*b) Test de comparaison de proportions*

On dispose de deux « grands » échantillons, un échantillon $A$ de taille $n_A$ où la proportion est $p_A$ qu'on suppose extrait d'une population $P$ où la proportion est $p$ et un échantillon $B$ de taille $n_B$ où la proportion est $p_B$ qu'on suppose extrait d'une population $P'$ où la proportion est $p'$.

La proportion du caractère dans la population notée $p$ est connue ou inconnue, si elle est inconnue on l'estime par $\hat{p} = \dfrac{n_A p_A + n_B p_B}{n_A + n_B}$.

La variable aléatoire $F_A$ qui à chaque échantillon de taille $n$ associe la proportion du caractère dans cet échantillon suit approximativement la loi normale $\mathcal{N}\left(p\,;\ \dfrac{p(1-p)}{n_A}\right)$.

La variable aléatoire $F_B$ qui à chaque échantillon de taille $n$ associe la proportion du caractère dans cet échantillon suit approximativement la loi normale $\mathcal{N}\left(p'\,;\ \dfrac{p'(1-p')}{n_B}\right)$.

Les variables $F_A$ et $F_B$ étant indépendantes la variable $F_A - F_B$ suit la loi normale :

$$\mathcal{N}\left(p - p'\,;\ \frac{p(1-p)}{n_A} + \frac{p'(1-p')}{n_B}\right).$$

On construit le test.

L'hypothèse nulle est $H_0$ : alors $p$ et $p'$ ne sont pas significativement différentes, l'hypothèse alternative est $H_1$ : alors $p$ et $p'$ sont significativement différentes.

Sous l'hypothèse $H_0$, $F_A - F_B$ suit la loi normale $\mathcal{N}\left(0\,;\ \dfrac{p(1-p)}{n_A} + \dfrac{p(1-p)}{n_B}\right)$.

On utilisera cette loi pour déterminer la région critique.