Human Demographic Contributions to the 2004 Death Rate by U.S. State and County

By Madelaine Brown, Jordyn Lucier, and Harley Clifton

I. Introduction

There is only one guarantee in life: death. Heart Disease has been the leading cause of death in the United States since 1950 (1). The number of car accident related deaths every year, and more recently, the high volume of COVID-19 caused deaths are widely reported. However, less is known about the influence of socioeconomic factors - such as income and rate of physicians by county - on the annual death rate in the US. When looking at the year 2004, based on all the filed death certificates in the US, the leading cause of death was heart disease, followed by malignant neoplasms (cancer), and cerebrovascular diseases (e.g. strokes, 2). For each cause of death, there exist differences amongst death rate by age, number of males per 100 females, race, and Hispanic origin were evident. For example, the leading cause of death for White, Black, and American Indian or Alaskan Native race categories was heart disease causing 27.5%, 25.8%, and 19.8% of all deaths, respectively (2). However, for the Asian or Pacific Islander individuals, the leading cause of death was malignant neoplasms at 24.6% of all deaths. While there is clear evidence that there are different causes of death are more prevalent among certain communities, there are few explanations as to why this is the case. Information like geographical location, income, and number of physicians in the area may help fill in some of these gaps.

United States Census Bureau data, categorized by States and Local Areas (3), was analyzed for this project. The open access data that was obtained focuses on the death rate in the U.S. during the year 2004. This data was last updated on October 1st, 2010 and was accessed in September of 2022. Within this dataset, there is information regarding the death rate by county and state in addition to predictor variables such as percent of county population in different age groups, number of males per 100 females, average income, and number of physicians for each county. Previous studies have focused on the relationship between income and death; however, there does not seem to be a strong consensus. For example, in 2010 the Center for Disease Control and Prevention stated that there was a higher association between premature mortality and low-income counties when compared to high-income counties (4). Another study conducted by Shi et al. found that women mortality rates were more impacted by income than men (5). Based on previous knowledge and the available data, this analysis focused on exploring which human demographic contributed to the death rate by county and state in the U.S. in 2004.

II. Data Overview

The dataset was sourced from the United States Census Bureau's 2008 Statistical Abstract of the United States (3). The Census Bureau compiled this data from the CDC's National Center of Health Statistics, which records mortality and physician data based on complete vital records filed in the registration offices of all states, independent cities, and the District of Columbia. Deaths of nonresident aliens in the United States and U.S. citizens outside the United States have been excluded from the data. Researchers used a multistage probability sample of 36−40,000 households (449 effective primary sampling units) to obtain the demographic data. The variables of primary interest for this analysis are per capita income and number of physicians per 100,000 people, and their influence on death rate by state and county in 2004. The effects of race, number of males per 100 females, population density, and age composition of the counties on the death rate were also considered.

When taking a closer look at the data, some missingness was found. Specifically, three counties - namely Yellowstone National Park county in Montana, Clifton Forge county in

Virginia, and South Boston county in Virginia - were missing data entries for all variables. After further investigation, it was discovered that this missingness was due to county consolidation prior to 2004. Yellowstone National Park County became incorporated with Gallatin County in 1997 and therefore was excluded without worry of misrepresenting the data. Clifton Forge independent city became incorporated with Alleghany County, VA in 2001 and South Boston was incorporated with Halifax County, VA in 1995, so they were also excluded from the dataset. Additionally, Kalawao County, HI was excluded due to the suspiciously high number of 0's in the data, including a 0 death rate. After the removal of these counties, 3110 observations remained. Interestingly, it was also discovered that Virginia had 29 missing values for the income variable, and these observations were excluded from any model that included the income variable. Virginia was the only state that had additional missingness within a specific variable.

The dataset contained information on the following variables:

- Death rate: number of people who died/ 1,000 people
- Income: average income in the county observed in USD
- Physician rate: number of physicians in the county observed per 100,000 people
- Population density: number of persons per sq. mile
- Sex: number of males per 100 females
- Age composition: percent of total population that belongs to each respective age group (under 5, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+)
- Racial composition: percent of total population that belongs to each respective racial group (White, African American, Hispanic or Latino, Native American or Native Alaskan, Native Hawaiian or other Pacific Islander, Asian)
- State: the state of the observation
- County: the county of the observation, nested inside State

III. Exploratory Data Analysis

Side-by-side boxplots for each state revealed there are some differences in observed death rate by state (Figure 1) and county (Figure 2). Due to the large number of counties and states observed, their inclusion in any model should only be as random effects. A variety of scatterplots were also created to illustrate the relationship between the 2004 death rate and multiple predictor variables. One such plot displays a negative relationship between the proportion of population between 5 and 14 years old and death rate in 2004 (Figure 3). This suggests that populations that have higher concentrations of people in this youthful age group are associated with decreased death rates. Another plot reveals a negative relationship between the percent of people aged 75 years or older and the death rate - meaning as levels of elderly population composition increase, death rate also increases (Figure 4). Based on these graphs, the relationship between percent of population that falls into each age group and the observed death rate should be explored.

Additionally, a tally of the number of counties within each state and district was created (Table 1). The table shows there are at least three counties in each individual state, with the exception of the District of Columbia which only has a singular county. Therefore, including State as a random effect is justified, but including county as a random effect may be overfitting the model due to the astronomical volume of counties observed in the data.

A scatterplot-correlation matrix was created using the ‘psych’ package in R (Figure 5) to explore and visualize potential relationships between quantitative variables in the data (6). There seems to be a slight positive relationship between the percentage of the population that is White and death rate. There also seems to be a slight negative relationship between scaled per capita income and the observed death rate. The plot shows no clear relationship between death rate and number of physicians per 100,00 people, the percentage of the population that is African

American, the percentage of the population that is Hispanic or Latino, the percentage of the population that is Asian, the percentage of the population that is Native American or Native Alaskan, and the percentage of the population that identifies as Native Hawaiian, other, or Pacific Islander.

IV. Methods

Exploratory data analysis was performed to visualize the data and get a preliminary idea of what the data looked like. The distribution of death rate was plotted against predictors, state, income, age, race, and physician rate using the 'ggplot2' package in R (7). After satisfactory exploratory data analysis, an initial linear model was fit with all predictors and all relevant interactions. Due to previous knowledge on healthcare for minority populations and their historically limited access to higher levels of education, suspicion arose regarding the possibility of important interactions between physician rate and race groups, and physician rate and income - all of which were included in the beyond optimal linear model (8). Next, a mixed effect model was fit with all the same predictors as the linear model, with an addition of a random effect of state using the 'nlme' package (9). While it is suspected that adding a random effect for county nested within state would be over-fitting the data, it was checked for it anyway. Since the data was not a result of a designed experiment, dropping county from the model was reasonable. The inclusion of county as a nested random effect did, in fact, overfit the data as expected and therefore was excluded. An anova test comparing the saturated model to the state random intercept model revealed that the mixed effect model was a superior fit (Output 1).

After the optimal random effect structure was determined, variable selection for fixed effects was performed. To compare models with different fixed effects structures, the mixed effect model had to be refit using maximum likelihood estimation. To perform variable selection, variables with the highest p-value were removed one at a time until we were satisfied with the final model. Following the stepwise backward variable selection process, the AIC and BIC values of all explored models were compared using an anova test to confirm the best fitting model (Output 2)(10). When multiple models had very similar AIC and BIC values, the principle of parsimony prompted the selection of the simplest model.

Next, diagnostics of the chosen model were analyzed using the model residuals (Figure 6). Model diagnostics displayed clear violations of the linearity and non-constant variance assumptions, so a log transformation of the response and non-constant variance structures were explored. Modeling a log transformation of death rate did not improve model diagnostics, so it was excluded from the model moving forward. Based on a plot of standardized residuals and the percent of the population between ages 25 to 34 (Figure 7), a var Power non-constant variance structure for this variable was introduced into the model. An anova model comparison revealed the inclusion of this variance structure was an improvement and therefore necessary to adequately model the data (Output 3). Following the addition of the non-constant variance structure to the model, fixed effect variable selection and assessment of final model diagnostics were redone.

After the updated backward stepwise variable selection was completed, the final model included fixed effects for number of physicians per 100,000 people, per capita income, age (percent of population between the ages 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75+ years), race (African American, Asian, Hispanic or Latino), interactions between physician rate and African American, physician rate and Hispanic or Latino, and physician rate in income. As previously discussed, the model also included a random effect for state and the non-constant variance structure is varPower for percent of people between ages 25 and 34 years old (Output 4). The final theoretical model written out is as follows:

$$DeathRate_{ij} = \beta_0 + \beta_1 PhysicianRate_{ij} + \beta_2 Income_{ij} + \beta_3 PercentAge5to14_{ij} + \beta_4 PercentAge15to24_{ij}$$
$$+\beta_5 PercentAge25to34_{ij} + \beta_6 PercentAge35to44_{ij} + \beta_7 PercentAge45to54_{ij} + \beta_8 PercentAge55to64_{ij}$$
$$+\beta_9 PercentAge65to74_{ij} + \beta_{10} PercentAge75plus_{ij} + \beta_{11} PercentAfricanAmerican_{ij} + \beta_{12} PercentAsian_{ij}$$
$$+\beta_{13} PercentHispanicOrLatino_{ij} + \beta_{14} PhysicianRate_{ij} : PercentAfricanAmerican_{ij}$$
$$+\beta_{15} PhysicianRate_{ij} : PercentHispanicOrLatino_{ij} + \beta_{16} PhysicianRate_{ij} : Income_{ij} + \underline{\beta_{17}} X_{State_{ij}} + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma^2 |PercentAge25to34_{ij}|^{2\delta})$$

Where i is the observation number and j = 1, 2, …, 51 to represent the 50 United States and the District of Columbia. Estimates for each of the coefficients can be found in the final model summary.

V. Results

Diagnostics for the model with the varPower non-constant variance structure were assessed to compare whether this model was an improvement over the original. There was only a slight increasing fanning pattern present in the Normalized Residual vs. Fitted plot, indicating little evidence against the constant variance assumption (Figure 8). Although there is still some evidence against this assumption, it showed notable improvement compared to the diagnostics for the original model and the one with the log-transformed response. The Residual vs. Fitted plot also suggests that there is very minor missed curvature, providing little evidence against the linearity assumption (Figure 8). Since little evidence against the linearity assumption was found in the Residual vs. Fitted plot, the Scale Location plot was used to further assess constant variance. The Scale Location plot showed some evidence against homoskedacity, but less than before, and also suggests the presence of missed curvature (Figure 9).

In the Normal Q-Q plot for the non-constant variance model, the residuals deviate from the 1:1 Q-Q line in a heavy tail pattern which is problematic and a clear violation of the normality of residuals assumption (Figure 10). However, this diagnostic did not seem to change much from the Q-Q plot for the original model and was a significant improvement compared to the Q-Q plot for the model with a log-transformed response.

Lastly, fitting a mixed effect model introduces the additional assumption that the random effects are normally distributed. The random effects for each state were extracted using the 'nlme' package in R, and a histogram was created to assess this assumption (Figure 11)(9, 10), The random effects for state were centered around zero and approximately normally distributed, providing little to no evidence against the assumption of normal random effects. This was a slight improvement over the histogram of random effects for the original model, and a vast improvement over the left skewed distribution of random effects for the log-transformed model.

Overall, the model with a varPower non-constant variance structure displayed improvement over both the original model's diagnostics and those for the model with a log transformed response. This model's estimated coefficients and 95% confidence intervals for each of these estimates can be found in the Final Model summary. Physician rate, income, all age categories except for 75 and over, Asain, and Hispanic or Latino are each associated with an estimated decrease in death rate (Final Model). The percent of population age 75 and over, African American, and the interactions between Physician rate and African American, physician rate and Hispanic or Latino, physician rate and income are all associated with an estimated increase in death rate (Final Model, Output 5).

VI. Conclusion

This analysis allowed for the identification of a subset of human demographics that contributed most to the death rate across the United States in 2004. The following predictors were important when modeling the death rate in 2004: physician rate, income, percent of the population of each county within 5 to 14, 15 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74, and 75+ age brackets, and the percent of the population of each county who are African American, Asian, and Hispanic or Latino. Model selection revealed a need to include an interaction between physician rate and African American population percentage, an interaction between physician rate and Hispanic or Latino population percentage, and the interaction between physician rate and income. Further, initial model selection revealed that including a random intercept for state drastically improves model fit.

Contrastingly, there were many variables explored that were not needed when modeling 2004 death rates across the United States. These include population density, number of males per 100 females, percent of the population under age 5, percent of population who are White, percent who are American Indian or Alaskan Native, percent who are Native Hawaiian, other, or Pacific Islander. Unimportant interactions were between physician rate and White population composition, the interaction between physician rate and American Indian or Alaskan Native population composition, the interaction between physician rate and Native Hawaiian, other, or Pacific Islander population composition. Although county is inherently nested within state via the data collection structure, fitting county nested within state for a random effect caused overfitting in the model. Due to the lack of degrees of freedom and the data not being collected from a designed experiment, the removal of the county random effect was appropriate.

After the important predictors were isolated, the scope of inference was determined. Since random assignment of race, age, number of males per 100 females, income, population density, and physician rate are impossible, we cannot conclude a causal relationship between the predictors and the response. Assuming that the multistage probability sample that the researchers used to collect the data resulted in a representative sample, the association - or lack thereof - between the predictors and 2004 death rate can be generalized to the entire United States.

It is reasonable to suspect temporal correlation may exist within this data due to increases in life expectancy in the United States over time (3). Ideally, this would have been explored and modeled in this analysis. Unfortunately, methods of data collection was changed in 2006. Therefore, it would be unreasonable to compare data collected using different methods. Also, due to the county consolidation, data prior to 2001 should not be compared to data collected afterwards. Ultimately, the United States Census Bureau only collects data once every 4 years, which limited to data to 2004. Therefore, it is recommended that temporal correlation be explored in future studies using data collected after 2006.

Appendix

References

1. Heron M, RN. (2016). Changes in the leading cause of death: Recent patterns in heart disease and cancer mortality. NCHS Data Brief, no 254. Hyattsville, MD: National Center for Health Statistics. 2016.

2. Heron M. (2007). Deaths: Leading Causes for 2004. Natl Vital Stat Rep. 2007 Nov 20;56(5):1-95. PMID: 18092547.

3. US Census Bureau. (2021). “Statistical Abstract of the United States: 2008” . Census.gov, 16 Dec. 2021, https://www.census.gov/library/publications/2007/compendia/statab/127ed.html.

4. “Preventing Chronic Disease.” (2012). Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, https://www.cdc.gov/pcd/issues/2012/11_0120.htm.

5. Shi, J. et al. (2021). “The Impact of Income Definitions on Mortality Inequalities.” SSM - Population Health, Elsevier, 7 Sept. 2021, https://reader.elsevier.com/reader/sd/pii/S2352827321001907?token=18223EDC1508B5C76DF567A36B1ECA6CA61EE180DD708FCA14E199C596A03E4D3F14E182A965102488BD3804EEA8A3E9&originRegion=us-east-1&originCreation=20221205224919.

6. Revelle, W. (2022). The psych Package: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 2.2.5.

7. Wickham, H. (2016). The ggplot2 Package: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org.

8. Benjamins MR, Silva A, Saiyed NS, De Maio FG. Comparison of All-Cause Mortality Rates and Inequities Between Black and White Populations Across the 30 Most Populous US Cities. *JAMA Netw Open.* 2021;4(1):e2032086. doi:10.1001/jamanetworkopen.2020.32086.

9. Pinheiro, J., Bates, D., R Core Team. (2022). The nlme Package: Linear and Nonlinear Mixed Effects Models. R package version 3.1-159, https://CRAN.R-project.org/package=nlme.

10. R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Figures

Figure 1: 2004 death rate by state (now in color!)

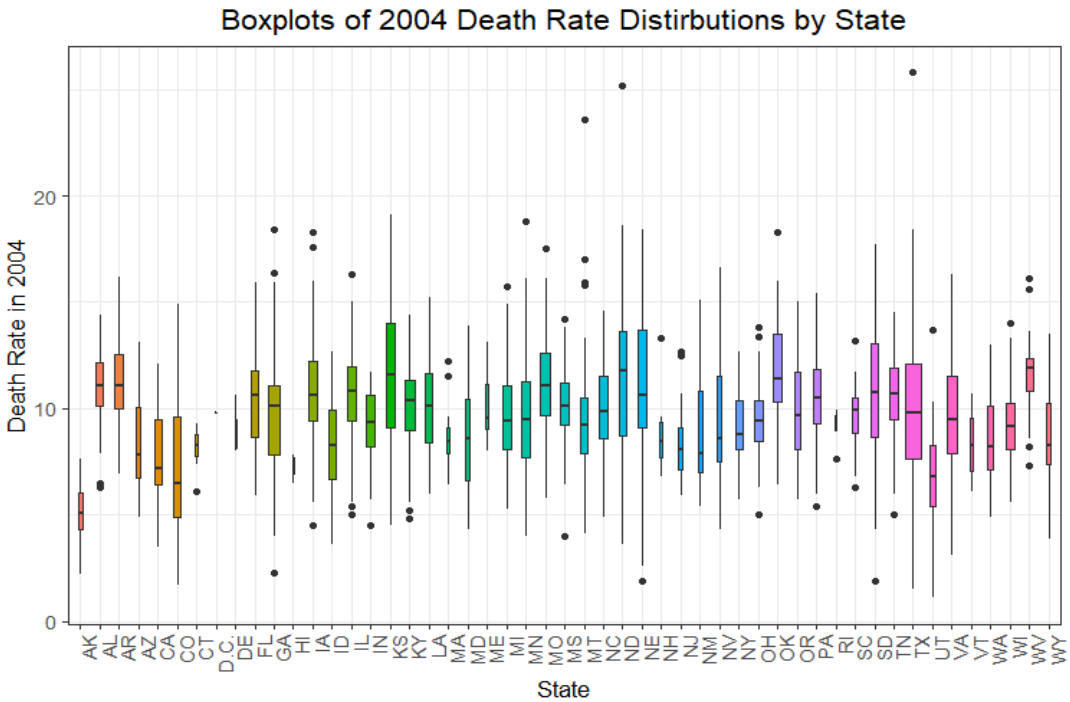


Figure 2: The death rate by county boxplot of your nightmares.

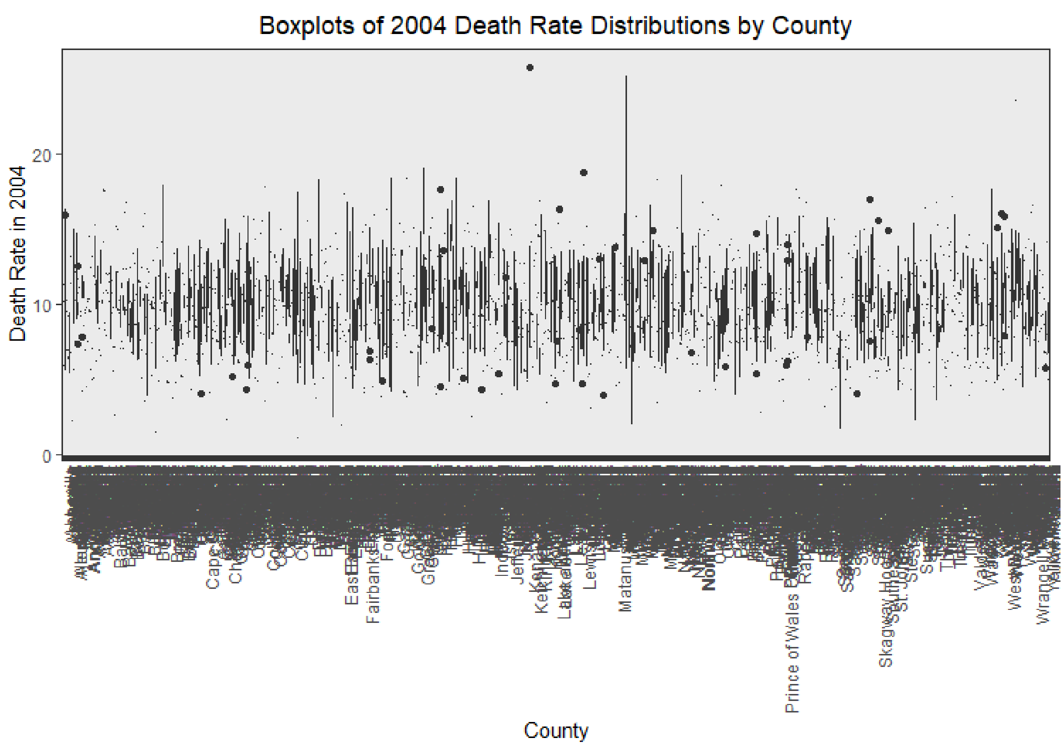


Figure 3: Youngsters by death rate.

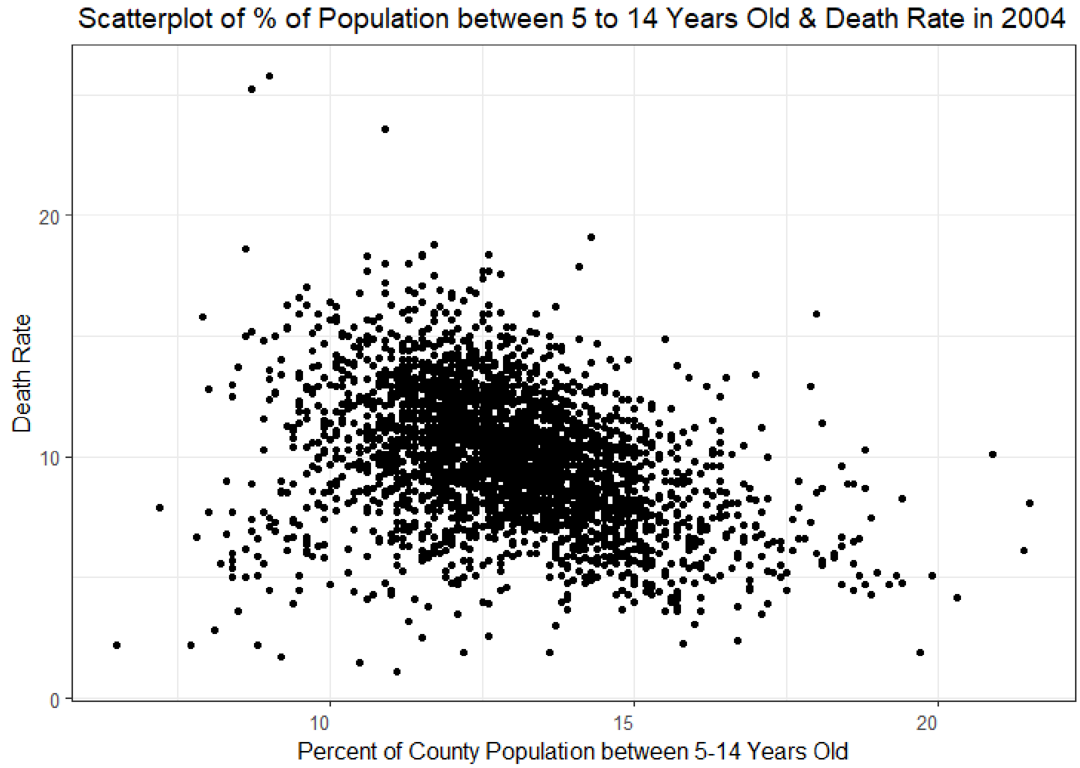


Figure 4: Percent of population that's old as f*ck by death rate.

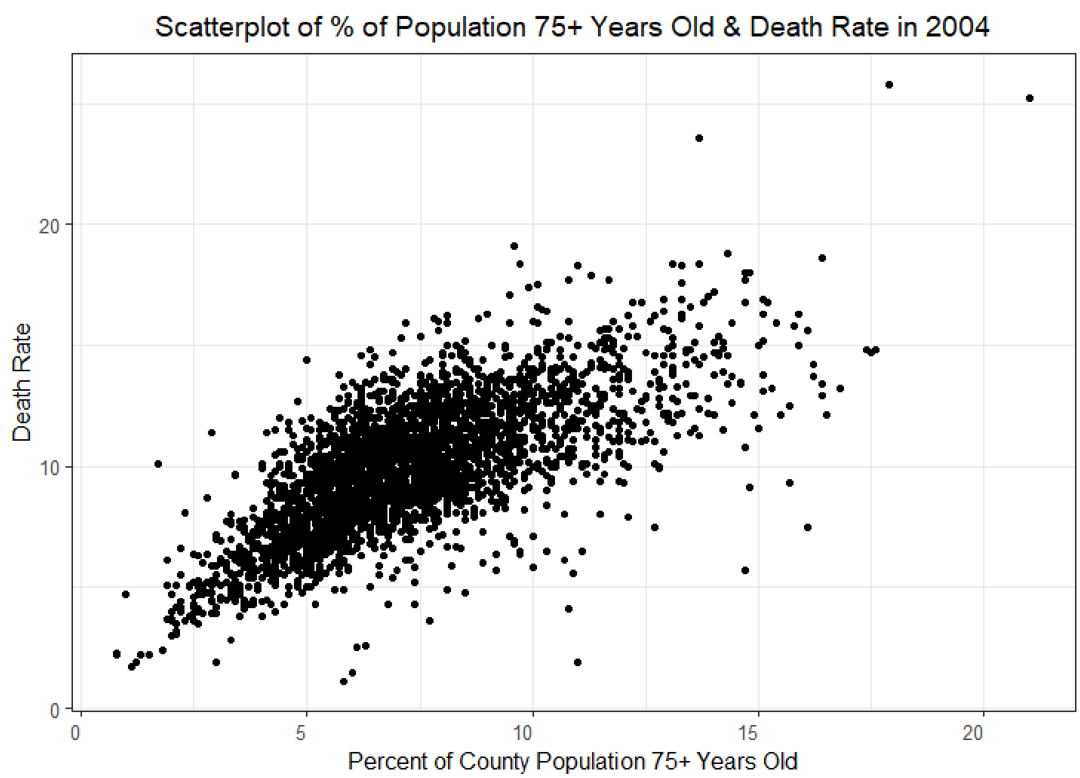


Figure 5: Scatterplot correlation matrix.

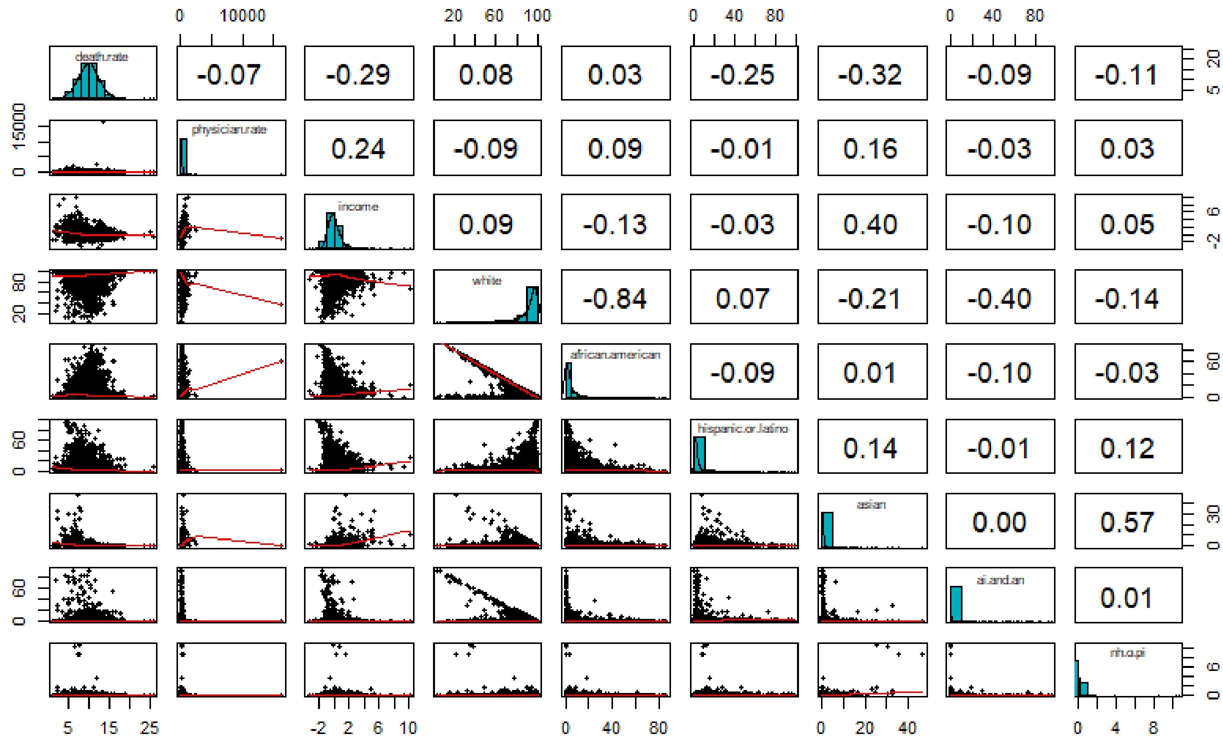


Figure 6: Diagnostic plots for the initial model post variable selection.

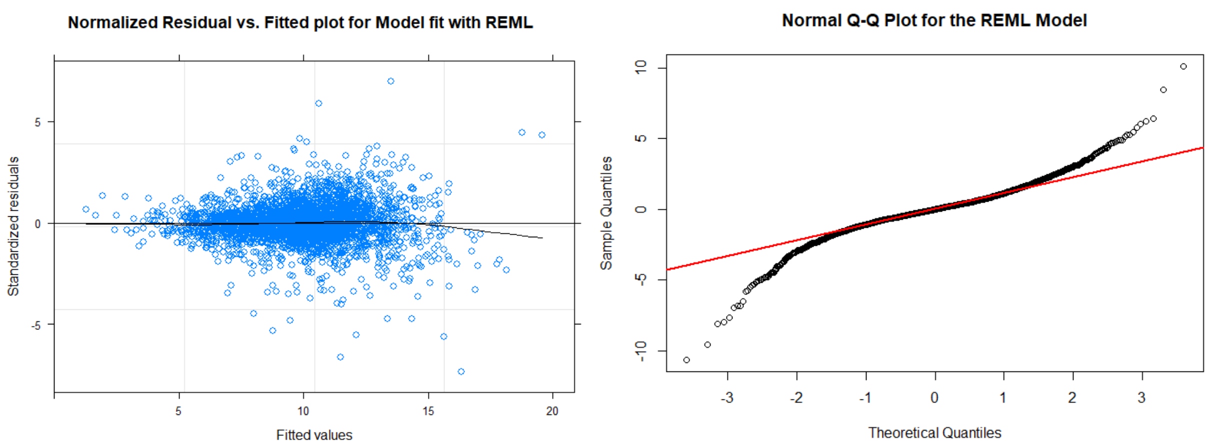


Figure 7: Residual Plot for Percent of Population Between Ages 25 to 34.

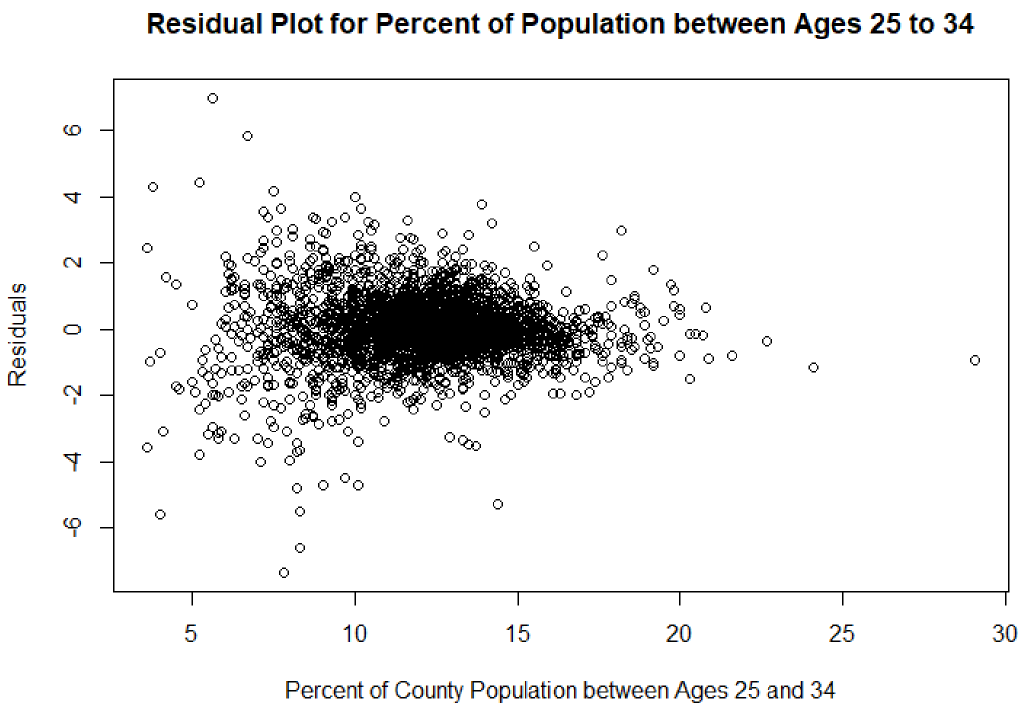


Figure 8: Residual vs. Fitted diagnostic plot for the final non-constant variance model.

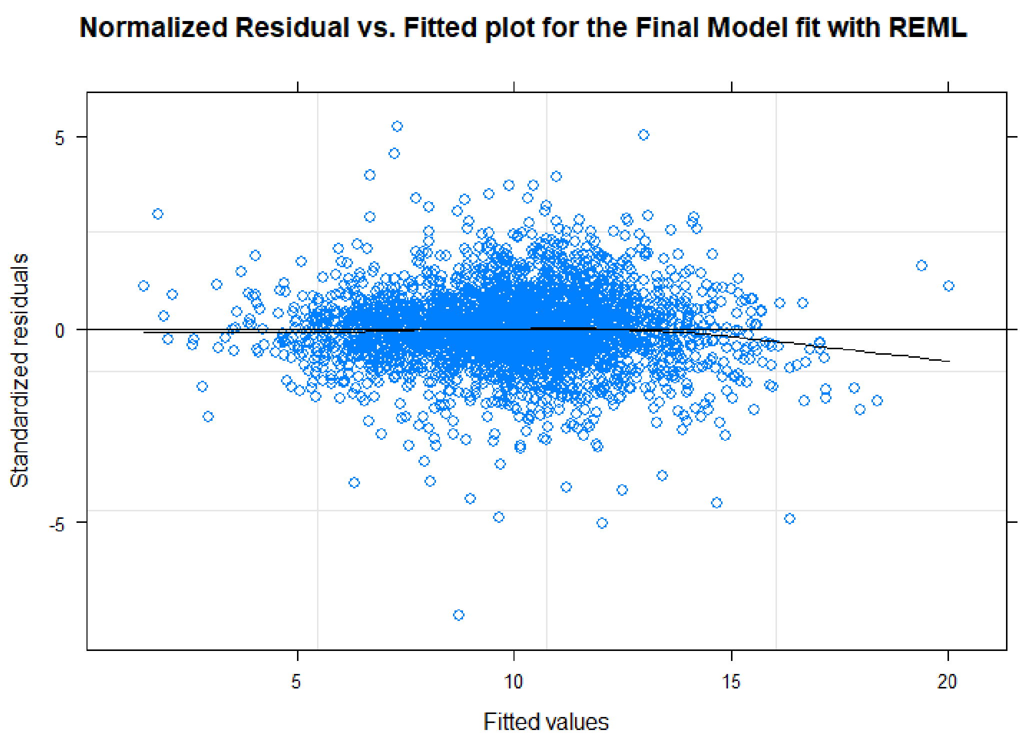


Figure 9: Scale-Location diagnostic plot for the final non-constant variance model.

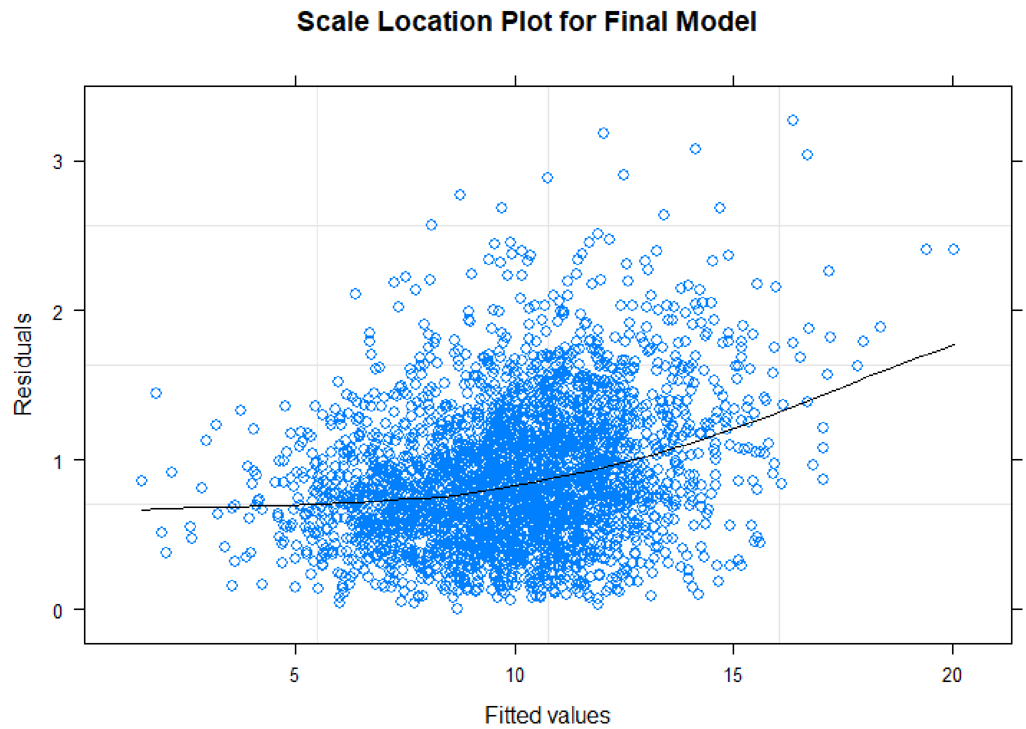


Figure 10: Normal Q-Q diagnostic plot for the final non-constant variance model

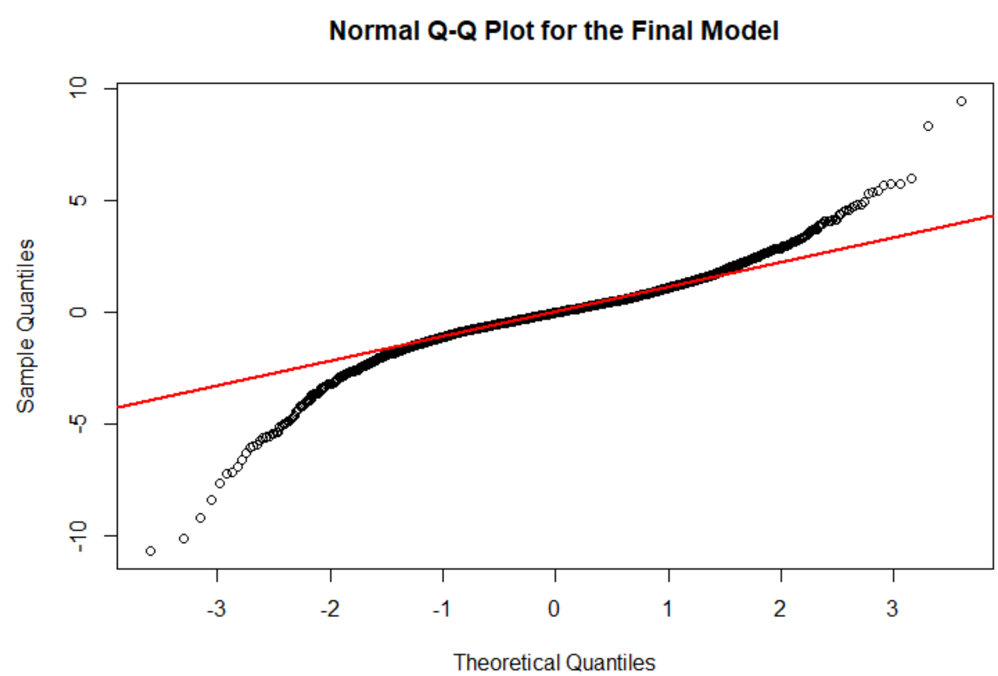
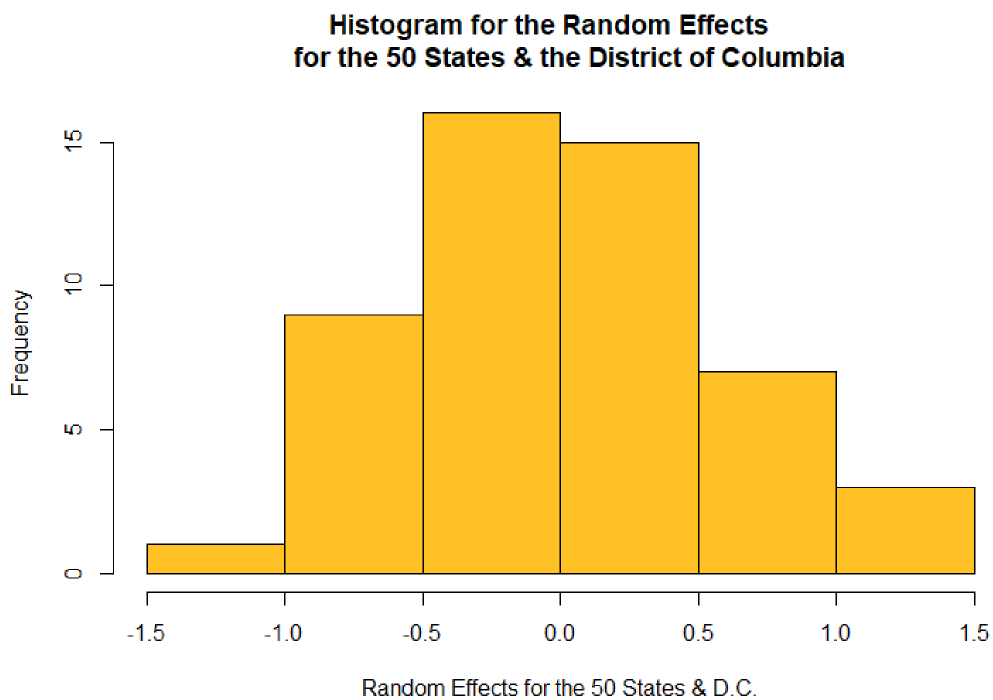


Figure 11: Histogram of random effects for the final non-constant variance model



Model Summaries

Final Model: Estimated coefficients for the final mixed effects model

Estimated Power (Delta) Parameter for Non-Constant Variance	-1.217987

	Estimated Coefficient	**Standard Error**
Intercept	39.71716	3.581595
Physician Rate	-0.00210	0.000410
Income	-0.00011	0.000006
Percent Age 5-14	-0.44548	0.055303
Percent Age 15-24	-0.39574	0.036103
Percent Age 25-34	-0.38688	0.043614
Percent Age 35-44	-0.33072	0.039150
Percent Age 45-54	-0.35441	0.042791
Percent Age 55-64	-0.15201	0.047168
Percent Age 65-74	-0.26820	0.050008
Percent Age 75+	0.39235	0.043165
African American	0.01682	0.002364
Asian	-0.05235	0.011271
Hispanic or Latino	-0.04246	0.003315
Physician Rate:African American	0.00001346	0.000005
Physician Rate:Hispanic or Latino	0.000003796	0.000016
Physician Rate:Income	0.000000061056	0.000000

95% Confidence intervals for the estimated model coefficients for fixed effects only

	Lower 95% CI	**Upper 95% CI**
Intercept	32.69	46.74
Physician Rate	-0.002903	-0.001295
Income	-0.0001186	-0.00009507
Percent Age 5-14	-0.5539	-0.3370
Percent Age 15-24	-0.4665	-0.3249
Percent Age 25-34	-0.4724	-0.3014
Percent Age 35-44	-0.4075	-0.2540
Percent Age 45-54	-0.4383	-0.2705
Percent Age 55-64	-0.2445	-0.5952
Percent Age 65-74	-0.3663	-0.1701
Percent Age 75+	0.3077	0.4770
African American	-0.01218	-0.02145
Asian	-0.07445	-0.03025
Hispanic or Latino	-0.04896	-0.03596
Physician Rate:African American	0.000004569	0.000022346
Physician Rate:Hispanic or Latino	0.000005975	0.000069937
Physician Rate:Income	0.00000004019	0.00000008192

Tables

Table 1: Number of counties and independent cities within each State & the District of Columbia, as recognized by the NCHS.

AK	AL	AR	AZ	CA	CO	CT	D.C.	DE	FL	GA	HI	IA
27	67	75	15	58	64	8	1	3	67	159	5	99

ID	IL	IN	KS	KY	LA	MA	MD	ME	MI	MN	MO	MS
44	102	92	105	120	64	14	24	16	83	87	115	82

MT	NC	ND	NE	NH	NJ	NM	NV	NY	OH	OK	OR	PA
56	100	53	93	10	21	33	17	62	88	77	36	67

RI	SC	SD	TN	TX	UT	VA	VT	WA	WI	WV	WY
5	46	66	95	254	29	134	14	39	72	55	23

Relevant R Output

Output 1: Anova Results Comparing the SLR Model and the Model with a State Random Effect.

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
m1.gls	1	28	11833.97	12002.91	-5888.986			
m2	2	29	11543.63	11718.61	-5742.815	1 vs 2	292.3417	<.0001

Output 2: Anova Results for Initial Variable Selection.

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
mem1	1	29	11280.94	11456.17	-5611.471			
mem2	2	28	11278.95	11448.13	-5611.473	1 vs 2	0.004590	0.9460
mem3	3	27	11277.19	11440.33	-5611.594	2 vs 3	0.240089	0.6241
mem4	4	26	11275.43	11432.53	-5611.715	3 vs 4	0.242481	0.6224
mem5	5	25	11274.56	11425.62	-5612.278	4 vs 5	1.127036	0.2884
mem6	6	24	11273.67	11418.69	-5612.837	5 vs 6	1.117148	0.2905
mem7	7	23	11271.77	11410.75	-5612.887	6 vs 7	0.100146	0.7517
mem8	8	22	11270.18	11403.11	-5613.090	7 vs 8	0.406108	0.5240
mem9	9	21	11268.22	11395.11	-5613.111	8 vs 9	0.042940	0.8358
mem10	10	20	11266.24	11387.08	-5613.118	9 vs 10	0.012361	0.9115
mem11	11	19	11265.28	11380.08	-5613.639	10 vs 11	1.041831	0.3074
mem11.5	12	18	11266.36	11375.12	-5615.180	11 vs 12	3.082713	0.0791
mem12	13	18	11266.40	11375.16	-5615.200			
mem13	14	17	11265.96	11368.68	-5615.981	13 vs 14	1.561574	0.2114
mem14	15	16	11267.61	11364.29	-5617.804	14 vs 15	3.646754	0.0562

Output 3: Anova Results comparing models with varPower Variance Structure for the different Age Groups.

	Model	df	AIC	BIC	logLik
varpower1	1	19	11158.42	11273.23	-5560.211
varpower2	2	19	10642.58	10757.39	-5302.290
varpower3	3	19	10993.44	11108.25	-5477.722
varpower4	4	19	11141.43	11256.23	-5551.713
varpower5	5	19	11079.86	11194.66	-5520.929
varpower6	6	19	10813.98	10928.79	-5387.992
varpower7	7	19	10932.42	11047.23	-5447.212

Output 4: Anova Results for Variable Selection after including varPower Variance Structure for the Percent of the Population Ages 25 to 34 years old.

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
mem11.5.1	1	30	10647.12	10828.39	-5293.558			
mem11.5.2	2	29	10645.12	10820.35	-5293.558	1 vs 2	0.0000256	0.9960
mem11.5.3	3	28	10643.52	10812.70	-5293.759	2 vs 3	0.4017465	0.5262
mem11.5.4	4	27	10641.72	10804.87	-5293.861	3 vs 4	0.2031967	0.6522
mem11.5.5	5	26	10640.26	10797.37	-5294.132	4 vs 5	0.5427403	0.4613
mem11.5.6	6	25	10638.29	10789.35	-5294.145	5 vs 6	0.0253878	0.8734
mem11.5.7	7	24	10636.42	10781.43	-5294.208	6 vs 7	0.1254151	0.7232
mem11.5.8	8	23	10634.74	10773.72	-5294.372	7 vs 8	0.3282264	0.5667
mem11.5.9	9	22	10632.86	10765.80	-5294.432	8 vs 9	0.1209613	0.7280
mem11.5.10	10	21	10633.12	10760.01	-5295.559	9 vs 10	2.2537742	0.1333
mem11.5.11	11	20	10634.21	10755.06	-5297.104	10 vs 11	3.0894839	0.0788

Output 5: Estimated Coefficients for the Final Non-Constant Variance Model

```
Linear mixed-effects model fit by REML
  Data: death.dat
      AIC      BIC    logLik
  10795.6 10916.34 -5377.799

Random effects:
 Formula: ~1 | State
        (Intercept) Residual
StdDev:   0.6156988 26.53938

Variance function:
 Structure: Power of variance covariate
 Formula: ~prc.25.34.years
 Parameter estimates:
    power
-1.217987
Fixed effects:  death.rate ~ physician.rate + income + prc.5.14.years + prc.15.24.years +      prc.25.34.ye
ars + prc.35.44.years + prc.45.54.years + prc.55.64.years +      prc.65.74.years + prc.75plus.years + afric
an.american + asian +      hispanic.or.latino + physician.rate * african.american +      physician.rate * h
ispanic.or.latino + income * physician.rate
                                  Value Std.Error   DF    t-value p-value
(Intercept)                    36.79257  3.553184 3043  10.354817  0.0000
physician.rate                 -0.00043  0.000236 3043  -1.810101  0.0704
income                         -0.70097  0.039356 3043 -17.811231  0.0000
prc.5.14.years                 -0.44548  0.055303 3043  -8.055157  0.0000
prc.15.24.years                -0.39574  0.036103 3043 -10.961313  0.0000
prc.25.34.years                -0.38688  0.043614 3043  -8.870584  0.0000
prc.35.44.years                -0.33072  0.039150 3043  -8.447644  0.0000
prc.45.54.years                -0.35441  0.042791 3043  -8.282449  0.0000
prc.55.64.years                -0.15201  0.047168 3043  -3.222704  0.0013
prc.65.74.years                -0.26820  0.050008 3043  -5.363207  0.0000
prc.75plus.years                0.39235  0.043165 3043   9.089682  0.0000
african.american                0.01682  0.002364 3043   7.115108  0.0000
asian                          -0.05235  0.011271 3043  -4.644898  0.0000
hispanic.or.latino             -0.04246  0.003315 3043 -12.810497  0.0000
physician.rate:african.american  0.00001  0.000005 3043   2.968693  0.0030
physician.rate:hispanic.or.latino  0.00004  0.000016 3043   2.327059  0.0200
physician.rate:income           0.00040  0.000070 3043   5.737000  0.0000
 Correlation: 
                                  (Intr) physc. income p.5.14 p.15.2 p.25.3 p.35.4 p.45.5 p.55.6 p.65.7
physician.rate                    -0.052                                                               
income                            -0.151 -0.156                                                        
prc.5.14.years                    -0.972  0.051  0.139                                                 
prc.15.24.years                   -0.979  0.013  0.170  0.939                                          
prc.25.34.years                   -0.941  0.008  0.176  0.922  0.886                                   
prc.35.44.years                   -0.792  0.113  0.111  0.724  0.818  0.594                            
prc.45.54.years                   -0.789 -0.016  0.034  0.754  0.748  0.799  0.413                     
prc.55.64.years                   -0.730  0.052  0.091  0.708  0.722  0.682  0.617  0.297              
prc.65.74.years                   -0.697  0.069  0.241  0.672  0.685  0.649  0.551  0.643  0.210       
prc.75plus.years                  -0.850  0.003  0.112  0.832  0.826  0.816  0.691  0.624  0.684  0.313
african.american                  -0.200  0.224  0.194  0.168  0.177  0.178  0.201  0.053  0.229  0.227
asian                             -0.049 -0.070 -0.140  0.083  0.020  0.067 -0.091  0.070  0.061  0.065
hispanic.or.latino                -0.134  0.146  0.151  0.065  0.123  0.137  0.081  0.145  0.107  0.135
physician.rate:african.american    0.067 -0.899  0.001 -0.059 -0.034 -0.027 -0.124  0.010 -0.068 -0.094
physician.rate:hispanic.or.latino -0.133 -0.226  0.009  0.119  0.147  0.098  0.135  0.113  0.109  0.089
physician.rate:income              0.067 -0.199 -0.582 -0.035 -0.057 -0.077 -0.098  0.024 -0.083 -0.118
                                  pr.75. afrcn. asian  hspn.. phy.:. ph.:..
physician.rate                                                             
income                                                                     
prc.5.14.years                                                             
prc.15.24.years                                                            
prc.25.34.years                                                            
prc.35.44.years                                                            
prc.45.54.years                                                            
prc.55.64.years                                                            
prc.65.74.years                                                            
prc.75plus.years                                                           
african.american                   0.092                                   
asian                              0.018 -0.026                            
hispanic.or.latino                 0.088  0.179 -0.008                     
physician.rate:african.american   -0.013 -0.296  0.067 -0.063              
physician.rate:hispanic.or.latino  0.097 -0.043 -0.103 -0.507  0.010       
physician.rate:income             -0.030 -0.240 -0.004  0.054  0.422 -0.362

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max 
-7.42192813 -0.54444122 -0.01696147  0.51369613  5.25140997 

Number of Observations: 3110
Number of Groups: 51 
```