire que la déviation des temps officiels est due à une erreur dans le calcul des rectifications de temps des départs différés.

Doit-on imputer cette erreur à l'organisation ou aux coureurs pour ne pas avoir respecté les différents horaires de départ ou simplement une erreur de chronométrage?

Une dernière chose à savoir : ce format de chronométrage a disparu dès l'année suivante et n'a jamais été utilisé ensuite...

