

Note: In providing this answer guide we've interleaved our R code. We normally don't do this, because it creates the false impression among students that this class is about demonstrating proficiency with a software package, rather than learning the underlying statistical concepts and how they relate to social-science problems. Nonetheless, we think you will find it helpful to see how we actually got the answers that we report.

1. Verify that $\hat{\theta}$ maximizes the likelihood function $\mathcal{L}(\theta; y)$ if and only if it also maximizes the log of the likelihood function, $\ln\mathcal{L}(\theta; y)$.

Answer: I'll show in general that maximizing the logarithm of a function is equivalent to maximizing the function itself; thus maximizing likelihood is equivalent to maximizing log-likelihood.

Let $g(x)$ be an increasing function ($\ln x$ has this property): $x > y \Leftrightarrow g(x) > g(y)$. Let $f(x)$ be some function with maximum at $x = x^*$. By definition $f(x^*) \geq f(x) \forall x$. Since $g(x)$ is increasing, $f(x^*) \geq f(x) \Leftrightarrow g(f(x^*)) \geq g(f(x))$. Then, $g(f(x))$ has maximum at $x = x^*$. To complete the proof, note that if we instead assume that $g(f(x))$ has maximum at $x = x^*$, so does $f(x)$ since $g(f(x^*)) \geq g(f(x)) \Rightarrow f(x^*) \geq f(x)$.

A less general (but for these purposes perfectly fine) proof uses the familiar differentiation approach to maximization. If we can assume differentiability of the likelihood function, then the problem is relatively trivial, relying again on the monotonicity of the logarithmic function. If $\hat{\theta}$ maximizes $\mathcal{L}(\theta)$ then

$$\left. \frac{\partial \mathcal{L}(\theta)}{\partial \theta} \right|_{\hat{\theta}} = 0. \quad (1)$$

Consider $\ln\mathcal{L}(\theta)$:

$$\left. \frac{\partial \ln\mathcal{L}(\theta)}{\partial \theta} \right|_{\hat{\theta}} = \frac{\partial \ln\mathcal{L}}{\partial \mathcal{L}} \left. \frac{\partial \mathcal{L}}{\partial \theta} \right|_{\hat{\theta}} = \frac{1}{\mathcal{L}(\theta)} \left. \frac{\partial \mathcal{L}}{\partial \theta} \right|_{\hat{\theta}} \quad (2)$$

By equation 1, the 2nd term is zero, and so provided $\mathcal{L}(\hat{\theta}) \neq 0$, we see that

$$\left. \frac{\partial \mathcal{L}(\theta)}{\partial \theta} \right|_{\hat{\theta}} = 0 \Rightarrow \left. \frac{\partial \ln\mathcal{L}(\theta)}{\partial \theta} \right|_{\hat{\theta}} = 0,$$

i.e., $\hat{\theta}$ maximizing the log-likelihood function is sufficient for showing that $\hat{\theta}$ maximizes the likelihood function. We now consider necessity. Suppose

$$\left. \frac{\partial \ln\mathcal{L}(\theta)}{\partial \theta} \right|_{\hat{\theta}} = 0.$$

From equation 2 we see that the only way that this assertion can be true is if

- $1/\mathcal{L}(\hat{\theta}) = 0 \iff \mathcal{L}(\hat{\theta}) = \infty$ or
- $\partial \mathcal{L} / \partial \theta|_{\hat{\theta}} = 0$, i.e., $\hat{\theta}$ maximizes $\mathcal{L}(\theta)$.

Ruling out cases that generate infinite likelihood, then we have established necessity.

2. A five member committee votes 3-2 in favor of a proposal. Assume voting is independent. Let π be the probability that a committee member votes for the proposal.

(a) We have no information with which to distinguish committee members (in the language of Bayesian statistics, we'd say that the committee members are *exchangeable*, but I digress). What is the maximum likelihood estimate (MLE) of π , the probability that any particular committee member votes for the proposal?

Answer: The MLE of π , $\hat{\pi} = 3/5 = .6$.

(b) What is the log-likelihood of $\pi = .5$? Compare this value of the log-likelihood function with that attained at the MLE with a likelihood ratio test. What does this say about the plausibility of $H_0 : \pi = .5$?

Answer: The likelihood function is just the binomial density

$$p(r; \pi, n) = \binom{n}{r} \pi^r (1 - \pi)^{n-r}$$

In this case we have $r = 3$, $n = 5$ and $\pi = .5$. The likelihood is .3125 and the log-likelihood is -1.16: i.e.,

```
1 > dbinom(3, 5, 0.5)
[1] 0.3125
```

```
1 > b0 <- dbinom(3, 5, 0.5, log = TRUE)
2 > b0
[1] -1.163151
```

```
1 > b1 <- dbinom(3, 5, 0.6, log = TRUE)
2 > b1
[1] -1.062473
```

```
1 > d <- 2 * (b1 - b0)
2 > d
[1] 0.2013551
```

```
1 > qchisq(0.95, df = 1)
[1] 3.841459
```

The log-likelihood evaluated at the MLE of $\hat{\pi} = .6$ is -1.06. Under the null hypothesis that $\pi = .5$, twice the difference in log-likelihoods is (asymptotically!) distributed χ^2 with one degree of freedom. In this case, the test statistic \mathcal{D} is .20, and is far smaller than the .95 quantile of the χ^2_1 distribution (3.84). We just don't have enough information to distinguish $\pi = .5$ from the MLE of .6.

(c) How would your conclusion about the plausibility of $H_0 : \pi = .5$ change if we observed

i. a 10 person committee splitting 6-4 in favor of the proposal?

```
1 > dbinom(6, 10, 0.5)
```

```
[1] 0.2050781
```

```
1 > b0 <- dbinom(6, 10, 0.5, log = TRUE)
2 > b1 <- dbinom(6, 10, 0.6, log = TRUE)
3 > 2 * (b1 - b0)
```

```
[1] 0.4027103
```

Answer: $\mathcal{D} = .40$, still far below the conventional critical value of 3.84.

ii. a 50 person assembly splitting 30-20 in favor of the proposal?

```
1 > dbinom(30, 50, 0.5)
```

```
[1] 0.04185915
```

```
1 > b0 <- dbinom(30, 50, 0.5, log = TRUE)
2 > b1 <- dbinom(30, 50, 0.6, log = TRUE)
3 > 2 * (b1 - b0)
```

```
[1] 2.013551
```

Answer: $\mathcal{D} = 2.01$, again, still below the conventional critical value.

i.e., what is happening to the likelihood function and/or the log-likelihood function in these cases relative to the case of a five person committee? In particular, what is happening to the 2nd derivative of the log-likelihood function in the neighborhood of the MLE?

Answer: The likelihood function (or log-likelihood function) is getting increasingly concentrated around the MLE, which (by design) is .6 in each case. This means we estimate the MLE more precisely, which is revealed clearly if we inspect the 2nd derivative of the log-likelihood function.

The binomial likelihood for r successes in n independent trials each with success probability π is

$$\mathcal{L}(\pi; r, n) \equiv p(r; \pi, n) = \binom{n}{r} \pi^r (1 - \pi)^{n-r}$$

with log-likelihood

$$\ln \mathcal{L}(\pi; r, n) \propto r \ln \pi + (n - r) \ln(1 - \pi),$$

i.e., terms not involving π are absorbed into the constant of proportionality. Differentiating wrt π :

$$\frac{\partial \ln \mathcal{L}}{\partial \pi} = \frac{r}{\pi} - \frac{n - r}{1 - \pi}$$

and the second derivative is

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \pi^2} = -\frac{r}{\pi^2} - \frac{n - r}{(1 - \pi)^2}.$$

Now evaluate the 2nd derivative at the MLE, i.e., setting $\pi = \hat{\pi} = r/n$. Note that at the MLE we have $r = n\hat{\pi}$ and $n - r = n(1 - \hat{\pi})$. Thus

$$\left. \frac{\partial^2 \ln \mathcal{L}}{\partial \pi^2} \right|_{\hat{\pi}} = -\frac{n}{\hat{\pi}} \cdot \frac{n}{(1 - \hat{\pi})} = -n \left(\frac{1}{\hat{\pi}} + \frac{1}{1 - \hat{\pi}} \right) = -n \left(\frac{1 - \hat{\pi} + \hat{\pi}}{\hat{\pi}(1 - \hat{\pi})} \right) = \frac{-n}{\hat{\pi}(1 - \hat{\pi})},$$

which is increasing in magnitude as function of n . In other words, the log-likelihood function is a concave function that is getting steeper around its maximum at the MLE as a function of n . Since the variance of the MLE is (asymptotically) minus the inverse of the 2nd derivative, i.e.,

$$V(\hat{\pi}) = \frac{\hat{\pi}(1 - \hat{\pi})}{n},$$

we also obtain the familiar result that the variance of the MLE is decreasing as a function of n . See Figure 1.

```

R Code
1 > xseq <- seq(.05, .95, by=.01)
2 > r <- c(3,6,30)
3 > n <- c(5,10,50)
4 > like <- matrix(NA, length(xseq), 3)
5 > logLike <- like
6 > for (j in 1:3){
7 +   like[,j] <- dbinom(x=r[j], size=n[j], prob=xseq)
8 +   logLike[,j] <- dbinom(x=r[j], size=n[j],
9 +                       prob=xseq, log=TRUE)
10 + }
11 > plot(x=c(0,1),
12 +      y=range(like),
13 +      xlab=expression(pi),
14 +      ylab="Likelihood",
15 +      axes=FALSE,
16 +      type="n")
17 > axis(1)
18 > lines(xseq, like[,1], col="red", lwd=2)
19 > lines(xseq, like[,2], col="blue", lwd=2)
20 > lines(xseq, like[,3], col="black", lwd=2)
21 > legend(x="topleft",
22 +       bty="n",
23 +       col=c("red", "blue", "black"),
24 +       lwd=2,
25 +       legend=paste("n=", n, sep=" "))
26 > plot(x=c(0,1),
27 +      y=range(logLike),
28 +      xlab=expression(pi),
29 +      ylab="Log-Likelihood",
30 +      axes=FALSE,
31 +      type="n")
32 > axis(1)
33 > lines(xseq, logLike[,1], col="red", lwd=2)
34 > lines(xseq, logLike[,2], col="blue", lwd=2)
35 > lines(xseq, logLike[,3], col="black", lwd=2)
36 > legend(x="center",
37 +       y=.5,
38 +       bty="n",
39 +       col=c("red", "blue", "black"),
40 +       lwd=2,
41 +       legend=paste("n=", n, sep=" "))

```

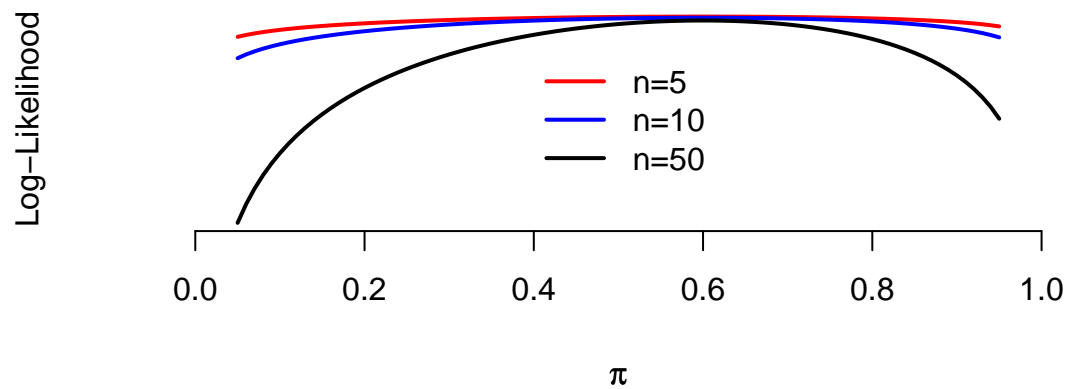
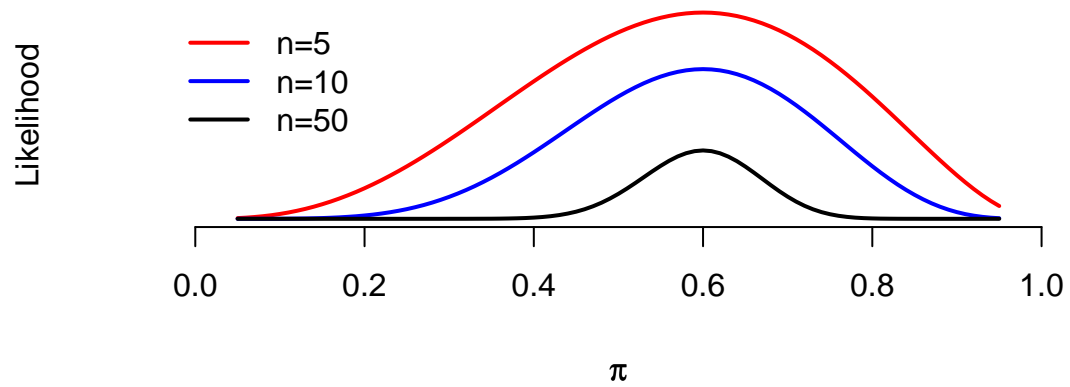


Figure 1: Binomial Likelihood and Log-Likelihood Functions with $r = .6n$ successes.

3. Download the data on Al Gore's share of the 2-party vote for president by congressional district in the 2000 election, [available from my web site](#). Denote this variable y_i . Assume (perhaps implausibly), $y_i \stackrel{iid}{\sim} N(\mu, \sigma^2)$.

- (a) What are the MLEs of μ and σ^2 ? **Answer:** $\hat{\mu} = .5150$ and $\hat{\sigma}^2 = .0214$. Note that the MLE is the sum of squared deviations around the mean divided by n , not $n - 1$. (That is, the MLE is biased.)

```

R Code
1 > data <- read.table(file="pvote00.asc",header=TRUE,sep=",")
2 > attach(data)
3 > n <- length(pvote00)
4 > mles <- c(mean(pvote00),
5 +           var(pvote00)*(n-1)/n)
6 > mles

```

```
[1] 0.51498851 0.02137810
```

- (b) Use graphical means to verify that your estimate of σ^2 is in fact a MLE. That is, over a grid of values for σ^2 that includes the MLE of σ^2 , compute the *conditional* MLE for μ and note the value of the conditional log-likelihood function at the MLE of μ (conditional, in that for the purposes of this exercise, we are treating each assumed value for σ^2 on the grid as a known constant, instead of a parameter to be estimated). Plot the values of the log-likelihood against the assumed grid of values for σ^2 . Comment briefly on the shape of this *profiled* log-likelihood in the neighborhood of the MLE for σ^2 . Hint: in R, you may need to write a loop over the grid of assumed values for σ^2 , and you'll also find the `dnorm` function helpful.

Some fancy footwork in R produces the picture we need, with the fancy part coming in with the code we use to find the exact width of a 95% confidence interval for σ^2 .

```

R Code
1 > ## a function for evaluating the log of a
2 > ## normal likelihood
3 > profilellhfunc <- function(mu,sigma,y){
4 +   sum(dnorm(y,mean=mu,sd=sigma,log=TRUE))
5 + }
6 > ## the maximum of the log-likelihood function
7 > ## for these data
8 > maxLogLike <- profilellhfunc(mu=mles[1],
9 +                             sigma=sqrt(mles[2]),
10 +                             y=pvote00)
11 > ## how much log-likelihood we could lose
12 > ## before rejecting a null hypothesis
13 > LogLikeCrit <- maxLogLike - 2*qchisq(.95,df=1)
14 > objFunc <- function(sigma,mu,y,k){
15 +   profilellhfunc(mu=mu,sigma=sqrt(sigma),y=y)-k
16 + }
17 > ## solve for the values of sigma^2 that generate the
18 > ## critical value 'rejection' value of the logLike
19 > bounds <- rep(NA,2)
20 > bounds[1] <- uniroot(objFunc,
21 +                     lower=.015,
22 +                     upper=.020,
23 +                     mu=mles[1],
24 +                     y=pvote00,
25 +                     k=LogLikeCrit,
26 +                     tol=.Machine$double.eps^(3/4))$root
27 > bounds[2] <- uniroot(objFunc,

```

```

28 +             lower=.025,
29 +             upper=.030,
30 +             mu=mles[1],
31 +             y=pvote00,
32 +             k=LogLikeCrit,
33 +             tol=.Machine$double.eps^(3/4))$root
34 > ## now evaluate the log-likelihood over a grid that includes these bounds
35 > sigmaSeq <- seq(bounds[1]-.002,
36 +               bounds[2]+.003,
37 +               length=501)
38 > m <- length(sigmaSeq)
39 > llh <- rep(NA,m) ## store log-likelihood in this
40 > ## loop over the grid
41 > for(j in 1:m){
42 +   llh[j] <- profilellhfunc(mu=mles[1], ## mean
43 +                           sigma=sqrt(sigmaSeq[j]), ## standard dev
44 +                           y=pvote00)
45 + }

```

Now we can do the actual plotting:

```

R Code
1 > par(las=1)
2 > plot(sigmaSeq,
3 +       llh,
4 +       ylim=c(min(llh),maxLogLike),
5 +       xlab=expression(sigma^2),
6 +       ylab="conditional log-likelihood",
7 +       type="l",
8 +       lwd=3,
9 +       axes=FALSE)
10 > axis(1)
11 > yTicks <- sort(c(pretty(llh,n=7),
12 +                 LogLikeCrit,maxLogLike))
13 > axis(2,
14 +       cex.axis=.65,
15 +       at=c(yTicks),
16 +       labels=as.character(round(yTicks,2)))
17 > abline(v=mles[2])
18 > abline(h=maxLogLike)
19 > abline(h=LogLikeCrit,lty=2)
20 > abline(v=bounds,lty=2)

```

See Figure 2. The log-likelihood is not symmetric around the MLE for σ^2 . This is an under-appreciated fact about log-likelihood functions for variance parameters. Note that we've also got a fairly large sample $n = 435$.

4. Consider the *Titanic* survival data. These data are part of R, or [available from my web site](#). The data set contains 2,201 observations (for each known passenger and crew member) on 4 variables:

- `class`: class of service, 0 = Crew, 1 = 1st class, 2 = 2nd class, 3 = 3rd class.
- `adult`: 1 if adult, 0 otherwise.
- `male`: 1 if male, 0 otherwise.
- `survive`: 1 if survive, 0 otherwise.

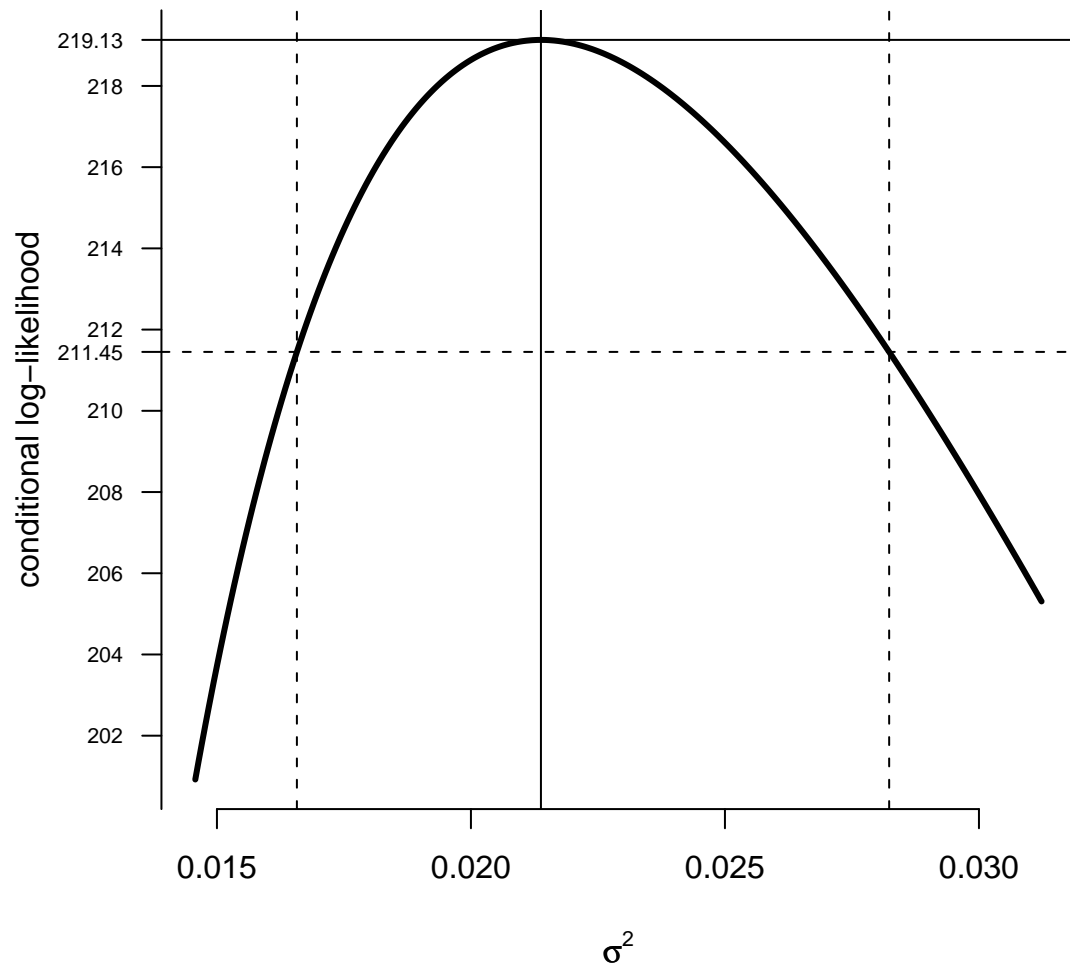


Figure 2: Profiled/Conditional Log-likelihood for Question 3.

- (a) Cross-tabulate survive and male. Test the hypothesis that there is no association between survival and gender. In R, use the `chisq.test` function.

Answer:

```
_____ R Code _____
1 > titanic <- read.table(file="titanic.dat",header=TRUE)
2 > tab <- table(titanic$survive,titanic$male)
3 > tab
```

```
      0      1
0  126 1364
1   344   367
```

```
_____ R Code _____
1 > chisq.test(tab)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: tab
X-squared = 454.4998, df = 1, p-value < 2.2e-16
```

- (b) Estimate the logistic regression of survive on male. In R, use the `glm` function, with the option, `family=binomial`. Use a z-test to test whether men were more or less likely to survive the disaster. Compare your results with the results from the cross-tabulation in the first part of this question.

Answer:

```
_____ R Code _____
1 > logit1 <- glm(survive ~ male,
2 +           data=titanic,
3 +           family=binomial)
4 > summary(logit1)
```

```
Call:
glm(formula = survive ~ male, family = binomial, data = titanic)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6226  -0.6903  -0.6903   0.7901   1.7613
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.0044     0.1041   9.645  <2e-16 ***
male        -2.3172     0.1196  -19.376 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 2769.5 on 2200 degrees of freedom
Residual deviance: 2335.0 on 2199 degrees of freedom
AIC: 2339
```

Number of Fisher Scoring iterations: 4

Note that we get the same result we obtained from the cross-tabulation: men were overwhelmingly less likely to survive the disaster than women.

- (c) Repeat the test you performed in the previous question, but via likelihood ratio test. That is, estimate a 2nd, restricted `glm` and compare it with the unrestricted model. In R, use the `anova` command, with the `test="Chi"` option. **Answer::**

```
R Code
```

```

1 > logit2 <- glm(survive ~ 1,
2 +             data=titanic,
3 +             family=binomial)
4 > anova(logit2,logit1,test="Chi")

```

Analysis of Deviance Table

```

Model 1: survive ~ 1
Model 2: survive ~ male
  Resid. Df Resid. Dev  Df Deviance P(>|Chi|)
1      2200    2769.46
2      2199    2334.99   1    434.47 1.731e-96

```

Observe that the χ^2 value is almost exactly the same as the value we obtained from the cross-tabulation. In either case we again see an overwhelmingly strong association between gender and survival (and in particular, that men were *far* less likely to survive the disaster than were women).

- Obtain the file `nagler.asc` from my website. This file contains 98,857 cases (welcome to large n research!) from the 1984 Current Population Survey, analyzed by Jonathan Nagler in two articles: “The Effects of Registration Laws and Education on Voter Turnout” *American Political Science Review*, 1991, 85:1393--1405; ”Scobit: an alternative estimator to logit and probit” *American Journal of Political Science*, 1994, 38:230--255.

The data in the file comprise the following variables (in column order):

- `turnout` 1 if the respondent reports turning out to vote in the 1984 presidential election, 0 otherwise.
- `educ` 1 for 0-4 yrs education; 2 for 5-7 yrs; 3 for 8 yrs; 4 for 9-11 yrs; 5 for 12 yrs; 6 for 1-3 yrs college; 7 for 4 yrs college; 8 for 5+ yrs college
- `age` age of respondent, in years
- `south` 1 if respondent line in the South, 0 otherwise.
- `govelec` 1 if a gubernatorial election coincided with the presidential election
- `closing` number of days before election day that voter registration closes.

```
R Code
```

```

1 > nagler <- scan(file="nagler.asc",
2 +             what=list(turnout=0,educ=0,age=0,south=0,
3 +             govelec=0,closing=0))

```

- Estimate the following model and provide a brief write-up of the parameter estimates (i.e., assess statistical significance and substantive implications) and the goodness-of-fit of the logistic regression model using `educ` and `age` and the square of each of these variables as predictors.

Answer:

```
R Code
```

```

1 > logit1 <- glm(turnout ~ educ + I(educ^2) + age + I(age^2),
2 +             data=nagler,
3 +             family=binomial)

```

We'll define a hitmiss function that will let us compute PCP (percent correctly predicted):

```

R Code
1 > ## first a function for PCP
2 > hitmiss <- function(glmobj,k=.5,print=TRUE){
3 +   ## check sane argument have been passed
4 +   if(!inherits(glmobj,"glm"))
5 +     stop("hitmiss only defined for objects of class glm\n")
6 +   if(family(glmobj)$family!="binomial")
7 +     stop("hitmiss only defined for glm objects with family=binomial\n")
8 +   y <- glmobj$y
9 +   if(!all(sort(unique(y))==c(0,1)))
10 +     stop("hitmiss only defined for glm objects with binary responses\n")
11 +   if(k<0 | k>1)
12 +     stop("classification threshold must be between zero and one\n")
13 +
14 +   p <- predict(glmobj,type="response")
15 +   tab <- matrix(NA,2,2)
16 +   tab[1,1] <- sum(p<k & y==0,na.rm=T)   ## true negative
17 +   tab[1,2] <- sum(p<k & y==1,na.rm=T)   ## false negative
18 +   tab[2,1] <- sum(p>k & y==0,na.rm=T)   ## false positive
19 +   tab[2,2] <- sum(p>k & y==1,na.rm=T)   ## true positive
20 +   pcp <- sum(diag(tab))/sum(tab) * 100   ## overall PCP measure
21 +   pcp0 <- tab[1,1]/sum(tab[1:2,1])*100   ## PCP | y = 0
22 +   pcp1 <- tab[2,2]/sum(tab[1:2,2])*100   ## PCP | y = 1
23 +
24 +   if(print){
25 +     dimnames(tab) <- list(c("yhat=0","yhat=1"),
26 +                           c("y=0","y=1"))
27 +     cat(paste("Classification Threshold =",k,"\n"))
28 +     cat("Classification Table:\n")
29 +     cat("-----\n")
30 +     print(tab)
31 +     cat("\n")
32 +     cat(paste("Percent Correctly Predicted = ",
33 +               format(pcp,digits=3,nsml=1),
34 +               "%\n",
35 +               sep=""))
36 +     cat(paste("Percent Correctly Predicted = ",
37 +               format(pcp0,digits=3,nsml=1),
38 +               "%, for y = 0\n",
39 +               sep=""))
40 +     cat(paste("Percent Correctly Predicted = ",
41 +               format(pcp1,digits=3,nsml=1),
42 +               "% for y = 1\n",
43 +               sep=""))
44 +     nullmodel <- max(c(sum(y==0),sum(y==1)))/sum(tab) * 100
45 +     cat(paste("Null Model Predicts = ",
46 +               format(nullmodel,digits=3,nsml=1),
47 +               "%\n",
48 +               sep=""))
49 +   }
50 +   out <- c(pcp=pcp,pcp0=pcp0,pcp1=pcp1)
51 +   out
52 + }

```

And we'll use the function:

```

R Code
1 > hitmiss(logit1)

```

```

Classification Threshold = 0.5
Classification Table:

```

```

-----
          y=0   y=1
yhat=0 11549  7799
yhat=1 21066 58443

Percent Correctly Predicted = 70.8%
Percent Correctly Predicted = 35.4%, for y = 0
Percent Correctly Predicted = 88.2% for y = 1
Null Model Predicts = 67.0%
      pcp    pcp0   pcp1
70.80126 35.41009 88.22650

```

We can get R to provide a nicely formatted table of the coefficients, using the `xtable` function:

```

R Code
-----
1 > library(xtable)
2 > foo <- summary(logit1)$coef
3 > dimnames(foo)[[1]][3] <- "educ$^2$"
4 > dimnames(foo)[[1]][5] <- "age$^2$"
5 > xtable(foo,digits=rep(4,5))

```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.4954	0.0786	-57.1715	0.0000
educ	0.2736	0.0247	11.0876	0.0000
educ ²	0.0254	0.0025	10.1360	0.0000
age	0.1129	0.0022	51.5682	0.0000
age ²	-0.0008	0.0000	-35.6501	0.0000

Write-up would come here. Goodness-of-fit.

- (b) How many unique predicted probabilities are produced by this model? Explain how you derived your answer.

Answer: There are 636 unique predicted probabilities from the model, one for each covariate class. Here is how we verify this in R:

```

R Code
-----
1 > length(unique(predict(logit1)))

```

[1] 636

```

R Code
-----
1 > ## convert each row of predictor matrix to a string,
2 > ## colon-delimited
3 > x <- update(logit1,x=TRUE)$x
4 > xString <- apply(x,1,paste,collapse=":")
5 > ## count unique strings
6 > length(unique(xString))

```

[1] 636

```

R Code
-----
1 > ## indices of first appearance of each covariate class
2 > ## we will use this later when extracting unique
3 > ## predicted probabilities
4 > indxUnique <- match(unique(xString),xString)

```

- (c) Compare the predicted probabilities from the logit model with the corresponding predicted probabilities from a probit model. How and why do they differ, if at all? Is there any statistical basis for preferring logit over probit or vice-versa?

Answer: The predicted probabilities are virtually identical, as Figure 3 makes clear.

```

R Code
1 > probit1 <- update(logit1,family=binomial(link=probit))
2 > plot(x=c(0,1),
3 +     y=c(0,1),
4 +     xaxs="i",yaxs="i",
5 +     type="l",
6 +     col="red",    ## thick red 45 degree line
7 +     lwd=4,
8 +     xlab="Predicted Probabilities, Logit Model",
9 +     ylab="Predicted Probabilities, Probit Model")
10 > points(x=predict(logit1,type="response")[indxUnique],
11 +        y=predict(probit1,type="response")[indxUnique],
12 +        cex=.5,
13 +        col=gray(.5))

```

There is also very little to distinguish the models in terms of goodness-of-fit: the log-likelihood of the logit model is -55564.84 while for the probit model we have -55607.27; note that this comparison of log-likelihoods is indicative only, since the logit and probit models do not “nest” in a way that would permit a likelihood ratio test.

- (d) Augment your logit model from the previous question with the following **additional** “contextual” predictors: south, govelec, and closing, and *interactions* between the two education variables (educ and educ²) and the closing date variable (i.e., make the effects of closing date quadratically conditional on the categorical education measure). Discuss the estimates and goodness-of-fit of this model in contrast with those obtained from the model for the previous question. Report a likelihood ratio test on the joint significance of the new predictors.

```

R Code
1 > logit2 <- update(logit1,
2 +                 ~ . + south + govelec + closing*(educ + I(educ^2)))
3 > anova(logit1,logit2,test="Chi")

```

Analysis of Deviance Table

```

Model 1: turnout ~ educ + I(educ^2) + age + I(age^2)
Model 2: turnout ~ educ + I(educ^2) + age + I(age^2) + south + govelec +
  closing + educ:closing + I(educ^2):closing
  Resid. Df Resid. Dev    Df Deviance P(>|Chi|)
1     98852     111130
2     98847     110663     5      467 9.796e-99

```

- (e) Using the estimates from the second model, plot the implied coefficient for closing as a function of education, given the interaction effects estimated above. Overlay 95% confidence intervals around the point estimates. Offer a substantive interpretation of what this plot reveals.

```

R Code
1 > xseq <- 1:8      ## range of values on educ
2 > theCoefficients <- grep("closing",names(coef(logit2)))
3 > b <- coef(logit2)[theCoefficients]      ## relevant coefs
4 > print(b)

```

```

      closing      educ:closing I(educ^2):closing
0.0006794623 -0.0046216211      0.0003523177

```

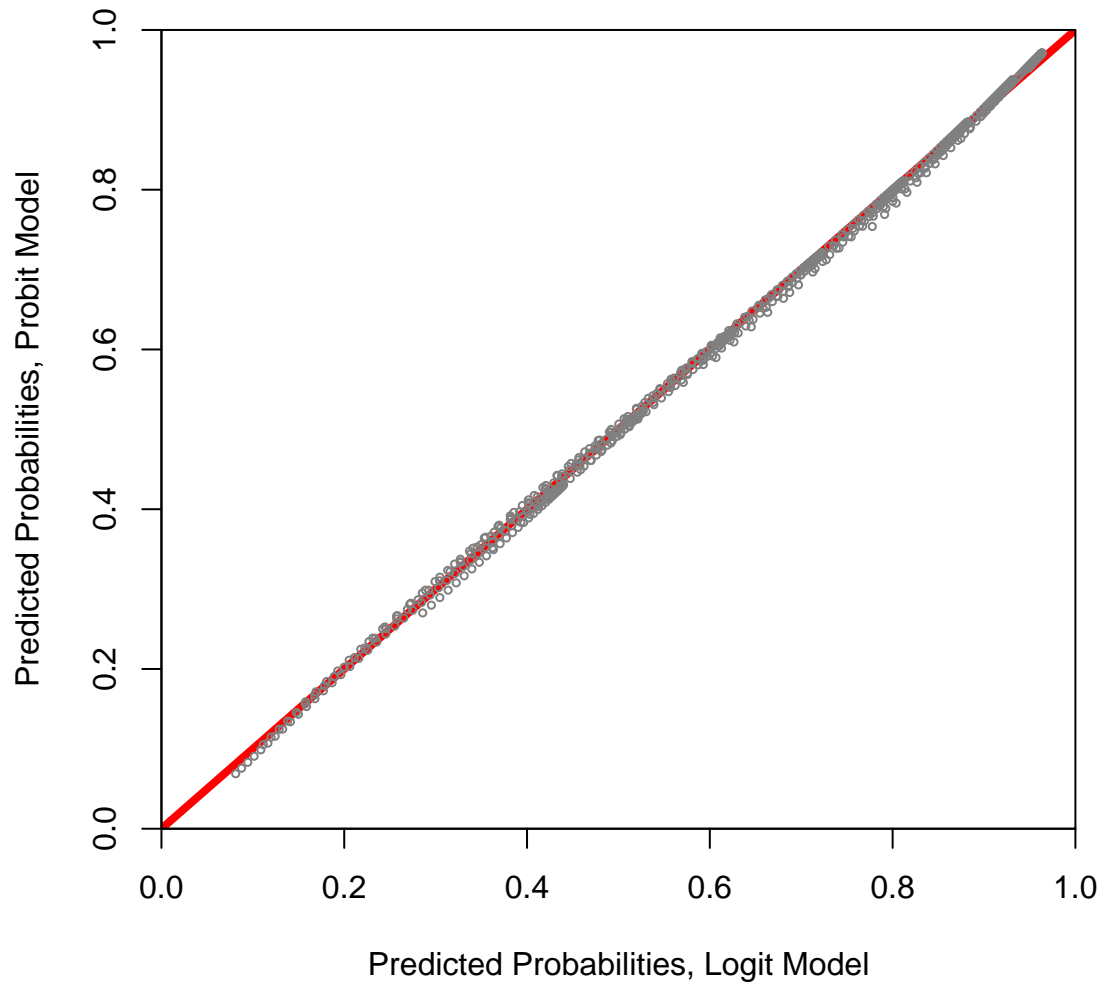


Figure 3: Comparison of Predicted Probabilities, Logit and Probit Models for Nagler Turnout Data. The red line (almost completely obscured by the plotted points) is a 45 degree line.

	Estimate	Std. Error	Estimate	Std. Error
Intercept	-4.5	0.079	-4.4	0.18
educ	0.27	0.025	0.36	0.071
educ ²	0.025	0.0025	0.019	0.0073
age	0.11	0.0022	0.11	0.0022
age ²	-0.00081	0.000023	-0.00082	0.000023
south			-0.18	0.017
govelec			0.0063	0.019
closing			0.00071	0.0062
educ × closing			-0.0046	0.0026
educ ² × closing			0.00035	0.00027
Deviance	111129.7		110662.5	
PCP	70.8%		71.1%	

Table 1: Logit Estimates, Two Models of Voter Turnout, Nagler Turnout Data ($n = 98,857$).

```

R Code
1 > bhat <- b[1] + b[2]*xseq + b[3]*(xseq^2) ## implied coef on closing
2 > varb <- vcov(logit2)[theCoefficients,theCoefficients] ## vars, covs
3 > print(varb)

      closing  educ:closing I(educ^2):closing
closing      3.847094e-05 -1.538043e-05      1.461037e-06
educ:closing -1.538043e-05  6.673732e-06     -6.741485e-07
I(educ^2):closing 1.461037e-06 -6.741485e-07  7.157603e-08
R Code
1 > vb <- varb[1,1] + (xseq^2)*varb[2,2] + (xseq^4)*varb[3,3] +
2 + 2*xseq*varb[1,2] + 2*(xseq^3)*varb[2,3] + 2*(xseq^2)*varb[1,3]
3 > ci <- sqrt(vb)*1.96 ## half width of 95% confidence interval
4 > require(plotrix) ## R package with plotCI function

[1] TRUE
R Code
1 > par(las=1,mar=c(5,5,.5,.5))
2 > plotCI(x=1:8,y=bhat,uiw=ci,
3 + pch=1,
4 + sfrac=0,
5 + xaxt="n", ## suppress axis labeling
6 + yaxt="n",
7 + xlab="Education (Years)",
8 + ylab="Implied Coefficient on Closing Day")
9 > abline(h=0,lty=2)
10 > axis(1,at=1:8,
11 + labels=c("0-4","5-7","8",
12 + "9-11","12","1-3yrs\ncollege",
13 + "4yrs\ncollege","5+yrs\ncollege"),
14 + cex.axis=.75)
15 > axis(2,cex.axis=.75)

```

See Figure 4.

- (f) Using the estimates from the second model, consider a hypothetical non-southerner, in a state without a gubernatorial election, who has 12 years of education and has the median age of a non-southerner with 12 years of education.

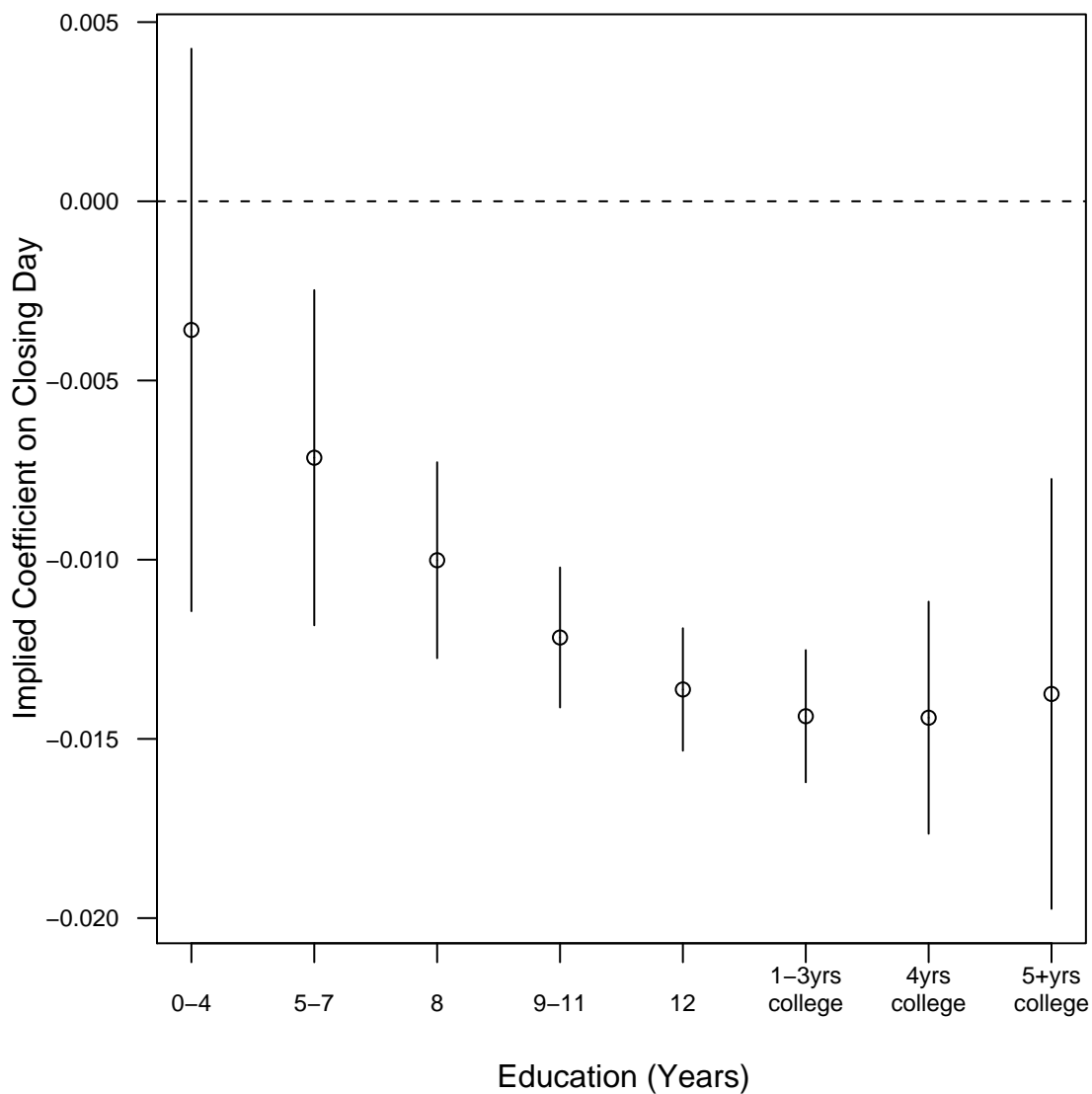


Figure 4: Implied Coefficient on Closing Day Requirement, as a function of education, from the Nagler turnout data. Vertical lines indicate a 95% confidence interval around each implied coefficient.

Plot the predicted probability of turnout for this person, as the closing date requirement varies over the range of closing date requirements observed in non-southern states. **Challenge question:** Can you add 95% confidence intervals about your point estimates?

Answer: See Figure 5.

```
                                R Code
1  > ## (5f) prob of turnout as a function of closing date
2  > closerange <- sort(unique((nagler$closing[nagler$south==0])))
3  > ok <- nagler$south==0 & nagler$educ==5
4  > newdata <- expand.grid(educ=5,
5  +                       govelec=0,
6  +                       south=0,
7  +                       closing=closerange,
8  +                       age=median(nagler$age[ok]))
9  > yhatd <- predict(logit2,
10 +                 newdata=newdata,
11 +                 type="link",
12 +                 se.fit=TRUE)
13 > yhatd <- cbind(lower=yhatd$fit-1.96*yhatd$se.fit, ## 95% CI
14 +               upper=yhatd$fit+1.96*yhatd$se.fit,
15 +               fit=yhatd$fit)
16 > yhatd <- apply(yhatd,2,binomial())$linkinv) ## convert to prob
```

- (g) Using the estimates from the second model, consider a hypothetical non-southerner, in a state without a gubernatorial election, who has 5+ years of college and has the median age of a non-southerner with 5+ years of college. Plot the predicted probability of turnout for this person, as the closing date requirement varies over the range of closing date requirements observed in non-southern states. Overlay 95% confidence intervals around the point estimates. Briefly compare the answers from this question with those from the previous question.

Answer: See Figure 5.

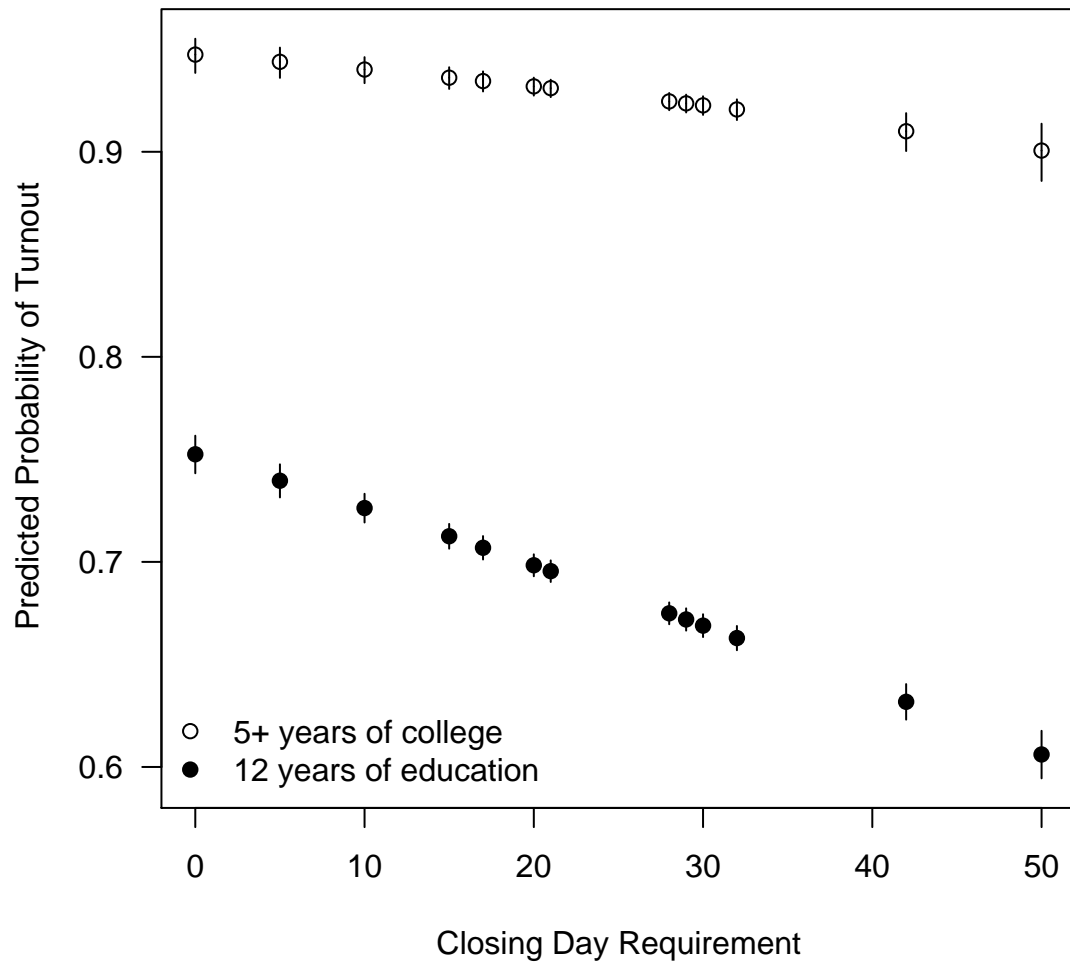


Figure 5: Predicted probabilities and 95% confidence intervals, voter turnout as a function of closing day requirement and education levels, for a non-southerner, in a state without a gubernatorial election, and with age set to the observed median age in each scenario. Vertical lines cover 95% confidence intervals.