

# simulate\_BMI-1.R

*danny*

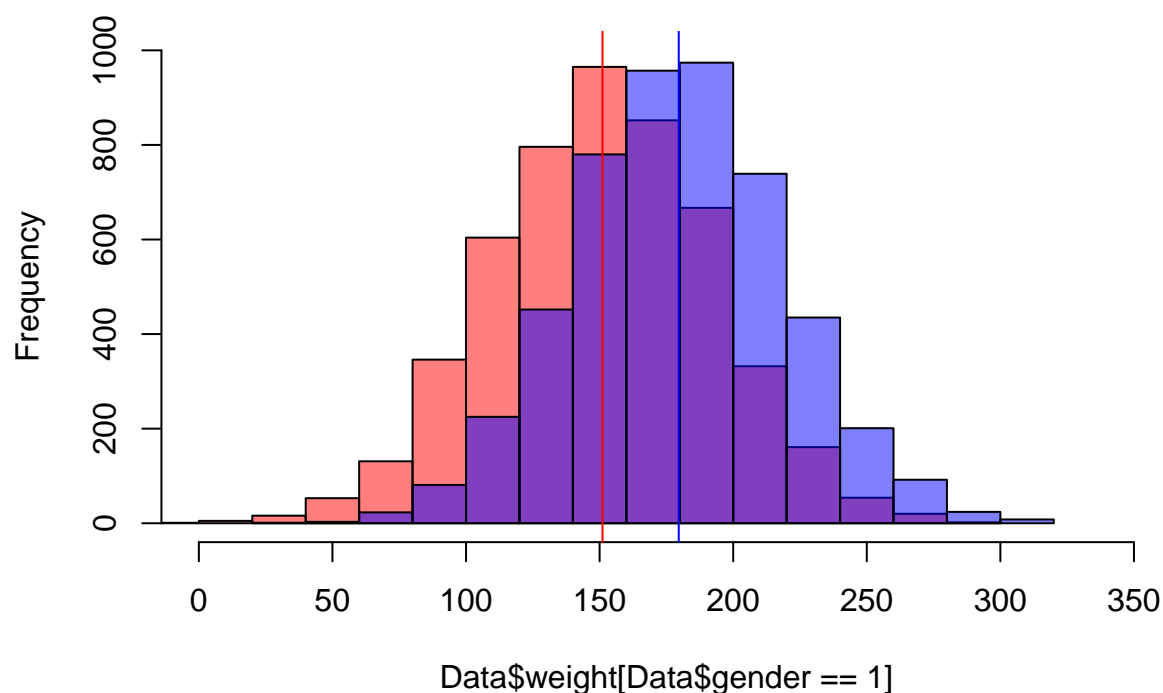
*2020-02-05*

```
# BMI example  
# Code below is used to generate simulated dataset  
  
set.seed(1000)  
gender=sample(0:1,10000,TRUE)  
#weight in lb for men and women  
weight=rnorm(10000, mean = 180, sd = 40) + gender*rnorm(10000,mean=-30,sd=10)  
#height in in for men and women  
height=rnorm(10000, mean = 60, sd = 1) + gender*rnorm(10000,mean=-6,sd=0.5)  
#using general formula  
bmi=(weight*703)/(height*height)+rnorm(10000, mean=0, sd = 3)  
  
Data<-data.frame(gender, weight, height, bmi)  
head(Data)
```

```
##  gender  weight  height    bmi  
## 1      1 233.7009 53.42704 58.05625  
## 2      1 137.1929 55.02970 29.92295  
## 3      0 242.2719 60.70448 43.60021  
## 4      1 108.3572 53.18934 25.30988  
## 5      0 200.4341 59.05139 42.82229  
## 6      1 169.1344 55.21242 42.28387
```

```
hist(Data$weight[Data$gender==1],col=rgb(1,0,0,.5),ylim=c(0,1000),xlim=c(0,350))  
hist(Data$weight[Data$gender==0],col=rgb(0,0,1,.5),add=T)  
abline(v=mean(Data$weight[Data$gender==1]),col="red")  
abline(v=mean(Data$weight[Data$gender==0]),col="blue")
```

# Histogram of Data\$weight[Data\$gender == 1]



```
# check averages
mean(weight)
```

```
## [1] 165.3459
```

```
summary(lm(weight~1)) ## run a regression to calculate the mean (run the regression against 1)
```

```
##
## Call:
## lm(formula = weight ~ 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -170.462  -28.729    0.436   28.656  148.521
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 165.3459    0.4296   384.9   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42.96 on 9999 degrees of freedom
```

```
summary(lm(weight~gender)) #average weight
```

```
##
## Call:
## lm(formula = weight ~ gender)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -156.211  -27.464    0.365   27.439  137.001
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  179.6254     0.5734   313.3  <2e-16 ***
## gender       -28.5304     0.8104   -35.2  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40.52 on 9998 degrees of freedom
## Multiple R-squared:  0.1103, Adjusted R-squared:  0.1102
## F-statistic: 1239 on 1 and 9998 DF, p-value: < 2.2e-16
```

```
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 179.6254     0.5734   313.3  <2e-16 ***
## gender      -28.5304     0.8104   -35.2  <2e-16 ***
## Men's average weight is 179 lbs.
## Women's average weight is 28 lbs lighter than men.
```

```
summary(lm(height~gender)) #average height
```

```
##
## Call:
## lm(formula = height ~ gender)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##  -4.2445  -0.7291  -0.0108   0.7098   3.5688
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  60.00062    0.01515  3959.2  <2e-16 ***
## gender       -5.98404    0.02142  -279.4  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.071 on 9998 degrees of freedom
## Multiple R-squared:  0.8864, Adjusted R-squared:  0.8864
## F-statistic: 7.804e+04 on 1 and 9998 DF, p-value: < 2.2e-16
```

```
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 60.00062    0.01515  3959.2  <2e-16 ***
## gender      -5.98404    0.02142  -279.4  <2e-16 ***
## Men's average height is 60 inches
```

```

## Women's average height is ~6 inches shorter.

# BMI as a function of weight and height #

summary(lm(log(bmi)~log(weight)+log(height))) #general formula

## Warning in log(weight): NaNs produced

##
## Call:
## lm(formula = log(bmi) ~ log(weight) + log(height))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.67898 -0.05467  0.00309  0.05895  1.50471
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.624136   0.072464   91.41  <2e-16 ***
## log(weight)  1.015946   0.003547  286.45  <2e-16 ***
## log(height) -2.038372   0.018753 -108.70  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09997 on 9996 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.8921, Adjusted R-squared:  0.8921
## F-statistic: 4.131e+04 on 2 and 9996 DF, p-value: < 2.2e-16

## Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.624136   0.072464   91.41  <2e-16 ***
## log(weight)  1.015946   0.003547  286.45  <2e-16 ***
## log(height) -2.038372   0.018753 -108.70  <2e-16 ***

## Intercept: Not meaningful (geometric mean for people with 0 weight and 0 height)

## When we have logs on both sides, we can interpret the coefficient like this:
## for every 1% increase in weight (lbs), we increase BMI by 1%
## for every 10% increase in height (in), we decrease BMI by 18%

summary(lm(log(bmi)~log(weight)*gender+log(height)*gender)) #adjusted for men and women

```

```

## Warning in log(weight): NaNs produced

##
## Call:
## lm(formula = log(bmi) ~ log(weight) * gender + log(height) *
##      gender)
##
## Residuals:
##      Min       1Q   Median       3Q      Max

```

```
## -1.67662 -0.05460 0.00326 0.05919 1.50986
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.733596   0.344313  19.557 <2e-16 ***
## log(weight)       1.012185   0.005982 169.212 <2e-16 ***
## gender           -0.226349   0.438641  -0.516  0.606
## log(height)      -2.060376   0.083729 -24.608 <2e-16 ***
## log(weight):gender 0.005958   0.007460   0.799  0.424
## gender:log(height) 0.048578   0.107759   0.451  0.652
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09998 on 9993 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.8921, Adjusted R-squared:  0.892
## F-statistic: 1.652e+04 on 5 and 9993 DF, p-value: < 2.2e-16
```

```
## Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.733596   0.344313  19.557 <2e-16 ***
## log(weight)       1.012185   0.005982 169.212 <2e-16 ***
## gender           -0.226349   0.438641  -0.516  0.606
## log(height)      -2.060376   0.083729 -24.608 <2e-16 ***
## log(weight):gender 0.005958   0.007460   0.799  0.424
## gender:log(height) 0.048578   0.107759   0.451  0.652
```

```
## gender: Essentially 0, there is no relationship with gender and BMI
## log(weight): gender: The interaction between log(weight) and gender is insignificant
## gender: log(height): The interaction between log(height) and gender is insignificant
```