Research on Using Synthetic Microdata to Protect Economic Data: Utility and Privacy Protection

Katherine Jenny Thompson

Workshop on Challenges and New Approaches for Protecting Privacy in Federal Statistical Programs

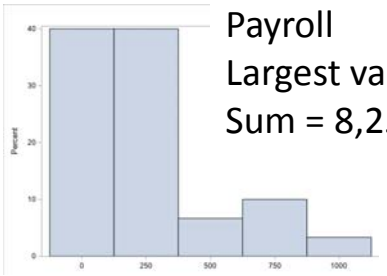
June 7, 2019 (SPEED Session)

United States Census Bureau

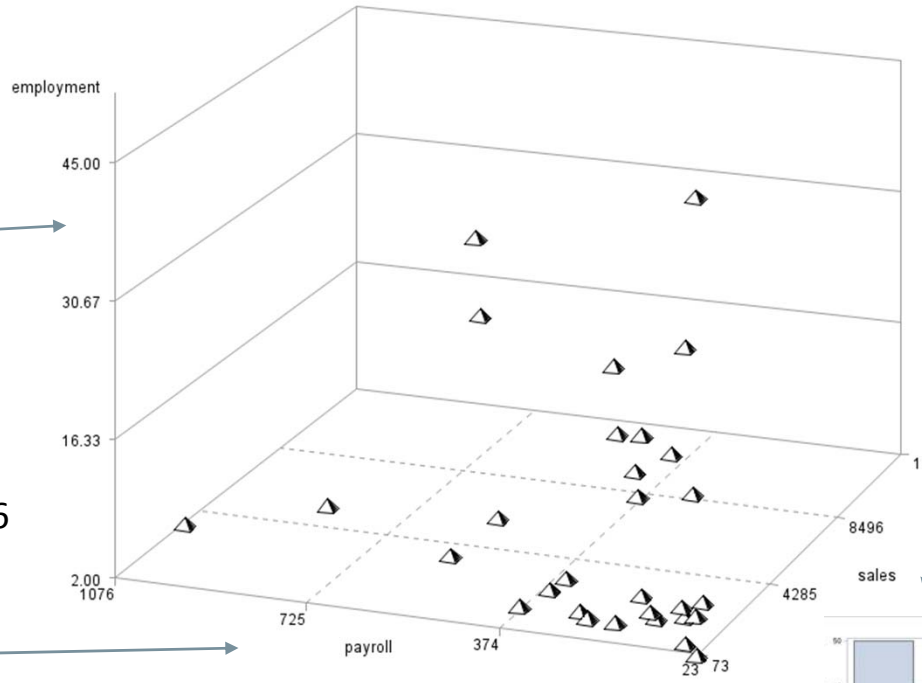
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

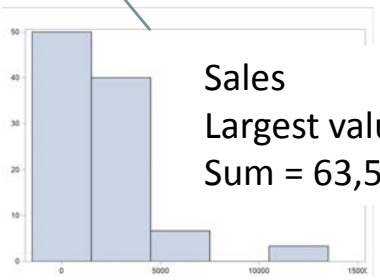
Original Data from One Fictional Industry

Employment
Largest value = 45
Sum = 325



Payroll
Largest value = 1,076
Sum = 8,257



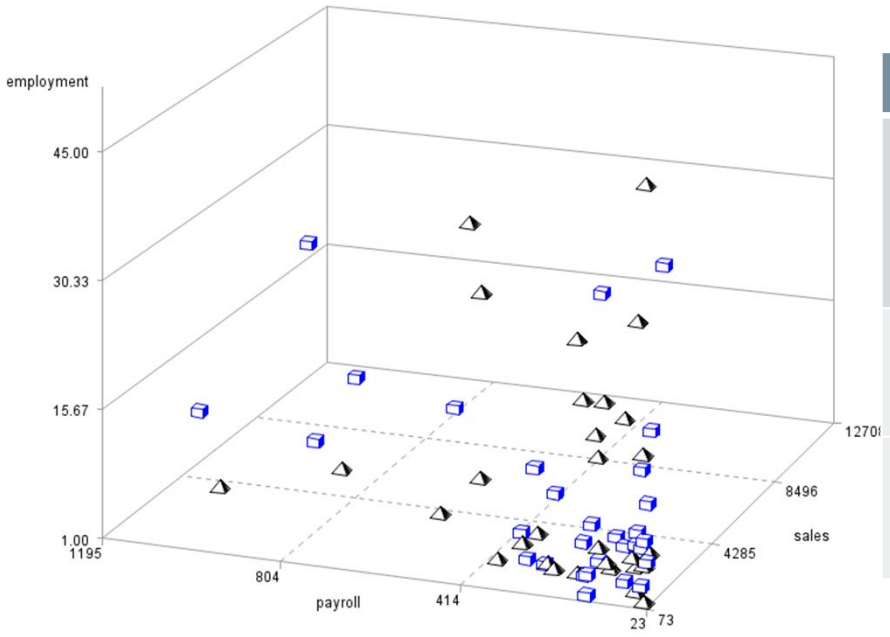


Sales
Largest value = 12,708
Sum = 63,504

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Synthetic Data from One Fictional Industry

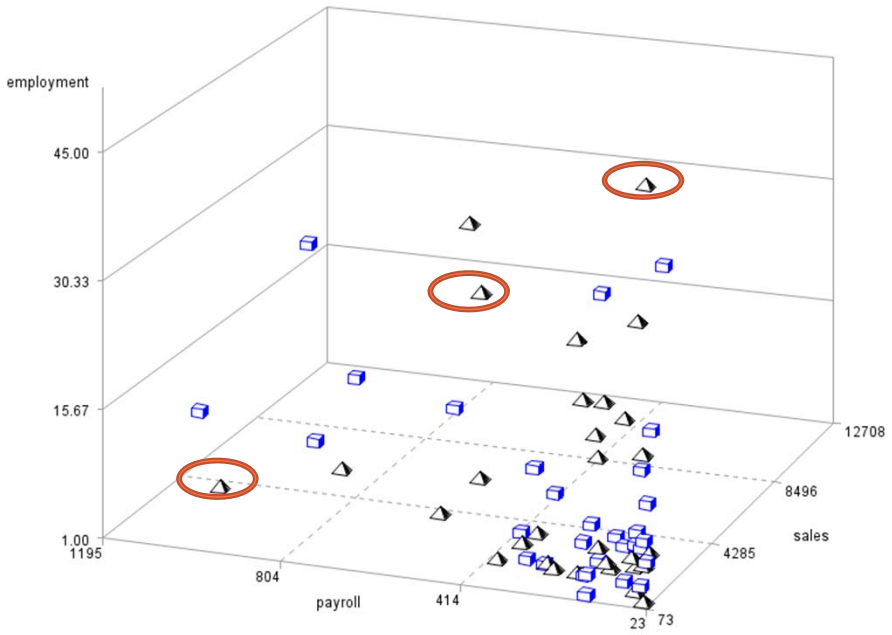


Utility

Metric		Original	Synthetic
Totals	Payroll	8,257	8,278
	Sales	63,504	57,803
	Employment	325	275
Industry Average	Payroll/Employment	25.41	30.10
	Sales/Payroll	7.69	6.98
Multiple Regression	log(Sales) = $\beta_0 + \beta_1$log(Payroll) + $\beta_2$log(Employment)	$\beta_0 = 2.10$ $\beta_1 = 0.78$ $\beta_2 = 0.53$	$\beta_0 = 2.11$ $\beta_1 = 0.79$ $\beta_2 = 0.51$

Note: Illustration! More than one synthetic data set produced via proposed method

Synthetic Data from One Fictional Industry



Disclosure Risk

Intruder Scenario	Maximum Value	Original	Synthetic	
			Metric 1	Metric 2
Intruder only has access to synthetic data *	Payroll	1,076	1,195	n/a
	Sales	12,708	8,903	
	Employment	45	33	
Intruder has access to true value of 2nd largest establishment in industry (by item)**	Payroll	1,076	1,195	97,436
	Sales	12,708	1,195	50,773
	Employment	45	33	245

* Most "revealing" scenario with synthetic data

1. Developed univariate (marginal distribution) risk metrics
2. Multivariate distribution risk metric under development
3. Need to decide on "global" measure

** Metric 1 = maximum(largest synthetic value, true 2nd largest value)
Metric 2 = Synthetic data total – true 2nd largest value (from p-percent rule)

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov