La fitxa estadística

Centre, dispersió i forma de `games_clean`

Noms _____

Superset → **SQL → SQL Lab**, base de dades `Dades classe`, esquema `public`, taula `games_clean` (26 685 jocs). Per parelles: un tecleja, l'altre escriu; canvieu a la Tasca 3. Escriviu els nombres aquí.

Referència · la cua

Poseu les dades en una cua, el més petit primer. Cinc amics guanyen 1 200, 1 300, 1 400, 1 500 i 30 000: la **mitjana** (sumar-ho tot i repartir-ho a parts iguals) és 7 080; la **mediana** (el del mig de la cua) és 1 400.

Pregunta	A la cua	SQL
Centre	On és el mig?	`AVG(x)` · `PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY x)`
Dispersió	Quant de lluny són el primer i l'últim? I la meitat central?	`MIN(x)` · `MAX(x)` · `STDDEV(x)` · percentils 0.25 i 0.75, els **quartils** Q1 i Q3
Forma	On s'amuntega la cua?	`width_bucket(x, from, to, n)` + `COUNT(*)`

Llegiu `WITHIN GROUP (ORDER BY price)` com "feu la cua per preu" i `PERCENTILE_CONT(0.5)` com "camineu fins al 50% de la cua i llegiu-hi el valor". La **desviació estàndard** és la distància típica a la mitjana. `width_bucket` diu en quina de les `n` caixes iguals entre `from` i `to` cau un valor; tot el que passa de `to` va a la caixa `n + 1`.

Tasca 1 · Centre i extrems de `price` (8 min)

1.1 Abans d'executar res: la mitjana del preu serà superior o inferior a la mediana? _____ Després executeu:

```
SELECT COUNT(*) AS n, MIN(price) AS lowest, MAX(price) AS highest,
       ROUND(AVG(price), 2) AS mean,
       PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY price) AS median
FROM games_clean;
```

`n` _____ `lowest` _____ `highest` _____ `mean` _____ `median` _____ Ho heu encertat? _____

1.2 Qui és al final de tot de la cua? Trobeu les tres files més cares (`ORDER BY price DESC LIMIT 3`).

L'últim de la cua: _____ És un joc? _____

1.3 Un company diu: "La mediana és a mig camí entre el més baix i el més alt." Calculeu aquest nombre.

A mig camí: _____ És la mediana? Per què no? _____

Tasca 2 · Dispersió (7 min)

```
SELECT ROUND(STDDEV(price), 2) AS std_dev,
       PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY price) AS q1,
       PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY price) AS q3
FROM games_clean;
```

`std_dev` _____ `q1` _____ `q3` _____ La meitat dels jocs costen entre _____ i _____

2.2 Compteu els jocs que costen més que la mitjana (`WHERE price >` la vostra mitjana).

Jocs per sobre de la mitjana: _____ de 26 685. És la meitat? _____

Tasca 3 · Forma de `price` (9 min)

Sis caixes de 10 de 0 a 60; tot el que passa de 60 va a la caixa 7. L'última columna dibuixa la barra: un `#` per cada 500 jocs.

```
SELECT width_bucket(price, 0, 60, 6) AS bucket,
       MIN(price) AS lowest, MAX(price) AS highest, COUNT(*) AS games,
       REPEAT('#', (COUNT(*) / 500)::int) AS bar
FROM games_clean GROUP BY bucket ORDER BY bucket;
```

Copieu la columna `games` a la taula de l'esquerra de més avall i dibuixeu les set barres. Després marqueu al dibuix on cauen la mitjana (**M**) i la mediana (**Md**).

Tasca 4 · Una segona columna: `metascore` (9 min)

Repetiu les Tasques 1 a 3 per a `metascore`, en dues consultes. Canvis: `ROUND(AVG(metascore)::numeric, 1)`; caixes `width_bucket(metascore, 20, 100, 8)`; un `#` per cada 25 jocs.

4.1 Executeu la consulta de les caixes tal com està. Surt una fila de més.

La seva caixa: _____ Els seus jocs: _____ Qui són? _____

4.2 Afegiu `WHERE has_metascore` a totes dues consultes i ompliu la fitxa i la taula de la dreta.

`n` _____ mínim _____ màxim _____ mitjana _____ mediana _____ desv. est. _____ Q1 _____ Q3 _____

Caixa	Jocs	`price`: la vostra barra
1		
2		
3		
4		
5		
6		
7		

Caixa	Jocs	`metascore`: la vostra barra
1		
2		
3		
4		
5		
6		
7		
8		

4.3 Mireu els dos dibuixos de costat.

Quina columna té mitjana ≈ mediana? _____ Quina forma té el seu dibuix? _____

I l'altra? _____

Heu acabat quan Tots els espais estan omplerts · els dos dibuixos tenen **M** i **Md** marcades. **Parella ràpida?** Afegiu `GROUP BY owners_tier` a la consulta de la Tasca 1: els jocs Hit costen més? Després, la fitxa estadística d'una columna numèrica del vostre conjunt de dades obert.

Encallats?

- **`function round(double precision, integer) does not exist`**: convertiu primer el tipus, `ROUND(metascore::numeric, 1)`.
- **El `lowest` d'una caixa no és un nombre rodó**: és correcte. És el joc més barat **de dins** la caixa.
- **`relation "games_clean" does not exist`**: demaneu `games_clean.sql` (18/11) al professor i torneu-lo a executar.

A casa Acabeu la Tasca 4 si cal. Guardeu aquest full: dilluns es converteix en la secció de "xifres clau" del vostre informe.