Cannlytics

Cannabis Data Science

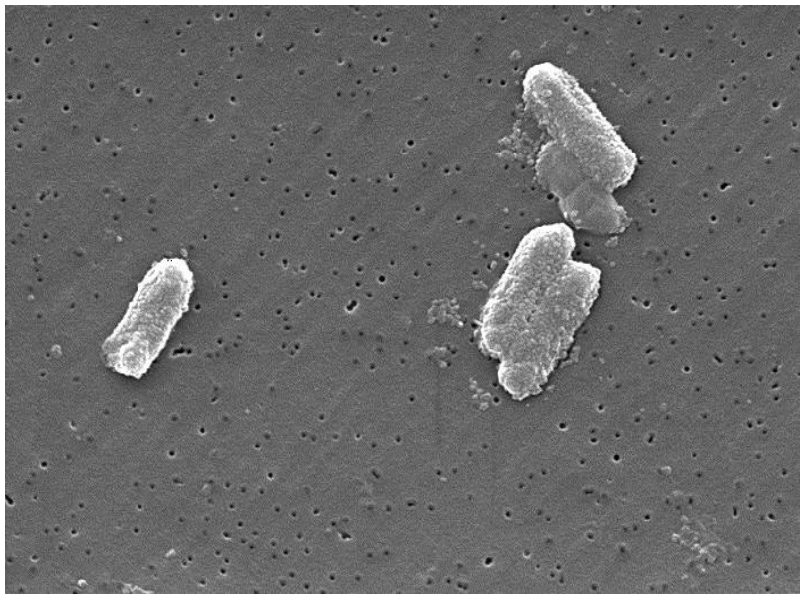
Saturday Morning Statistics #18

April 2nd, 2022

Question and Hypothesis

Question of the day.

- What can producers do to minimize their risk of quality control testing?



Citrobacter freundii, one member of the family **Enterobacteriaceae** – bile tolerant gram negative bacteria (BTGN).

Methodology: Probit Models

Given a latent variable representation of the **probit model**:

$$z_i = x_i\beta + \epsilon_i, \qquad \epsilon_i \overset{iid}{\sim} \mathcal{N}(0,1),$$

$$y_i = \begin{cases} 1 \text{ if } z_i > 0 \\ 0 \text{ if } z_i \leq 0 \end{cases}$$

You can estimate the parameters using the **likelihood function**

$$L(\beta) = \prod_{i=1}^{n} \Phi(x_i\beta)^{y_i}[1 - \Phi(x_i\beta)]^{1-y_i}.$$

Methodology: Tobit Models

Tobit models incorporate the unequal sampling probability for each observation depending on whether the latent dependent variable fell above or below the determined threshold.

- Censored from below at y_L when the latent variable $y_j^* \leq y_L$

$$I(y) = \begin{cases} 0 \text{ if } y \leq y_L, \\ 1 \text{ if } y > y_L \end{cases}$$

- Censored from above at y_U when the latent variable $y_j^* \geq y_U$

$$I(y) = \begin{cases} 0 \text{ if } y \geq y_U, \\ 1 \text{ if } y < y_U \end{cases}$$

Estimating Tobit Models

A **Tobit model** modifies the likelihood function to reflect the unequal sampling probability

$$\mathcal{L}(\beta, \sigma) = \prod_{j=1}^{N} \left(\frac{1}{\sigma} \phi \left(\frac{y_j - X_j \beta}{\sigma} \right) \right)^{I(y_j)} \left(1 - \Phi \left(\frac{X_j \beta - y_L}{\sigma} \right) \right)^{1 - I(y_j)},$$

where ϕ is the **standard normal PDF** and Φ is the **standard normal CDF**.

Interpreting Tobit Models

- Tobit models allow for **consistent** parameter estimation. If y_i is regressed on x_i alone, then the parameter β is **inconsistent** ($\hat{\beta}$ does not converge in probability to β).

- The parameter β is interpreted as

  1. The change in y_i of those above the limit, weighted by the probability of being above the limit;

  2. The change in the probability of being above the limit, weighted by the expected value of y_i if above the limit.

Types of Tobit Models

Type 1 Tobit models censor at a value different from zero

$$y_i = \begin{cases} y_i^* \text{ if } y_L < y_i^* < y_U, \\ y_L \text{ if } y_i^* \leq Y_L, \\ y_U \text{ if } y_i^* \geq y_U. \end{cases}$$

Type 2 Tobit models (**Heckman selection models**), introduce a second latent variable to allow the process of participation (selection) and the outcome of interest to be independent

$$y_{2i} = \begin{cases} y_{2i}^* \text{ if } y_{1i}^* > 0, \\ 0 \text{ if } y_{1i}^* \leq 0. \end{cases}$$

Type 3, **Type 4**, **Type 5** Tobit models introduce additional cases for additional latent variables.

🙏 Thank you for coming.

Insight of the Day

- Asking the right question is of utmost importance.