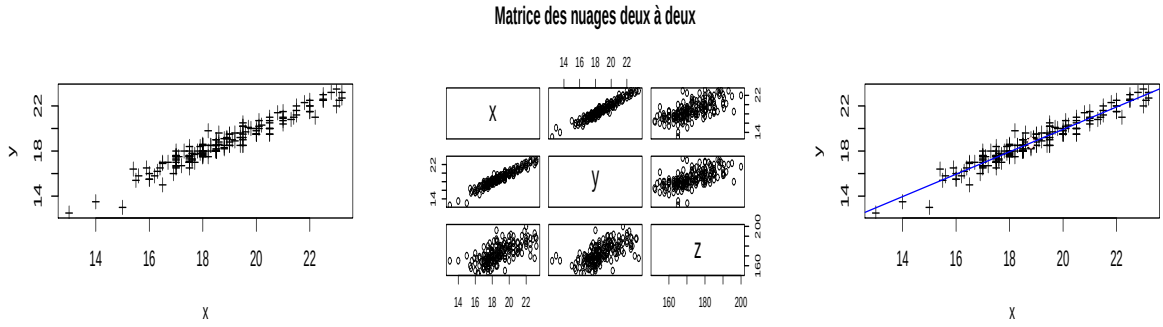
Cours 05
Régression linéaire

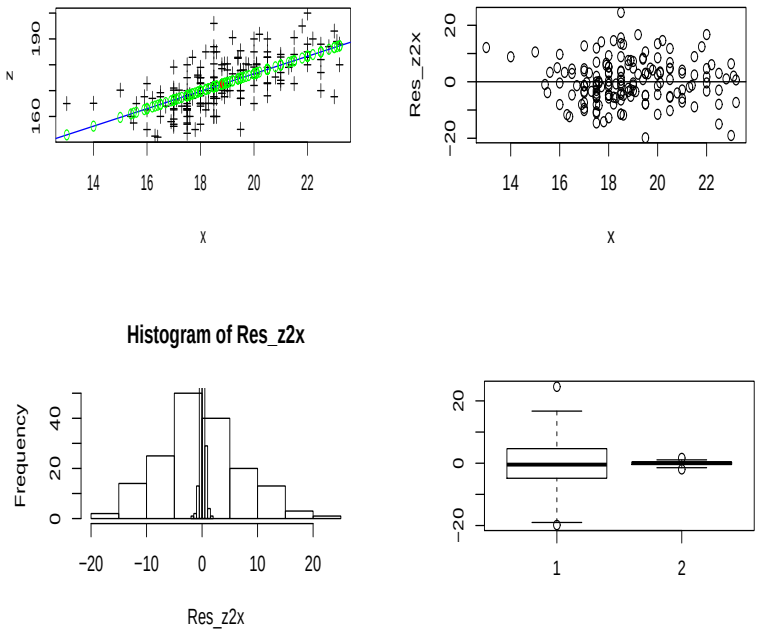
1 Nuage et droite



Le centre de gravité du nuage est la moyenne de ses points. Pour le nuages des `x` et `y` c'est donc le point $(\bar{x}, \bar{y}) =$`(mean(x),mean(y))`.

2 Résidus

On a choisi de représenter les résidus de la régression des `z` sur les `x` car ils sont plus grands que ceux de la régression des `y` sur les `x` : cette différence se voit aussi sur leurs histogramme et leurs boite à moustaches.



Les *valeurs prédites* par le modèle linéaire : `y2x[i]`$= \hat{y}_i = ax_i + b$. Elles sont données par la commande `fitted(`*`DteReg_y_sur_x`*`)` dans le premier exemple du cours.
Les *résidus* : `Res[i]`$= \epsilon_i = y_i - \hat{y}_i$
Les a et b sont choisis de manière à minimiser la somme des carrés des résidus $\sum_{i=1}^{n} \epsilon_i^2$.

3 Résumé

Le code distribué comporte à la fois des commandes `R` déjà connues, des commandes nouvelles, qui elles-même se divisent en commande de type gestion graphique, et des commandes spécifiques à la notion de statistique qui sont l'objet de ce cours : la régression linéaire ; il y a aussi quelques commandes de manipulation de données qui seront approfondies plus tard : nous pensons que vous pouvez commencer à la utiliser en mimant les exemples. Il y a enfin des nom de variables, choisi de façon à être facilement mémorisées, mais dont le choix est totalement arbitraire. C'est un bon exercice de distinguer ces trois derniers types, en soulignant/stabilotant de façon distincte ces trois derniers types (réservez le stabilo à la régression linéaire).

Regression linéaire

`lm` Exemple : `DteReg_y_sur_x=lm(y x,data=MonTableau)`. Effectue la régression desy sur les `x`.
`coef` Exemple : `coef(DteReg_y_sur_x)`. Retournes les coefficients $b =$`Intercept` et a.
`abline` Exemple : `abline(coef(DteReg_y_sur_x),col="blue")`. Ajoute la droite de régression.
`fitted` Exemple : `y2x=fitted(DteReg_y_sur_x)`. Calcule les valeurs du caractère y que prédirait le modèle pour l'échantillon `x` ; les différences avec les valeurs mesurées `y` s'appellent les résidus du modèle.

Commandes graphiques

`plot` Exemple : `plot(x,z,type="p",pch=8)` : dessine le nuage de points.
`points` Exemple : `points(mean(x),mean(y),col="white",pch=19)` : ajoute un point.
`windows` Exemple : `windows()` : ouvre et active une nouvelle fenêtre.
`dev.cur` Exemple : `dev.cur()` : récupère le numéro de la fenêtre active.
`dev.set`

Commandes qui seront revues plus tard

`data.frame` Exemple : *`MonTableau`*`=data.frame(x,y,z)`
`pairs` Exemple : `pairs( x+y+z,data=`*`MonTableau`*`, main="`*`Matrice des nuages deux à deux`*`")`. Donne le scatter-plot des échantillons *x*, *y* et *z*

Les noms de variables utilisés

Ce sont *x*, *y*, *z* qui sont les noms des échantillons utilisés, *MonTableau* qui est le tableau (data.frame) constitué de ces trois échantillons, *DteReg_y_sur_x* qui est le résultat de la régression linéaire des `y` sur les `x`, *DteReg_z_sur_x*, idem pour les `z`, *y2x* et `z2x` sont, pour l'échantillon `x`, les valeurs qui seraient prédites par le modèle linéaire $y = a_1x + b_1$ et $z = a_2x + b_2$. *Res_z2x*, et *Res_y2x* sont les résidus des deux régressions.