

# Analyse en composante principales

Choix du nombre de composantes

---

Véronique Tremblay

“ “

## Proportion de la variation expliquée

---

La **proportion de variation expliquée** la composante principale  $Y_i$  est

$$\frac{\lambda_i}{\lambda_1 + \dots + \lambda_p}.$$

## Choix du nombre de composantes

---

## La règle des 80%

On garde les composantes qui représentent 80% de la variance.

	Variance	% Variance	% Cumulé
comp 1	4.05	50.57	50.57
comp 2	1.76	21.98	72.55
comp 3	0.77	9.57	82.12
comp 4	0.65	8.15	90.27
comp 5	0.38	4.71	94.98
comp 6	0.24	3.05	98.03
comp 7	0.11	1.33	99.36
comp 8	0.05	0.64	100.00

## La règle de Kaiser

On garde les composantes qui correspondent aux  $\lambda > 1$ .

	Variance	% Variance	% Cumulé
comp 1	4.05	50.57	50.57
comp 2	1.76	21.98	72.55
comp 3	0.77	9.57	82.12
comp 4	0.65	8.15	90.27
comp 5	0.38	4.71	94.98
comp 6	0.24	3.05	98.03
comp 7	0.11	1.33	99.36
comp 8	0.05	0.64	100.00

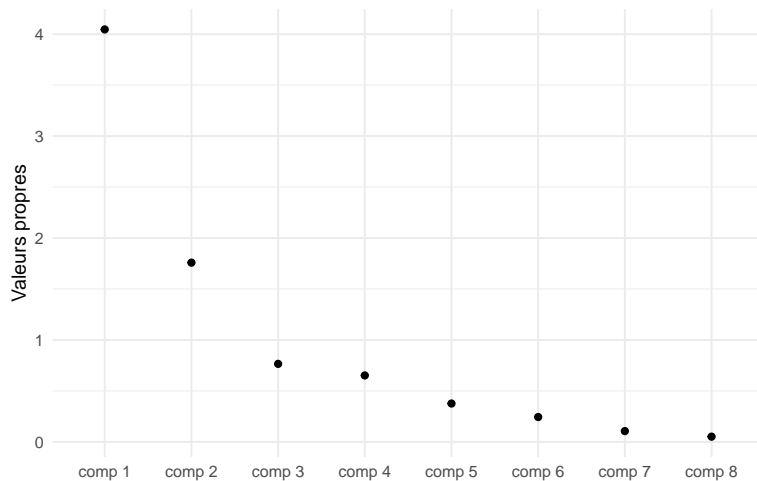
## La règle de Joliffe

On garde les composantes qui correspondent aux  $\lambda > 0.70$ .

	Variance	% Variance	% Cumulé
comp 1	4.05	50.57	50.57
comp 2	1.76	21.98	72.55
comp 3	0.77	9.57	82.12
comp 4	0.65	8.15	90.27
comp 5	0.38	4.71	94.98
comp 6	0.24	3.05	98.03
comp 7	0.11	1.33	99.36
comp 8	0.05	0.64	100.00

# La règle de Cattell

On conserve les composantes qui précèdent le pied de l'éboulis.





Les composantes de variance maximale ne sont pas nécessairement les plus importantes pour la prédiction.

Les composantes de variance maximale ne sont pas nécessairement les plus intéressantes.