

# Executive Master Statistique et Big Data

## Modèle linéaire généralisé

### Devoir maison 1

Robin Ryder

Pour le 4 juillet 2017

Devoir maison optionnel, uniquement susceptible d'augmenter votre note finale. Comptez-le et code R à envoyer par courriel avant le cours du 4 juillet.

Vous traiterez au choix l'un ou l'autre de ces deux sujets. Vous devrez importer les données sous R, les mettre sous un format utilisable, et appliquer la ou les méthodes qui vous semblent pertinentes pour les analyser. N'hésitez pas à réfléchir également aux limites de l'approche choisie.

## 1 Présidents américains

Vous chercherez à expliquer le succès à l'élection présidentielle américaine des candidats en fonction des caractéristiques données dans les deux onglets du fichier `US_presidents.xls`. Pour chaque candidat majeur (élu ou battu) depuis 1896, vous disposez de :

- la taille en cm ;
- l'année de naissance ;
- le nombre d'années d'expérience (au jour de l'élection) en tant que : président, vice-président, gouverneur, sénateur, représentant à la chambre, élu dans une législature d'état, général d'armée, membre du gouvernement. Une case vide correspond à la valeur 0 (et non à un NA) ;
- le parti politique (Républicain ou Démocrate).

*Source des données : [www.electoral-vote.com](http://www.electoral-vote.com) ; Sienna Research Institute ; Robin Ryder*

## 2 Comestibilité des champignons

Vous chercherez à expliquer si des espèces de champignons sont comestibles ou vénéneux. Les données du fichier `mushroom.csv` contiennent, pour différents agarics et lépiotes, tout d'abord la classe (e=edible=comestible, p=poisonous=vénéneux), puis les covariables suivantes, décrites en anglais :

- cap-shape : bell=b, conical=c, convex=x, flat=f, knobbed=k, sunken=s
- cap-surface : fibrous=f, grooves=g, scaly=y, smooth=s
- cap-color : brown=n, buff=b, cinnamon=c, gray=g, green=r, pink=p, purple=u, red=e, white=w, yellow=y
- bruises : bruises=t, no=f
- odor : almond=a, anise=l, creosote=c, fishy=y, foul=f, musty=m, none=n, pungent=p, spicy=s
- gill-attachment : attached=a, descending=d, free=f, notched=n
- gill-spacing : close=c, crowded=w, distant=d
- gill-size : broad=b, narrow=n
- gill-color : black=k, brown=n, buff=b, chocolate=h, gray=g, green=r, orange=o, pink=p, purple=u, red=e, white=w, yellow=y
- stalk-shape : enlarging=e, tapering=t
- stalk-root : bulbous=b, club=c, cup=u, equal=e, rhizomorphs=z, rooted=r, missing=?
- stalk-surface-above-ring : fibrous=f, scaly=y, silky=k, smooth=s
- stalk-surface-below-ring : fibrous=f, scaly=y, silky=k, smooth=s
- stalk-color-above-ring : brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y
- stalk-color-below-ring : brown=n, buff=b, cinnamon=c, gray=g, orange=o, pink=p, red=e, white=w, yellow=y
- veil-type : partial=p, universal=u
- veil-color : brown=n, orange=o, white=w, yellow=y
- ring-number : none=n, one=o, two=t
- ring-type : cobwebby=c, evanescent=e, flaring=f, large=l, none=n, pendant=p, sheathing=s, zone=z
- spore-print-color : black=k, brown=n, buff=b, chocolate=h, green=r, orange=o, purple=u, white=w, yellow=y
- population : abundant=a, clustered=c, numerous=n, scattered=s, several=v, solitary=y
- habitat : grasses=g, leaves=l, meadows=m, paths=p, urban=u, waste=w, woods=d

*Source des données : Audubon Society Field Guide to North American Mushrooms (1981); Kaggle.*