HUDM6052 Psychometric II Homework_01

Chenguang Pan (cp3280@tc.columbia.edu)

2023-09-28

Contents

Q1-Part(a)

Plot the following two items using the normal ogive model over the range..

My Solution:

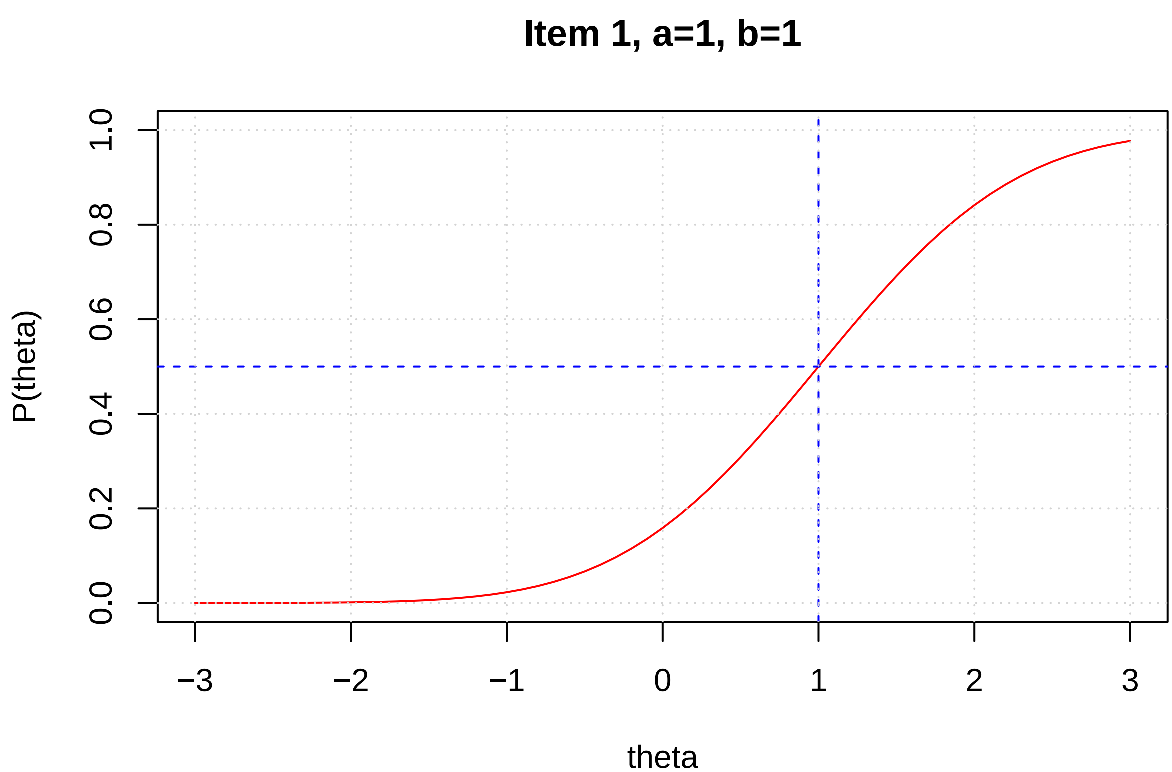
First, I define the normal ogive model.

```
> # define the theta
> theta <- seq(-3,3, by=0.1)
>
> # define the normal ogive model
> nom_2pl <- function(theta, a, b){
+   # get the probit vector
+   Z <- a*(theta-b)
+   func_ <- function(probit){(1/sqrt(2*pi))*exp(-0.5*probit^2)}
+   p_list <- c()
+   # using a for loop to get the p iteratively
```

```
+   for (i in 1:length(Z)) {
+     # get the CDF
+     p_i <- integrate(func_, -Z[i], Inf)
+     # store the result
+     p_list[i] <- p_i
+   }
+   return(p_list)
+ }
```

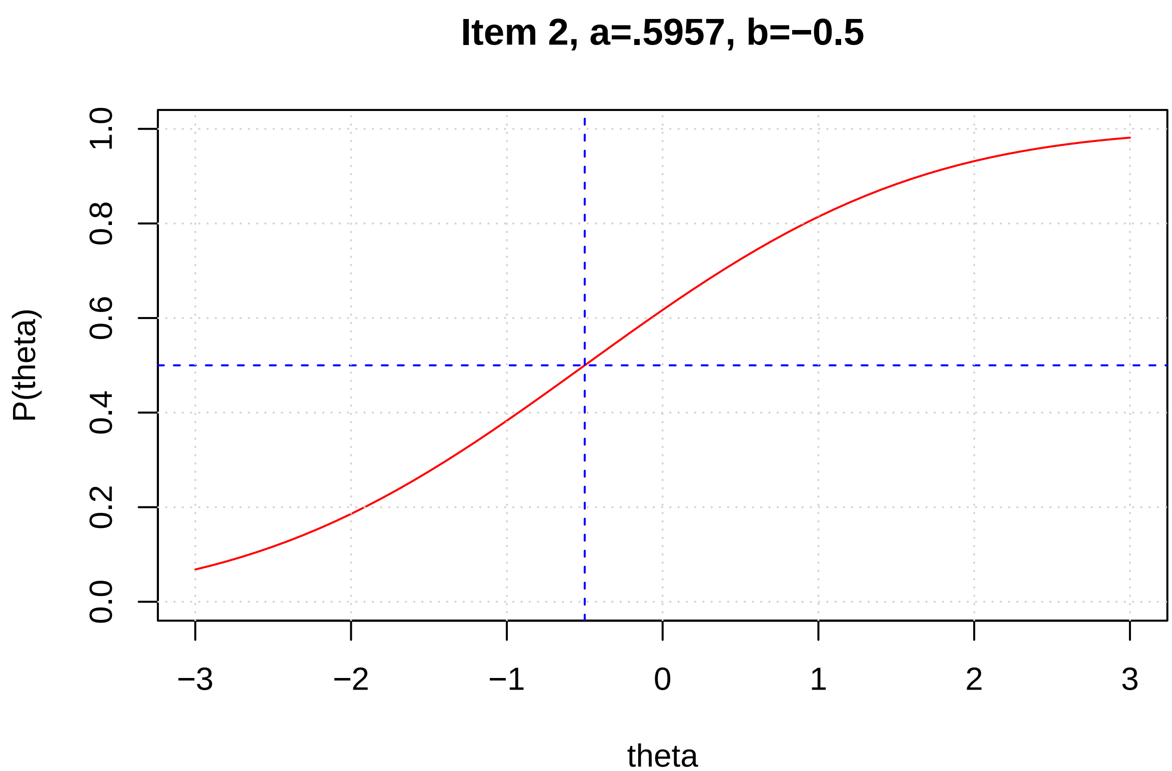
Then, using the function above to plot the ICCs for two items.
For the item 1 with $\alpha = 1$ and $\beta = 1$.

```
> # Item 1: a=1, b=1
> # define the xlim
> xlim_min <- -3
> xlim_max <- 3
> P_item1_nom <- nom_2pl(theta=theta,a=1,b=1)
> plot(theta, P_item1_nom, type = "l",
+      col="red", xlim = c(xlim_min,xlim_max), ylim = c(0,1),
+      main = "Item 1, a=1, b=1", ylab = "P(theta)")
>
> # add a horizontal line at p=0.5 to see the b value
> abline(h=0.5, col="blue", lty=2)
> abline(v=1, col="blue", lty=2)
> grid()
```



For the Item 2 with $\alpha = .5957$ and $\beta = -0.5$.

```
> # Item 1: a=.5957, b=-.5
> # par(mfrow=c(1,2))
> P_item2_nom <- nom_2pl(theta=theta,a=.5957,b=-0.5)
> plot(theta, P_item2_nom, type = "l",
+      col="red", xlim = c(xlim_min,xlim_max), ylim = c(0,1),
+      main = "Item 2, a=.5957, b=-0.5", ylab = "P(theta)")
>
> # add a horizontal line at p=0.5 to see the b value
> abline(h=0.5, col="blue", lty=2)
> abline(v=-0.5, col="blue", lty=2)
> grid()
```



Q1-Part(b)

Plot the following two items using a logistic model over the range..

My Solution:

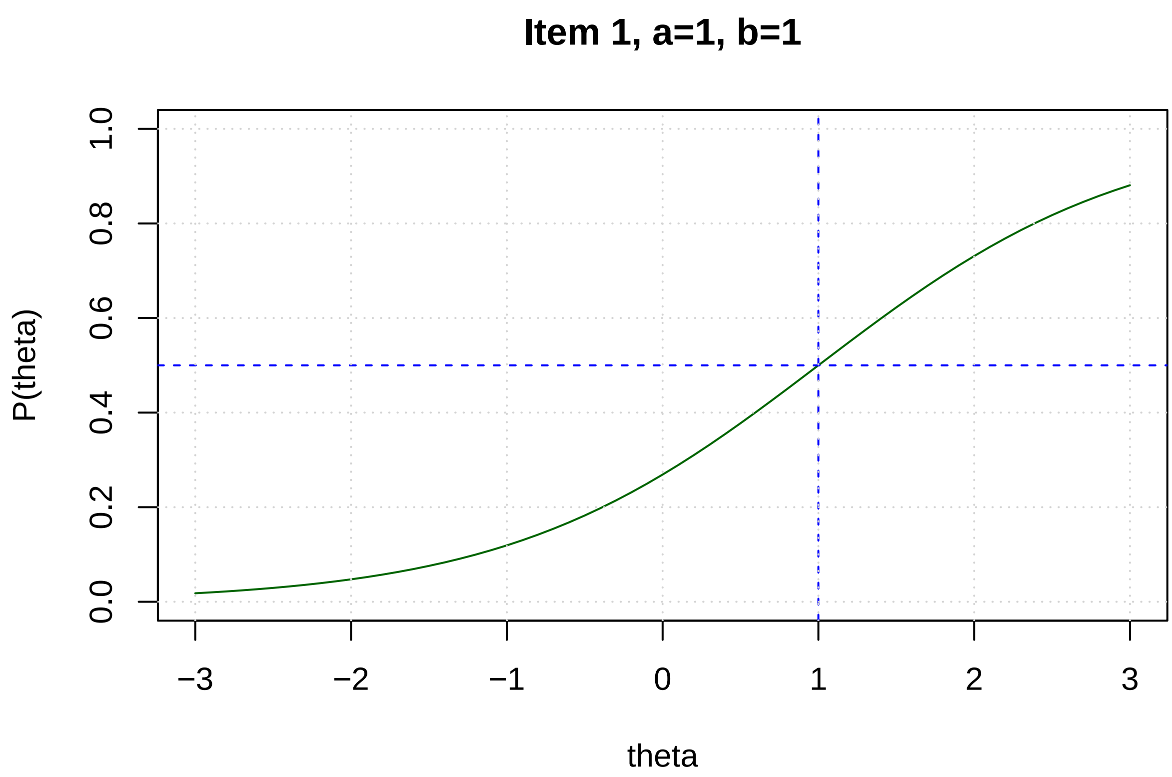
Like before, I defined the model first. This time, the model is much simpler.

```
> logit_2pl <- function(theta, a, b){
+   # get the logit
+   Z <- a*(theta-b)
```

```
+   # get the probability
+   P <- 1/(1 + exp(-Z))
+   return(P)
+ }
```

Then, using the function above to plot the ICCs for two items.
For the item 1 with $\alpha = 1$ and $\beta = 1$.

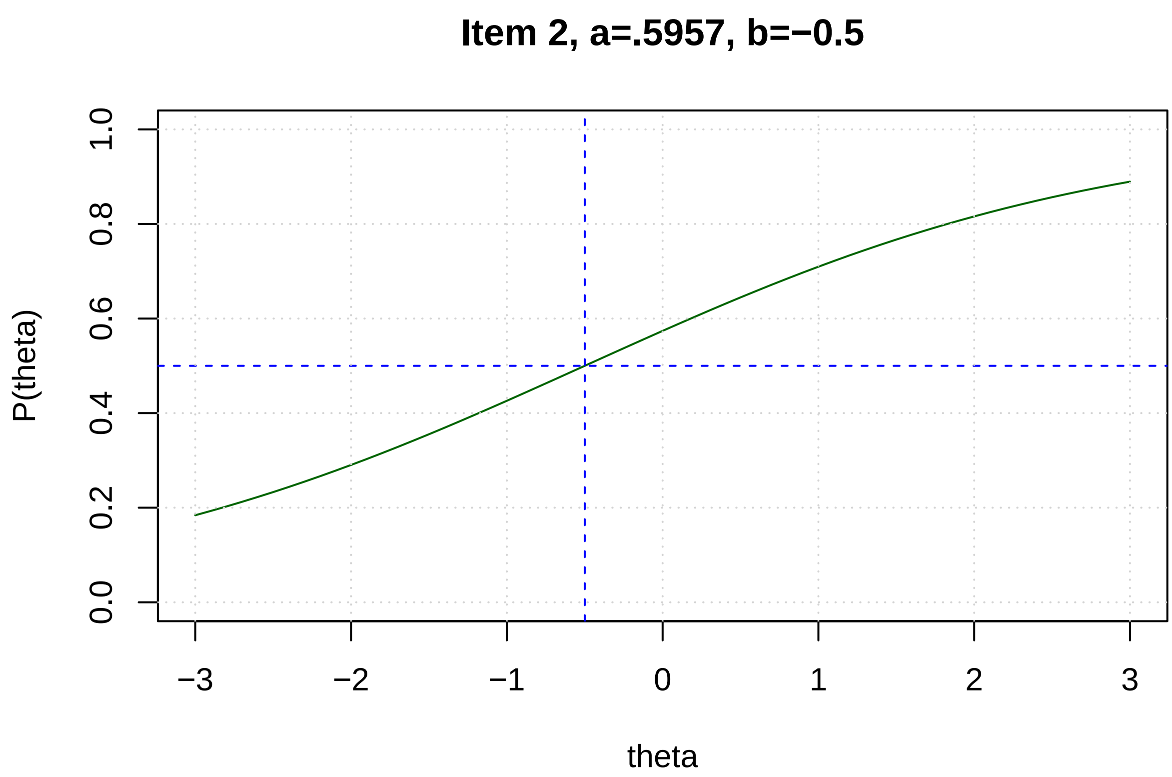
```
> P_item1_logit <- logit_2pl(theta=theta,a=1,b=1)
> plot(theta, P_item1_logit, type = "l",
+      col="darkgreen",  xlim = c(xlim_min,xlim_max), ylim = c(0,1),
+      main = "Item 1, a=1, b=1", ylab = "P(theta)")
>
> # add a horizontal line at p=0.5 to see the b value
> abline(h=0.5, col="blue", lty=2)
> abline(v=1, col="blue", lty=2)
> grid()
```



For the Item 2 with $\alpha = .5957$ and $\beta = -0.5$.

```
> # Item 1: a=.5957, b=-.5
> # par(mfrow=c(1,2))
> P_item2_logit <- logit_2pl(theta=theta,a=.5957,b=-0.5)
> plot(theta, P_item2_logit, type = "l",
+      col="darkgreen", xlim = c(xlim_min,xlim_max), ylim = c(0,1),
```

```
+       main = "Item 2, a=.5957, b=-0.5", ylab = "P(theta)")
>
> # add a horizontal line at p=0.5 to see the b value
> abline(h=0.5, col="blue", lty=2)
> abline(v=-0.5, col="blue", lty=2)
> grid()
```



Q1-Part(c)

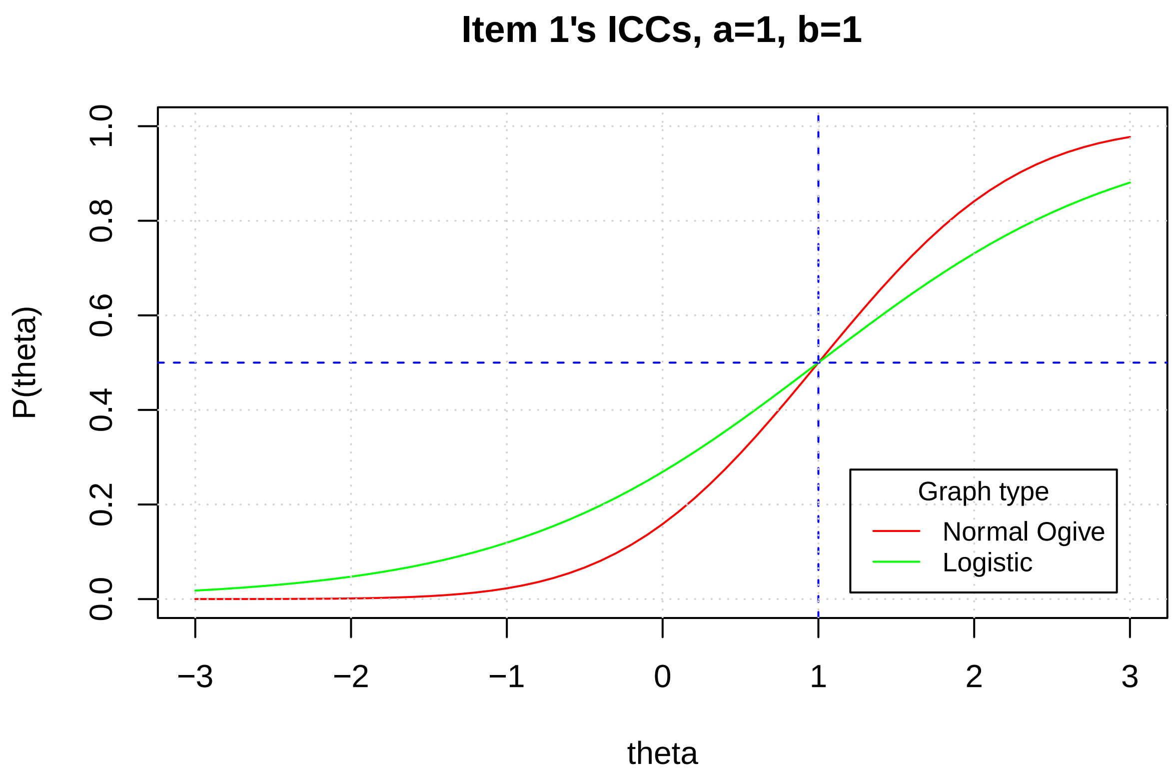
Compare the two plots in (b) with the two plots in part (a). What do you find?..

My Solution:

For better comparison, I combined the item 1's ICCs from two model, and so did item 2.
ICCs for item 1.

```
> plot(theta, P_item1_nom, type = "l",
+       col="red", xlim = c(xlim_min,xlim_max), ylim = c(0,1),
+       main = "Item 1's ICCs, a=1, b=1", ylab = "P(theta)")
> lines(theta, P_item1_logit, col="green")
> abline(h=0.5, col="blue", lty=2)
> abline(v=1, col="blue", lty=2)
> grid()
> # add legend
> legend('bottomright',inset=0.05,
```

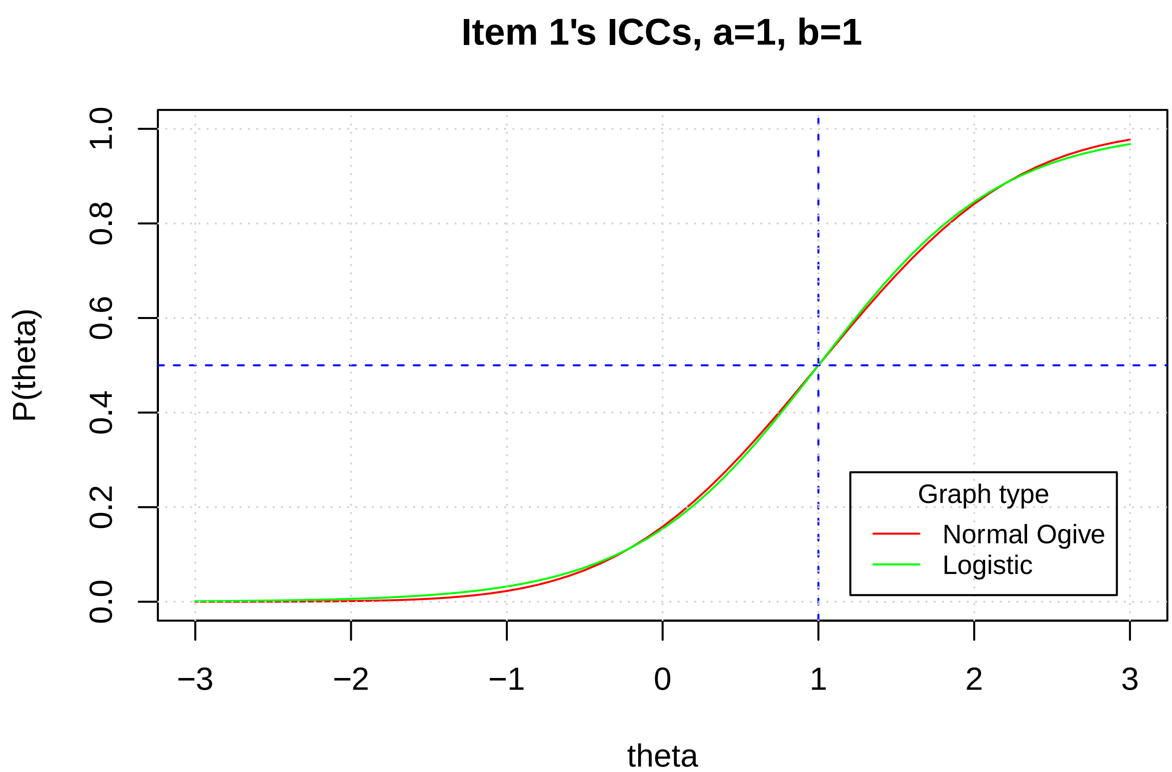
```
+         c("Normal Ogive","Logistic"),lty=1,
+         col=c("red","green"),title="Graph type",
+         cex=0.8)
```



The plot shows that ICCs intersect at the $P(\theta) = 0.5$ with a same corresponding $\beta = 1$. However, the slope of logistic model's ICC is lighter than it from the normal ogive model.

Next, I use the converted $\alpha_c = 1.702\alpha$ to re-plot the ICCs.

```
> # adjust the alpha
> P_item1_logit <- logit_2pl(theta=theta,a=1*1.702,b=1)
> # replot the two ICCs
> plot(theta, P_item1_nom, type = "l",
+      col="red", xlim = c(xlim_min,xlim_max), ylim = c(0,1),
+      main = "Item 1's ICCs, a=1, b=1", ylab = "P(theta)")
> lines(theta, P_item1_logit, col="green")
> abline(h=0.5, col="blue", lty=2)
> abline(v=1, col="blue", lty=2)
> grid()
> # add legend
> legend('bottomright',inset=0.05,
+         c("Normal Ogive","Logistic"),lty=1,
+         col=c("red","green"),title="Graph type",
+         cex=0.8)
```



Now, the two ICCs are pretty close to each other.
Item 2 will have the same phenomenon, I skip plotting the item 2 for space saving.

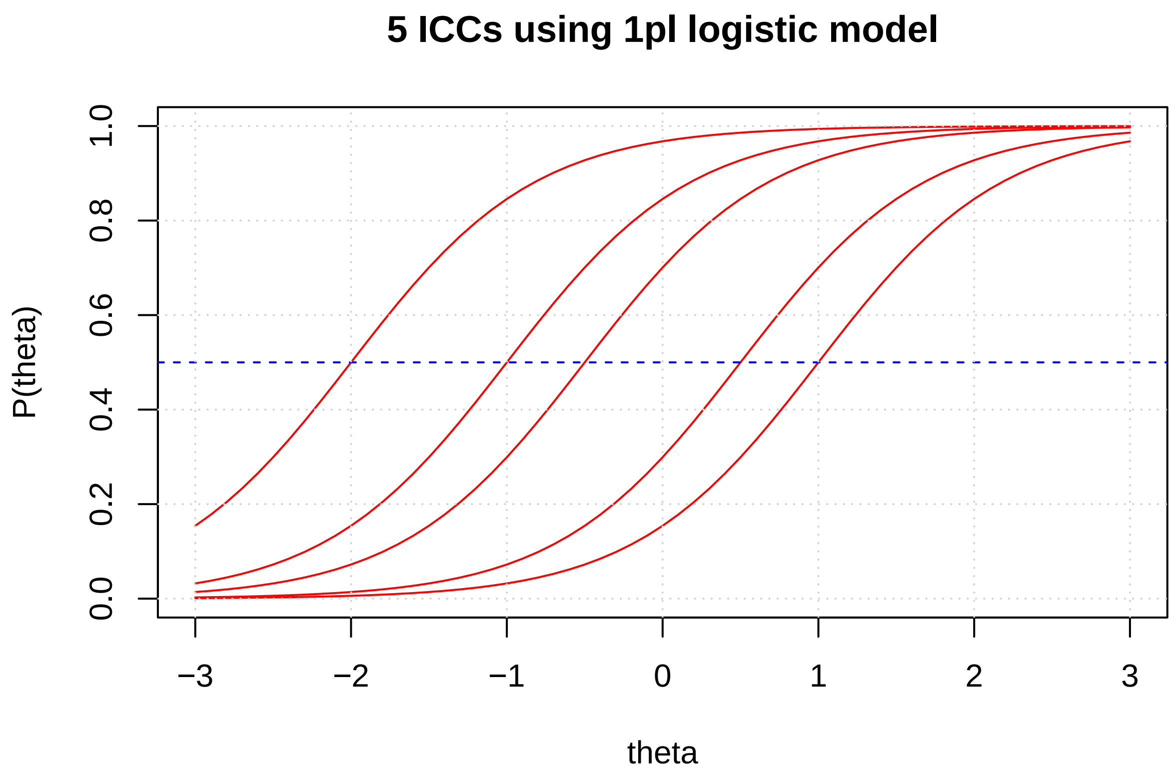
Q2-Part(a)

Plot the ICCs of the five items in the . . .

My Solution:
Using the function created above to plot five ICCs on one plot. Logistic model with converted α is used. I use 1PL model since the discrimination information is not mentioned.

```
> # define the beta vector as input
> B <- c(-2, -1, -0.5, 0.5, 1)
>
> # using a for loop to get all results
> Ps <- list()
> for (i in 1:length(B)){
+   # note a is a constant with value of 1.702 in 1PL model.
+   P_item1_logit <- logit_2pl(theta=theta,a=1.702,b=B[i])
+   # store the result into a list
+   Ps[[i]] <- P_item1_logit
+ }
>
> # plot all five ICCs on same plot
> # replot the two ICCs
> plot(theta, Ps[[1]], type = "l",
```

```
+       col="red", xlim = c(xlim_min,xlim_max), ylim = c(0,1),
+       main = "5 ICCs using 1pl logistic model", ylab = "P(theta)")
> # using a for loop to plot the rest 4 ICCs
> for (i in c(2:5)) {
+   lines(theta, Ps[[i]], col="red")
+ }
> abline(h=0.5, col="blue", lty=2)
> grid()
```



Q2-Part(b)

Compare the probabilities of correct response for Items ...

My Solution:

For this 1PL model, item 1 ($\beta = -2$) is easier than item 4 ($\beta = 0.5$). For example, for an examinee with trait level of 0 (i.e., $\theta = 0$), the probabilities to correctly answer the item 1 is around .96 and around .3 for item 4. Same phenomenon can be also found for the examinees with a shared trait level across the trait range $[-3, 3]$. However, at the two extreme ends (i.e., the negative infinity and the positive infinity), the probabilities should be same, either 0 or 1.

Q2-Part(c)

Which of the five items is easiest ...

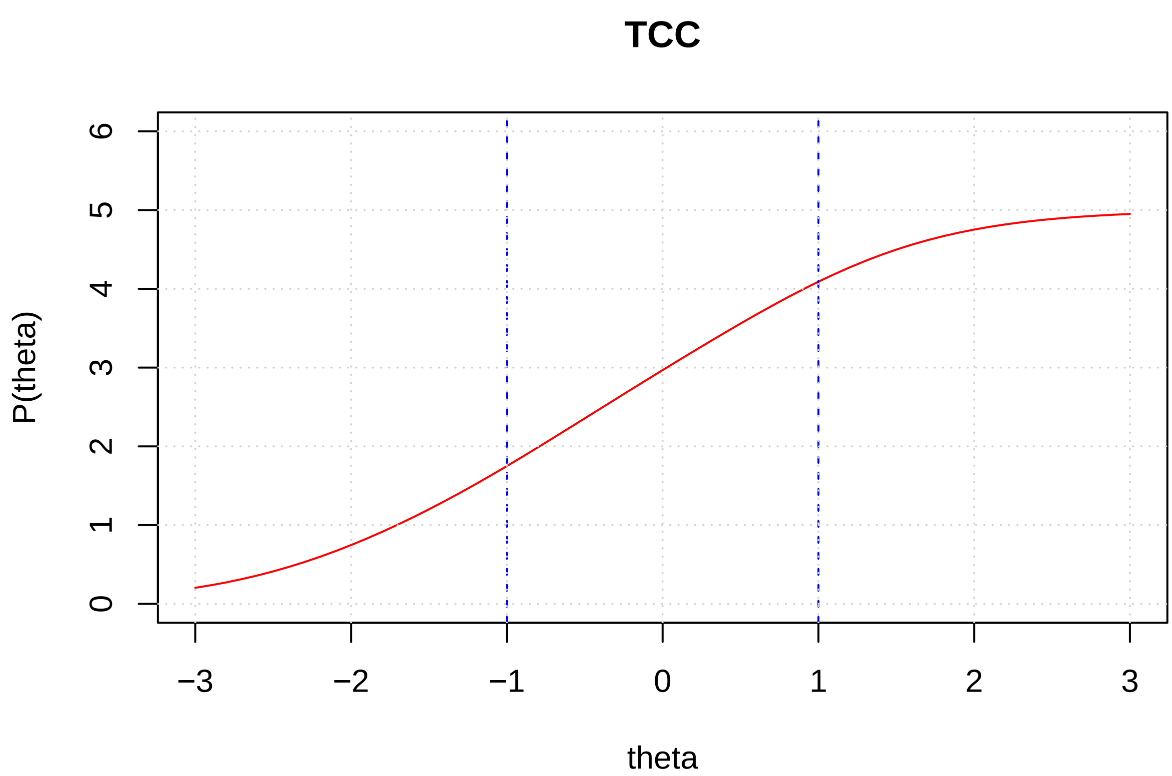
My Solution:
The easiest item is item 1, and the item 5 is the most difficult. Yes. By definition, the diffculty is an item's characteristic conditional on θ. Therefore, Are these statements true for all θ.

Q2-Part(d)

Find TCC and Plot TCC ...

My Solution:
By defintion, the TCC is sum of the $P(\theta_j)$ for all items.

```
> # get the sum of all P_thetas
> P_tcc <- Ps[[1]] + Ps[[2]] + Ps[[3]] + Ps[[4]] + Ps[[5]]
>
> # plot the TCC across thetas
> plot(theta, P_tcc, type = "l",
+      col="red", xlim = c(xlim_min,xlim_max), ylim = c(0,6),
+      main = "TCC", ylab = "P(theta)")
> abline(v=-1, col="blue", lty=2)
> abline(v=1, col="blue", lty=2)
> grid()
```



Q2-Part(e)

TCC also represents the expected number of correct responses ...

My Solution:
From the plot above, one might expect an examinee with the trait $\theta = -1.0$ to correctly answer around 1.9 items, and above 4 items for an examinee with the trait at $\theta = 1$.

Q3-Part(a)

Find the mean of Item 6 and its correlation with ...

My Solution:

For group 1, the difficulty(easiness) is .510 and the discrimination (using point bi-serial) is .276.
For group 2, the difficulty(easiness) is .670 and the discrimination is .553.

Q3-Part(b)

Do the indices remain the same across the two groups ...

My Solution:
The indices are not the same across the groups. Since one limitation of CTT is sample-dependent. Different sample will return different item characteristics. CTT does not guarantee a indices invariance.

Q3-Part(c)

Which group has higher ability ...

My Solution:

The Group 2's ability is greater than Group 1 since the average total score is higher in Group 2.

Q4

Responses to three items at a fixed ability level can be ...

My Solution:
The contingency tables for each pair as followed:

		Item 2	
		correct	incorrect
Item 1	correct	54	26
	incorrect	3	17

		Item 3	
		correct	incorrect
Item 1	correct	38	42
	incorrect	8	12

		Item 3	
		correct	incorrect
Item 2	correct	36	21
	incorrect	10	33

Next, I use the built-in function in R to conduct the χ^2 test. Note, the null hypothesis of χ^2 test is that the variables are independent.

```
> # load the data
> df <- read.csv("~/Desktop/PhD_Learning/HUDM6052 Psychometric II/assignment 1/hw1_4.csv")
>
> # chi-square test for item 1 and 2
> result_12 <- chisq.test(table(df[,c(1,2)]))
> result_13 <- chisq.test(table(df[,c(1,3)]))
> result_23 <- chisq.test(table(df[,c(2,3)]))
> (pvalue_12 <- result_12$p.value)
[1] 6.627178e-05
> (pvalue_13 <- result_13$p.value)
[1] 0.7254943
> (pvalue_23 <- result_23$p.value)
[1] 0.0001692519
```

The p-values from both tests of pair 1 (item 1 and item 2) and pair 3 (item 2 and item 3) are significant ($\alpha =$ .05), which means there are associations in both two pairs. Therefore, the local independent assumption does not hold for item 1 and 2 and for item item 2 and item 3. Only item 1 and item 3 are locally independent.

Q5

Using the logistic model, plot the following three items over the range ...

My Solution:
I continue to use the function `logit_2pl()` created in the first question since it can fit for both 1PL and 2PL model. I write a new function for 3PL model as followed.

```
> # write 3PL function
> logit_3pl <- function(theta, a, b, c){
+   # get the logit
+   Z <- a*(theta-b)
+   # get the probability
+   P <- c + (1-c)/(1 + exp(-Z))
+   return(P)
+ }
```

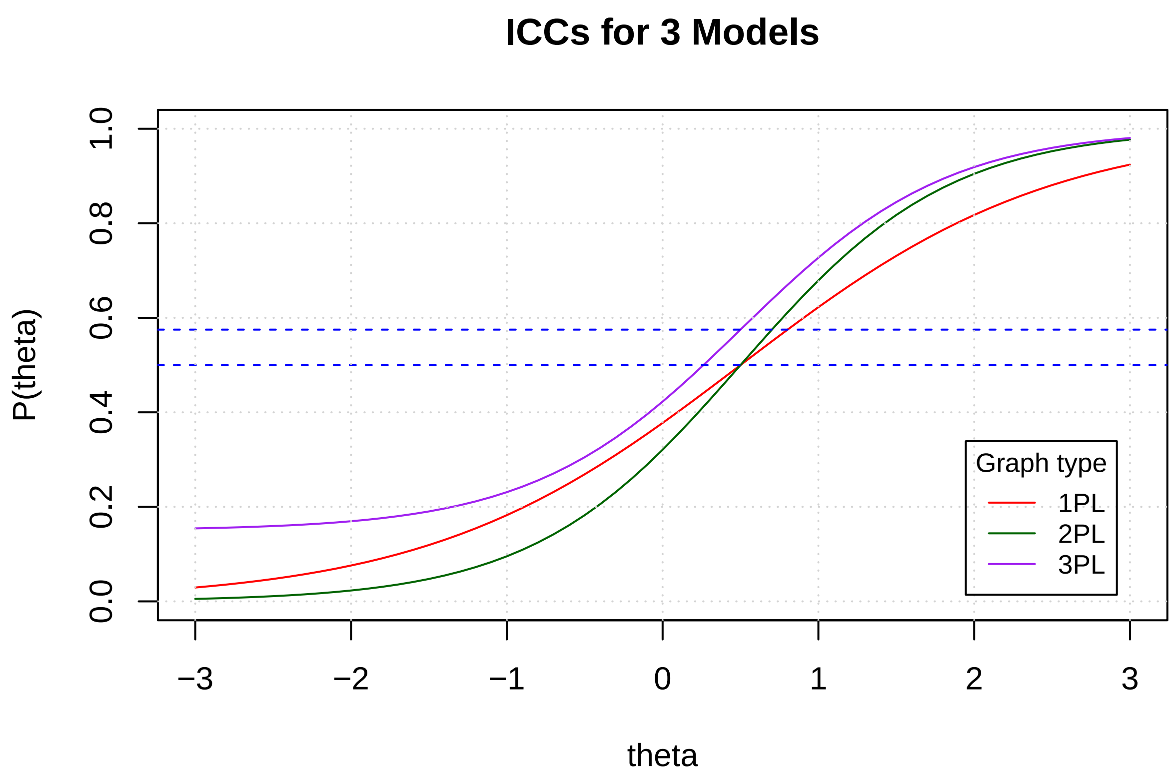
Then, I calculate the probability vector for each model.

```
> # for the 1Pl model, b=0.5
> P_1pl <- logit_2pl(theta=theta, a=1, b=0.5)
> # for the 2PL model, a=1.5, b=0.5
> P_2pl <- logit_2pl(theta = theta, a=1.5, b=0.5)
> # for the 3PL model, a=1.5, b=0.5, c=0.15
> P_3pl <- logit_3pl(theta = theta, a=1.5, b=0.5, c=0.15)
```

Next, plot all 3 ICCs on same plot.

```
> plot(theta, P_1pl, type = "l",
+      col="red", xlim = c(xlim_min,xlim_max), ylim = c(0,1),
+      main = "ICCs for 3 Models", ylab = "P(theta)")
```

```
> lines(theta, P_2pl, col="darkgreen")
> lines(theta, P_3pl, col="purple")
> abline(h=0.5, col="blue", lty=2)
> abline(h=0.575, col="blue", lty=2)
> grid()
> # add legend
> legend('bottomright',inset=0.05,
+        c("1PL","2PL","3PL"),lty=1,
+        col=c("red","darkgreen","purple"),title="Graph type",
+        cex=0.8)
```



Q6

For the three-parameter logistic model ...

My Solution:
By definition, the 3PL model is

$$P(\theta) = c + (1-c)\frac{1}{1+e^{-\alpha(\theta-\beta)}}.$$

Set the $\theta = \beta$, we have

$$P(\theta = \beta) = c + (1-c)\frac{1}{1+e^{0}}.$$

Since the $e^0 = 1$, finally, we have

$$P(\theta = \beta) = c + (1 - c)\frac{1}{2} = \frac{1}{2}(1 + c).$$