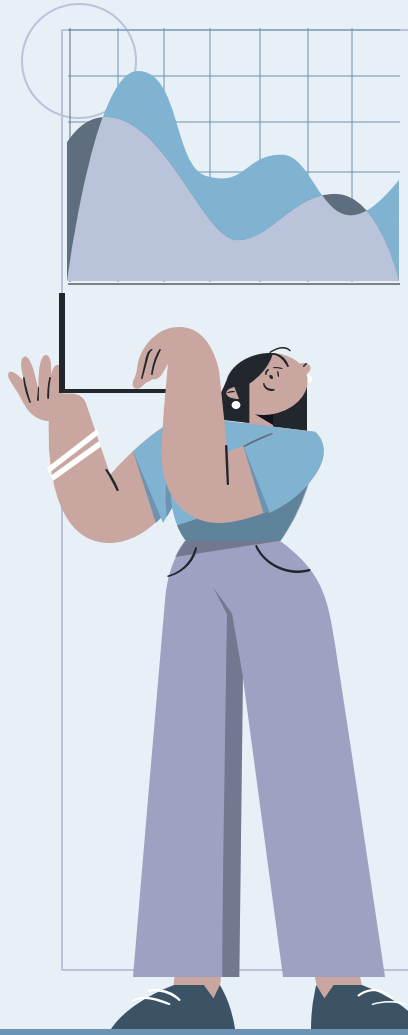


645: Business application with SAS

Project Presentation

Agenda

01 Dataset Introduction

02 Variables

03 Approach

04 Analysis

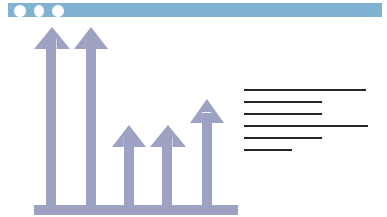
05 Conclusion

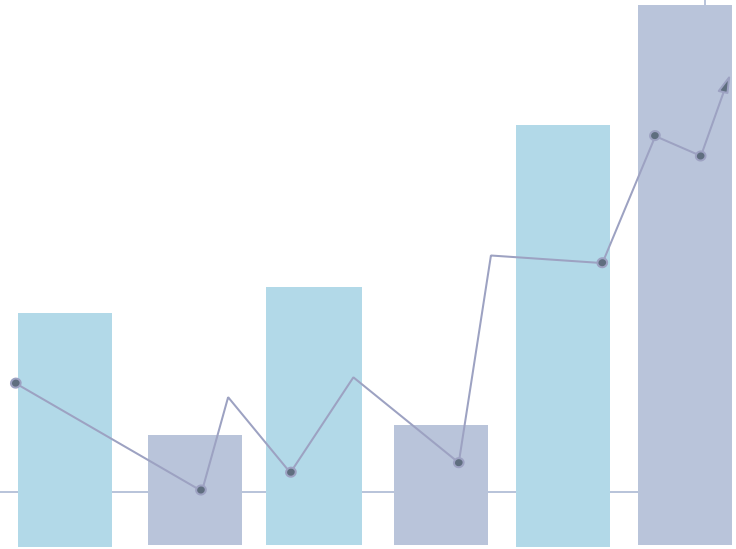
06 Questions?

01

Introduction

United States County Dataset





Introduction to Dataset

- Dataset that unveils insights into our society and its dynamics
- We aim to find stories that shape our communities
- Our Dataset helps us understand how different demographics like economic, educational, health and urbanization interacts with each other.

02

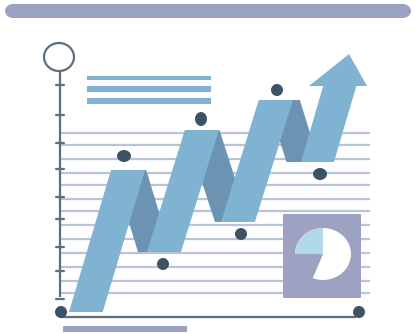
Variables

- Poverty, Unemployment Rate, Home Ownership, Median Household Income and Per-Capita Income
- Multi-Unit apartment and Metropolitan
- Population
- Smoking ban

Name	state	pop2017	poverty	homeownership	multi_unit	unemployment_rate	metro	median_edu	per_capita_income	median_hh_income	smoking_ban
Autauga County	Alabama	55504	13.7	77.5	7.2	3.86	yes	some_college	27841.7	55317	none
Baldwin County	Alabama	212628	11.8	76.7	22.6	3.99	yes	some_college	27779.85	52562	none
Barbour County	Alabama	25270	27.2	68	11.1	5.9	no	hs_diploma	17891.73	33368	partial
Bibb County	Alabama	22668	15.2	82.9	6.6	4.39	yes	hs_diploma	20572.05	43404	none
Blount County	Alabama	58013	15.6	82	3.7	4.02	yes	hs_diploma	21367.39	47412	none
Bullock County	Alabama	10309	28.5	76.9	9.9	4.93	no	hs_diploma	15444.16	29655	none
Butler County	Alabama	19825	24.4	69	13.7	5.49	no	hs_diploma	17014.95	36326	none
Calhoun County	Alabama	114728	18.6	70.7	14.3	4.93	yes	some_college	23609.64	43686	none
Chambers County	Alabama	33713	18.8	71.4	8.7	4.08	no	hs_diploma	21079.51	37342	none
Cherokee County	Alabama	25857	16.1	77.5	4.3	4.05	no	hs_diploma	23067.93	40041	none
Chilton County	Alabama	44067	19.4	75.1	4.4	4.05	yes	hs_diploma	22793.82	43501	none
Choctaw County	Alabama	12945	22.3	85.6	3.9	6.39	no	hs_diploma	20363.87	32122	none
Clarke County	Alabama	24083	25.3	80	6.3	8.48	no	hs_diploma	20099.21	33827	none
Clay County	Alabama	13367	19.1	72.8	11.2	4.37	no	hs_diploma	20879.67	37287	none
Cleburne County	Alabama	14900	19.1	74.9	5.3	4.46	no	hs_diploma	20158.98	37396	none
Coffee County	Alabama	51874	16.1	69.7	13.6	4.39	no	some_college	25627.63	49821	none
Colbert County	Alabama	54500	16.8	73.5	12.3	5.21	yes	some_college	22915.45	45477	none
Conecuh County	Alabama	12469	26.4	81.6	6	6.14	no	hs_diploma	14814.23	30434	none
Coosa County	Alabama	10754	14.4	83.7	1.9	4.62	no	hs_diploma	19147.01	34792	none

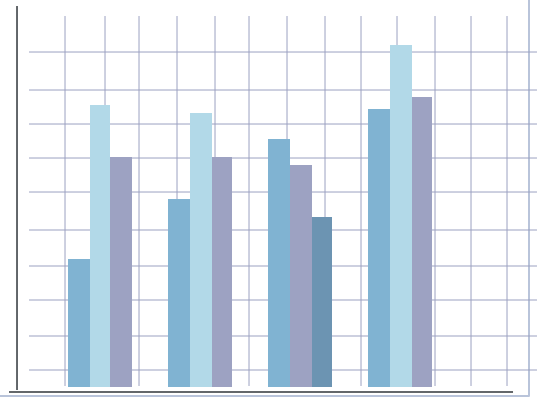
What are we trying to Analyze?

- How do the variables correlate with per_capta_income?
- How do the variables correlate with median_hh_income?
- How does smoking_ban moderate the relationship between per_capta_income and median_hh_income?
- what is the relationship between unemployement_rate and median_hh_income?



03 & 04

Approach & Analysis



Anova

Dependent Variable: per_capita_income per_capita_income

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	52985692956	17661897652	792.84	<.0001
Error	3131	69748813853	22276848.883		
Corrected Total	3134	122734506809			

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

median_edu	per_capita_income LSMEAN	LSMEAN Number
bachelors	46465.1804	1
below_hs	11217.0500	2
hs_diploma	22158.1708	3
some_college	28783.2522	4

Least Squares Means for effect median_edu
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: per_capita_income

i/j	1	2	3	4
1		<.0001	<.0001	<.0001
2	<.0001		0.0059	<.0001
3	<.0001	0.0059		<.0001
4	<.0001	<.0001	<.0001	

- As the p-value is less than alpha, we reject the null and conclude that not all means are equal.
- From the Anova table, we can say that group mean of per capita income of counties having bachelor degree, college degree, high school diploma and below high school are different from each other.

Correlation Analysis

1 With Variables:	per_capita_income
4 Variables:	poverty homeownership unemployment_rate median_hh_income

Pearson Correlation Coefficients, N = 3135 Prob > |r| under H0: Rho=0				
	poverty	**homeownership**	**unemployment_rate**	**median_hh_income**
per_capita_income	-0.72183	0.03108	-0.44823	0.86237
per_capita_income	<.0001	0.0819	<.0001	<.0001

- Except home ownership all other variables correlate with per-capita income
- Poverty and unemployment has negative relationship with per-capita income while the rest have positive relationship with per-capita income.

Correlation Analysis

1 With Variables:	median_hh_income
5 Variables:	poverty homeownership multi_unit unemployment_rate per_capita_income

Pearson Correlation Coefficients, N = 3135
Prob > |r| under H0: Rho=0

	poverty	homeownership	multi_unit	unemployment_rate	per_capita_income
median_hh_income	-0.75009	0.07527	0.32927	-0.40865	0.86237
median_hh_income	<.0001	<.0001	<.0001	<.0001	<.0001

- All variables correlate with median_hh_income
- Poverty and unemployment has negative relationship with median_hh_income while the rest have positive relationship.

Linear Regression

Dependent Variable: unemployment_rate unemployment_rate

Number of Observations Read	3135
Number of Observations Used	3135

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1405.59049	1405.59049	628.07	<.0001
Error	3133	7011.50754	2.23795		
Corrected Total	3134	8417.09803			

Root MSE	1.49598	R-Square	0.1670
Dependent Mean	4.60663	Adj R-Sq	0.1667
Coeff Var	32.47447		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	7.14219	0.10464	68.25	<.0001
median_hh_income	median_hh_income	1	-0.00005098	0.00000203	-25.06	<.0001

- Independent and dependent variables are numerical.
- We have a significant model.
- R-square value is 0.167.
- Per unit increase in median household income , unemployment rate of a county is decreasing by 0.00005 units.

Multiple Linear Regression

Least Squares Model (No Selection)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	51386	12847	183.49	<.0001
Error	3130	219144	70.01399		
Corrected Total	3134	270530			

Root MSE	8.36744
Dependent Mean	12.31764
R-Square	0.1899
Adj R-Sq	0.1889
AIC	16462
AICC	16462
SBC	13355

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	14.219850	0.450071	31.59	<.0001
unemployment_rate	1	0.173043	0.096904	1.79	0.0742
median_edu bachelors	1	18.100033	1.266541	14.29	<.0001
median_edu below_hs	1	-3.410754	5.928890	-0.58	0.5651
median_edu hs_diploma	1	-6.635768	0.319606	-20.76	<.0001
median_edu some_college	0	0	.	.	.

- We have a significant model.
- Adj R-square value is 18.89.
- Unemployment rate and below high school degrees are not significant.
- When compared to counties with some college degree, counties with bachelor's degree has positive relation of 18.1 with multi unit. While counties with high school diploma as median education has negative relation of –6.63.

Moderation Analysis using Linear Regression

Least Squares Model (No Selection)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4.039599E11	1.346533E11	3079.16	<.0001
Error	3131	1.369205E11	43730582		
Corrected Total	3134	5.408803E11			

Root MSE	6612.91024
Dependent Mean	49739
R-Square	0.7469
Adj R-Sq	0.7466
AIC	58297
AICC	58297
SBC	55184

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	7619.172764	1122.692914	6.79	<.0001
per_capita_income	1	1.602147	0.040606	39.46	<.0001
smoking_ban none	1	-6534.047320	1258.465414	-5.19	<.0001
smoking_ban partial	0	0	.	.	.
per_capit*smoking_ba none	1	0.269358	0.045889	5.87	<.0001
per_capit*smoking_ba partial	0	0	.	.	.

- Model is good fit.
- Adj R-square value is 0.74
- Smoking ban moderates the relationship between per-capita income and median household income.
- When compared to partial smoking ban, non-smoking_ban counties have more household income of 26%

Moderation analysis using logistic regression

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	4137.941	3985.793
SC	4143.991	4009.994
-2 Log L	4135.941	3977.793

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	158.1480	3	<.0001
Score	136.5493	3	<.0001
Wald	122.8778	3	<.0001

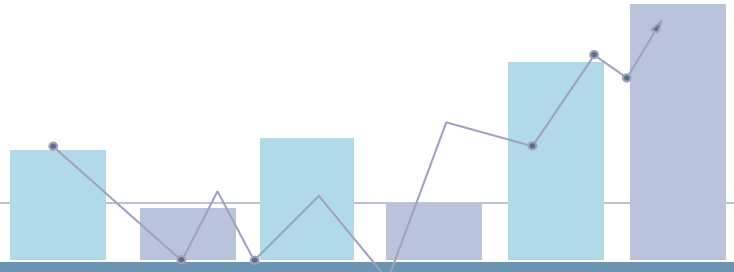
Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.5770	0.3272	3.1093	0.0778
poverty	1	-0.00789	0.0208	0.1442	0.7042
unemployment_rate	1	0.2848	0.0749	14.4741	0.0001
poverty*unemployment	1	-0.0150	0.00419	12.7721	0.0004

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	62.6	Somers' D	0.252
Percent Discordant	37.4	Gamma	0.252
Percent Tied	0.0	Tau-a	0.118
Pairs	2294244	c	0.626

- This model is significant as -2 Log L of "intercept only" is greater than "intercept and covariances".
- unemployement does moderate the relationship between Poverty and metro politan city
- Per unit increase in poverty*unemployement the chances of existence of metropolitan city decreases by 1.5%
- Model accuracy: 62.6 %

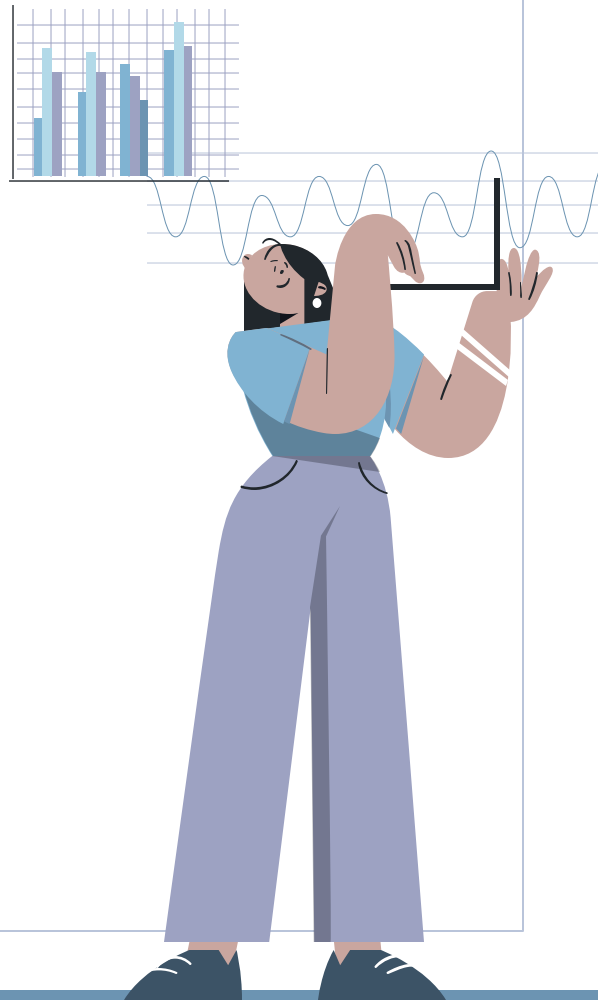
05

Conclusion



06

Questions?



Whoa!