Theroux: Sunday lunch, after church, to which I was invited on the spur of the moment by Mr. and Mrs. H.B. Williams in Monroeville, Alabama—me, a perfect stranger from the distant North, to share chicken, mac and cheese, sweet potatoes, and pie. I recall the peacefulness of it, the good humor, the great food, the hospitality. It seemed to characterize all the hospitality and candor I met in the South.

Biographical Sketch

Michael Ray Taylor is a frequent reviewer and literary writer for Chapter16.org, the website of Humanities Tennessee, a nonprofit agency sponsoring literary events and providing free reviews to Tennessee newspapers. He is a co-author, with Randy Duncan and David Stoddard, of *Creating Comics as Journalism, Memoir and Nonfiction*, a textbook published by Routledge Press in 2015.

Arkansas Lottery Spending

Michael Lloyd
Professor of Mathematics

Abstract. The Arkansas Lottery per capita spending by county is surprisingly large. It will be regressed with demographic data to see how much of the spending can be explained and also which variables appear to be important.

Introduction

Only data from 2014 was used because the latest census demographic data was only for the years 2010 and 2014. Here is a boxplot of the per capita spending:

Arkansas Counties 2014

Lottery Sales per Capita

http://www.arcounties.org/counties/
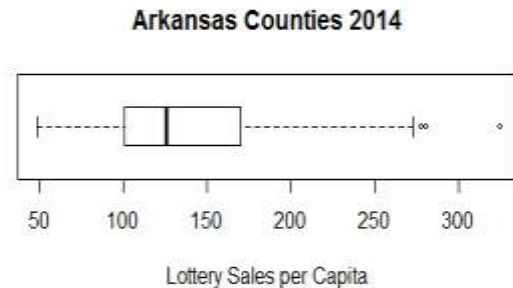
The smallest spending per person was $48 in Montgomery County, which has been attributed to a lack of ticket sellers.[3] The median spending was $125 per person, and the three outliers were Arkansas ($324/person), Nevada ($281/person), and Conway ($280/person) counties. Nevada is a poorer county, but the high ticket sales have been attributed to interstate traffic between Hot Springs and Shreveport.[3] The poverty rate in Nevada County was 25%, which was in the top 20% of all poorer Arkansas counties.
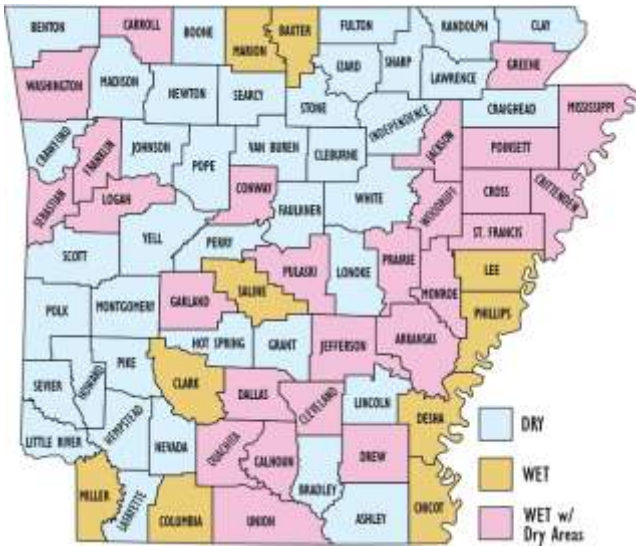
Here were the forty-six predictor variables obtained from the census website in the order of decreasing coefficient of determination:

IndianPctJul14, BlackPctJul14, WhitePctJul14, WhiteNotHispanicPctJul14, TotalRetailSalesPerCapita, NoHealthInsurancePct, FemalePctJul14, MeanTravelTimeWork, MedianMonthlyOwnerCostsWithoutMortgage, TwoRacesPctJul14, OwnerOccupiedHousingPct, PersonsPerHousehold, PopPctChg, MedianValueOwnerHousing, PopPctUnder5Jul14, ForeignersPct, HispanicPctJul14, AsianPctJul14, NonEnglishSpeakerPct, PovertyPct, LandArea, PacificPctJul14, MedianHouseholdIncome, BuildingPermits, MedianMonthlyOwnerCostsWithMortgage, PopPctUnder18Jul14, PopPctOver64Jul14, LaborForceTotalPct, LaborForceFemalePct, TotalNonemployerExtablishments, PopulationDensity, PopEstJul14, TotalEmploymentPctChange, Veterans, Households, HousingUnitsJul14, TotalAnnualPayroll, MedianGrossRent, LivingSameHousePct, HighSchoolOrHigherPct, BSorHigherPct, DisabilityPct, TotalRetailSales, PerCapitaIncome, TotalEmployers, TotalEmployment

The Arkansas Democratic Gazette article indicated that truckers may buy a large number of lottery tickets, so an Interstate variable was added. This variable is True if and only if a county contained one of the following interstates: I-30, I-40, I-540, I-530, I-55.

http://www.sitesatlas.com/Flash/USCan/static/ARFH.htm

http://www.encyclopediaofarkansas.net/media/gallery/Map/wet_dry_map_f.gif

Also, a preliminary study indicated LotterySalesPerCapita outliers on the edge of the state, so the following variables were also included: East, West, North, South. The variable Wet was added, which indicates if a county is wet or dry. So far, there is a total of 52 predictor variables.

Applying a logit transformation to BlackPctJul2014 increased its coefficient of determination with LotterySalesPerCapita from 0.201 to 0.250. Please see the accompanying graph. Similarly, logit transformations were applied to the following: IndianPctJul14, WhitePctJul14, WhiteNotHispanicPctJul14, OwnerOccupiedHousingPct.
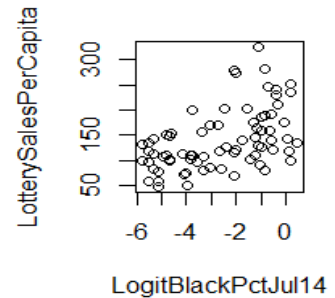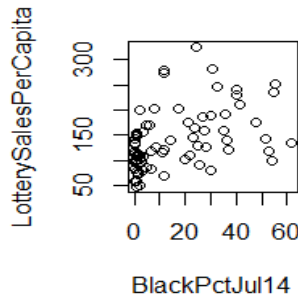
Here are plots for the response variable and these eight predictor variables:
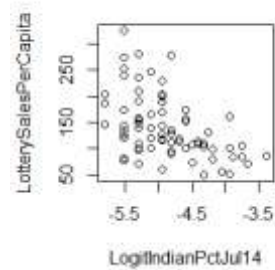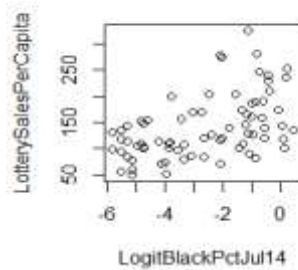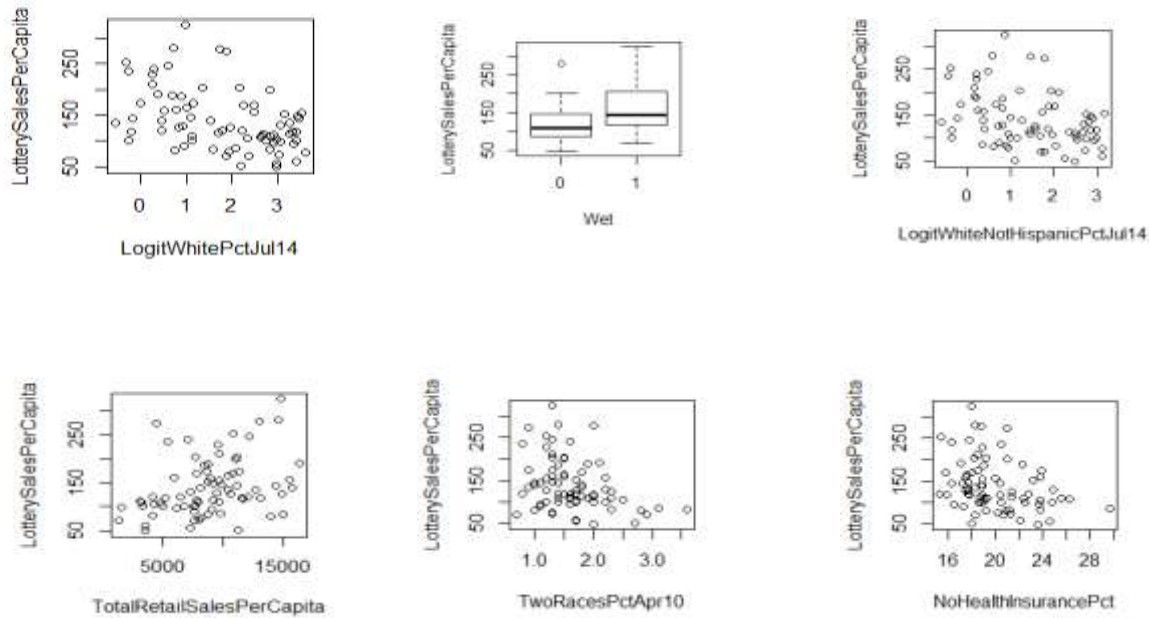
Here were the variables with a coefficient of determinations at least 0.100.

Race appears to be relatively important because they have the largest r^2 values. We do not know why per capita lottery spending would be negatively correlated with the percent of native Americans. The NoHealthInsurancePct correlation being negative indicates that a higher fraction of the population not having health insurance is associated with more lottery spending.

Variable	r	r^2
LogitBlackPctJul14	0.500	0.250
LogitIndianPctJul14	-0.485	0.235
LogitWhitePctJul14	-0.450	0.202
Wet	0.391	0.153
LogitWhiteNotHispanicPctJul14	-0.374	0.140
TotalRetailSalesPerCapita	0.351	0.123
TwoRacesPctApr10	-0.337	0.114
NoHealthInsurancePct	-0.327	0.107

Best Predictive Model

Backwards elimination based on the Akaike Information Criterion (AIC) selected 31 predictor variables with an R^2 of 86% and an adjusted R^2 of 76%:

LotterySalesPerCapita ~ LogitIndianPctJul14 + LogitBlackPctJul14 + LogitWhitePctJul14 + TotalRetailSalesPerCapita + NoHealthInsurancePct + FemalePctJul14 + MedianMonthlyOwnerCostsWithoutMortgage + TwoRacesPctJul14 + PopPctUnder5Jul14 + HispanicPctJul14 + NonEnglishSpeakerPct + MedianHouseholdIncome + MedianMonthlyOwnerCostsWithMortgage + PopPctUnder18Jul14 + LaborForceTotalPct + LaborForceFemalePct + PopulationDensity + Veterans + Households + HousingUnitsJul14 + TotalAnnualPayroll + HighSchoolOrHigherPct + BSorHigherPct + DisabilityPct + PerCapitaIncome + Interstate + East + West + North + South + Wet
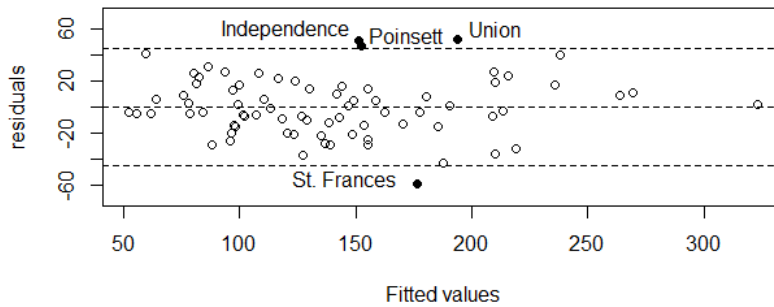
The dashed lines are two standard deviation units away from zero. There were four suspected outliers: Independence, Poinsett, St. Francis, and Union Counties.

The residuals were approximately normally distributed, so with 75 counties, we would expect about 3.4 counties with absolute residuals of more than two standard deviations.

Histogram of residuals
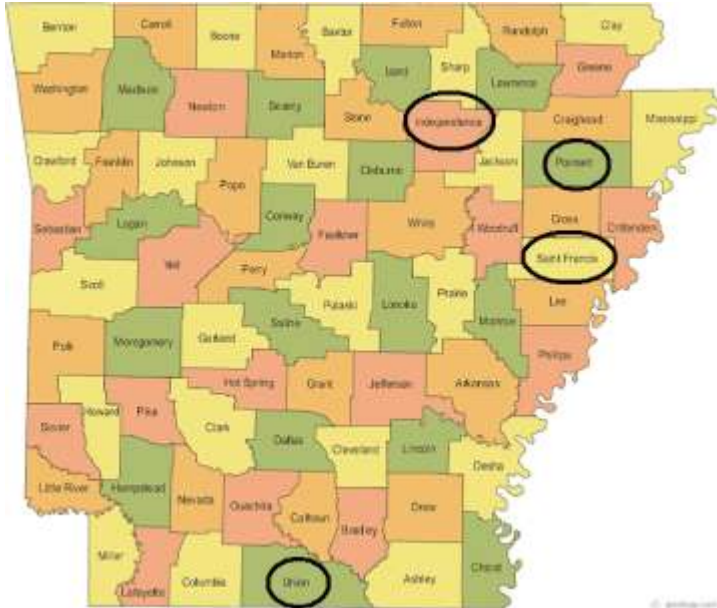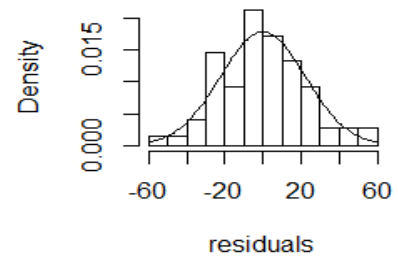
Here is a map of the counties with outliers based on the AIC model.

http://www.arcounties.org/counties/

Best Model with Significant Coefficients

Here is a model with 12 variables where all the variables are significant. It was obtained by manually doing trial and error using forward selection and backwards elimination. The AIC statistics was 769, R^2 was 67%, and the adjusted R^2 was 63%:

```
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                          -1.585e+02  1.656e+02  -0.957  0.34218
LogitIndianPctJul14                  -6.359e+01  1.219e+01  -5.218 2.22e-06 ***
LogitWhitePctJul14                   -1.830e+01  5.666e+00  -3.229  0.00199 **
TotalRetailSalesPerCapita             5.716e-03  1.744e-03   3.278  0.00172 **
NoHealthInsurancePct                  6.526e+00  2.756e+00   2.368  0.02102 *
LogitOwnerOccupiedHousingPct          7.976e+01  2.365e+01   3.373  0.00129 **
PersonsPerHousehold                  -1.353e+02  5.023e+01  -2.693  0.00909 **
PopPctUnder5Jul14                     2.863e+01  8.869e+00   3.229  0.00199 **
LandArea                              8.405e-02  3.681e-02   2.284  0.02583 *
MedianHouseholdIncome                 4.062e-03  1.646e-03   2.468  0.01636 *
MedianMonthlyOwnerCostsWithMortgage  -3.239e-01  7.542e-02  -4.294 6.28e-05 ***
Interstate                            3.116e+01  1.188e+01   2.624  0.01093 *
Wet                                   2.653e+01  1.132e+01   2.345  0.02226 *
```

The Arkansas Democrat Gazette indicated that there might be a positive correlation between ticket sales per person and the poverty and black percentages.[3] (The negative coefficient for the white percentage is consistent with a negative correlated black percentage.) Counties containing an interstate spend, on average, about $31 more per person on the lottery than those not containing an interstate. Wet counties spend, on average, about $27 more per person on the lottery than dry counties. It is left to the reader to ponder the sign of some of the other coefficients. For example, for each additional square mile, on average, about 8 cents more per person is spent on the lottery. Why do counties with a higher percentage of children under the age of 5 years tend to spend more on the lottery?

Here is a residual plot versus the fitted values for this last model. The dashed lines are two standard deviation units away from zero. The suspected outliers were Ashley, Conway, and Randolph counties:
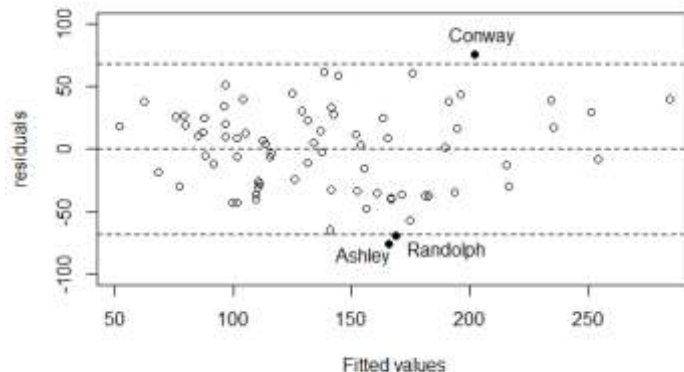
Histogram of residuals
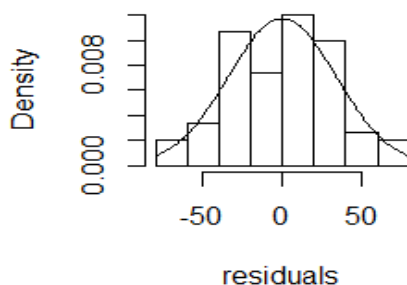
The distribution of residuals was approximately normal, so, again, it was reasonable that there were three outliers.

These counties with large residuals have been circled:

http://www.arcounties.org/counties/

Acknowledgement

References

1. Arkansas Scholarship Lottery Sales by Month and County (http://www.pulseofconway.com/lotterypage.php).
2. Demographic data (http://www.census.gov).
3. "Poorer counties top lottery lists," Arkansas Democratic Gazette, August 14, 2011.

Biographical Sketch

Michael Lloyd graduated cum laude and in the honors program in Chemical Engineering with a B.S. in 1984. He accepted a position at Henderson State University in 1993 shortly after earning his Ph.D. in Mathematics (Probability Theory) from Kansas State University. He has presented papers at meetings of the Academy of Economics and Finance, the American Mathematical Society, the Arkansas Conference on Teaching, and the Southwest Arkansas Council of Teachers of Mathematics. He has been an active member of the Mathematical Association of America since 1993, earned 18 hours in computer science, and has been an Advanced Placement statistics consultant since 2002.