hw3

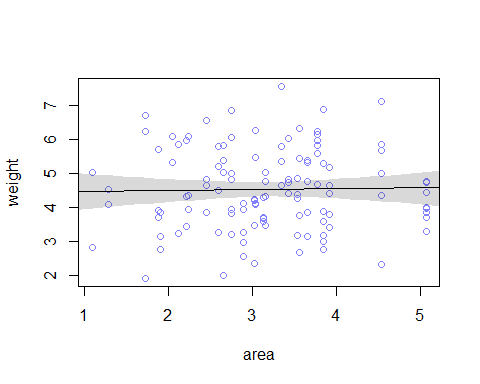
GISH

2019撟3<9c><88>27

### Q1

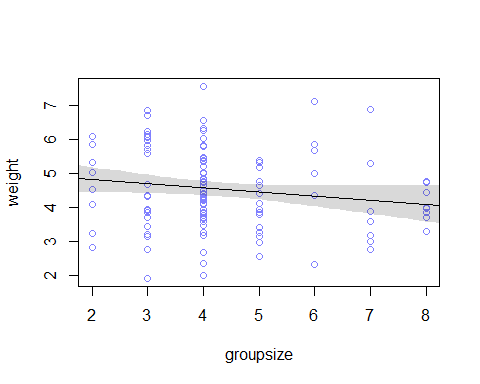
## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]

## Warning in abline(modelA): only using the first two of 3 regression  
## coefficients



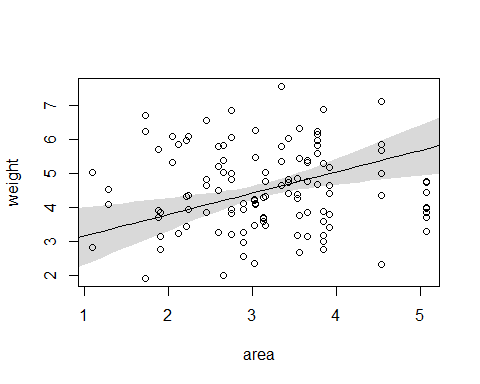
## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]

## Warning in abline(modelB): only using the first two of 3 regression  
## coefficients

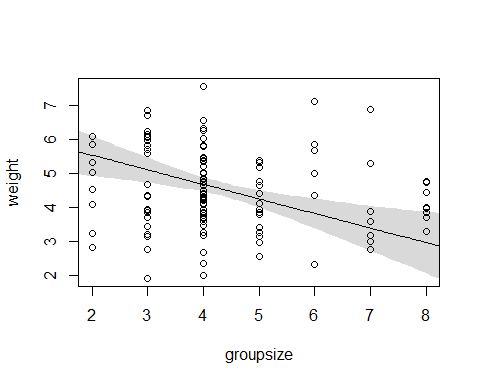
 # 從兩張圖來看，資料都非常散佈，所畫出的回歸與區間只包含到一點點的資料 # 以area來說，資料幾乎看不出什麼關係，所求的回歸線斜率(bArea幾乎為0，得出area與weight從此張圖來看沒有顯著線性關係 # 以groupsize來說，groupsize為discrete資料，且看起來呈現散狀，看起來也與weight沒什麼關係，所求回歸線斜率(bgroupsize)也接近0，得出groupsize與weight從此張圖來看沒有顯著線性關係

### Q2

## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]



## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]

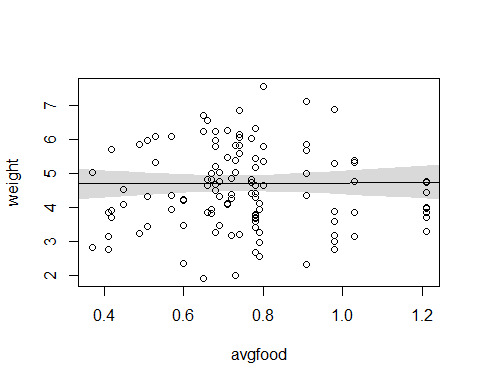


# 從結果可以發現兩者回歸斜率的絕對值變大了，把另一個預測變數設為不變的情況下，該變數對weight的影響會比較顯著

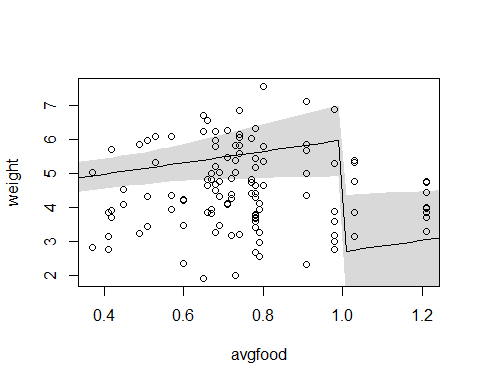
# 可以猜測兩者(area與groupsize)之間可能有很大的相關(correlation)，兩者變數對weight的影響卻為一正一負，剛好抵銷單一變數對weight的影響，使得模型只放單一變數的時候，斜率會非常小

### Q3

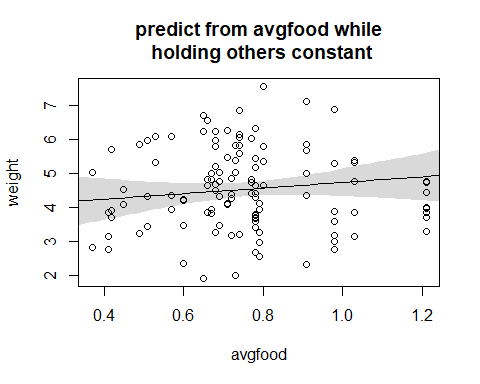
## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]



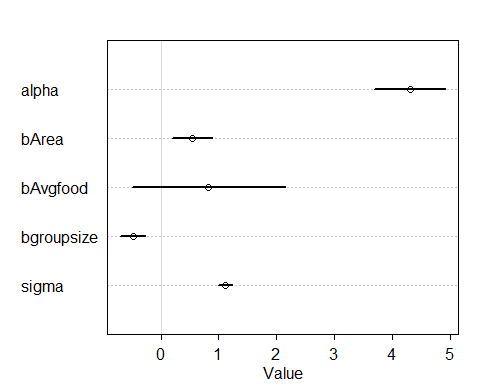
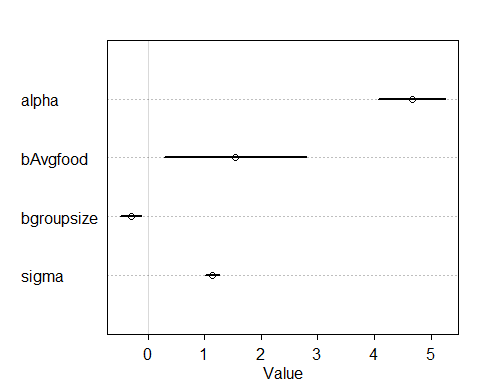
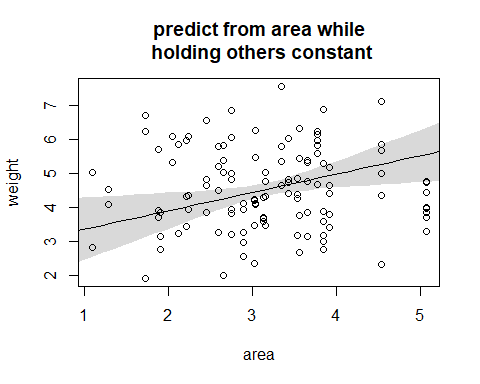
## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]



## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]



## [ 100 / 1000 ]  
[ 200 / 1000 ]  
[ 300 / 1000 ]  
[ 400 / 1000 ]  
[ 500 / 1000 ]  
[ 600 / 1000 ]  
[ 700 / 1000 ]  
[ 800 / 1000 ]  
[ 900 / 1000 ]  
[ 1000 / 1000 ]



# (a)我認為avgfood比較能預測weight,在沒有手動控制其他變數的情況下，從敘述統計來看，在三個變數的模型中bAvgfood的值較bArea來得大，表示average food與weight的比較有關係

# 而且從第二小題可以看出area與groupsize高度正相關且會互相抵銷影響，如果模型只放入其中之一的變數但沒有控制住另一個變數，常會讓人誤以為該變數對weight的影響不大

# (b)與area跟groupsize會互相抵銷影響一樣，avgfood與area也有高度正相關，發生線性重合問題

# 但如果用人為方法控制變數(像是上面我畫的)，把其他變數控制在平均，可以發現其實area這個變數斜率反而比較大，在控制了其他的變數情況下，其實area對於weight比較顯著