Tests d'hypothèse

2 variables explicatives β_0, β_1
Test: $\mathcal{H}_0 = 0$
Si $\mathcal{H}_0$ est vraie, alors l'estimateur est sans biais $\mathbb{E}[\beta_1] = \beta_1$
Un problème commun c'est quand plusieurs variables explicatives sont corrélées (ex: les talents d'un joueur de baseball). Le test de Student n'est pas efficace.
Pour parer à ce problème on fait un test de Fisher.
$SCE = \Sigma_i u_i^2 = u'u$
SCE est la variable d'ajustement.
On impose 2 contraintes: $\beta_1 = 0$, $\beta_2 = 0$ puis on calcule la statistique de Fisher:

$$F = \frac{(SCE_c - SCE_{NC})/2}{SCE_{NC}/(n-3)}$$

Numérateur: diviser par le nombre de contraintes
(Rapport de 2 Chi-2 divisées par leur degrés de liberté respectifs).

Si hypothèse vraie, la contrainte ne change rien => stat de Fisher proche de zéro. Pour savoir si la stat est assez proche de zéro, on fixe un seuil: si on est à droite du quantile à 95%, on rejette l'hypothèse que les 3 paramètres sont égaux à zéro. Par Fisher, on en conclut que **les variables sont donc significatives par Fisher**. On voit qu'on ne peut pas regarder uniquement les variables individuelles pour savoir si on garde les variables.

On remarque que $F \sim T_{n-k}^2$

Cas variables qualitatives ou discrètes
Problème: les variables ne sont pas continues. Ex: si on travaille sur des salaires, le niveau minimum est minoré. Problème de **censure**: le modèle va considérer une gaussienne globale. On aura un biais d'estimation.

En mettant une variable binaire pour le genre, c'est équivalent à faire l'hypothèse que la pente du salaire des hommes et femmes est constant. La différence de salaire se fait uniquement sur l'ordonnée à l'origine.
Lorsque la somme de k variables est égale à 1, on supprime une variable pour éviter la colinéarité.

Autre exemple:
$MBR = \beta_0 + \delta_1 CR_1 + \delta_2 CR_2 + \delta_3 CR_3 + \delta_4 CR_4 + autresfacteurs$
δ_1 est la différence entre avoir un CR_1 et CR_5 (CR_5 omis pour ne pas avoir la colinéarité (voir exemple précédent).

Autre exemple:
$log(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 female * educ + u$
Ici on a rajouté la variable d'intéraction pour ne plus avoir une croissance constante. En revanche le test de significativité devient plus compliqué, il faut désormais faire un test de Fisher.
Le test de Fisher est rejeté, on conclut donc que l'hypothèse d'une variation constante des salaires (homme ou femme) était réaliste.