

# Onderzoeksvragen

---

## Onderzoeksvraag 1: Fruit Consumptie & Diabetes

### Onderzoeksvraag:

"Heeft het regelmatig eten van fruit ( $\geq 1x$  per dag) invloed op de kans op diabetes?"

### Variabelen

- **Oorzaak (X):** Fruits (0 =  $< 1x$  per dag, 1 =  $\geq 1x$  per dag)
- **Gevolg (Y):** Diabetes\_binary (0 = geen diabetes, 1 = diabetes/prediabetes)

### Hypotheses

- **H0 (Nulhypothese):** Fruit consumptie heeft geen effect op de kans op diabetes. De diabetes prevalentie is gelijk tussen mensen die wel en niet dagelijks fruit eten. ( $p_{\text{fruit}} = p_{\text{geen\_fruit}}$ )
- **H1 (Alternatieve hypothese):** Dagelijkse fruit consumptie verlaagt de kans op diabetes. ( $p_{\text{fruit}} < p_{\text{geen\_fruit}}$ )

## Onderzoeksvraag 2: Lifestyle Patronen & Diabetes (Associatieregels)

### Onderzoeksvraag:

"Welke lifestyle patronen komen vaak samen voor bij mensen met diabetes, en wat is de kans op diabetes bij specifieke combinaties zoals roken zonder fysieke activiteit?"

### Variabelen

#### Oorzaken (X1, X2, X3, X4, X5):

- Smoker (0 = nee, 1 = ja)
- PhysActivity (0 = nee, 1 = ja)
- Fruits (0 =  $< 1x/\text{dag}$ , 1 =  $\geq 1x/\text{dag}$ )
- Veggies (0 =  $< 1x/\text{dag}$ , 1 =  $\geq 1x/\text{dag}$ )
- HvyAlcoholConsump (0 = nee, 1 = ja)

**Gevolg (Y):** Diabetes\_binary (0 = geen diabetes, 1 = diabetes/prediabetes)

### Hypotheses

- **H0 (Nulhypothese):** Er is geen associatie tussen specifieke lifestyle patronen en diabetes. De combinatie {Smoker=1, PhysActivity=0} heeft geen hogere diabetes kans dan verwacht op basis van individuele factoren. (Observed = Expected)
- **H1 (Alternatieve hypothese):** Specifieke ongezonde lifestyle patronen (bijv. {Smoker=1, PhysActivity=0}) komen significant vaker voor bij mensen met diabetes dan verwacht. (Observed > Expected, Lift > 1.5)

## Onderzoeksvraag 3: Socio-economische Status & Diabetes

### Onderzoeksvraag:

"Heeft een lagere socio-economische status (gemeten aan de hand van inkomen en opleidingsniveau) een negatief effect op de kans op diabetes?"

### Variabelen

#### Oorzaken (X1, X2):

- Income (ordinaal, 1-8, waarbij 1 = < \$10,000 en 8 = \$75,000+)
- Education (ordinaal, 1-6, waarbij 1 = geen opleiding en 6 = universitair)

**Gevolg (Y):** Diabetes\_binary (0 = geen diabetes, 1 = diabetes/prediabetes)

### Hypotheses

- **H0 (Nulhypothese):** Socio-economische status heeft geen effect op de kans op diabetes. Inkomen en opleiding zijn niet geassocieerd met diabetes prevalentie. ( $\beta_{\text{Income}} = 0$  EN  $\beta_{\text{Education}} = 0$ )
- **H1 (Alternatieve hypothese):** Een lagere socio-economische status (lager inkomen en/of lager opleidingsniveau) verhoogt de kans op diabetes. ( $\beta_{\text{Income}} < 0$  OF  $\beta_{\text{Education}} < 0$ )

## Onderzoeksvraag 4: Cumulatief Effect Lifestyle Factors

### Onderzoeksvraag:

"Hoe verandert het diabetes risico wanneer meerdere ongezonde lifestyle factors gecombineerd worden, en is dit effect additief of synergistisch?"

### Variabelen

#### Oorzaken (X1, X2, X3, X4):

- Smoker (0 = nee, 1 = ja)
- PhysActivity (0 = nee, 1 = ja) → omgekeerd als risicofactor
- Fruits (0 = < 1x/dag, 1 =  $\geq$  1x/dag) → omgekeerd als risicofactor
- HvyAlcoholConsump (0 = nee, 1 = ja)

#### Afgeleide variabele:

Unhealthy\_Lifestyle\_Score = Smoker + (1 - PhysActivity) + (1 - Fruits) + HvyAlcoholConsump  
Range: 0-4 (0 = geen risicofactoren, 4 = alle risicofactoren)

**Gevolg (Y):** Diabetes\_binary (0 = geen diabetes, 1 = diabetes/prediabetes)

### Hypotheses

- **H0 (Nulhypothese):** Het cumulatieve effect van meerdere ongezonde lifestyle factors is additief. Het diabetes risico stijgt lineair met het aantal risicofactoren. ( $\beta_{\text{linear}}$  significant,  $\beta_{\text{quadratic}} = 0$ )
- **H1 (Alternatieve hypothese):** Het cumulatieve effect is synergistisch/exponentieel. Meerdere risicofactoren samen verhogen het diabetes risico meer dan de som van individuele effecten. ( $\beta_{\text{quadratic}} > 0$  OF exponentieel model past beter)

## Onderzoeksvraag 5: Cholesterol + Bloeddruk & Diabetes

### Onderzoeksvraag:

"Wat is de relatie tussen hoog cholesterol, hoge bloeddruk en diabetes, en verhoogt de combinatie van beide condities het diabetes risico synergistisch?"

### Variabelen

#### Oorzaken (X1, X2):

- HighBP (0 = normale bloeddruk, 1 = hoge bloeddruk)
- HighChol (0 = normaal cholesterol, 1 = hoog cholesterol)
- Interactie: HighBP x HighChol (synergy term)

**Gevolg (Y):** Diabetes\_binary (0 = geen diabetes, 1 = diabetes/prediabetes)

### Hypotheses

- **H0 (Nulhypothese):** Er is geen synergistisch effect tussen hoge bloeddruk en hoog cholesterol op diabetes risico. De interactieterm is niet significant. ( $\beta_{\text{HighBP} \times \text{HighChol}} = 0$ )
- **H1 (Alternatieve hypothese):** De combinatie van hoge bloeddruk en hoog cholesterol heeft een synergistisch effect dat het diabetes risico meer verhoogt dan de som van de individuele effecten. ( $\beta_{\text{HighBP} \times \text{HighChol}} > 0$ )