ED- Christel Vrain

Alexandre Masson

14 Janvier 2013

Table des matières

1 Motivation **3**

2 Langage R **4**

3 Compléments sur R **6**
3.1 Un peu de terminologie 6

4 Graphique sous R **7**

5 Rappels de probabilité **9**
5.1 Notion de base 9
5.2 20 Février 2013 10
5.3 Variable aléatoire continue 11

6 6 Mars 2013 **11**

7 13 Mars 2013 **12**
7.1 Analyse de régression 12

8 20 Mars 2013 **14**
8.1 Approche Matériel de la régression linéaire 14
8.2 Intervalle de confiance, test d'hypothèse 15
8.2.1 Intervalle de confiance d'une estimation 15

9 4 Avril 2013 **16**
9.1 Intervalle de confiance d'une estimation 16
9.2 Test d'hypothese 17

10 8 avril 2013 **19**
10.1 Feuille d'exercice 19

les TDs sont en anglais.

1 Motivation

Plein de données, expansion de l'univers digital. différentes étapes :

- statistique descriptive
- Régression : analyser la relation d'une var vec d'autres
- Fouille de données : Classification supervisée ou non, prédiction , recherche de motifs fréquents.

univ lyon enseignements R

2 Langage R

Langage et environnement pour le calcul statistique et les graphiques.
il tes libre.

commande de bases **Affectation** -> u <- ou assign("x",valeur).

Des Variables : pas de chiffre ou caractère spécial en premier dans le nom, sensible à la casse, certains noms sont déjà pris.

structure de données : vecteur numérique la structure élémentaire est i, vecteur, un nombre est un vecteur, on affecte a une variable un vecteur, avec l'opérateur c(...) en mettant dans c, le contenu du vecteur. l'opérateur [] donne l'élément du tableau , on peux passer à [] un vecteur de données, il renvoi les valeurs placées à tous les index présents dans le tableau entre [].

Arithmétique vectorielle toutes les opérations telles que sin, et cos, sont appliquées point à point sur tous les variables d'un vecteur. les tableaux sont indexés de 1 à N. la fonction sort(vecteur) est implémentée, ainsi que min et max, et range, length, sum, prod, mean.

NaN signifie not a number.

Vecteur Logique les valeurs sont TRUE,FALSE,NA(not available). Nous avons a disposition des opérateurs logiques tels que <, <=, >, >=, ==, !=. nous avons aussi c1&c2, c1|c2 et !c1. is.na() teste si X est NA ou NAN, is.nan(x) teste uniquement le nan.

Vecteur de caractères les valeurs d'un vecteur doivent etre de meme type : logical , numéric, complex, character (ou raw). la fonction mode(x) renvoie le type des valeurs présentent dans x. on utlise as.character(x) pour transformer les valeurs de x en characters.

Index des vecteurs un vecteur, l'opérateur x :y construit un vecteur remplis avec toutes les valeurs entre x et y. l'opérateur - sur un vecteur renvoie le complémentaire du vecteur. il est possible avec names(vecteur) <- c("chaine", "chaine2",...] pour nommer les index des tableaux.

Séquences fonction seq avec au plus 5 arguments :
- from =
- to =
- by =
- length =
- along = vecteur(seul argument si utilisé, créer une séquence de 1 à length(vecteur).

Factor Utilisé pour étiqueter les données de vecteurs de même longueur., il est ensuite possible d'effectuer des opération en discriminant les valeurs par les étiquettes., tapply permet d'appliquer une fonction (3em param) sur un vecteur(1er param), en utilisant les étiquettes un vecteur d'étiquettes (2nd param).

Fonctions définies de la forme nom <- function(arg1,...,argn). les instructions sont séparées par des ;. On a aussi accès aux conditionnelles, et au boucle, mais il faut éviter les boucles, les boucles for, repeat, while(condition) expr, et le break pour terminer les boucles.

Tableau - Matrices Un vecteur peut être utilisé comme un tableau à plusieurs dimensions si on lui associe un vecteur de dimension. La fonction dim(vect) <- c(x,y) ; vect, transforme le vecteur en tableau de lignes et de y colonnes, il remplie ensuite le tableau avec les données du vecteur, en remplissant la première colonne avec autant de valeurs que de lignes, puis la seconde colonne avec la suite, etc...

Tableaux d'indices il peux être construit par la fonction array(vecteur, dimension). on a aussi des opérations sur les matrices, soient a et b deux matrice et f une fonction n alors outer applique la fonction et renvoi une matrice de taille la concaténation des taille des deux vecteurs.

Listes Collection ordonnée d'objets , appelés composant. Un composant peut être désigne par :

- son numéro : Liste [[x]]
- un nom : List $name
- ATTENTION : List[i :J] retourne une sous liste (avec les noms des composants).

Les listes sont extensibles, il est possible de rajouter des champs en mettant : List$nomChamp<-valeur.

Data frame c'est comme une matrice lais avec des modes et attributs différents, les chaines de caractères sont transformées en facteurs.

Lecture des fichiers première ligne doit avoir une chaine de caractère par attribut du dataframe, tout les reste est ensuite lu comme attribut, mais les chaines sont considérées comme facteur.

Réseau de neurones Introduit dans les 60's. Les observations sont décrites par n variables et une étiquette 1-,1 ou 0,1. Les entrées sont reliées au neurones avec des poids, et l'étiquette c'est le produit scalaire des entrées et des poids, et si elle est plus grande qu'un poids on retourne un sinon zéro. Le neurone sépare donc l'espace en deux demi plan, un ou c'est vrai et un ou c'est faux. on note O pour output. Le perceptron est donc un ensemble de neurones qui sont connectés et où les sorties de neurones sont les entrées d'autres neurones.

Apprendre le perceptron c'est apprendre les poids, on constate que les poids et le seuil ne sont pas du même coté de l'équation, on remplace le b par une entrée x0 toujours égale à 1, et un poids w0 qui devra être égale à b.

apprentissage par correction d'erreurs

```
Entrées : un ensemble d'exemples S de R^{n}*0.1
Sortie : P un perceptron défini par (w0,w1,...,wn)
Initialiser aléatoirement les poids wi
Repeter{
Pour tout exemple (x1,x2,...,xn) dans S{
calculer la sortie o
pour tout i, wi <- wi + (c-o)xi
}
jusqu'a convergence }
```

Première question : converge-t-il ? il a été montré que si les exemples sont linéairement séparable, il existe donc un hyperplan, il va le trouver. Dans la pratique, on va limiter le nombre de répétition des tours de boucle.
2 critiques pour cet algorithme :

- : algorithme ne converge que si les données sont linéairement séparables. Donc il converge s'il doit converger.
- : pas de garantie sur la droite que l'on trouve, pas de distance maximum.
- Cet algorithme s'appelle algorithme par descente de gradient.

3 Compléments sur R

3.1 Un peu de terminologie

- Population : ensemble d'éléments sur lesquels se porte l'analyse
- individu : élément de la population
- variable : sert à décrire la population
- modalité d'une variable : valeur de la variable

plot plot(x,y), ou plot(u) avec liste de vecteur, ou matrice 2D.

variables qualitatives Soucis de commit, penser à regarder le pdf graphique sous R.

4 Graphique sous R

6 Février 2013

Continuons la partie représentation summary nous donne des infos sur une dataframe telles que le min, le 1st.quarter, la median, la mean, le 3rd. Quarter ainsi que le max. Faire attention, car le choix de découpage influe beaucoup sur la perception, et peut donner une fausse interprétation. faire plusieurs diagramme pour voir les données sous différents angles.
Regarder d'un œil critique

- Mesures de tendances centrales ; moyenne, médiane, etc...
-

Peut être un cours de rappel de proba pas des proba pour faire des probas, mais par exemple revoir les lois gaussienne car on utilise des échantillons gaussiens

Propriété de la moyenne Toutes les observations ont le même poids : les valeurs extrêmes influencent autant que les autres.
Aussi la somme des écarts à la moyenne est nulle $\Sigma(x_i - mean(x)) = 0$.
La somme des carrés des distances à la moyenne est plus faible que la somme des carrés des distances à toute autre valeur.
-> la moyenne est la mesure qui minimise la somme des carrés des écarts à elle-même.

autres moyennes

- moyenne géométrique : racine n_{ieme} du produit des n facteurs : x1*x2*...*xn
- moyenne géométrique pondérée racine k_{ieme} des x_n^k
- moyenne harmonique : k éléments, divisé par la somme des $1/x^k$
- moyenne quadratique : Q= $\sqrt{1/k(x_1^2 + x_2^2 + ... + x_n^2)}$

Toutes ces moyennes se généralisent en : le nombre M tel que
$f(M) = 1/n * (n_1.f(x_1 + ... + n_n.f(x_n))$

Médiane point qui partage la distribution d'une série d'observations en deux parts égales -> partage l'histogramme(densité) en deux parties de surface égales.
Ne s'applique que pour des observations pouvant être ordonnées.
Minimise la somme des distances absolues de toutes les observations à ce point.

Estimation Classement des observations par ordre croissant
Calcul : si le nombre d'observations est impair, la médiane est la (k+1)/2 valeur
Si c'est paire, toute observation entre k/2 et(k/2)+1, ou la moyenne des observations k/2 et (k/2)+1.
si les observations sont regroupées par classe, une valeur plus précise sous l'hypothèse d'une répartition uniforme des observations.

Mode Valeur qui possède la fréquence la plus élevée
n'est pas toujours une valeur centrale de la distribution
une distribution peut avoir plusieurs modes, unimodable, bimodale, asymétrique...

Comparaison

- moyenne : tiens compte des valeurs de la distribution
- mode : indique une seule valeur de la distribution
- médiane : indique un rang
- si la distribution est unimodale et symétrie, moyenne , mode, et médiane sont confondus.
- Si la distribution est bimodale et symétrique, moyenne et médiane sont confondues.
- si la distribution est asymétrique, ils peuvent avoir des valeurs différentes.
  - tirée a droit, mode, puis médiane , puis moyenne
  - tirée a gauche : moyenne , médiane, mode

Variance Elle donne la répartition des valeurs autour de la moyenne, si elle est faible, l'ensemble des valeurs est proche de la moyenne, sinon, cela nous indique que la moyenne c'est pas représentative des valeurs des observations.
Propriété
la variance est toujours positive
si elle est nulle , toutes les observations sont identiques,
si on ajoute la même constante a toutes les observations, alors la moyenne change, mais pas la variance.
si on multiplie les observations par une même constante positive ou négative, on modifie la variance à un facteur multiplicatif(le carré de la constante, vu que la variance fait un carré)

Autre façons de calculer la variance variable quantitative avec k observations, x1,...,xn = 1/k*$\Sigma(x_i^2 - mean(x)^2)$ Différence entre var de R ,et la vrai variance, car R se base sur un échantillon pour trouver la moyenne pour effecteur son calcul. On parle d'estimateur sans biais, si quand on change d'échantillon, on retrouve la même valeur qu'avec toue la population, donc la variance ne l'est pas.

Autres mesures écart moyen : prend la valeur absolue plutôt que le carré
écart médian : prend la médiane au lieu de la moyenne pour le calcul de variance.
intervalle interquartile :

- Intervalle comprenant 50 % des observations le plus au ventre de la distribution.
- Se définit à partir des quantiles qui sont des positions particulières (ex :médiane)
- quartile : coupe en quatre, déciles : amis en 10, centiles : idem mais en 100.

– IQR : intervalle inter quantile : il fait la différence entre le 75% et le 25%

Dispersion pour les variables qualitatives

– variables dichotomique : même définitions que pour les variables qualitatives

– variable multicatégorielles : $\sigma^2(X) = p_1 * ... * p_j$ ou p_i est la proportion d'apparition de la classe.

Coefficient de Pearson 13 février 2013

Moment d'ordre r, en langage R $sum[(x - mean(x))^3)/length(x)]^2$

mesures d'aplatissement coefficient de Pearson $\beta_2 = \frac{u^4}{u_2^2}$
où $\mu_r = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^r$ (moment centré d'ordre r)
d'autant plus faible que la courbe est platicurtique.
coefficient de Fisher $\gamma 2 = \beta_2 - 3$

5 Rappels de probabilité

5.1 Notion de base

Expérience aléatoire : on connaît l'ensemble des résultats possible, on ne peut prédire le résultat
Ensemble fondamental Ω de tout les résultats possibles de l'expérience
Événement : résultat , événement vide $= \emptyset$

exemple : lancer de deux dés successifs

– événement vide $\emptyset$:somme des deux lancer $= 0$
– événement certain Ω somme < 13

Opérations

– Négation
– conjonction
– disjonction

Propriétés : Axiome

– $p(\Omega) = 1.$
– $p(\emptyset) = 0$
– $p(!A) = 1 - p(A)$
– $p(A \cup B) = p(A) + p(B)$

Exemple : deux billets pour quatre enfants, 2 filles(b et d), deux garçons (a et c), quelle probabilité de choisir un gars et une fille ?

– événement Ω : on choisit deux personnes a b, a c, a d, b c, b d, c d. Donc chaque couple a une probabilité de $\frac{1}{6}$
– E = a b, a d, b c, b d. donc $\frac{4}{6}$

probabilités conditionnelles probabilité de B sachant A
$P(B|A) = \frac{P(A \cap B)}{P(A)}$

Théorème de Bayes $P(B|A).P(A) = P(A|B).P(B)$

Exemple concret classification supervisée
Utilisation du classifier bayésien.
Il nous dit que les données sont décrites par des attributs $A_1, A_2, ..., A_n$.
Et sont étiquetées par leur classe.
Étant donnée une observation $A_1 = v_1, A_2 = v_2, ..., A_n = v_n$
déterminer la classe.

$$argmax_{C \in Classes} P(C|A_1 = v_1, A_2 = v_2, ..., A_n = v_n) =$$
$$argmax_{C \in Classes} \frac{P(A_1=v_1, A_2=v_2, ..., A_n=v_n|C) * P_r(C)}{P(A_1=v_1, A_2=v_2, ..., A_n)}$$
$$= argmax_{C \in Classes} P(A_1 = v_1, A_2 = v_2, ..., A_n = v_n|C) * P_r(C)$$

2 classes, 2 attributs A_1, A_2 qui ont 2 valeurs
$2 + 4 * 2 = 10$ probabilités à calculer, c'est beaucoup
Hypothèse : attributs indépendants/à la classe
$P(A_1 = v_1, A_2 = v_2, ..., A_n = v_n|C) = P(A_1 = v_1|C) * ... * P(A_n = v_n|C)$

Indépendance $P(A \cap B) = P(A) * P(B)$

Indépendance/à la classe $P(A \cap B|C) = P(A|C) * P(B|C)$

5.2 20 Février 2013

On rappelle que l'ensemble Ω est l'ensemble des événements.

fonction de répartition $F(x) = P(X \leq x)$
- F est croissante
- $\forall x, F(x) \in [0, 1]$

Formule Espérance : $E(X) = \sum x_i p(x_i)$
Variance $\sigma^2 = Var(X) = \sum (x_i - \mu)^2 . p(x_i) = E(X - \mu)^2$

Propriété de l'espérance et de la variance
- E(a.X + B) = a.E(X) +b
- si X et Y sont indépendants : E(X.Y) = E(X).E(Y)
- Var(a.X+b) = a^2.Var(X)

loi conjointe , loi marginale X : nombre de piles, Y : nombre de pile dans les deux premiers essais. On considère donc les probabilités d'avoir les événements X et Y, à savoir quel(s) événement(s) satisfont Y et X. si x et y sont indépendants $p(x, y) = p(x) * p(y)$. la covariance de deux ev indep est 0 , donc on calcule souvent cette donnée sur nos observations pour savoir si des ev sont

indépendants.
E(X) $=0 * \frac{1}{8} + 1 * \frac{3}{8} + 2 * \frac{3}{8} + 3 * \frac{1}{8} = 12/8 = 3/2$

E(Y) $= 0 * \frac{2}{8} + 1 * \frac{4}{8} + 2 * \frac{2}{8} = \frac{8}{8} = 1$

loi binomiale : somme d'une série de Bernoulli (ensemble d'ev suivant la même loi)
Probabilité d'avoir x succès sur n tirage : $P(X = x) = C_n^x * p^x * q^{n-x}$ avec $C_n^x = \frac{n!}{x!(n-x)!}$
$\mu = n * p$ (somme de n variables de Bernoulli indépendantes)
$\sigma^2 = n * p * q$
Exemple avec R
rXXXX : génération de nombres aléatoires suivant la loi XXXX
dXXXX étant donné un ensemble de valeurs, retourne la hauteur de la probabilité de distribution à chaque point.
pXXXX : probabilité P(x≤x) qu'un nombre tiré aléatoirement suivant cette loi soit inférieur à un nombre x donné.
qXXXX : nombre X tel que P(X≤x) soit égale à n , qui est donné.

5.3 Variable aléatoire continue

Machine avec 24 composants
probabilité qu'un composant tombe en panne : 0.2
la machine fonctionne quand au moins deux tiers des composants sont en marche.
Probabilité que la machine fonctionne.
X nombre de composants en marche.
P(X≥16)$=\sum_{i=16}^{24} P(X = i) = \sum_{i=16}^{24} C_{24}^i (0.8)^i (0.2)^{24-i}$

6 6 Mars 2013

fonction de densité densité f de la variable aléatoire X de fonction de répartition F, c'est la dérivée de la fonction de répartition.
Propriété : $P(a \le X \le b) = \int_a^b f(x)dx$. pour Espérance et variance, on utilise al fonction de densité sur un intervalle I

E : $\int_I x.f(x)dx = \int_0^{360} \frac{x}{360} dx$

$\frac{1}{360} \int_0^{360} x dx$

$\frac{1}{360} [\frac{x^2}{2}]_0^{360}$

Var $= \int_0^{360} (x - \frac{1}{180})^2 . \frac{1}{360} dx$

loi uniforme loi uniforme sur intervalle [a,b], densité f(x) = $\frac{1}{b-a}$.
$F(x) = \int_a^x \frac{dx}{b-a} = \frac{x-a}{b-a}$, pour $a \leq x \leq b$

loi normale Soit X une variable aléatoire à valeur réelles. la loi normale est définie par 2 paramètres, l'espérance μ et la variance σ^2, notée $X \sim N(\mu, \sigma^2)$

Loi binomiale loi binomiale de paramètre (probabilité) p de taille (nombre de lancés) n.
$E(X) = n * p$
$Var(X) = n * p * (1 - p)$.
2 expériences
- n fixé, faire varier p
- p fixé, on fait croître n.

On considère $Z = \frac{X-np}{\sqrt{n(1-p)}}$. Lorsque n tend vers l'infini, loi binomiale tend vers $N(0,1)$ (la loi normale centrée).
-> Cas particulier du théorème central limite. Soient $X_1, ..., X_n$ suite de V.A indépendantes de loi de moyenne μ et de variance σ^2.
Soit $\overline{X_n}$ la moyenne arithmétique de $X_1, ..., X_n$
alors $E(\overline{X_n}) = \mu$, $Var(\overline{X_n}) = \frac{\sigma^2}{n}$
$Z_n = \frac{\overline{X_n} - \mu}{\frac{\sigma^2}{\sqrt{n}}}$ tend vers $N(0,1)$ quand n->∞

7 13 Mars 2013

A savoir connaître loi jointe de deux variables, connaître les rappels de statistiques, un peu de dev R, un peu d'analyse de données. Controle lundi après midi

Analyse de régression et corrélation analyse : recherche d'une relation stochastique qui lie 2 ou plusieurs variables. Corrélation : indice mesurant l'intensité de la relation entre 2 variables.

7.1 Analyse de régression

Estimer la valeur d'une des variables à l'aide des valeurs des autres. $Y = f(X)$ avec Y : variable expliquée et X variable explicative.
déterminer le modèle, c'est à dire l'équation de f(x) qui relie X et Y.
estimer le degré de fiabilité de l'estimation.

régression linéaire Entrées : 2 variables X et Y, Sortie : si Y est une fonction de X, notée : $Y = f(X)$.
généralement en premier on regarde le diagramme de dispersion de Y et X, c'est a dire tracer les points d abscisse X et d'ordonnée Y, pour voir si cela suit une fonction linéaire, avec plot(X,Y), si on constate Y=aX+b.
On s'intéresse au cas où l'influence d'une variable sur une autre peut être représentée par une droite de régression.

premier Cas : $Y_i = a + bX_i$. relation déterministe peu probable.
Second cas : $Y_i = a + bX_i + \epsilon_i$ où ϵ_i est un terme d'erreur qui prend en compte les influences aléatoire sur Y.

On cherche un modèle (d'ordre 1) $Y_i = a+bX_i+\epsilon_i$. les inconnues sont a,b,ϵ_i. On cherche à estimer a et b à partir des observations $(X_1Y_1), ..., (X_nY_n)$ de manière à minimiser les erreurs $\epsilon_i, \forall i$.
on pose $D = \sum \epsilon_i^2 = \sum (Y_i - a - bX_i)^2$. on peut choisir de minimiser $\sum |\epsilon_i|$ au lieu du carré ou minimiser le max des epsilon.
On veux trouver â et b̂ , qui sont $argmin_{a,b} D = \sum argmin(Y_i - a - bX_i)^2$.
â , b̂ vérifient $\frac{\delta D}{\delta a} = 0$ et $\frac{\delta D}{\delta b} = 0$.
$\frac{\delta D}{\delta a} = 0 <=> (f(x)^2)' = 2f(x)f'(x) <=> \sum 2(Y_i - \hat{a} - \hat{b}X_i)(-1) = 0 <=> \sum Y_i - n\hat{a} - \hat{b}\sum X_i = 0$ (1)

$\frac{\delta D}{\delta b} = 0 <=> (f(x)^2)' = 2f(x)f'(x) <=> \sum 2(Y_i - \hat{a} - \hat{b}X_i)(-X_i) = 0 <=> \sum Y_iX_i - n\hat{a}\sum X_i - \hat{b}\sum X_i^2 = 0.$ (2)

$\bar{Y} = \frac{\sum Y_i}{n}$, $\bar{X} = \frac{\sum X_i}{n}$.
(1) s'écrit $\bar{Y} - \hat{a} - \hat{b}\bar{X} = 0$.
$\bar{Y} = \hat{a} + \hat{b}\bar{X} => \hat{a} = \bar{Y} - \hat{b}\bar{X}$.
on remplace â par $= \bar{Y} - \hat{b}\bar{X}$ dans 2).

On trouve $\hat{b} = \frac{\sum X_iY_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{Covar(X,Y)}{Var(X)}$.

X=année	Y=inclinaison
1975	624
1976	644
1977	656
1978	667

sous R lm(Y ∼ X). on dois calculer en premier $\bar{X}$ et $\bar{Y}$, ensuite b̂ à partie de la formule $\hat{b} = \frac{\sum X_iY_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{Covar(X,Y)}{Var(X)}$ et ensuite â.

Évaluation de la droite de régression estimée Calcul de l'écart entre la valeur observée et la valeur prédite.
$\sum(Y_i - \bar{Y})^2 = \sum(Y_i - \hat{Y}_i)^2 + \sum(\hat{Y}_i - \bar{Y})^2$. On appelle cela la variation totale : SC_{totale} = Variation inexpliquée (somme des carrés des résidus) + variation expliquée(sommé des carrés de la régression).
Coefficient de détermination $R^2 = \frac{Varitationexplique}{Variationtotale}$

8 20 Mars 2013

8.1 Approche Matériel de la régression linéaire

en générale on ne cherche pas la régression linéaire sur une seule variable, mais plutôt sur plusieurs variables telles que $Y = a_0 + a_1X_1 + ... + a_nX_n$
Observations
y_1 x_1
y_2 x_2
y_3 x_3
recherche d'une relation $Y = \beta_0 + \beta_1 X + \epsilon$, on l'instancie un y et un epsilon par variable, que l'on écrit sous forme de matrice

$$Y = X\beta + \epsilon$$

$$\begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ ... & \\ 1 & x_n \end{pmatrix} \text{ x } \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ ... \\ \epsilon_n \end{pmatrix}$$

Minimiser $\epsilon_1^2 + ... + \epsilon_n^2 =^t \epsilon * \epsilon$ où ${}^t\epsilon = \epsilon_1 ... \epsilon_n$
$=^t (Y - X\beta)(Y - X\beta)$
$= ({}^tY -^t \beta^t X)(Y - X\beta)$
$=^t YY -^t X\beta -^t \beta^t XY +^t \beta^t XX\beta$
$=^t YY - 2\beta^t XY +^t \beta^t XX\beta$

dériver par rapport à β
la solution est la machine β où la dérivée s'annule.
$-2^tXY + 2^tXX\beta = 0$ donc ${}^tXX\beta =^t XY$
$\beta = {}^{-1}({}^tXX)({}^tXY)$

l'interet est que l'on peux étendre ce procédé a la régression multiples, puisque cela change juste notre matrice X et celle des β, de plus la formule de β ne change pas.

8.2 Intervalle de confiance, test d'hypothèse

8.2.1 Intervalle de confiance d'une estimation

Paramètre valeur inconnue de la population à estimer à partir d'un échantillon
Estimation Utilisation des informations sur un échantillon pour déduire des résultats sur toute une population.
le paramètre(caractéristique de la population (exprimé avec les lettres grecques)) est estimé partir d'une statistique(carac de l'échantillon) de l'échantillon.
on cherche à exprimer

$$\begin{pmatrix} & parametre & statistique \\ moyenne & \mu & \bar{x} \\ ecart-type & \sigma & s \\ variance & \sigma^2 & s^2 \\ pourcentage & \pi & p \end{pmatrix}$$

On cherche a connaître la relation entre le paramètre et la statistique
si θ est un paramètre et si $(x_1, ...x_n)$ est un échantillon.
un estimateur de θ est une fonction G$(x_1, ..., x_n)$ utilisé pour trouver une valeur estimée de θ
exemple :
$\bar{x} = \frac{x_1 + ... + x_n}{n}$

$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})$
Il existe donc une valeur de $\bar{x}$ pour échantillon, on a donc $\bar{x}$ qui est une variable aléatoire , dont on peux calculer l'espérance, sur tous les échantillons possibles.
Si on a une population de taille N et si on choisit des échantillons de taille n ,

on a comme échantillons possibles $C_N^n = \begin{pmatrix} N \\ n \end{pmatrix} = \frac{N!}{n!(N-n)!}$

NO parle de distribution d'échantillonnage = distribution de toutes les moyennes obtenues en considérant tous les échantillons de taille n possibles . Il est dit sans biais si l'ensemble des toutes les moyennes des estimateur est en moyenne égale à la valeur souhaités, if(mean(mean(∀ échantillon)) = param recherché.
$\bar{x} = \frac{x_1+...+x_n}{n}$ est un estimateur dans biais de μ
$E(\bar{x}) = \mu$.
En revanche la variance $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})$ n'est pas un estimateur sans biais de la variance (σ^2), car $E(s^2) = \frac{n-1}{n}\sigma^2$
La distribution d'échantillonnage des moyennes , si n est suffisamment grand suit approximativement une loi normale de moyenne μ et de variance $\frac{\sigma^2}{n}\frac{N-n}{N-1}$
Si la population suit une loi normale, la distribution d'échantillonnage aussi, et quelque soit la taille de l'échantillon aussi.
Intervalle de confiance d'un estimation : Soi tθ un paramètre à estimer et T son estimateur à partir d'un échantillon, on détermine un intervalle [T-e,T+e] tq $P(T - e \leq \theta \leq T + e) = 1 - \alpha$ où $1 - \alpha$ désigne le niveau de confiance.
Exemple pour la moyenne :
Hypothèse : la variance est connue, comme la population suit une loi normale, je sais que la distribution suit aussi une loi normale.
donc $\bar{X} \sim N(\mu, \sigma_{\bar{x}})$
Soit $z_{\frac{\alpha}{2}}$ le réel tel que P($Z \leq z_{\frac{\alpha}{2}}$) = $1 - \frac{\alpha}{2}$ (table de Gauss)
P($-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}$) = $\Phi(z_{\frac{\alpha}{2}}) - \Phi(-z_{\frac{\alpha}{2}})$
$= \Phi(z_{\frac{\alpha}{2}}) - (1 - \Phi(z_{\frac{\alpha}{2}}))$
$= 2\Phi(z_{\frac{\alpha}{2}}) - 1 = 1 - \alpha$ car $\Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$

$P(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \leq z_{\frac{\alpha}{2}})$
$P(\bar{X} - z_{\frac{\alpha}{2}}\sigma_{\bar{X}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}}\sigma_{\bar{X}})$

9 4 Avril 2013

9.1 Intervalle de confiance d'une estimation

Quelle est la confiance que l'on peut avoir dans l'estimation d'un paramètre.
deux loi :
population :avec une certaine distribution de probabilité
distribution d'échantillonnage, on calcul la moyenne sur un échantillon, cette moyenne n'est pas la même sur un autre échantillon, ne pas confondre
exemple Estimation de la moyenne μà partir d'un échantillon $x_1, ..., x_n$

$$\bar{x} = \frac{x_1+...+x_n}{n}$$

Si n est suffisamment grand, la distribution d'échantillonnage de la moyenne est approximativement normale de moyenne μ et de variance $\frac{\sigma^2}{n}\frac{N-n}{N-1}$

Trouver e tel que $P(\bar{X} - e \leq \mu \leq \bar{X} + e) = 1 - \alpha$ où α est le paramètre donné (par exemple 5 pourcent)
2 cas :
si σ connu : déjà fait
pour σ inconnu, il faut l'estimer $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$
Trouver e tel que $P(|\bar{X} - \mu| \leq \epsilon) = 1 - \alpha$
$P(\frac{|\bar{X}-\mu|}{S\sqrt{n}} \leq e') = 1 - \alpha$ où e' $= \frac{e}{S\sqrt{n}}$
Ceci est i,e distribution de Student, elle est symétrique , centrée en 0 , dépend de la taille de l'échantillon, elle est plus plate que la normale, se rapproche de la loi normale quand $n_{->\infty}$. On dit qu'elle a n-1 degrés de liberté où n est la taille de l'échantillon.

$\bar{X} - t(\frac{\alpha}{2}, n-1)\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t(\frac{\alpha}{2}, n-1)\frac{S}{\sqrt{n}}$

9.2 Test d'hypothese

Tester des hypothèses sur les caractéristiques de la population
2 hypotheses
$H_0, \mu = 25$ hypothèse nulle
$H_1, \mu > 25$ hypothèse alternative
Si H_0 est vraie , pas de changement par rapport à l'année dernière

deux types d'erreurs

$$\begin{pmatrix} & Hypothese & vraie \\ Decision & H_0 & H_1 \\ accepter H_0 & 1-\alpha & \beta \\ rejetre H_0 & \alpha & 1-\beta \end{pmatrix}$$

α est appelée erreur de première espace, rejeter H_0 alors qu'elle est vraie
β est appelée erreur de seconde espèce, accepter H_0 alors qu'H_1 est vraie.

Exemple :
$H_0, \mu = 45$
$H_1, \mu > 45$
soit $\bar{x}$ la moyenne des heures d'études d'un échantillon
$E(\bar{X}) = \mu_{\bar{X}} = \mu$
$\sigma(\bar{X}) \approx \frac{\sigma}{\sqrt{n}}$ (en réalité, c'est légèrement différent)
hyp : N est assez grand pour corriger le facteur correctif

Hyp : n assez grand (n> 30) car $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ $\sigma' = \frac{\sigma^2}{n}$
$P(\bar{X} - \mu > \epsilon) = 1 - \alpha$
$<=> 1 - P(\bar{X} - \mu \leq \epsilon) = 1 - \alpha$
$<=> P(\bar{X} - \mu \leq \epsilon) = \alpha$

<=> $\Phi(\epsilon) = \alpha$
table de Gauss -> trouver ϵ tel que $\Phi(\epsilon) = \alpha$
si la moyenne de la population est 45 , la probabilité que la moyenne d'échantillonnage soit plus grande que $45+\epsilon\sigma'$ est de 5%.

Application à la régression linéaire $\hat{Y} = \hat{a} + \hat{b}X$
test d'hypothèse sur la pente
$H_0, \hat{b} = 0$
$H_1, \hat{b} > 0$

10 8 avril 2013

Au planning aujourd'hui, correction du CC, ensuite une feuille d'exercice,

Correction du CC Partie Christel

premier exercice On avais pour chaque question le nombre de vote
1) $P(X_1 = oui)$ on regarde le nombre de personne qui on voté oui à la première question, divisé par le nombre totale soit $\frac{10+30}{200} = \frac{1}{5}$
2) $P(X_1 = oui|X_2 = oui) = \frac{P(X_1=oui,X_2=oui)}{P(X_2=oui)} = \frac{\frac{10}{200}}{\frac{10+40}{200}} = \frac{10}{50} = \frac{1}{5}$
3) $P(X_1 = oui|X_2 = oui) = P(X_1 = oui)$
$\frac{P(X_1=oui,X_2=oui)}{P(X_2=oui)} = P(X_1 = oui)$
donc $P(X_1 = oui, X_2 = oui) = P(X_1 = oui) * P(X_2 = oui)$
les événements $X_1 = oui$ et $X_2 = oui$ sont indépendants

Remarque Si on regarde les autres cas, (e.g $P(X_1 = oui|X_2 = non)$)

Second Exercice **1)** $E(A) = E(2X+3Y) = 2E(X)+3E(Y) = 2*1+3*2 = 8$
$E(B) = E(X - 1.5Y) = E(X) - 1.5(Y) = 1 - 1.5 * 2 = -2$
2) $var(X) = E(X^2) - (E(X))^2$
$E(X^2) = Var(X) + (E(X))^2 = 1 + 1 = 2$
3) $E(AB) = E((2X + 3Y)(X - 1.5Y))$
$= E(2X^2 - 3XY + 3XY - 4.5Y^2)$
$E(2X^2 - 4.5Y^2)$
$2 * 2 - 4.5 * 9$
=-36.5

Exercice 3 $X \sim N(\mu, \sigma^2$
Posons $Z = \frac{X-\mu}{\sigma}, Z \sim N(0, 1)$
$\Phi(Z \leq z)$
1) $P(X \leq \mu) = P(\frac{X-\mu}{\sigma} \leq \frac{\mu-\mu}{\sigma})$
$= P(Z \leq 0) = \Phi(0)$

2) $P(X \leq \mu + 2\sigma) = P(\frac{X-\mu}{\sigma} \leq \frac{\mu+2\sigma-\mu}{\sigma})$
$P(Z \leq 2) = \Phi(2)$

3) $P(|X - \mu| > 2\sigma) = 1 - P(|X - \mu| \leq 2\sigma) = 1 - P(\frac{|X-\mu|}{\sigma} \leq \frac{2\sigma}{\sigma}) = 1 - P(-2 \leq Z \leq 2 = 1 - (\Phi(2) - \Phi(-2)) = 1 - (\Phi(2) - (1 - \Phi(2))$(on remplace $\Phi(-2)$ par ça car loi normale est symétrique soit f(-x) = f(x)))

10.1 Feuille d'exercice

Exercice 1 **Question 1** $\bar{X} = \frac{x_1+x_2+...+x_n}{n}$ où X_i est un variable aléatoire, montrons que $E(\bar{X}) = \mu$ et $Var(\bar{X} = \frac{\sigma^2}{n}$

$E(\bar{X}) = \frac{1}{n}(E(X_1) + ... + E(X_n)) = \frac{n*\mu}{n} = \mu$
et pour VAR : $Var(\bar{X} = Var(\frac{1}{n}(X_1 + ... + X_n))$