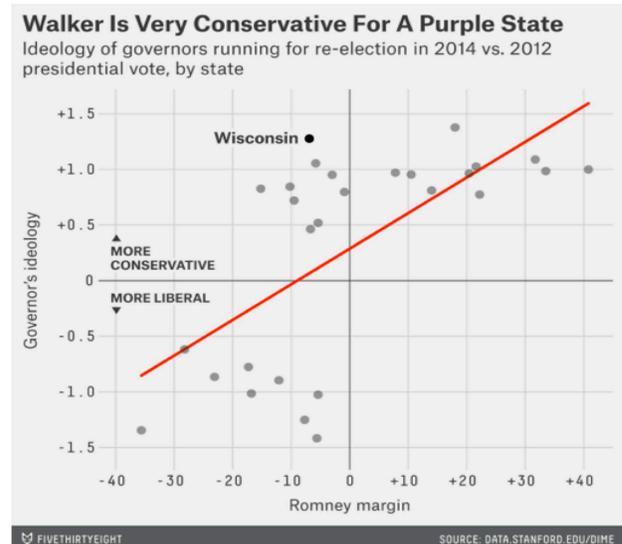
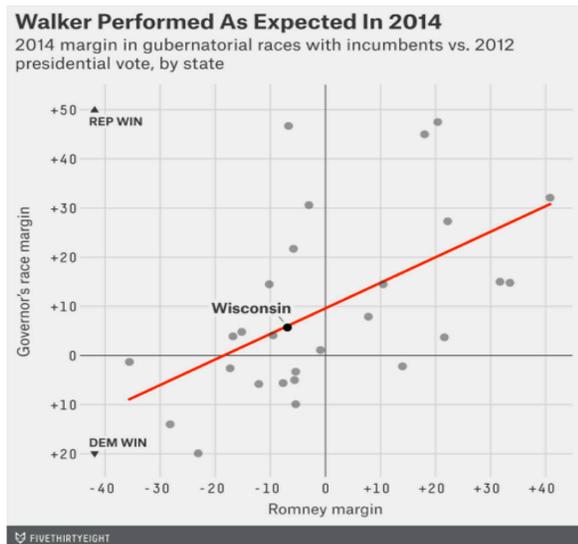


American University, Department of Government

Ph.D. Methodology Qualifying Exam, Summer 2018, 4 Hours, Answer All Questions

1. LINEAR MODEL APPLICATION.

For each of the regression plots below identify a major problem and give a solution for this problem.



2. LINEAR MODEL ASSUMPTIONS.

Give the 5 Gauss-Markov Assumptions that underlie the linear regression model.

3. LITERATURE ANALYSIS.

Read the attached article: "Heightening Comparativists' Concern for Model Choice: Voting Behavior in Great Britain and the Netherlands" by Whitten and Palmer

(<http://links.jstor.org/sici?sici=0092-5853%28199602%2940%3A1%3C231%3AHCCFMC%3E2.0.CO%3B2-F>)

and answer the following questions:

- What is the authors' main methodological point about model choice?
- Is it correct? Specifically argue why it is or is not.
- What do their empirical examples show?
- Suppose that you were reviewing this article for a leading political science journal. What specific criticisms would you have? Be specific and defend your claims.

4. LOGISTIC REGRESSION.

The following is the result from a logistic regression of church attendance per week on support for abortion rights (1 = support abortion for any reason) from the 2006 GSS and some summary statistics on the two variables involved.

```

                coef.est coef.se
(Intercept)  0.37      0.12
attend      -0.22      0.03
---
Log-Likelihood = -485.6831
n = 769, k = 2

> summary(D$abort)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000 0.0000  0.0000  0.4018  1.0000  1.0000
> table(D$abort)
 0   1
460 309

> summary(D$attend)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000  1.000  3.000  3.663  7.000  8.000
> table(D$attend)
 0  1  2  3  4  5  6  7  8
167 49 99 70 67 76 29 158 54

```

- (a) Comment on the statistical significance of the parameter estimates.
- (b) What is the maximal percentage change in $p(y = 1)$ for a unit change in attend? (approximations welcome)
- (c) Find the scaled Deviance, AIC, and BIC.
- (d) Compute z -values for the intercept and attend; how would you get p values if you had a computer?
- (e) Set up a predicted probability (you don't have a computer so you do not have to produce an actual value) for a "central" value of attend. Discuss what might be meant by "central."

5. NONLINEAR TESTING.

Fully describe the Wald test, the likelihood ratio test, and the Lagrange multiplier test. Explain how they are different, how they are the same, and the advantages of each one. Illustrate with a graph.

6. PROBABILITY AND COUNTING.

You are trying to form a coalition cabinet with three parties. There are six senior members of the Liberal Party, five senior members of the Christian Democratic Party, and four senior members of the Green Party

ying for positions in the cabinet. How many ways could you choose a cabinet composed of three Liberals, two Christian Democrats, and three Greens?

7. SOFTWARE.

What does the following R code do?

```
mi <- function(data.mat) {  
  for (i in 1:ncol(data.mat)) {  
    if (sum(is.na(data.mat[,i])) > 0) {  
      print(paste("column", i, "has missing data"))  
      mean.col <- mean(data.mat[,i], na.rm = TRUE)  
      for (j in 1:nrow(data.mat)) {  
        if (is.na(data.mat[j,i]) == TRUE)  
          data.mat[j,i] <- mean.col  
      }  
    }  
  }  
  return(data.mat)  
}
```

8. SHORT ANSWER QUESTIONS. One sentence only.

- (a) Define covariation.
- (b) Delineate between an empirical and a normative theory.
- (c) What is “inference”?
- (d) What is a random variable?
- (e) What is multicollinearity in empirical models?
- (f) What is a unit of analysis?

9. **PROBABILITY PREDICTIONS.** A public agency is investigating a short-term, high-interest “payday” lender. Commonly, payday loans are not repaid at the end of their terms, but instead are rolled over into new loans with the same lender. These new loans are often larger than the original loans, since they include the original principal and some of the interest as well. These rollovers can lead to a spiral of increasing burdens on the borrower. The agency is interested in the true underlying probability π that a new loan gets rolled over. They collect data on 50 new loans that were to end in a given week, and they record that 38 of these loans were rolled over.

- (a) Calculate an estimate of π , the probability of a rollover, to two decimal places.
- (b) What *estimator* did you use to estimate π , the parameter of interest?
- (c) Calculate the standard error around your estimate, to two decimal places.

10. **CROSS-TABULATION.** Consider the following R code and resulting analysis:

```
> data(Titanic)
> titanic <- as.data.frame(Titanic)
> xtabs(Freq ~ Sex + Survived, data = titanic)
```

	Survived	
Sex	No	Yes
Male	1364	367
Female	126	344

```
> summary(xtabs(Freq ~ Sex + Survived, data = titanic))
```

```
Call: xtabs(formula = Freq ~ Sex + Survived, data = titanic)
```

```
Number of cases in table: 2201
```

```
Number of factors: 2
```

```
Test for independence of all factors:
```

```
Chisq = 456.9, df = 1, p-value = 2.302e-101
```

Use the evidence here to decide whether there is a male/female effect in survival (perhaps a chivalry test?). State your hypothesis, give the exact steps for testing this hypothesis, make a decision, and explain the evidence for your decision.