

Scusate il file Word, ma mi era molto più comodo per fare un riassunto veloce

Piccola chiarezza su come funziona il procedimento fisico dell'esercizio in sé: vengono utilizzate delle taniche con una certa quantità di una determinata sostanza per un tempo stabilito. Alla fine di questo tempo si va a controllare quanti organismi sono sopravvissuti al test.

RIASSUMENDO:

$X_{\text{concentrazioni}} + Y_{\text{tempi}} \rightarrow Z_{\text{organismi_sopravvisuti}}$

Noi vogliamo la concentrazione per la quale, in un fissato tempo, muore esattamente la metà degli organismi iniziali. Possiamo quindi pulire il dataset da tutti gli altri EndPoint.

In realtà potremmo anche vedere altro, in termini di EndPoint, ma quella è in caso facciamo poi.

Possiamo inoltre scartare tutto quello che riguarda gli embrioni, nella parte "life Stage". Per quanto riguarda invece il gruppo di esseri con il quale partire possiamo utilizzare quello riguardante i pesci (attenzione, i pesci appaiono in più di un gruppo eco\_tox).

Una cosa che si può fare, è di utilizzare la nomenclatura per raggruppare il tutto.

Per quanto riguarda invece gli elementi chimici: CAS\_VALUES and CHIMICAL\_NAME. Dal CAS possiamo incrociare i composti chimici ma è molto difficile risalire poi ad un csv, si potrebbe andare sul sito del EPI SUITE per poi fare scraping. La cosa migliore sarebbe quella di avere la formulazione SMILES che contiene sia le informazioni per quanto riguarda le quantità ma anche la struttura del composto chimico.

Altre informazioni utili sulle sostanze in generale:

Idrofobo → Lipidico → si attacca bene → peggiore per noi  
composto lungo non si associa bene  
I gruppi -OH meno tossici perché si scompongono

Per prima cosa urge fare un csv completa da cui partire. Dopo di che abbiamo la necessità di fare delle belle statistiche a riguardo. Procediamo così: Pulizia, Encoding e Imputazioni.

Domande nostre:

Per quanto riguarda la threshold non esiste una universalmente accettata, possiamo utilizzare 1mg/litro per il momento.

EXPOSURE, ci sono 89 unità di misura, tocca vedere bene. La lunghezza standard di un'esperimento è di circa 4 giorni.

Come results possiamo utilizzare sicuramente LC50 ma se vogliamo estendere anche EC50 è ok, giusto per avere qualche dato in più.