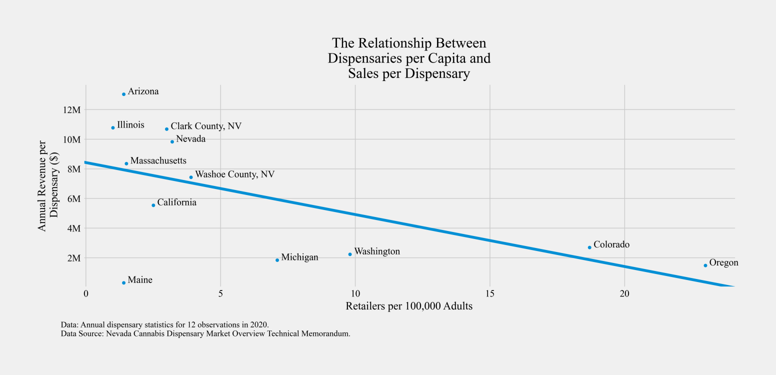
Cannlytics

Cannabis Data Science

Saturday Morning Statistics

November 27th, 2021

What is the relationship between the number of dispensaries and profitability?



Research Question: What is the relationship between dispensaries per capita and revenue per dispensary?

Table 2: Basic Analysis Data & Calculations, by State: 2020

State	Year Adult-Use Legalized	21+ Population	Revenue	Dispensaries (2021)	Dispensaries / 100,000	Revenue / Dispensary
Alaska	2014	542,152		145	26.7	
Arizona	2020	5,486,239	$1,030,000,000	79	1.4	$13,037,975
California	2016	30,131,464	$4,101,540,638	740	2.5	$5,542,622
Colorado	2012	4,347,644	$2,191,091,679	812	18.7	$2,698,389
Illinois	2019	9,570,274	$1,034,790,099	96	1.0	$10,779,064
Maine	2016	1,062,570	$4,706,160	15	1.4	$313,744
Massachusetts	2016	5,426,264	$702,407,378	84	1.5	$8,361,993
Michigan	2018	7,513,559	$984,700,000	535	7.1	$1,840,561
Montana	2020	811,418				
Nevada	2016	2,524,682	$786,479,410	80	3.2	$9,830,993
Oregon	2014	3,249,774	$1,111,027,558	748	23.0	$1,485,331
Vermont	2018	491,627		7	1.4	
Virginia	2021	6,473,738				
Washington	2012	5,783,473	$1,266,224,177	566	9.8	$2,237,145
Nevada Counties						
Clark	2016	1,878,401	$609,167,054	57	3.0	$10,687,141
Washoe	2016	387,180	$111,585,750	15	3.9	$7,439,050

Note: Blanks denote data not available.
Sources: Various

Source: Nevada Cannabis Dispensary Market Overview Technical Memorandum

Panel Data

A panel has the form

$$X_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

where i is the individual dimension and t is the time dimension.

A general panel data regression model is written as

$$y_{it} = \alpha + \beta' X_{it} + u_{it}$$

where

$$u_{it} = \mu_i + v_{it}$$

Estimation with a **fixed effects** or **random effects** model depends on assumptions about μ_i, the individual-specific, time-invariant effects.

Fixed Effects Models

Consider the linear unobserved effects model for N observations and T time periods:

$y_{it} = X_{it}\beta + \alpha_i + u_{it}$ for $t = 1, \ldots, T$ and $i = 1, \ldots, N$

Where:

- y_{it} is the dependent variable observed for individual i at time t.
- X_{it} is the time-variant $1 \times k$ (the number of independent variables) regressor vector.
- β is the $k \times 1$ matrix of parameters.
- α_i is the unobserved time-invariant individual effect. For example, the innate ability for individuals or historical and institutional factors for countries.
- u_{it} is the error term.

Unlike X_{it}, α_i cannot be directly observed.

Unlike the random effects model where the unobserved α_i is independent of X_{it} for all $t = 1, \ldots, T$, the fixed effects (FE) model allows α_i to be correlated with the regressor matrix X_{it}. Strict exogeneity with respect to the idiosyncratic error term u_{it} is still required.

Random Effects Models

Random effect models assist in controlling for unobserved heterogeneity when the heterogeneity is constant over time and not correlated with independent variables. This constant can be removed from longitudinal data through differencing, since taking a first difference will remove any time invariant components of the model.[5]

Two common assumptions can be made about the individual specific effect: the random effects assumption and the fixed effects assumption. The random effects assumption is that the individual unobserved heterogeneity is uncorrelated with the independent variables. The fixed effect assumption is that the individual specific effect is correlated with the independent variables.[5]

If the random effects assumption holds, the random effects estimator is more efficient than the fixed effects model.

Thank you for coming.