

# Regression Interpretation Activity

## Purpose

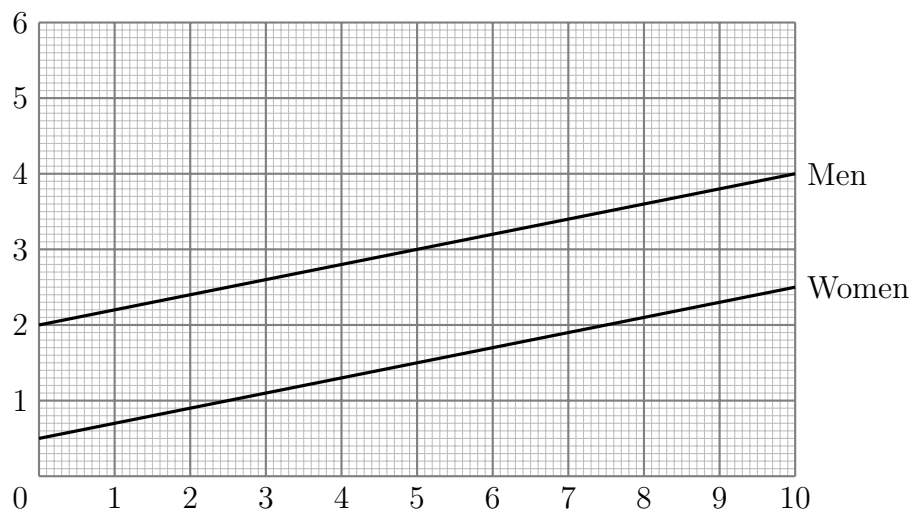
This activity is intended to improve your ability to interpret regression estimates, fitted values, and the marginal effect of variables included in complex regression models. For each of the following regression equations, please draw (by hand) the fitted regression surface according to the equation and instructions provided.

## 1 Ordinary Least Squares

In this equation, we are predicting life satisfaction as a function of years of education and gender. The estimated model is as follows:

$$\hat{y} = 2 + 0.2x + (-1.5z) \quad (1)$$

where  $x$  is years of post-secondary education (from 0 to 8), and  $z$  is an indicator for being female. Please draw this fitted regression surface in the space provided below.



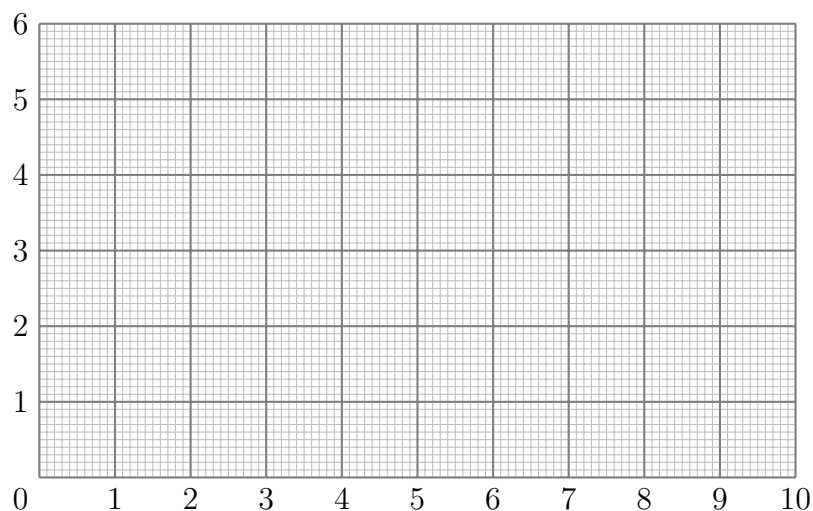
Then, for the values  $x \in \{0, 4, 8\}$ , calculate the **marginal effect** of  $x$  (education) on  $y$  and the **marginal effect** of  $z$  (being female) on  $y$ .

## 2 Ordinary Least Squares (with Interaction)

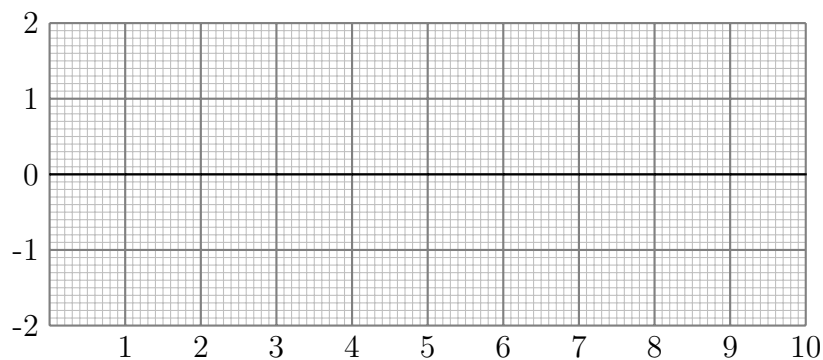
In this equation, we are predicting life satisfaction as a function of years of education, separately for men and women. The estimated model is as follows:

$$\hat{y} = 2 + 0.2x + (-1.5z) + 0.25x * z \quad (2)$$

where  $x$  is years of post-secondary education (from 0 to 8) and  $z$  is an indicator for being female. Please draw this fitted regression surface in the space provided below.



Then, for the values  $x \in \{0, 2, 4, 6, 8\}$  and  $z \in \{0, 1\}$ , calculate the **marginal effects** of both  $x$  and  $z$  on  $y$ . Draw both marginal effect in the space below.

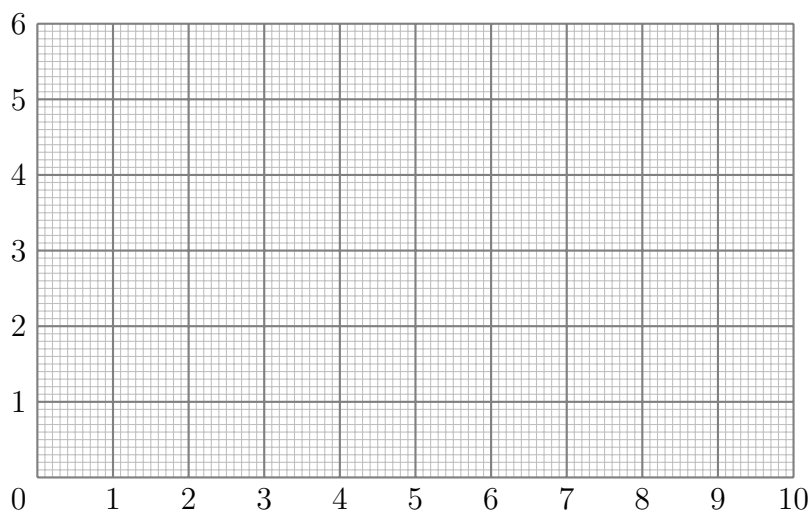


### 3 Ordinary Least Squares (with Non-Linear Term)

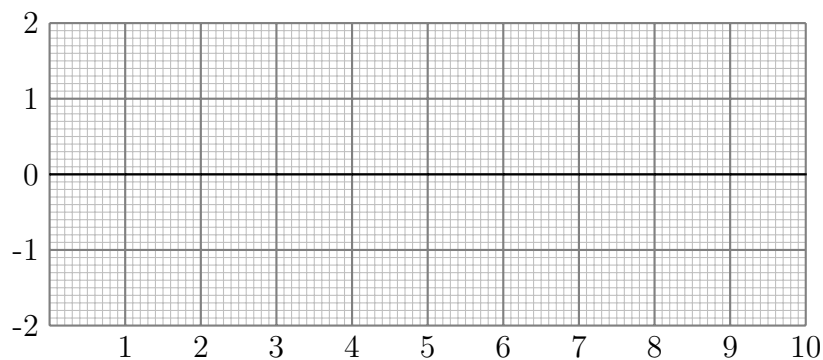
In this equation, we are predicting life satisfaction as a function of years of education. The estimated model is as follows:

$$\hat{y} = 5 - x + 0.1x^2 \quad (3)$$

where  $x$  is years of post-secondary education (from 0 to 8). Please draw this fitted regression surface in the space provided below.



Then, for the values  $x \in \{0, 1, \dots, 10\}$ , calculate the **marginal effect** of  $x$  on  $y$ . Draw this marginal effect in the space below.

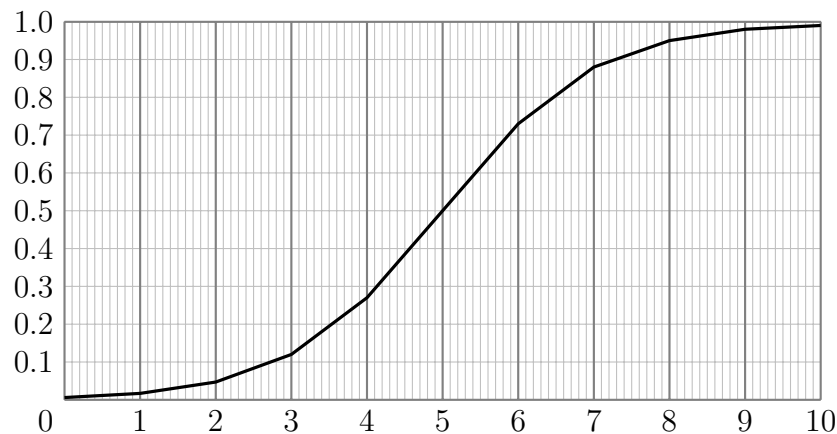


## 4 Logistic Regression with One Variable

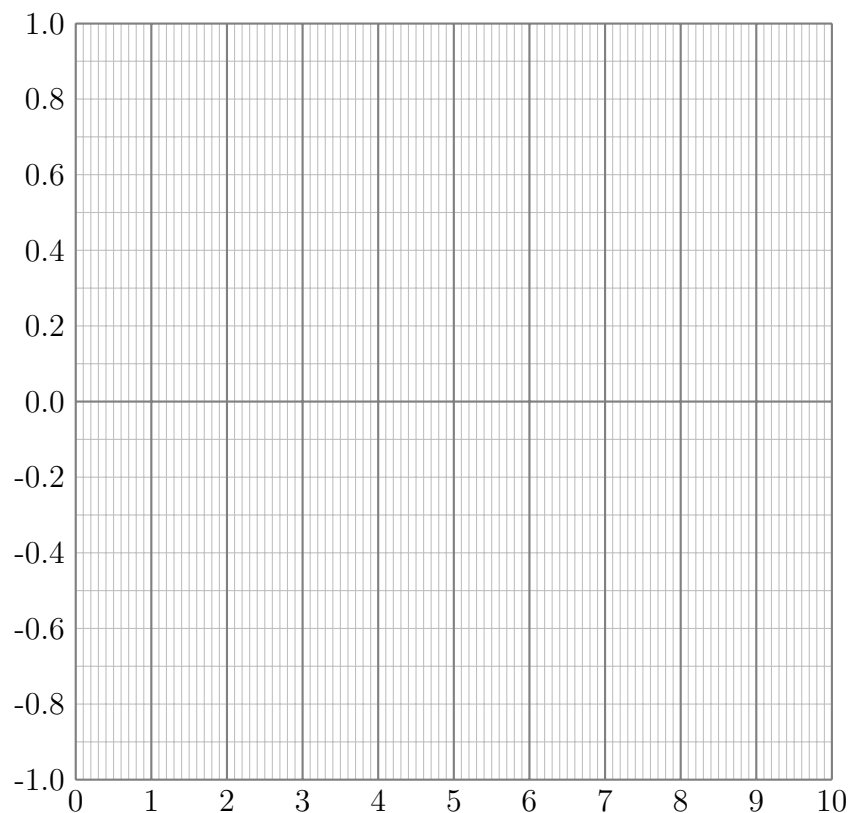
In this equation, we are predicting the probability of expressing satisfaction with life (life satisfaction is 1 rather than 0) as a function of years of education. The estimated model is as follows:

$$Pr(y = 1) = f(\hat{y}^*) = f(-5 + x) \quad (4)$$

where  $x$  is years of post-secondary education (from 0 to 8) and  $f()$  is the logit function. This fitted regression surface is provided in the space below.



For the values  $x \in \{0, 1, \dots, 10\}$ , calculate the **marginal effect** of  $x$  on  $y$ . Draw this marginal effect curve in the space below.

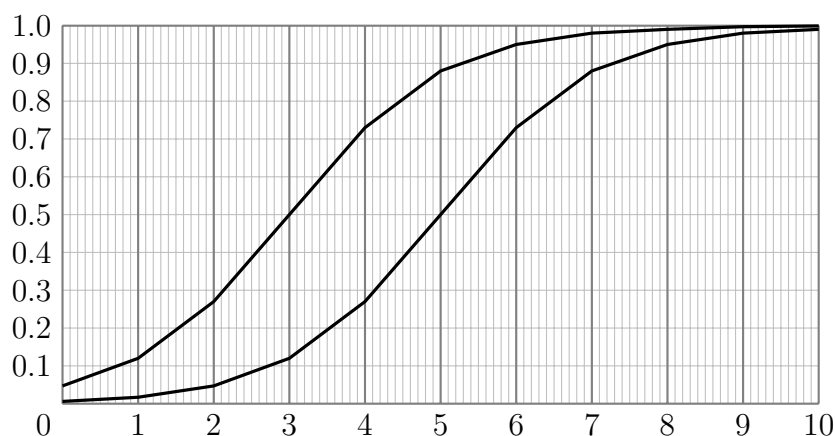


## 5 Logistic Regression with Discrete Effect

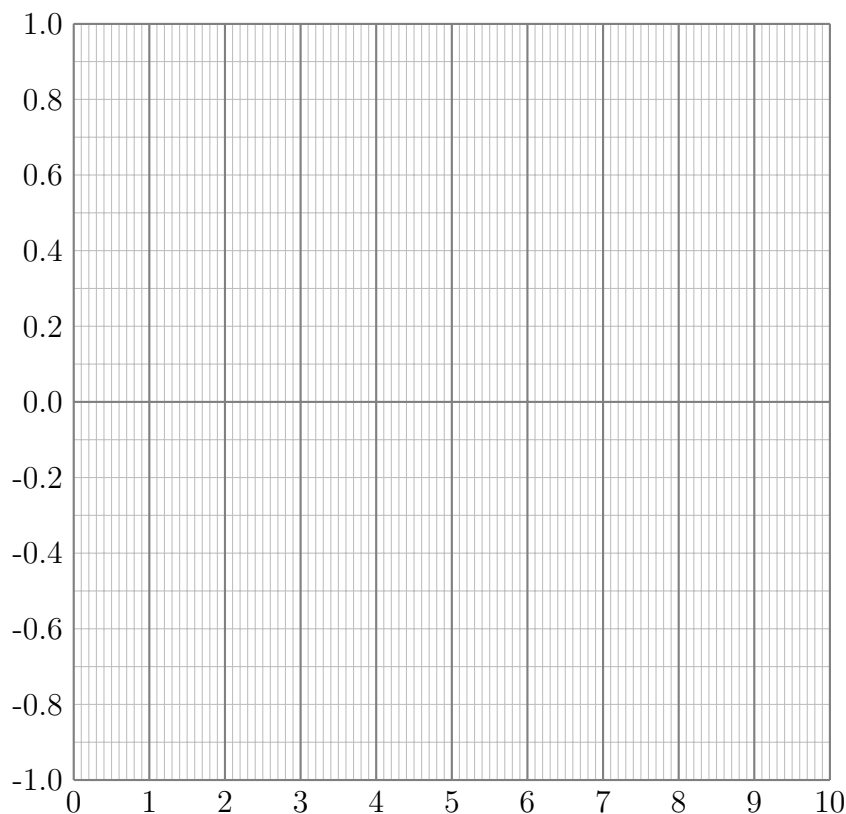
In this equation, we are predicting the probability of expressing satisfaction with life (life satisfaction is 1 rather than 0) as a function of years of education and gender. The estimated model is as follows:

$$Pr(y = 1) = f(\hat{y}) = f(-5 + x + 2z) \quad (5)$$

where  $x$  is years of post-secondary education (from 0 to 8),  $z$  is an indicator for being female, and  $f()$  is the logit function. This fitted regression surface is provided in the space below. Which line represents men and which line represents women?



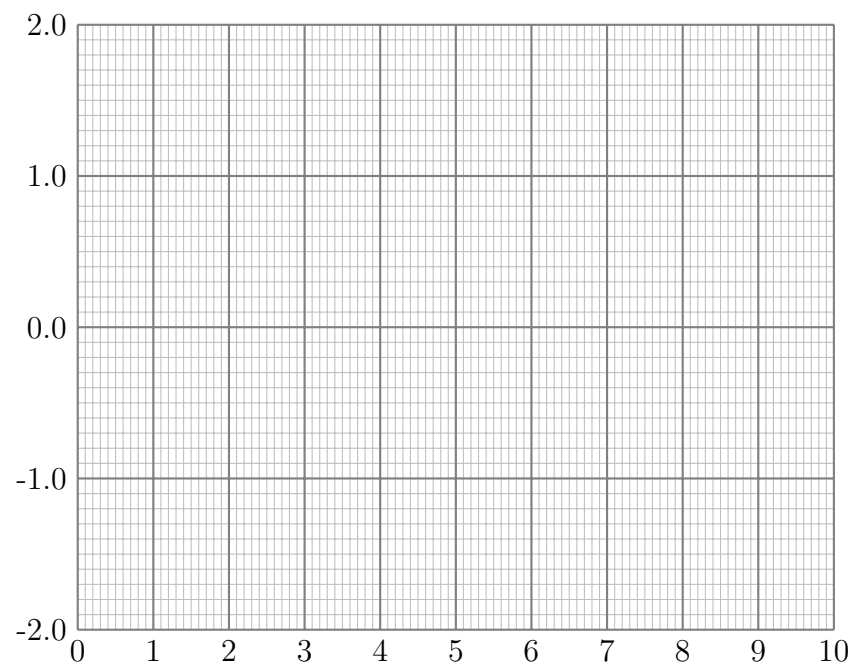
For the values  $x \in \{0, 1, \dots, 10\}$ , calculate the **marginal effect** of  $z$  on  $y$ . Note that for a discrete variable, the we often calculate a *discrete* change between categories rather than the instantaneous marginal effect. Draw this discrete effect curve in the space below.



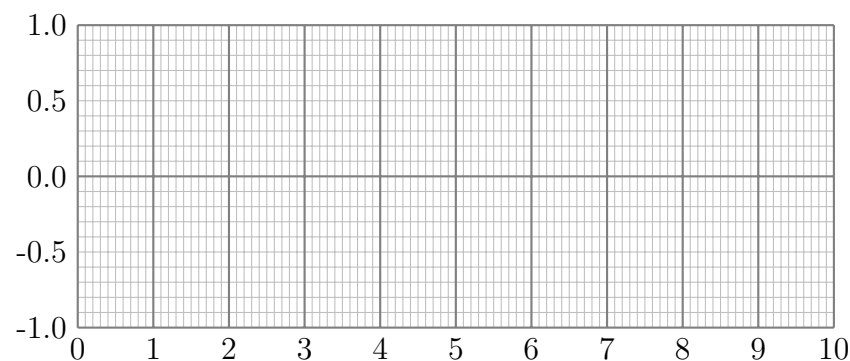
## 6 Logistic Regression with Covariates

An important difference between marginal effects in linear models (OLS) compared to generalized linear models relates to how covariates in the model affect the marginal effect estimate. In OLS, the marginal effect of  $x$  does not depend on the values of  $z$  unless there is an explicit interaction terms. In generalized linear models, this is not the case.

Using the same example regression equation as the previous example ( $f(-5 + x + 2z)$ ), draw the marginal effects of  $x$  when  $z = 0$  and when  $z = 1$ . In the first space below, assume that  $f()$  is the identity function:



Now assume that  $f()$  is the logistic function (i.e., use the predicted probability plots on the previous page to estimate the marginal effects)

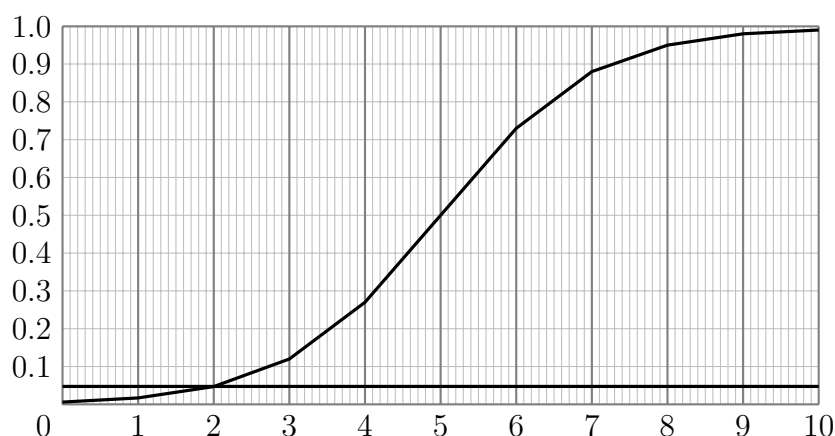


## 7 Logistic Regression with Interaction

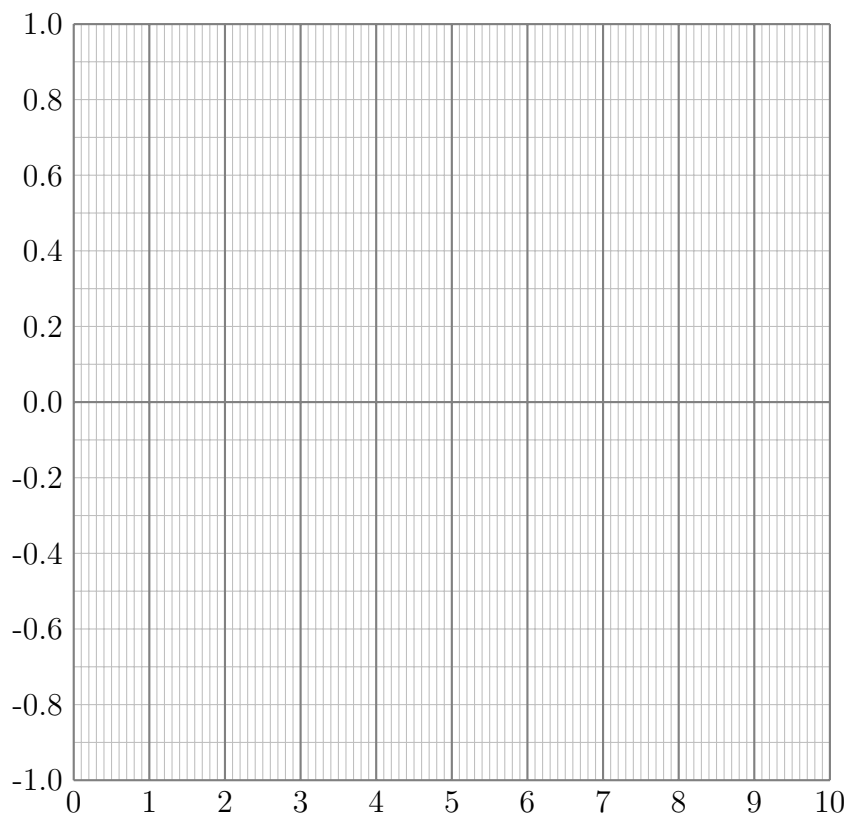
In this equation, we are predicting the probability of expressing satisfaction with life (life satisfaction is 1 rather than 0) as a function of years of education and gender. The estimated model is as follows:

$$Pr(y = 1) = f(\hat{y}^*) = f(-5 + x + 2z - x * z) \quad (6)$$

where  $x$  is years of post-secondary education (from 0 to 8),  $z$  is an indicator for being female, and  $f()$  is the logit function. This fitted regression surface is provided in the space below. Which line represents men and which line represents women?



For the values  $x \in \{0, 1, \dots, 10\}$  and  $z \in \{0, 1\}$ , calculate the **marginal effects** of  $x$  and  $z$  on  $y$ . Draw these curves below

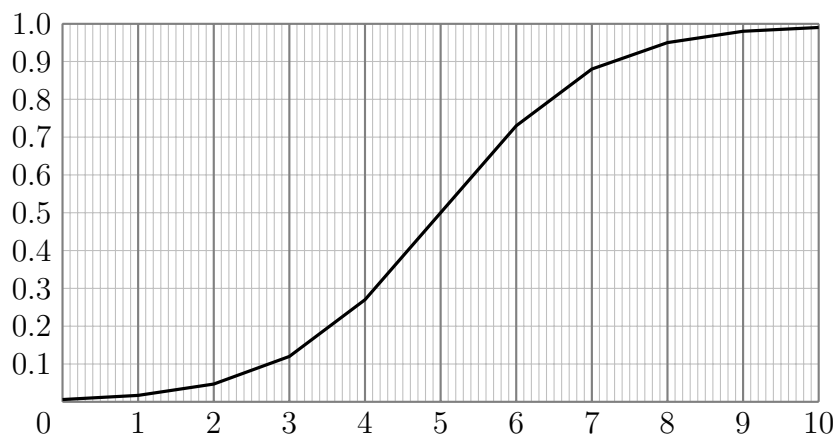


## 8 Logit versus Probit

The numeric values of coefficients in the logit and probit models differ from one another considerably, but tend to produce very similar inferences. Consider, again, the one-variable regression from above:

$$Pr(y = 1) = f(\hat{y}^*) = f(-5 + x) \quad (7)$$

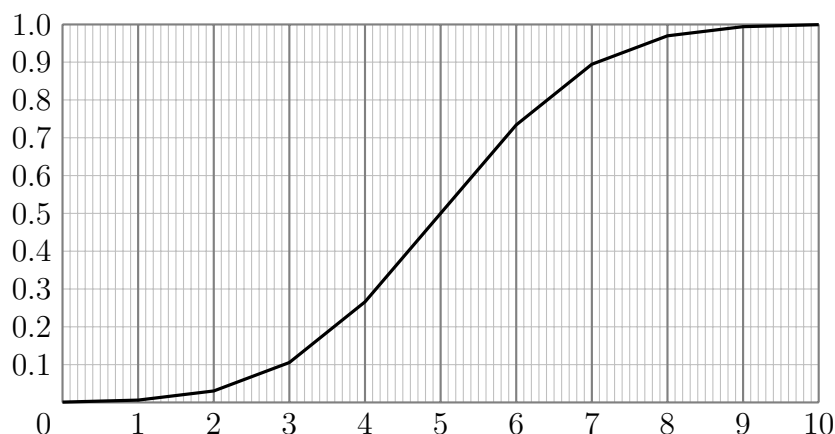
where  $x$  is years of post-secondary education (from 0 to 8) and  $f()$  is the logit function. This fitted regression surface is provided in the space below:



Here is another version of the same model, where  $g()$  is the cumulative normal distribution function:

$$Pr(y = 1) = g(\hat{y}^*) = g(-3.125 + 0.625x) \quad (8)$$

This regression surface is provided below:



How do the marginal effects of  $x$  in the logit and probit specifications compare?