**Opisna napovedna spremenljivka** ima **dve vrednosti** - naredimo **eno umetno spremenljivko**

$$w_{1i} = \begin{cases} 0, & x_i = a_1 \qquad \text{referenčna vrednost/skupina} \\ 1, & x_i = a_2 \end{cases}$$

Model

$$y_i = \beta_0 + \beta_1 w_{1i} + \varepsilon_i$$

Pričakovana vrednost $y_i$:

$$\mathbb{E}(y_i) = \begin{cases} \beta_0, & x_i = a_1 \\ \beta_0 + \beta_1, & x_i = a_2 \end{cases}$$

Pomen parametrov:

- $\beta_0$ je povprečje $y$ za referenčno vrednost $x = a_1$, $\mu_{a_1}$
- $\beta_1$ je razlika $\mu_{a_2} - \mu_{a_1}$

Testiramo ničelni domnevi

$$
\begin{aligned}
H_0 :& \beta_0 = \mu_{a_1} = \beta \\
H_0 :& \beta_1 = \mu_{a_2} - \mu_{a_1} = \delta
\end{aligned}
$$

**Primer:** zveza med `FEV` in `Smoke`

***t*-test za primerjavo dveh povprečij** (*t*-test za dva neodvisna vzorca)

Predpostavke:

- imamo dva neodvisna vzorca
- analiziramo slučajno spremenljivko $y$, ki je v prvi populaciji porazdeljena $N(\mu_1, \sigma^2)$, v drugi populaciji pa $N(\mu_2, \sigma^2)$
- varianci obeh normalnih porazdelitev sta enaki

Zanima nas, ali sta povprečni vrednosti spremenljivke $y$ v obeh populacijah enaki.

Testiramo ničelno domnevo:

$$\begin{aligned} H_0 : &\mu_1 = \mu_2 \quad \text{ali} \quad \delta = \mu_1 - \mu_2 = 0 \\ H_1 : &\mu_1 \neq \mu_2 \quad \text{ali} \quad \delta = \mu_1 - \mu_2 \neq 0 \end{aligned}$$

Testna statistika

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sqrt{s_{sk}^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1 + n_2 - 2}$$

**Primer:** vpliv kajenja `Smoke` na povprečno `FEV`

Opisna napovedna spremenljivka $x$ ima $I$ vrednosti $(a_1, a_2, ..., a_I)$.
Taka spremenljivka podatke deli v $I$ skupin.
V model vključimo $I - 1$ regresorjev z vrednostmi 0 in 1.

$$w_1 = \begin{cases} 0, & x_i = a_1 \\ 1, & x_i = a_2 \\ 0, & x_i = a_3 \\ ... & \\ 0, & x_i = a_I \end{cases} \qquad w_{I-1} = \begin{cases} 0, & x_i = a_1 \\ 0, & x_i = a_2 \\ 0, & x_i = a_3 \\ ... & \\ 1, & x_i = a_I \end{cases}$$

Model

$$y_i = \beta_0 + \beta_1 w_{1i} + \beta_2 w_{2i} + ... + \beta_{I-1} w_{Ii} + \varepsilon_i,$$

Pomen parametrov:

- $\beta_0 = \mu_{a_1}$ povprečje odzivne spremenljivke pri referenčni vrednosti opisne napovedne spremenljivke $a_1$
- razlike med povprečji $j$-te skupine in referenčne skupine: $\beta_j = \mu_{a_j} - \mu_{a_1}$, $j = 1, ..., I - 1$

$$\mathbb{E}(y_i) = \begin{cases} \beta_0, & x_i = a_1 \\ \beta_0 + \beta_1, & x_i = a_2 \\ ... & \\ \beta_0 + \beta_{I-1}, & x_i = a_I. \end{cases}$$

**Primer:** `FEV`, spremenljivka `Gender.Smoke`

**Varianta 1: Model brez interakcije opisnih in številske spremenljivke**

Kakšna je zveza med `FEV`, `Ht`, `Gender` in `Smoke`?
Geometrijsko: štiri vzporedne premice (presečišča so različna, nakloni so enaki).

$$y_i = \beta_0 + \beta_1 \texttt{GenderM}_i + \beta_2 \texttt{SmokeDa}_i + \beta_3 \texttt{Ht}_i + \varepsilon_i$$

# Opisne in številske napovedne spremenljivke

## Dve opisni in ena številska napovedna spremenljivka

$$E(y_i) = \begin{cases} \beta_0 + \beta_3 \mathtt{Ht}_i, & \mathtt{Gender} = Z, \quad \mathtt{Smoke} = Ne \\ (\beta_0 + \beta_1) + \beta_3 \mathtt{Ht}_i, & \mathtt{Gender} = M, \quad \mathtt{Smoke} = Ne \\ (\beta_0 + \beta_2) + \beta_3 \mathtt{Ht}_i, & \mathtt{Gender} = Z, \quad \mathtt{Smoke} = Da \\ (\beta_0 + \beta_1 + \beta_2) + \beta_3 \mathtt{Ht}_i, & \mathtt{Gender} = M, \quad \mathtt{Smoke} = Da \end{cases}$$

Pomen parametrov:

- $\beta_0$ predstavlja povprečni FEV žensk nekadilk pri $\mathtt{Ht}=0$
- $\beta_1$ je razlika povprečja FEV za moške in povprečja FEV za ženske nekadilce pri konstantni $\mathtt{Ht}$ na $[\mathtt{Ht}_{min}, \mathtt{Ht}_{max}]$
- $\beta_2$ je razlika povprečja FEV za kadilke/kadilce in povprečja FEV za nekadilke/nekadilce pri konstantni $\mathtt{Ht}$ na $[\mathtt{Ht}_{min}, \mathtt{Ht}_{max}]$
- $\beta_3$ je naklon vzporednih premic

## Varianta 2: Model z interakcijo dveh opisnih in številske spremenljivke

V model vključimo `Ht`, `Smoke` in `Gender` ter **interakcijske člene**:

- **tri dvojne interakcije**: `Gender : Smoke`, `Gender : Ht` in `Smoke : Ht`
- **ena trojna interakcija**: `Gender : Smoke : Ht`

Predpostavimo, da je zveza med `FEV` in `Ht` različna pri kadilcih in nekadilcih, ta razlika je različna pri moških in pri ženskah.

Geometrijsko: **štiri različne premice** (presečišča so različna, nakloni so različni).