Assignment 1

February 22, 2018

For grader: Please comment in red. Either "Everything is correct!" or commenting to help your fellow classmate out if you think anything can be improved. If you use Markdown, bolded text is fine.

The cross-grading deadline is the end of the same week as the homework. We TAs will review your cross-grading and assign final numerical grade.

Corruption and Wealth (3 points)

Regression of corruption on GDP per capita

```r
df_corruption <- haven::read_dta("corruption.dta")
```

Running regression using OLS

```r
m1 <- lm(ti_cpi ~ undp_gdp, data = df_corruption)
stargazer(m1)
```

Running regression using MLE
Note that the likelihood function is

$$Y \sim N(X\beta, \sigma^2) \tag{1}$$

$$LL = \sum_i \log N(X\beta, \sigma^2) \tag{2}$$

```r
X <- cbind(1, df_corruption$undp_gdp)
y <- df_corruption$ti_cpi
N <- length(y)
#' The loglikelihood is a function of beta and sigma2. This is a direct
#' implementation of the formula on slide MLE week 3, p 30
#' @param params is a vector of length 3,
#' the 1st element is beta for the intercept,
#' the 2nd element is beta for GDP per capita,
```

Table 1:

	Dependent variable:
	ti_cpi
undp_gdp	0.0002***
	(0.00001)
Constant	2.502***
	(0.124)
Observations	170
R^2	0.673
Adjusted R^2	0.671
Residual Std. Error	1.207 (df = 168)
F Statistic	346.398*** (df = 1; 168)
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
#' the 3rd element is sigma2
negative_log_likelihood <- function(params) {
  beta <- c(params[1], params[2])
  sigma2 <- params[3]
  return(N / 2 * log(sigma2) +
         1 / (2 * sigma2) * t(y - X %*% beta) %*% (y - X %*% beta))
}

# MLE estimate
m2 <- optim(par = c(0, 0, 1), fn = negative_log_likelihood)$par
names(m2) <- c("beta0", "beta1", "sigma2")
print(m2)

##        beta0        beta1       sigma2
## 2.5027995972 0.0001729247 1.4381806067
```

We see that the MLE for β from `optim` result are 2.5027996, 1.7292473×10^{-4}, exactly as the OLS result above.

The intercept means that when a country has a GDP per capita of 0, its predicted corruption level (in terms of CPI score) is 2.503.

The coefficient for GDP per capita means that an increase of USD 1000 in GDP per capita is associated with a change of 0.173 in corruption level (in terms of CPI score).

Interpret the substantive relevance

```
new_data <- data.frame(undp_gdp = quantile(df_corruption$undp_gdp,
                                            c(0.25, 0.75)))
pred_m1 <- predict(m1, newdata = new_data)
print(pred_m1)

##      25%      75%
## 2.843991 4.381464
```

The level of corruption in countries with GDP per capita that corresponds to the 25th percentile is 2.844 and a country with GDP per capita in the 75th percentile is 4.381.

Maximum Likelihood I (3 points)

1.

Logit and probit coefficients always differ because the error terms in the two models are normalized to have a different scale.

More specifically, recall that both probit and logit are derived from the following random utility model:

$$U_i^* = \beta_0^* + \beta_1^* X + \epsilon^* \tag{3}$$

where ϵ^* is the unobserved component of the utility. Call the standard deviation of this unobserved component σ (which is unknown, because we don't observe ϵ^*).

In probit, we want to normalize so that the error term is normally distributed with a standard deviation of 1. We do this by dividing everything by σ:

$$U_i = \frac{\beta_0^*}{\sigma} + \frac{\beta_1^*}{\sigma} X + \frac{\epsilon^*}{\sigma} \tag{4}$$

$$\tag{5}$$

In logit, we want to normalize so that the error term is follows standard Gumbel distribution, with a standard deviation of $\sqrt{\pi^2/6} \sim \sqrt{1.6}$. We do this by dividing everything by $\frac{\sigma}{\sqrt{1.6}}$

$$U_i = \frac{\sqrt{1.6}\beta_0^*}{\sigma} + \frac{\sqrt{1.6}\beta_1^*}{\sigma} X + \frac{\sqrt{1.6}\epsilon^*}{\sigma} \tag{6}$$

$$\tag{7}$$

As you can see, the coefficient we get from probit is $\frac{\beta_1^*}{\sigma}$, while the coefficient we get from logit is $\frac{\sqrt{1.6}\beta_1^*}{\sigma}$. So the logit's coefficient is always $\sqrt{1.6}$ times larger than probit's.

2.

Predicted values from a binary model are either 0 or 1. The expected value is the expectation of the predicted value, which for binary model is the probability that the outcome is 1.

3.

A binomial distribution is usually used to model the number of successes in N trials. Here we can consider each student as a trial, and a pass as a success.

Define y as the number of passes, we have:

$$y \sim Binom(13, \theta) \tag{8}$$

$$lik = \binom{13}{y} \theta^y (1-\theta)^{n-y} \tag{9}$$

Maximum Likelihood II (4 points)

```r
df_turnout <- haven::read_dta("turnout.dta")
m_turnout <- glm(voted ~ informed + age + edu + female, data = df_turnout,
                 family = binomial)
summary_turnout <- summary(m_turnout)
stargazer(m_turnout)
```

1.

The β for `female` is 0.427, meaning that the odd of voting is $\exp(\beta) = 1.533$ times larger for female.

2.

Predicted probability of turnout for woman, age = 25, edu = 3, and average political information is

```r
predict(m_turnout,
        newdata = data.frame(female = 1, age = 25, edu = 3,
                             informed = mean(df_turnout$informed)),
        type = "response")

##         1
## 0.8065287
```

Table 2:

	Dependent variable:
	voted
informed	0.522***
	(0.103)
age	0.032***
	(0.005)
edu	0.343***
	(0.088)
female	0.427***
	(0.162)
Constant	$-$2.495***
	(0.428)
Observations	1,078
Log Likelihood	$-$514.638
Akaike Inf. Crit.	1,039.276
Note:	*p<0.1; **p<0.05; ***p<0.01

3.

To show the effect of political information at low [0] and high [3] education level, we calculate the probability of voting. We use simulation to demonstrate uncertainty

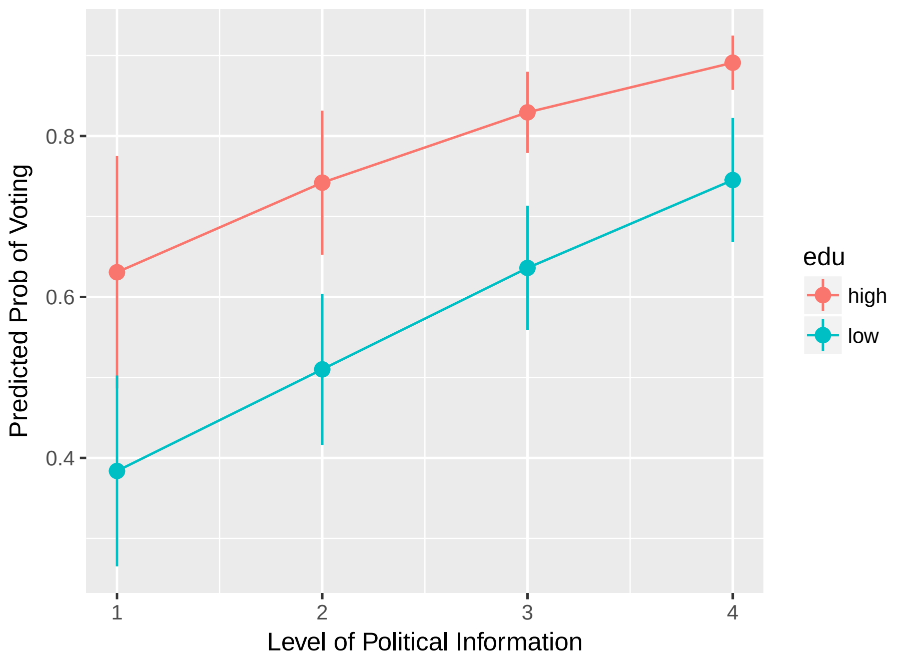
```r
S <- 1000 # Number of simulations
beta_hat <- coef(m_turnout)
V_beta_hat <- summary_turnout$cov.unscaled

# Draw an estimate from N(beta_hat, V_beta_hat)
beta <- mvtnorm::rmvnorm(S, mean = beta_hat, sigma = V_beta_hat)
```

```r
# Set other values
X_edulow <- data.frame(intercept = 1,
                       informed = 1:4,
                       age = mean(df_turnout$age),
                       edu = 0,
                       female = 0) %>% as.matrix()
XB <- beta %*% t(X_edulow)
predicted_probs <- exp(XB) / (1 + exp(XB))
dat_edulow <- as.data.frame(predicted_probs) %>%
  tidyr::gather(key = "informed", value = "predicted_prob") %>%
  mutate(informed = recode(informed, "V1" = 1, "V2" = 2, "V3" = 3, "V4" = 4),
         edu = "low")

# Set values and calculate predicted_probs
X_eduhigh <- data.frame(intercept = 1,
                       informed = 1:4,
                       age = mean(df_turnout$age),
                       edu = 3,
                       female = 0) %>% as.matrix()
XB <- beta %*% t(X_eduhigh)
predicted_probs <- exp(XB) / (1 + exp(XB))
dat_eduhigh <- as.data.frame(predicted_probs) %>%
  tidyr::gather(key = "informed", value = "predicted_prob") %>%
  mutate(informed = recode(informed, "V1" = 1, "V2" = 2, "V3" = 3, "V4" = 4),
         edu = "high")


dat <- bind_rows(list(dat_edulow, dat_eduhigh))
```

Below we plot the predicted probability of voting across levels of political information. For respondents with low education, the probability of voting increases from 0.4 ± 0.1 to 0.7 ± 0.08 as the level of political information increases from 1 to 4. For respondents with high education, the probability of voting increases from 0.6 ± 0.1 to 0.9 ± 0.03 as the level of political information increases from 1 to 4. (The error bars reflect the 95% confidence interval. The uncertainty is highest for respondents where `informed = 1` because we have the least amount of data for them.)

Overall, the level of political information has a positive effect on the probability of voting. The size of the effect is also consistent across both high and low education respondents.

```r
ggplot(data = dat, aes(x = informed, y = predicted_prob, color = edu)) +
  stat_summary(fun.y = mean,
               fun.ymin = function(x) mean(x) - 1.96 * sd(x),
               fun.ymax = function(x) mean(x) + 1.96 * sd(x),
               geom = "pointrange") +
  stat_summary(fun.y = mean, geom = "line") +
  labs(x = "Level of Political Information", y = "Predicted Prob of Voting")
```


Predicted Prob of Voting
Level of Political Information
edu
high
low