Binary Response

Kei Sakamoto

LPM , Logit , Probit regression (後者２つは Maximum Liklihood estimation)

```
load("~/計量経済学演習/R data sets for 5e/mroz.RData")
mroz<-data
```

LPM

```
linprob <- lm(inlf~nwifeinc+educ+exper+I(exper^2)+age+kidslt6+kidsge6,data=mroz)
```

t-test using heteroscedasticity-robust SE(homoskedasticには構造上なり得ないのでhetero-robust seを使ってt-testやるか、そもそもweighted least squared とかで推定すべき)

```
library(lmtest);library(car)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: carData
```

```
coeftest(linprob,vcov=hccm)
```

```
##
## t test of coefficients:
##
##              Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)  0.58551922  0.15358032  3.8125  0.000149 ***
## nwifeinc    -0.00340517  0.00155826 -2.1852  0.029182 *
## educ         0.03799530  0.00733982  5.1766 2.909e-07 ***
## exper        0.03949239  0.00598359  6.6001 7.800e-11 ***
## I(exper^2)  -0.00059631  0.00019895 -2.9973  0.002814 **
## age         -0.01609081  0.00241459 -6.6640 5.183e-11 ***
## kidslt6     -0.26181047  0.03215160 -8.1430 1.621e-15 ***
## kidsge6      0.01301223  0.01366031  0.9526  0.341123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

prediction for 2 extreme women

```
xpred <- list(nwifeinc=c(100,0),educ=c(5,17),exper=c(0,30),
           age=c(20,52),kidslt6=c(2,0),kidsge6=c(0,0))
predict(linprob,xpred,type = "response")
```

```
##          1          2
## -0.4104582  1.0428084
```

response は確率なのに**0~1**の間に収まっていないのはおかしすぎるので**LPM**は**observaition**がはじの方では合わない。この欠点を次の２つで克服しにいく。

Logit model

```
summary(logitres<-glm(inlf~nwifeinc+educ+exper+I(exper^2)+age+kidslt6+kidsge6,
                      family=binomial(link=logit),data=mroz))
```

```
##
## Call:
## glm(formula = inlf ~ nwifeinc + educ + exper + I(exper^2) + age +
##     kidslt6 + kidsge6, family = binomial(link = logit), data = mroz)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1770  -0.9063   0.4473   0.8561   2.4032
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.425452   0.860365   0.495  0.62095
## nwifeinc    -0.021345   0.008421  -2.535  0.01126 *
## educ         0.221170   0.043439   5.091 3.55e-07 ***
## exper        0.205870   0.032057   6.422 1.34e-10 ***
## I(exper^2)  -0.003154   0.001016  -3.104  0.00191 **
## age         -0.088024   0.014573  -6.040 1.54e-09 ***
## kidslt6     -1.443354   0.203583  -7.090 1.34e-12 ***
## kidsge6      0.060112   0.074789   0.804  0.42154
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1029.75  on 752  degrees of freedom
## Residual deviance:  803.53  on 745  degrees of freedom
## AIC: 819.53
##
## Number of Fisher Scoring iterations: 4
```

Log likelihood value

```
logLik(logitres)
```

```
## 'log Lik.' -401.7652 (df=8)
```

McFadden’s pseudo R-squared

```
1 - logitres$deviance/logitres$null.deviance
```

```
## [1] 0.2196814
```

prediction(just same extreme women as LPM)

```
predict(logitres, xpred,type = "response")
```

```
##           1           2
## 0.005218002 0.950049117
```

extreme な**observation** だがちゃんと**0~1**に収まっている。

Probit model

```
summary(probitres<-glm(inlf~nwifeinc+educ+exper+I(exper^2)+age+kidslt6+kidsge6,
                       family=binomial(link=probit),data=mroz))
```

```
##
## Call:
## glm(formula = inlf ~ nwifeinc + educ + exper + I(exper^2) + age +
##     kidslt6 + kidsge6, family = binomial(link = probit), data = mroz)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2156  -0.9151   0.4315   0.8653   2.4553
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.2700736  0.5080782   0.532  0.59503
## nwifeinc    -0.0120236  0.0049392  -2.434  0.01492 *
## educ         0.1309040  0.0253987   5.154 2.55e-07 ***
## exper        0.1233472  0.0187587   6.575 4.85e-11 ***
## I(exper^2)  -0.0018871  0.0005999  -3.145  0.00166 **
## age         -0.0528524  0.0084624  -6.246 4.22e-10 ***
## kidslt6     -0.8683247  0.1183773  -7.335 2.21e-13 ***
## kidsge6      0.0360056  0.0440303   0.818  0.41350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1029.7  on 752  degrees of freedom
## Residual deviance:  802.6  on 745  degrees of freedom
## AIC: 818.6
##
## Number of Fisher Scoring iterations: 4
```

Log likelihood value

```
logLik(probitres)
```

```
## 'log Lik.' -401.3022 (df=8)
```

McFadden’s pseudo R-squared

```
1 - probitres$deviance/probitres$null.deviance
```

```
## [1] 0.2205805
```

glmでは**coef** テストに確率変数**z**を使っていることも１つの特徴。つまり**t-**分布でなく**Standard Normal**使っている。というか**t-test**では**large sample** でも**t-**分布使っていたことに驚き。**large sample**なら **Standard Normal**使っていいってのは手計算の時。

standard normalに近似はするけどやはり正確には**t-**分布だから**lm**ではあくまで**t-**分布使ってたっぽい。

```
predict(probitres,xpred,type = "response")
```

```
##           1           2 
## 0.001065043 0.959869044
```

logit とは若干違うが**0~1**の間には同様に収まっている。

Likelihood Ratio Test for probit model

restricted modelは**defaoult**では**constant**のみ。

```
library(lmtest)
lrtest(probitres)
```

	#Df <dbl>	LogLik <dbl>	Df <dbl>	Chisq <dbl>	Pr(>Chisq) <dbl>
1	8	-401.3022	NA	NA	NA
2	1	-514.8732	-7	227.142	2.008673e-45

2 rows

```
#LR-stat は probitres$null.deviance - probitres$deviance でも計算できる
```

流石に**constant**以外全て**0**説はない。

exper and age are irrelevant 説**(restricted model**は自分で作る。**exper**と**age**を抜けばいい**)**

```
restr <- glm(inlf~nwifeinc+educ+ kidslt6+kidsge6,
                    family=binomial(link=logit),data=mroz)
lrtest(restr,probitres)
```

	#Df <dbl>	LogLik <dbl>	Df <dbl>	Chisq <dbl>	Pr(>Chisq) <dbl>
1	5	-464.9249	NA	NA	NA
2	8	-401.3022	3	127.2454	2.121545e-27

2 rows

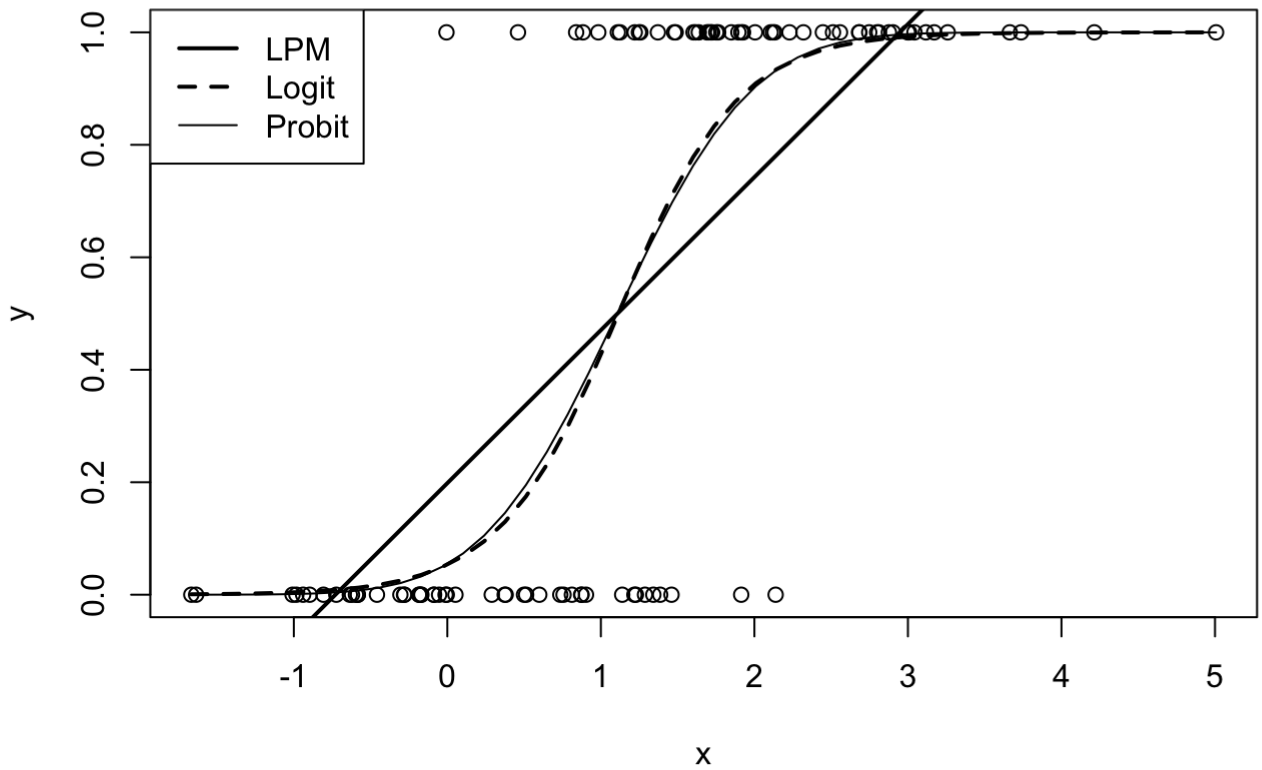
experも**age**も**relevant**と言えそう

regressorsが２個以上あるのでregression line(prediction)は描画はできないが、説明変数１つならMonte Carlo Simulation で作って描画できる

```
set.seed(8237445)
y<-rbinom(100,1,0.5)
x<-rnorm(100)+2*y
LPMres<-lm(y~x)
Logitres<-glm(y~x,family=binomial(link=logit))
Probitres<-glm(y~x,family=binomial(link=probit))

xlim<-seq(from=min(x),to=max(x),length=50)
LPM.p<-predict(LPMres,list(x=xlim),type="response")
Logit.p<-predict(Logitres,list(x=xlim),type="response")
Probit.p<-predict(Probitres,list(x=xlim),type="response")

plot(x,y)
lines(xlim,LPM.p,lwd=2,lty=1)
lines(xlim,Logit.p,lwd=2,lty=2)
lines(xlim,Probit.p,lwd=1,lty=1)
legend("topleft",c("LPM","Logit","Probit"),lwd=c(2,2,1),lty=c(1,2,1))
```



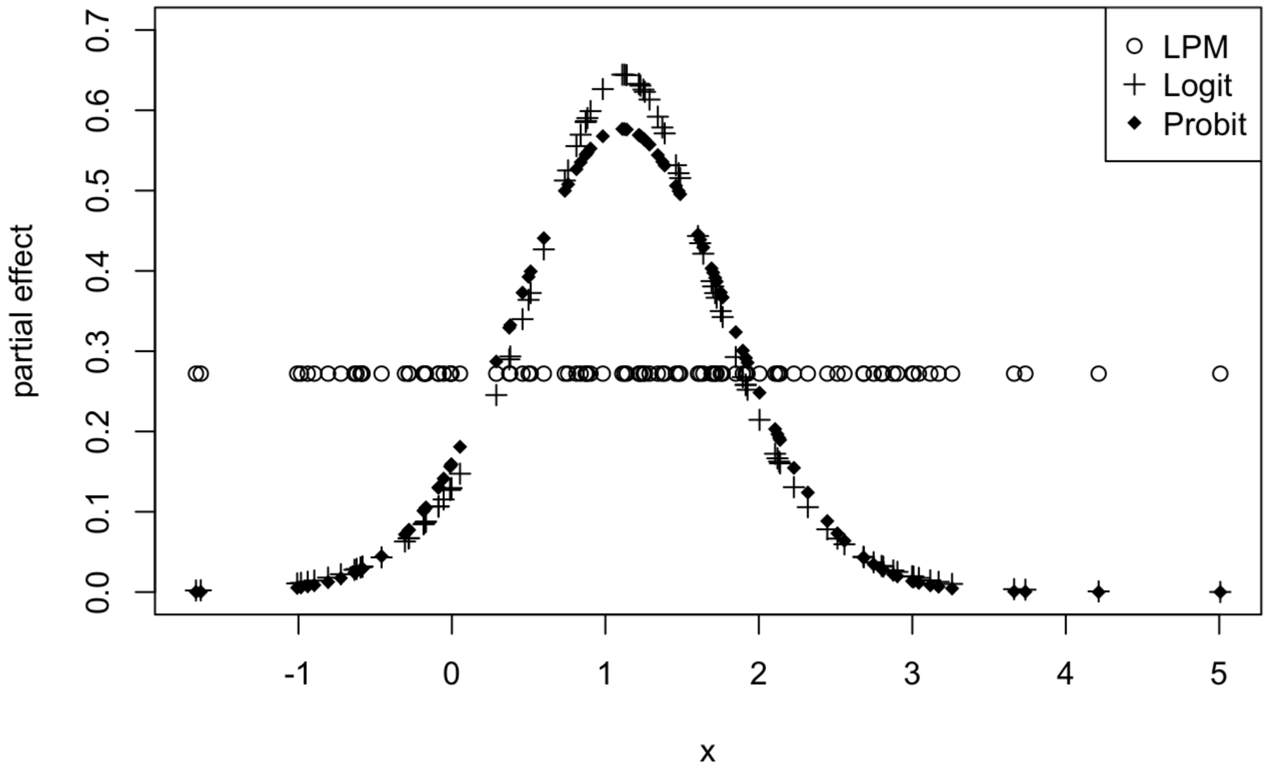
LogitとProbitはほとんど同じ。

ついでにmarginal(partial) effect も描画。

LPMはyはxに対してもパラメータに関しても線形だからxで一階微分して出てくるmarginal effectは横一線。Logit Probithがもともとmodelが CDFだからmarginal effect はpdfっぽくなるの当たり前。

```
LPM.eff<-coef(LPMres)["x"]*rep(1,100)#1を100個生成。なくてもいいけど。
Logit.eff<-coef(Logitres)["x"]*dlogis(predict(Logitres))
Probit.eff<-coef(Probitres)["x"]*dnorm(predict(Probitres))

plot(x,LPM.eff,pch=1,ylim=c(0,0.7),ylab="partial effect")
points(x,Logit.eff,pch=3)
points(x,Probit.eff,pch=18)
legend("topright",c("LPM","Logit","Probit"),pch=c(1,3,18))
```



Logit のAPEの計算(Average Partial Effect。automaticの方のみ)

```
library(mfx)
```

```
## Loading required package: sandwich
```

```
## Loading required package: MASS
```

```
## Loading required package: betareg
```

```
logitmfx(inlf~nwifeinc+educ+exper+I(exper^2)+age+kidslt6+kidsge6,
                              data=mroz, atmean=FALSE)
```

```
## Call:
## logitmfx(formula = inlf ~ nwifeinc + educ + exper + I(exper^2) +
##     age + kidslt6 + kidsge6, data = mroz, atmean = FALSE)
##
## Marginal Effects:
##                  dF/dx   Std. Err.       z     P>|z|
## nwifeinc   -0.00381181  0.00153898 -2.4769  0.013255 *
## educ        0.03949652  0.00846811  4.6641 3.099e-06 ***
## exper       0.03676411  0.00655577  5.6079 2.048e-08 ***
## I(exper^2) -0.00056326  0.00018795 -2.9968  0.002728 **
## age        -0.01571936  0.00293269 -5.3600 8.320e-08 ***
## kidslt6    -0.25775366  0.04263493 -6.0456 1.489e-09 ***
## kidsge6     0.01073482  0.01339130  0.8016  0.422769
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#atmean=TRUEにすればPEA(Partial Effect at Average)
```

コマンド１つは強すぎる…。