

## SAE 2.04 – Partie analyse de données :

- En termes de participation à un nombre de jeux olympiques Ian Millar est l'athlète ayant participé à le plus de différents jeux, pour avoir ce résultat nous avons fait cette requête :

```
SELECT name, COUNT(DISTINCT id_jeux) FROM athlete INNER JOIN participe
USING(id_athlete) GROUP BY id_athlete ORDER BY count DESC LIMIT 20;
```

name	count
Ian Millar	10
Afanasij's Kuzmins	9
Hubert Raudaschl	9
Piero D'Inzeo	9
Durward Randolph Knowles	9
Raimondo D'Inzeo	9
Paul Bert Elvstrm	9
Josefa Idem-Guerrini	8
Franois Jacques Marie Grard "Frans" Lafortune	8
Nino Salukvadze (-Machavariani)	8
Rajmond Debevec	8
Francisco Boza Dibos	8
Lesley Allison Thompson-Willie	8
Andrew James Hoy	7
Noriaki Kasai	7
Seiko Hashimoto (-Ishizaki)	7
Aladr Gerevich (-Gerei)	7
Henry William Furse "Bill" Hoskyns	7
Yekaterina Anatolyevna Khodotovich-Karsten	7
James Robert "Jim" Elder	7
(20 lignes)	

- Nous avons choisi les Jeux Olympiques de Rio 2016, et nous avons les résultats suivants

Pays	Moyenne d'âge	Nb sportif	Min âge	Max âge
203	26.255	10 956	13	62

Voici les requêtes que nous avons utilisé pour obtenir ces valeurs :

**Pour le nombre de pays :**

```
SELECT COUNT(DISTINCT NOC) FROM participe INNER JOIN jeux USING(id_jeux) WHERE
jeux.year=2016;
```

**Pour Moyenne âge :**

```
SELECT AVG(age) FROM participe INNER JOIN jeux USING(id_jeux) WHERE jeux.year=2016;
```

**NB sportif :**

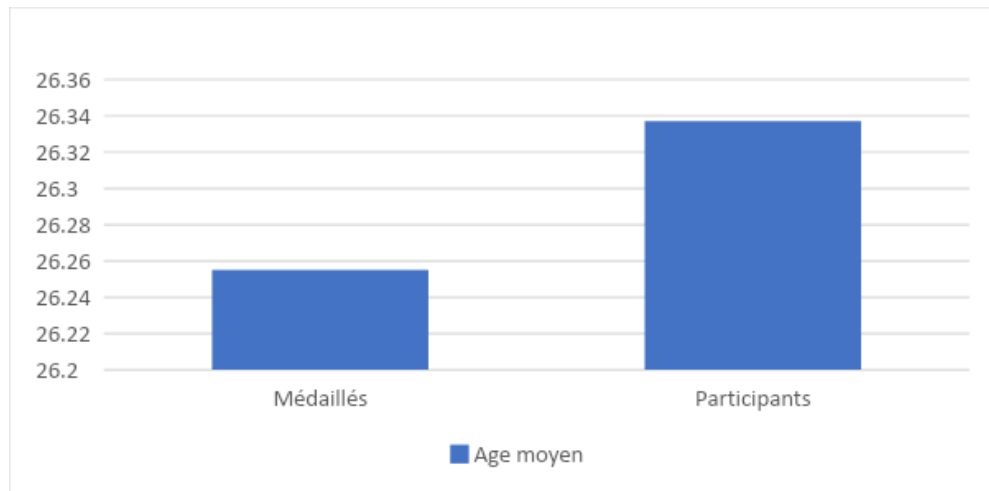
```
SELECT COUNT(DISTINCT id_athlete) FROM participe INNER JOIN jeux USING(id_jeux)
WHERE jeux.year=2016;
```

**Min âge :**

```
SELECT MIN(age) FROM participe INNER JOIN jeux USING(id_jeux) WHERE jeux.year=2016;
```

**Max âge :**

```
SELECT MAX(age) FROM participe INNER JOIN jeux USING(id_jeux) WHERE
jeux.year=2016;
```



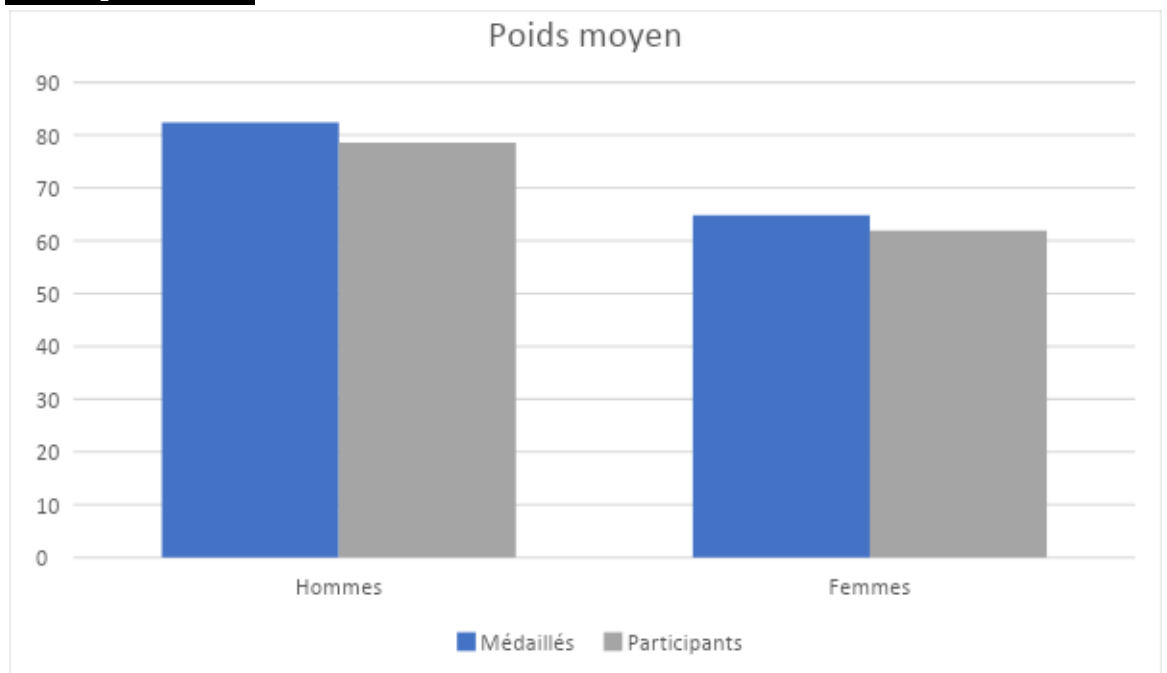
b.

Le graphique ci-dessus représente l'âge moyen de tous les participants comparé à l'âge moyen des médaillés. L'âge moyen des participants étant légèrement plus élevé, on peut donc dire qu'en moyenne les médaillés font partie des plus jeunes participants.

Nous avons obtenu ces valeurs en utilisant ces 2 requêtes :

```
SELECT AVG(age) FROM participe INNER JOIN jeux USING(id_jeux)
WHERE year=2016 AND medal IS NOT NULL;
```

```
SELECT AVG(age) FROM participe INNER JOIN jeux USING(id_jeux)
WHERE year=2016;
```



c.

Le graphique ci-dessus démontre la différence de poids moyen entre les hommes et les femmes, total ou médaillés.

On constate que les hommes sont généralement plus lourds que les femmes. En sachant que les hommes médaillés font partie des participants les plus lourds.

On peut aussi dire que les femmes médaillés font également partie des participantes les plus lourdes.

Voici les requêtes utilisées pour obtenir ces valeurs :

```
SELECT AVG(weight) FROM participe INNER JOIN athlete USING(id_athlete)
INNER JOIN jeux USING(id_jeux) WHERE year=2016 GROUP BY sex;
SELECT AVG(weight) FROM participe INNER JOIN athlete USING(id_athlete)
INNER JOIN jeux USING(id_jeux) WHERE year=2016 AND medal IS NOT NULL;
```

3.

a. Voici la liste des 15 premiers pays qui ont gagné le plus de médailles cumulés (jeux d'été et d'hiver confondus) entre 1992 et 2016 :

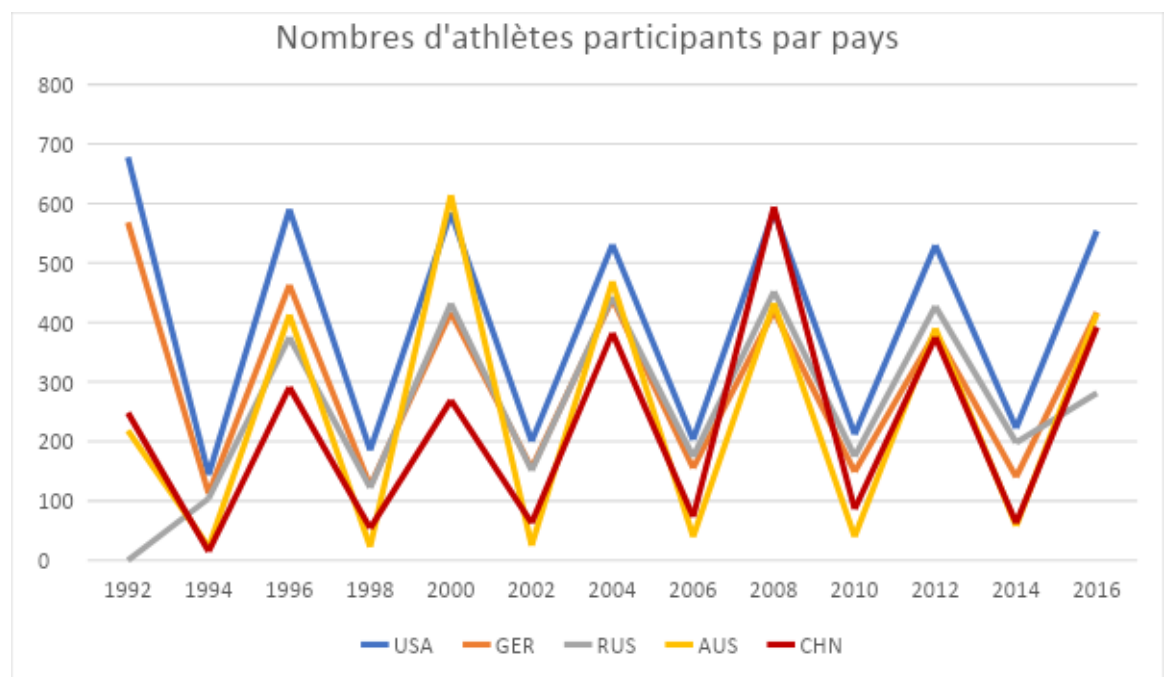
noc	count
USA	2157
GER	1263
RUS	1128
AUS	881
CHN	857
CAN	739
ITA	594
FRA	581
GBR	558
NED	506
JPN	478
KOR	419
SWE	361
ESP	358
NOR	351
(15 lignes)	

La requête utilisée pour trouver ces valeurs :

```
SELECT NOC, COUNT(medal) FROM participe WHERE id_jeux IN (SELECT id_jeux FROM jeux
WHERE year>1991 AND year<2017) GROUP BY NOC ORDER BY count DESC LIMIT 15;
```

b.

Nous avons choisi les 5 pays suivants : les Etats-Unis (USA), l'Allemagne (GER), la Russie (RUS), l'Australie (AUS) et la Chine (CHN). Il est important de souligner que la Russie ne participait pas en 1992, les athlètes russes participaient sous le nom d' « Équipe Unifiée » avec d'autres athlètes des anciennes Républiques soviétiques.



i.

En moyenne, sur les différents jeux ainsi que les différents pays, chaque délégation envoie **285** athlètes environ. De plus, nous pouvons voir grâce au graphique que les jeux d'hiver sont des jeux où il y a peu d'athlètes.

Requete :

```
SELECT COUNT(DISTINCT id_athlete) FROM participe INNER JOIN jeux  
USING(id Jeux) WHERE year>1991 AND year<2017 AND NOC='*PAYS*' GROUP BY year  
ORDER BY year;
```

Avec les calculs ci-dessous, nous pouvons affirmer qu'en moyenne les Etats-Unis sont la nation qui envoie le plus d'athlètes aux Jeux Olympiques, 25% de ces participations se sont déroulées avec au plus de 201.5 athlètes.

Quelques calculs à l'aide d'excel :

Etendu = Valeur Max - Valeur Min = 678 - 0 = 678

Médiane des participants pour chaque pays :

USA = 530, GER = 383, RUS = 198, AUS = 218, CHN = 248

Exemple pour USA : =mediane(B2;B14)

Moyenne pour chaque pays :

USA = 402, GER = 303, RUS = 257, AUS = 243, CHN = 224

Exemple pour USA : =moyenne(B2;B14)

Premier quartile :

USA = 201.5, GER = 144.5, RUS = 136.5, AUS = 32.5, CHN = 63

Exemple pour USA : =quartile.exclure(B2;B14;1)

Troisième quartile :

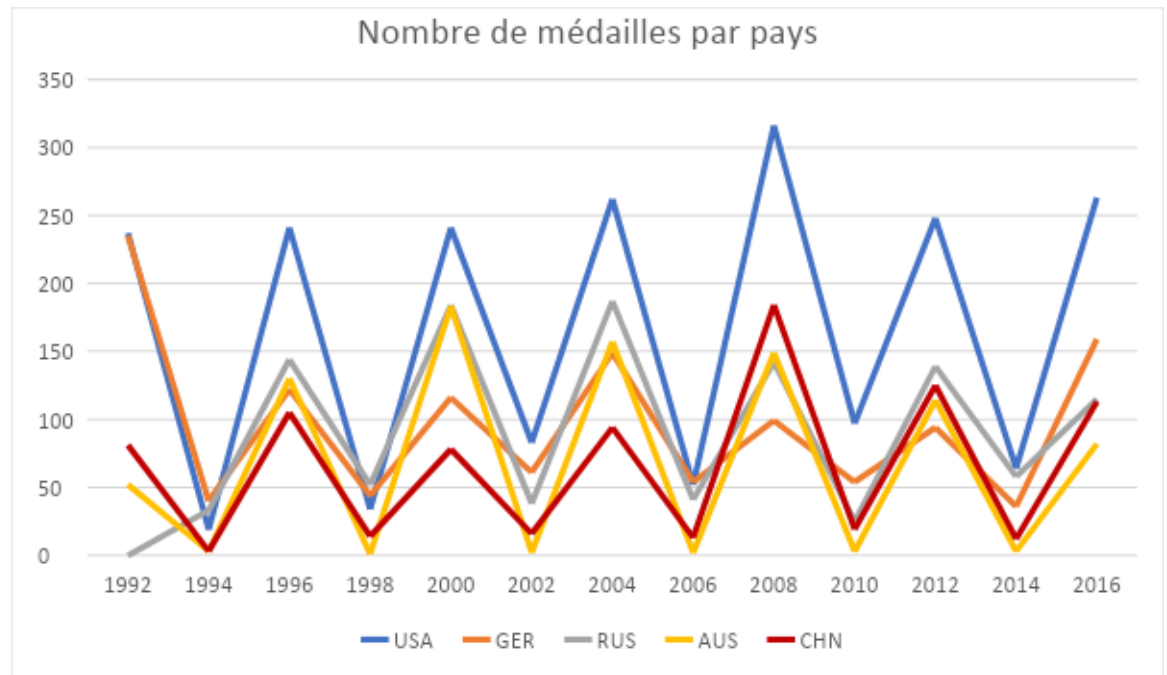
USA = 585.5, GER = 429, RUS = 429.5, AUS = 423, CHN = 378.5

Exemple pour USA : =quartile.exclure(B2;B14;3)

Ecart-type :

USA = 203.62, GER = 163.92, RUS = 153.48, AUS = 217.29, CHN = 179.41

Exemple pour USA : =ecartype(B2;B14)



ii.

On voit sur ce graphique que les Etats-Unis ont souvent été au-dessus des autres grandes nations sur le nombre de médailles gagnées. En moyenne, ces 5 grandes nations gagnent 97 médailles par jeu.

Requete :

```
SELECT COUNT(medal) FROM participe INNER JOIN jeux USING(id jeux) WHERE
year>1991 AND year<2017 AND NOC='*PAYS*' GROUP BY year ORDER BY year;
```

Notre estimation est bonne car les Etats-Unis ont la moyenne de médailles la plus haute, suivie par l'Allemagne. L'Allemagne a au moins 135.5 médailles dans 75% de ces participations.

Quelques calculs à l'aide d'excel :

Etendu = Valeur Max - Valeur Min = 316-0 = 316

Médiane des médailles pour chaque pays :

USA = 237, GER = 94, RUS = 58, AUS = 52, CHN = 78

Exemple pour USA : =mediane(B2;B14)

Moyenne de médaille pour chaque pays :

USA = 166, GER = 98, RUS = 89, AUS = 68, CHN = 66

Exemple pour USA : =moyenne(B2;B14)

Premier quartile :

USA = 58, GER = 49, RUS = 35.5, AUS = 2.5, CHN = 13.5

Exemple pour USA : =quartile.exclure(B2;B14;1)

Troisième quartile :

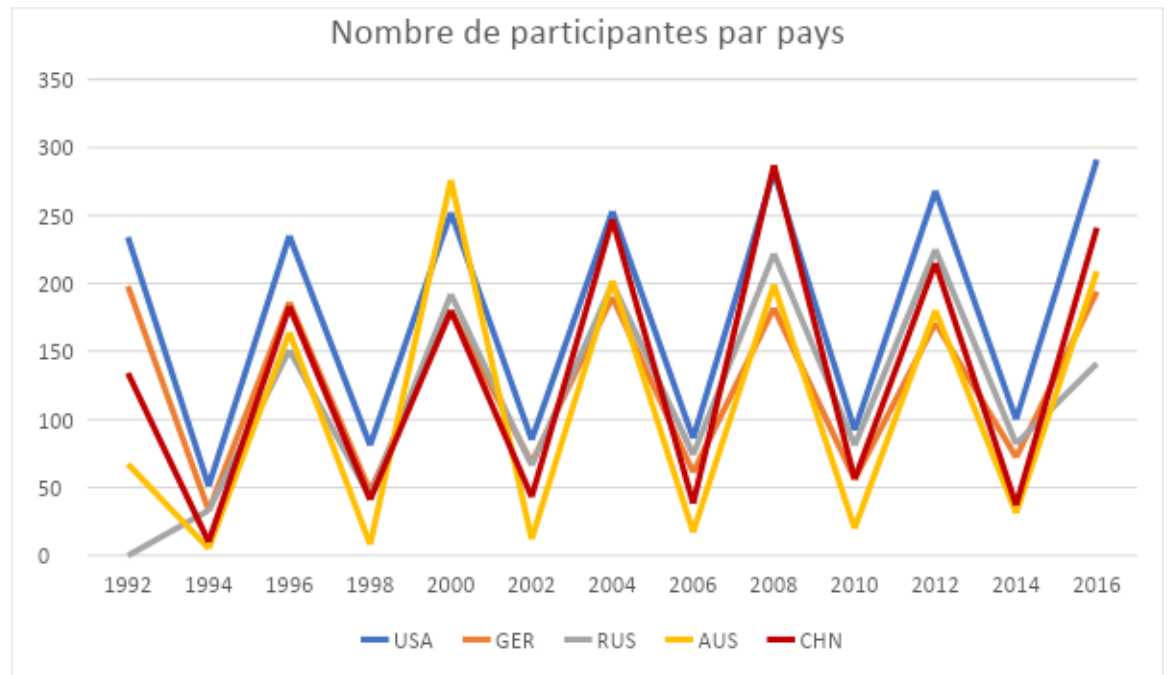
USA = 255, GER = 135.5, RUS = 143, AUS = 139.5, CHN = 109

Exemple pour USA : =quartile.exclure(B2;B14;3)

Ecart-type :

USA = 107, GER = 59, RUS = 65, AUS = 70, CHN = 58

Exemple pour USA : =ecartype(B2;B14)



iii.

On voit que la présence des femmes a augmenté légèrement au fil des années, l'Australie a été souvent le pays avec le moins de participantes sauf pour les Jeux de 2000, qui se déroulait à Sydney. La moyenne de femmes participantes est de 132 femmes.

Requête :

```
SELECT COUNT(DISTINCT id_athlete) FROM participe INNER JOIN jeux
USING(id Jeux) INNER JOIN athlete USING(id_athlete) WHERE year>1991 AND
year<2017 AND NOC='*PAYS*' AND sex='F' GROUP BY year ORDER BY year;
```

La Chine est la 2e nation des 5 nations que nous avons choisies, au niveau du nombre de femmes participantes. En effet, en moyenne lors de ces participations, elle envoie 132 athlètes féminines, et même dans 25% de ces participations, il y a au moins 228 femmes.

Quelques calculs à l'aide d'excel :

Etendu = Valeur Max - Valeur Min = 291-0=291

Médiane des participants pour chaque pays :

USA = 178, GER = 126, RUS = 116, AUS = 107, CHN = 132

Exemple pour USA : =mediane(B2;B14)

Moyenne pour chaque pays :

USA = 178, GER = 126, RUS = 116, AUS = 107, CHN = 132

Exemple pour USA : =moyenne(B2;B14)

Premier quartile :

USA = 85.5, GER = 58.5, RUS = 54.5, AUS = 14.5, CHN = 39.5

Exemple pour USA : =quartile.exclure(B2;B14;1)

Troisième quartile :

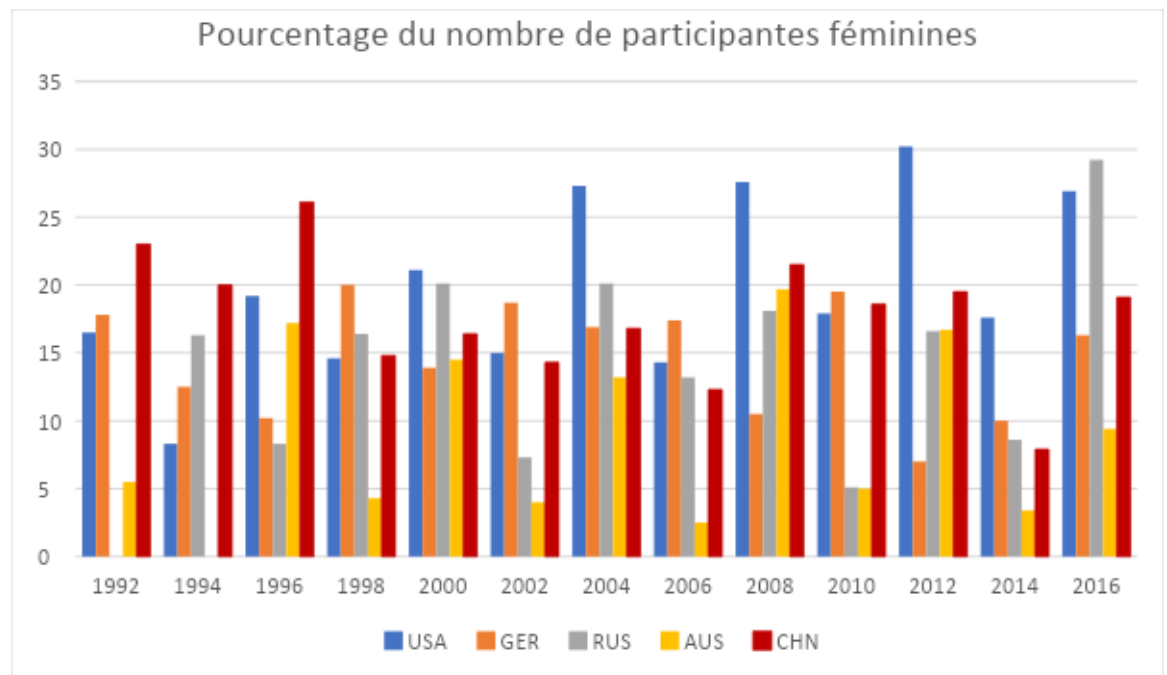
USA = 260.5, GER = 188, RUS = 196, AUS = 200.5, CHN = 228

Exemple pour USA : =quartile.exclure(B2;B14;3)

Ecart-type :

USA = 94, GER = 68, RUS = 77, AUS = 99, CHN = 98

Exemple pour USA : =ecartype(B2;B14)



iv.

Cet histogramme permet de se rendre compte que les femmes restent sous représentées même si nous avons vu que les femmes participaient de plus en plus, elles restent à la même proportion. Pour ces 5 nations, le pourcentage moyen de femmes participantes sur tous les athlètes est de 15%.

Pas de requête SQL cette fois ci, nous avons utilisé les autres graphiques, et grâce aux formules excel, nous avons obtenu cette proportion:

**=ARRONDI((NBFemmes/NBParticipants)\*100;1)**

L'Australie a un pourcentage très faible vis-à-vis de ses participantes féminines, en effet en moyenne compte parmi tous ses participants, seulement 9% de femmes. Et dans 25% des cas, la proportion de femmes n'excède pas 3.7%

Quelques calculs à l'aide d'excel :

Etendu = Valeur Max - Valeur Min = 30.2-0=30.2%

Médiane du pourcentage de participation féminine pour chaque pays :

**USA = 17.9%, GER = 16.3%, RUS = 16.3%, AUS = 5.5%, CHN = 18.6%**

**Exemple pour USA : =mediane(B2;B14)**

Moyenne pour chaque pays :

**USA = 20%, GER = 15%, RUS = 14%, AUS = 9%, CHN = 18%**

**Exemple pour USA : =moyenne(B2;B14)**

Premier quartile :

**USA = 14.8%, GER = 10.4%, RUS = 7.8%, AUS = 3.7%, CHN = 14.6%**

**Exemple pour USA : =quartile.exclure(B2;B14;1)**

Troisième quartile :

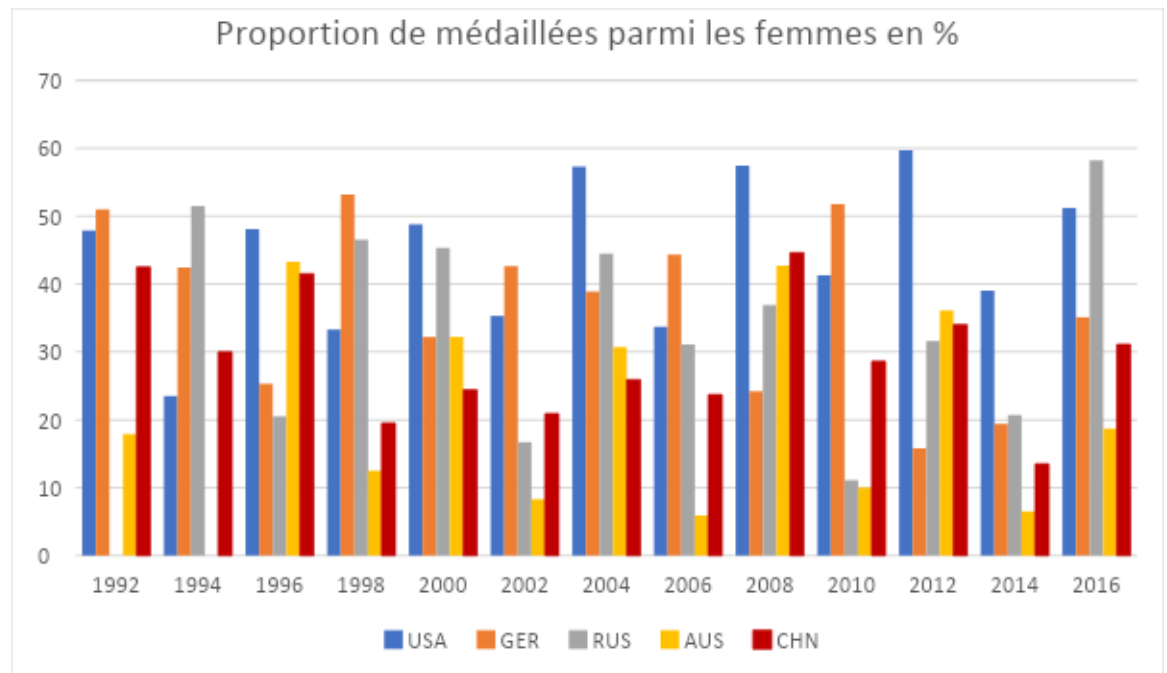
**USA = 27.1%, GER = 18.25%, RUS = 19.1%, AUS = 15.6%, CHN = 20.75%**

**Exemple pour USA : =quartile.exclure(B2;B14;3)**

Ecart-type :

**USA = 6.51%, GER = 4.24%, RUS = 7.75%, AUS = 6.57%, CHN = 4.76%**

**Exemple pour USA : =ecartype(B2;B14)**



V.

Les femmes sont souvent autrices d'exploit et comme on peut le voir beaucoup de participantes deviennent médaillées, le pourcentage moyen de médailles parmi les femmes est de 32%.

Pas de requête SQL cette fois ci, nous avons utilisé les autres graphiques, et grâce aux formules excel, nous avons obtenu cette proportion :

**=ARRONDI((NBMédaillées/NBFemmes)\*100;1)**

En moyenne, 44% des médailles américaines proviennent de femmes, et pour la moitié de ces participations de 1992 à 2016, les femmes ramènent au moins 47.9% de médailles sur le total de médailles américaines.

Quelques calculs à l'aide d'excel :

Etendu = Valeur Max - Valeur Min = 59.7-0=59.7%

Médiane de la proportion de médaillés parmi les femmes pour chaque pays :

**USA = 47.9%, GER = 38.9%, RUS = 31.6%, AUS = 20.36%, CHN = 29.24%**

**Exemple pour USA : =mediane(B2;B14)**

Moyenne pour chaque pays :

**USA = 44%, GER = 37%, RUS = 32%, AUS = 20%, CHN = 29%**

**Exemple pour USA : =moyenne(B2;B14)**

Premier quartile :

**USA = 34.5%, GER = 24.75%, RUS = 18.6%, AUS = 7.4%, CHN = 22.3%**

**Exemple pour USA : =quartile.exclure(B2;B14;1)**

Troisième quartile :

**USA = 54.25%, GER = 47.65%, RUS = 45.9%, AUS = 34.15%, CHN = 37.75%**

**Exemple pour USA : =quartile.exclure(B2;B14;3)**

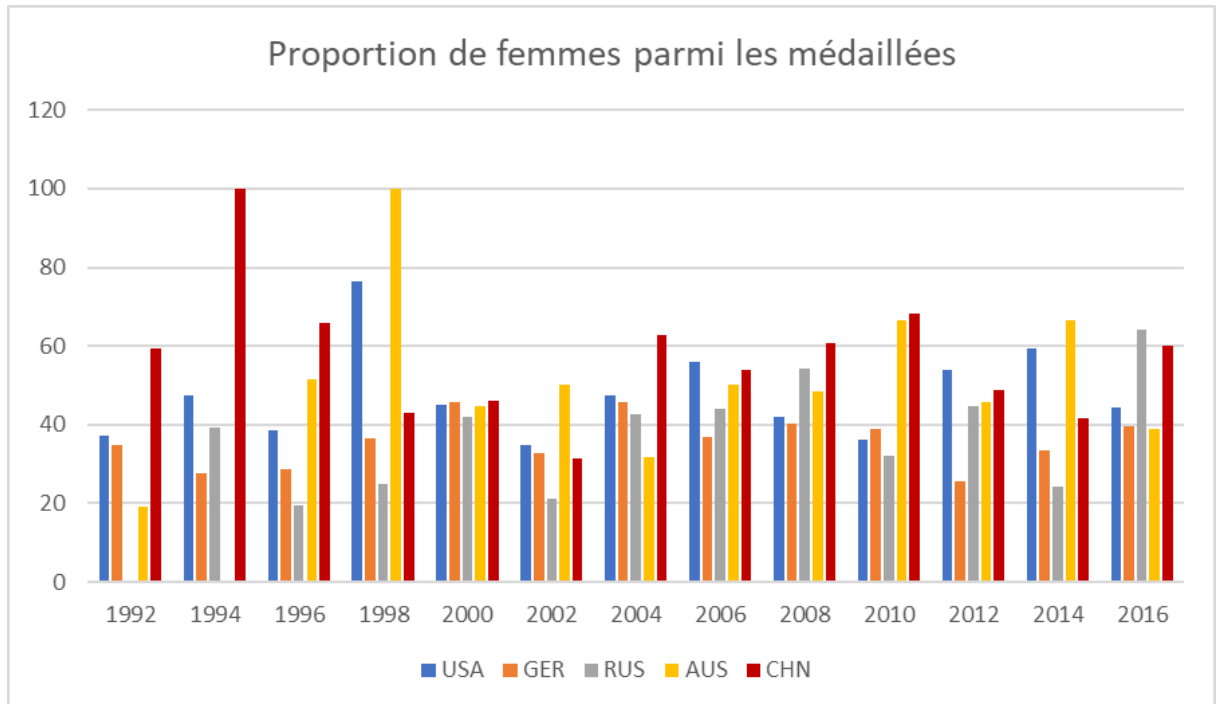
Ecart-type :

**USA = 11%, GER = 12.5%, RUS = 17.3%, AUS = 14.9%, CHN = 9.4%**

**Exemple pour USA : =ecartype(B2;B14)**



vi.



On peut voir que pour certaines délégations les femmes ont été les seules sources de médailles. En moyenne la proportion de femmes parmi les médaillés est de 44.5%.

Requête : Pas de requête SQL cette fois ci, nous avons utilisé les autres graphiques, et grâce aux formules excel, nous avons obtenu cette proportion :

**=ARRONDI((NBFemmesMédaillées/NBMédaillés)\*100;1)**

La Russie a en moyenne une proportion de 35% de femmes parmi les médaillées et dans 75% des cas au plus de 22.6% de médailles féminines.

Quelques calculs à l'aide d'excel :

Etendu = Valeur Max - Valeur Min = 100-0=100%

Médiane de la proportion de femme parmi les médaillés pour chaque pays :

**USA = 45.2%, GER = 36.4%, RUS = 39.4%, AUS = 48.3%, CHN = 59.3%**

**Exemple pour USA : =mediane(B2;B14)**

Moyenne pour chaque pays :

**USA = 48%, GER = 36%, RUS = 35%, AUS = 47%, CHN = 57%**

**Exemple pour USA : =moyenne(B2;B14)**

Premier quartile :

**USA = 37.85%, GER = 30.75%, RUS = 22.6%, AUS = 35.4%, CHN = 44.55%**

**Exemple pour USA : =quartile.exclure(B2;B14;1)**

Troisième quartile :

**USA = 11.60%, GER = 6.34%, RUS = 16.9%, AUS = 24%, CHN = 17%**

**Exemple pour USA : =quartile.exclure(B2;B14;3)**

Ecart-type :

**USA = 12%, GER = 6%, RUS = 17%, AUS = 24%, CHN = 17%**

**Exemple pour USA : =ecartype(B2;B14)**