REDELVE INTERNATIONAL PUBLICATIONS

Journal of Business Management and Administration
Volume 2019, Issue 01
Research Article: RD-BMA-10004

Social Factors Associated with Incidence of Diabetes By County In The United States East Coast

[1]Kelly Xie, [1]Jinting Zhang, *[1]Paul D. Berger

Bentley University Waltham, Ma 02452, USA

[*]**Corresponding author:** *Paul D. Berge, Bentley University Waltham, Ma 02452, USA. Email: pdberger1@verizon.net*

Citation: *Kelly Xie, 1Jinting Zhang, Paul D. Berge (2019) Social Factors Associated with Incidence of Diabetes by County in the United States East Coast, Journal of Business Management and Administration. RD-BMA-10004*

Received Date: *08 March 2019;* **Accepted Date:** *18 March 2019;* **Published Date:** *20 March 2019*

Abstract

In this paper we consider several factors and examine the relationship of these factors to the incidence of diabetes. We use data from different counties in the Eastern United States. We study five core factors, education level, earnings, age, obesity rate, and unemployment rate, find that four of the five core factors (all except earnings) we study are related to incidence of diabetes, and that there are several significant two-factor interaction effects.

Keywords: Adult Obesity; Diabetes; Factorial Experimental Design; U.S. East Coast County Data.

Introduction

According to the National Diabetes Statistics Report released by the Center for Disease Control and Prevention [CDC] in 2017, more than 100 million U.S. adults are now living with diabetes or prediabetes. As diabetes was the seventh leading cause of death in the U.S. in 2015, it is useful to consider a different statistical analysis for this common chronic disease [1].

The report also showed that the incidence of diabetes is different across different areas of the U.S. Beside geographical differences, there may be differences in *social factors* such as education level, income and age that may have a significant impact on incidence of diabetes that can be seen at the county level. In this paper, we design an experiment to test our hypothesis about county differences. More specifically, we focus only on counties in states that touch the East Coast of the United States (Maine, New Hampshire, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Delaware, Maryland, Virginia, North Carolina, South Carolina, Georgia, and Florida).

Literature Review

Many scholars agree that incidence of diabetes is likely to increase with the aging of the population, and scientific experiments have already shown that type 2 diabetes is usually diagnosed in adults over the age of 45 [2].

Based on past studies, the effect of education and income on incidence of diabetes is not consistent. Wangdi and Tshering [3] conducted an analysis with data from Bhutan, Himalayas and found that participants with education levels of high school and diploma holders were at a higher risk of diabetes as compared to those without education as reported in other studies. However, Murad, Samia, Rahila and Bayan [2] analyzed a study from diabetes care centers in Jeddah and concluded that diabetic patients were more likely to be less educated; they were also more likely to have lower annual incomes. While West and other scholars [4] believed that low education and a higher annual income, while negatively correlated, were both associated with higher incidence of diabetes. As a result, we consider these as two of the factors in our experiment.

Mokdad and Ford [5] studied the incidence of obesity, diabetes, and obesity-related health risk factors and found that being overweight and obesity were significantly positively associated with incidence of diabetes. Other literature also shown that at the individual level, weight gain and body mass are central to the formation and rising incidence of type 1 and type 2 diabetes [6].

Singh and Siahpush [7] examined the association between unemployment and life expectancy in the United States during 1990-2010 and found that life expectancy was consistently lower in areas with higher unemployment rates. However, they didn't investigate the relationship between unemployment and diabetes specifically. An article by Rautio et al. [8] quantifies the relationship between varying levels of unemployment and impaired glucose metabolism, suggesting a relation between amount of unemployment and incidence of type 2 diabetes in men.

Methodology

We found a dataset [9] containing more than 90 variables for all 3,143 counties in the United States. Based on the previous literature review, we selected 5 variables/factors for our experiment, as shown in Table 1; each data value pertains to an individual county:

Factor A - Percentage of Population without High-School Diploma
Factor B - Median Earnings (measured in 2010 dollars)
Factor C - Median Age
Factor D - Adult Obesity Rate
Factor E - Unemployment Rate

Table 1: List of factors in our experiment.

We filtered the data by counties along the East Coast, as noted earlier. Then we split the levels of each factor into two levels (based on being above or below average; each factor is quantitative), and, as is usual for two-level experimentation, assigned label H ("High") or L ("Low") to the level of each factor. If an individual data value was above the average, it was labeled H; otherwise, it was labeled L. In the end, we had 749 data points representing the 5 factors and 32 combinations of factor levels ("treatment combinations".) Below, in Table 2, is a summary of counts of the 32 treatment combinations, in which the levels are listed in the A, B, C, D, E order as in the Table 1.

Labels	Count
HHHLH	1
HHHHH	3
LHHHH	5
HHLHL	6

LHLHH	6
LLLLH	6
HHLLL	7
HLLLL	8
HHLLH	8
HHLHH	8
HHHHL	10
HHHLL	11
LLLHH	11
LLLHL	12
HLHLL	15
LHHLH	17
LHLLH	17
LLHHL	18
LLHHH	20
LHHHL	20
HLHLH	20
HLLLH	21
LLHLH	21
LLLLL	22
LHLHL	23
HLHHL	26
HLLHL	31
LLHLL	50
HLHHH	70
LHHLL	75
LHLLL	79
HLLHH	102
Grand Total	749

Table 2: Frequencies for each of 32 treatment combinations

In Table 3, we performed Yates' Algorithm to calculate the (estimated) main effect of each factor and all the interaction effects. We decided to use only one randomly-chosen data value from each of the treatment combinations, making this an un-replicated 2^5 experimental design. For ease of calculation, we multiplied the original diabetes incidence data by 100 (which put it into percentages.)

YIELD	Original	Original*	1ST COL	2ND COL	3RD COL	4TH COL	5TH COL	Estimate
1	0.0720	7.2	17.3	35.2	77.1	174.6	362.7	22.669
a	0.1010	10.1	17.9	41.9	97.5	188.1	22.3	1.394
b	0.0710	7.1	22.9	51.0	88.2	14.2	-2.5	-0.156
ab	0.1080	10.8	19.0	46.5	99.9	8.1	15.1	0.944
c	0.1300	13.0	25.7	39.3	7.1	-2.4	18.3	1.144
ac	0.0990	9.9	25.3	48.9	7.1	-0.1	-7.3	-0.456
bc	0.0770	7.7	22.6	46.7	-0.6	12.8	-7.3	-0.456
abc	0.1130	11.3	23.9	53.2	8.7	2.3	5.9	0.369
d	0.1240	12.4	18.3	6.6	-3.3	2.2	32.1	2.006
ad	0.1330	13.3	21.0	0.5	0.9	16.1	9.3	0.581
bd	0.1190	11.9	26.1	2.4	-0.6	-3.8	5.3	0.331
abd	0.1340	13.4	22.8	4.7	0.5	-3.5	-0.3	-0.019
cd	0.1130	11.3	23.6	1.1	7.5	-2.8	-14.3	-0.894
acd	0.1130	11.3	23.1	-1.7	5.3	-4.5	10.5	0.656
bcd	0.0960	9.6	26.1	4.7	0.2	10.0	13.7	0.856
abcd	0.1430	14.3	27.1	4.0	2.1	-4.1	3.7	0.231
e	0.0950	9.5	2.9	0.6	6.7	20.4	13.5	0.844
ae	0.0880	8.8	3.7	-3.9	-4.5	11.7	-6.1	-0.381
be	0.0960	9.6	-3.1	-0.4	9.6	0.0	2.3	0.144
abe	0.1140	11.4	3.6	1.3	6.5	9.3	-10.5	0.656
ce	0.1290	12.9	0.9	2.7	-6.1	4.2	13.9	0.869
ace	0.1320	13.2	1.5	-3.3	2.3	1.1	0.3	0.019
bce	0.1240	12.4	0.0	-0.5	-2.8	-2.2	-1.7	-0.106
abce	0.1040	10.4	4.7	1.0	-0.7	1.9	-14.1	-0.881
de	0.1080	10.8	-0.7	0.8	-4.5	-11.2	-8.7	-0.544
ade	0.1280	12.8	1.8	6.7	1.7	-3.1	9.3	0.581
bde	0.1020	10.2	0.3	0.3	-6.0	8.4	-3.1	-0.194
abde	0.1290	12.9	-2.0	4.7	1.5	2.1	4.1	0.256
cde	0.1240	12.4	2.0	2.5	5.9	6.2	8.1	0.506
acde	0.1370	13.7	2.7	-2.3	4.1	7.5	-6.3	-0.394
bcde	0.1220	12.2	1.3	0.7	-4.8	-1.8	1.3	0.081
abcde	0.1490	14.9	2.7	1.4	0.7	5.5	7.3	0.456

Table 3: Yates' algorithm results.

Analysis and Discussion of Results

Our Hypothesis are:

For Factor A - Less Than High School Diploma Population Percentage

H_0: Counties' incidence of diabetes **is no different** for H vs. L level of "less than high school diploma population percentage."
H_1: Counties' incidence of diabetes **is different** for H vs. L level of "less than high school diploma population percentage."

For Factor B - Median Earnings (measured in 2010 dollars)

H_0: Counties' incidence of diabetes **is no different** for H vs. L median earnings (measured in 2010 dollars).
H_1: Counties' incidence of diabetes **is different** for H vs. L level of median earnings (measured in 2010 dollars).

For Factor C - Median Age

H_0: Counties' incidence of diabetes **is no different** for H vs. L median age.
H_1: Counties' incidence of diabetes **is different** for H vs. L median age.

For Factor D - Adult Obesity Rate

H_0: Counties' incidence of diabetes **is no different** for H vs. L adult obesity rate.
H_1: Counties' incidence of diabetes **is different** for H vs. L adult obesity rate.

For Factor E - Unemployment Rate

H0: Counties' incidence of diabetes **is no different** for H vs. L unemployment rate.
H1: Counties' incidence of diabetes **is different** for H vs. L unemployment rate.

We did not form any hypotheses about interaction effects.

Then we built an ANOVA table based on the Yates' Algorithm results. This is presented in Table 4; as is most commonly done, we considered all three-way and higher-order interactions to be zero and used their sum-of-squares values to form the error [10].

	SSQ	**df**	**MSQ**	**F**
A	0.060704	1	0.060704	136.6299
B	0.000763	1	0.000763	1.71718
C	0.04088	1	0.04088	92.01065
D	0.125782	1	0.125782	283.104
E	0.022247	1	0.022247	50.0728
AB	0.027833	1	0.027833	62.64549
AC	0.006505	1	0.006505	14.64137
AD	0.010558	1	0.010558	23.76303

AE	0.004542	1	0.004542	10.22341
BC	0.006505	1	0.006505	14.64137
BD	0.003429	1	0.003429	7.717696
BE	0.000646	1	0.000646	1.453421
CD	0.024962	1	0.024962	56.18339
CE	0.023585	1	0.023585	53.08423
DE	0.00924	1	0.00924	20.79574
Error	0.007109	16	0.000444	
Total		31		
F(1,6)	4.49			

Table 4: Analysis of Variance results.

From the ANOVA table and assuming a 5% significance level,

1. We conclude that there are differences (reject H_0) in counties' incidence of diabetes depending on the level of Factor A: less than high school population percentage. (p-value = .000)
2. We conclude that we cannot say that there are differences (accept H_0) in counties' incidence of diabetes depending on the level of Factor B: median earnings (measured in 2010 dollars). (p-value = .209)
3. We conclude that there are differences (reject H_0) in counties' incidence of diabetes depending on the level of Factor C: median age. (p-value = .000)
4. We conclude that there are differences (reject H_0) in counties' incidence of diabetes depending on the level of Factor D: adult obesity rate. (p-value = .000)
5. We conclude that there are differences (reject H_0) in counties' incidence of diabetes depending on the level of Factor E: unemployment rate. (p-value = .000)
6. We reject H0 and accept H1 for each hypothesis, except for Factor B, for which we accept H0.
7. We conclude that all two-factor interaction effects are significant, except the interaction effect between median earnings (measured in 2010 dollars) and unemployment rate.

The directions of the (main effect) significant results indicate that incidence of diabetes in a county is higher for a county with higher "less than high school population percentage," higher for counties with higher "median age," higher for counties with higher "adult obesity percentage," and higher for counties with higher "unemployment percentage." None of these results appear to be surprising. We cannot conclude that the incidence of diabetes in a county varies by the county's median income. This is consistent with past studies that indicated that earnings and incidence of diabetes are not related [11].

We noted in item 6 above that there are many significant two-factor interactions. While significant, all of them have effect sizes well below all the main effects, except for the main effect of E, unemployment rate, where 3 of the 9 significant two-way interactions are of the same order-of-magnitude as the Factor E effect.

Conclusions

In this paper, we explored the association between five factors (less-than-high-school- population percentage, median age, median income, adult obesity rate, and unemployment rate) and incidence of diabetes in counties in the United States. We concluded that all main effects and two-factor interaction effects are significant except for median earnings (measured in 2010 dollars) and interaction between median earnings and unemployment.

Limitations and Directions for Future Research

However, there are some limitations in our study. First of all, we select only 32 data points for our study. This resulted in each treatment combination having the same sample size, but perhaps greatly reduced the power of the hypothesis tests. Yet, given the great prevalence of significant results, we do not believe that this reduced power was particularly material to the study and its results.

Second, we covered counties only on the East Coast of the United States, and it is possible that the results could be different were the entire United States included. Still, the results were so dramatically strong, that we do not believe that the results would change, but cannot say this with certainty. Third, we used median income in this analysis; it is possible that some other measure of income would yield a different result – our result was not significant, but the p-value was around .2, not somewhere near 1.0. Thus, we must allow for the possibility that some other measure of income (e.g., the percent of people with income below some poverty level) might change the result.

References

1. National Diabetes Statistics Report (2017) Center for Disease Control, Atlanta, Georgia, U.S.A. Centers for Disease Control and Prevention 1-20.
2. Murad M, Abdulmageed S, Iftikhar R, Sagga B (2014) "Assessment of the Common Risk Factors Associated with Type 2 Diabetes Mellitus in Jeddah." International Journal of Endocrinology, January 1-9.
3. Wangdi K, Jamtsho T (2018) "Risk Factors for Self-Reported Diabetes among Bhutanese Adults: A Nationally Representative Survey Data Analysis." PLoS ONE 13(11); 1-13.
4. West S, Munoz B, Klein R, Broman A, Sanchez R, et al. (2002) "Risk Factors for Type II Diabetes and Diabetic Retinopathy in a Mexican-American Population: Proyecto VER," American Journal of Ophthalmology. 134(3); 390-398.
5. Mokdad A, Ford E, Bowman B, Dietz W, Vinicor F, et al. (2003) "Prevalence of Obesity, Diabetes, and Obesity-Related Health Risk Factors, 2001." JAMA 289 (1); 76-79.
6. Tsai A, Williamson D, Glick H (2011) "Direct Medical Cost of Overweight and Obesity in the USA: A Quantitative Systematic Review," Obesity Reviews: An Official Journal of The International Association for The Study of Obesity. 12(1); 50-61.
7. Singh G, and Siahpush G (2016) "Inequalities in US Life Expectancy by Area Unemployment Level, 19902010," Scientifica, March; 1-12.
8. Rautio N, Varanka-Ruuska T, Vaaramo E, Palaniswamy S, Nedelec R, et al. (2017) "Accumulated Exposure to Unemployment Is Related to Impaired Glucose Metabolism in Middle-Aged Men: A Follow-up of the Northern Finland Birth Cohort 1966," Primary Care Diabetes 11(4); 365-372.
9. US County Dataset from HAWKES Stat, 2018.
10. Berger P, Maurer R, and Celli G (2018) Experimental Design with Applications in Management, Engineering, And the Sciences, Second Edition. Springer, Inc.

11. López A, Cecchetto E, Aguirre A, Ontiveros M, Roitter Claudia V, et al. (2017) “Factors Associated with Quality of Life Related to Health in Persons with Diabetes Mellitus of the University Social Work of Cordoba.” Revista De La Facultad De Ciencias Medicas (Cordoba, Argentina) 74 (4); 306-312.
12. Al-Goblan A, Al-Alfi M, Khan M (2014) “Mechanism Linking Diabetes Mellitus and Obesity.” Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy 7, December: p. 587-591.
13. Wilkin T (2001) “The Accelerator Hypothesis: Weight Gain as the Missing Link between Type I and Type II Diabetes.” Diabetologia, 44(7); 914-22.
14. Naser K, Gruber A, Thomson G (2006) “The Emerging Pandemic of Obesity and Diabetes: Are We Doing Enough to Prevent A Disaster?” International Journal of Clinical Practice, 60(9); 1093-1097.