

# Regression Equations: HINTS 7 Diabetes Privacy Study

## 1. Main Regression Model

### Model Specification:

$$\begin{aligned} Y = & \beta_0 \\ & + \beta_1(\text{diabetic}) \\ & + \beta_2(\text{privacy\_caution\_index}) \\ & + \beta_3(\text{age\_continuous}) \\ & + \beta_4(\text{education\_numeric}) \\ & + \beta_5(\text{region\_numeric}) \\ & + \beta_6(\text{urban}) \\ & + \beta_7(\text{has\_insurance}) \\ & + \beta_8(\text{male}) \\ & + \varepsilon \end{aligned}$$

### Where:

$Y$  = WillingShareData\_HCP2 (Data sharing willingness, binary: 0/1)  
 $\text{diabetic}$  = Diabetes status (0 = No diabetes, 1 = Has diabetes)  
 $\text{privacy\_caution\_index}$  = Privacy caution index (0-1 scale, continuous)  
 $\text{age\_continuous}$  = Age in years (continuous)  
 $\text{education\_numeric}$  = Education level (1-6, ordinal)  
 $\text{region\_numeric}$  = Census region (1-4, categorical)  
 $\text{urban}$  = Urban/rural status (0 = Rural, 1 = Urban)  
 $\text{has\_insurance}$  = Health insurance status (0 = No, 1 = Yes)  
 $\text{male}$  = Gender indicator (0 = Female, 1 = Male)  
 $\beta_0$  = Intercept;  $\beta_1$  to  $\beta_8$  = Regression coefficients;  $\varepsilon$  = Error term

### Key Results:

Sample Size: 2,421 observations  
 $R^2 = 0.1736$

Diabetes Effect ( $\beta_1$ ): 0.0278 ( $p = 0.1608$ , not significant)

Notes: Both models use weighted least squares regression. Sample: HINTS 7 Public Dataset (2022), 2,421 valid observations.

Privacy Effect ( $\beta_2$ ): 2.3159 ( $p < 0.001$ , highly significant)

Age Effect ( $\beta_3$ ): 0.0024 ( $p < 0.001$ , highly significant)

### Model Specification:

$$\begin{aligned} Y = & \beta_0 \\ & + \beta_1(\text{diabetic}) \\ & + \beta_2(\text{privacy\_caution\_index}) \\ & + \beta_3(\text{diabetic} \times \text{privacy\_caution\_index}) \\ & + \beta_4(\text{age\_continuous}) \\ & + \beta_5(\text{education\_numeric}) \\ & + \beta_6(\text{region\_numeric}) \\ & + \beta_7(\text{urban}) \\ & + \beta_8(\text{has\_insurance}) \\ & + \beta_9(\text{male}) \\ & + \varepsilon \end{aligned}$$

### Where:

$Y$  = WillingShareData\_HCP2 (Data sharing willingness, binary: 0/1)  
 $\text{diabetic}$  = Diabetes status (0 = No diabetes, 1 = Has diabetes)  
 $\text{privacy\_caution\_index}$  = Privacy caution index (0-1 scale, continuous)  
 $\text{diabetic} \times \text{privacy\_caution\_index}$  = Interaction term (moderation effect)  
 $\text{age\_continuous}$  = Age in years (continuous)  
 $\text{education\_numeric}$  = Education level (1-6, ordinal)  
 $\text{region\_numeric}$  = Census region (1-4, categorical)  
 $\text{urban}$  = Urban/rural status (0 = Rural, 1 = Urban)  
 $\text{has\_insurance}$  = Health insurance status (0 = No, 1 = Yes)  
 $\text{male}$  = Gender indicator (0 = Female, 1 = Male)  
 $\beta_0$  = Intercept;  $\beta_1$  to  $\beta_9$  = Regression coefficients;  $\varepsilon$  = Error term

### Key Results:

Sample Size: 2,421 observations  
 $R^2 = 0.1753$

Diabetes Effect ( $\beta_1$ ): -0.1712 ( $p = 0.0810$ , marginally significant)

Privacy Effect ( $\beta_2$ ): -2.4409 ( $p < 0.001$ , highly significant)

Interaction Effect ( $\beta_3$ ): 0.4896 ( $p = 0.0383$ , significant)

Age Effect ( $\beta_4$ ): 0.0023 ( $p < 0.001$ , highly significant)

### Model Comparison:

- Interaction model adds:  $\text{diabetic} \times \text{privacy\_caution\_index}$

•  $R^2$  increases from 0.1736 to 0.1753

• Interaction effect is significant ( $p = 0.0383$ )

• Diabetes moderates the privacy-sharing relationship