```
getwd()
```

```
## [1] "C:/Users/Nupur Shrinet/Desktop/Market research &enginerring"
```

```
sales.data.raw<-read.csv("sales.csv")
```

```
summary(sales.data.raw)
```

```
##     acctAge       visitsMonth      spendToDate       spendMonth
##  Min.   : 1.00   Min.   : 1.000   Min.   :   6.0   Min.   :   4.0
##  1st Qu.: 8.00   1st Qu.: 6.000   1st Qu.:  28.0   1st Qu.:   9.0
##  Median :13.00   Median : 7.000   Median :  45.0   Median :  17.0
##  Mean   :12.84   Mean   : 7.181   Mean   : 125.9   Mean   :  34.2
##  3rd Qu.:18.00   3rd Qu.: 9.000   3rd Qu.: 100.5   3rd Qu.:  34.0
##  Max.   :24.00   Max.   :15.000   Max.   :1196.0   Max.   :1041.0
##
##     satSite         satQuality        satPrice        satOverall
##  Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
##  1st Qu.: 5.000   1st Qu.: 5.000   1st Qu.: 5.000   1st Qu.: 5.000
##  Median : 6.000   Median : 6.000   Median : 6.000   Median : 6.000
##  Mean   : 5.715   Mean   : 6.011   Mean   : 5.757   Mean   : 5.696
##  3rd Qu.: 7.000   3rd Qu.: 7.000   3rd Qu.: 7.000   3rd Qu.: 7.000
##  Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000
##
##        region        coupon          purchase
##  Mideast  :219   Min.   :0.0000   Min.   :0.00000
##  Midwest  :135   1st Qu.:0.0000   1st Qu.:0.00000
##  Mountains: 12   Median :0.0000   Median :0.00000
##  Northeast: 50   Mean   :0.3389   Mean   :0.03114
##  South    :247   3rd Qu.:1.0000   3rd Qu.:0.00000
##  Southwest:135   Max.   :1.0000   Max.   :1.00000
##  West     : 37
```

```
str(sales.data.raw)
```

```
## 'data.frame':    835 obs. of  11 variables:
##  $ acctAge    : int  21 9 16 4 16 19 23 6 16 17 ...
##  $ visitsMonth: int  9 9 7 8 6 2 7 4 12 13 ...
##  $ spendToDate: int  21 55 64 61 27 32 84 29 261 106 ...
##  $ spendMonth : int  21 55 17 8 9 32 35 9 50 10 ...
##  $ satSite    : int  6 4 6 6 5 6 4 6 6 5 ...
##  $ satQuality : int  5 5 6 6 5 8 4 6 7 4 ...
##  $ satPrice   : int  6 4 7 6 6 5 4 6 6 5 ...
##  $ satOverall : int  6 4 7 6 6 4 4 5 6 4 ...
##  $ region     : Factor w/ 7 levels "Mideast","Midwest",..: 1 6 1 6 2 5 5 3 1 2 ...
##  $ coupon     : int  0 0 1 1 0 0 0 0 0 0 ...
##  $ purchase   : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
#Answer #1
library(car)
```

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

```
mod.raw1 <- lm(spendMonth ~ ., sales.data.raw)
summary(mod.raw1)
```

```
##
## Call:
## lm(formula = spendMonth ~ ., data = sales.data.raw)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -174.02  -19.82   -7.00    7.85  833.86
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      27.36084   12.75432   2.145   0.0322 *
## acctAge           0.22165    0.37661   0.589   0.5563
## visitsMonth       0.76715    1.05809   0.725   0.4686
## spendToDate       0.10161    0.01062   9.568  < 2e-16 ***
## satSite          -3.11972    4.80385  -0.649   0.5162
## satQuality        3.74817    2.27209   1.650   0.0994 .
## satPrice          1.45999    4.97994   0.293   0.7695
## satOverall       -4.80922    3.18069  -1.512   0.1309
## regionMidwest    -3.08636    6.82112  -0.452   0.6510
## regionMountains  -1.10005   18.81345  -0.058   0.9534
## regionNortheast -13.61680    9.78545  -1.392   0.1644
## regionSouth      -4.69717    5.83054  -0.806   0.4207
## regionSouthwest  -8.06633    6.93934  -1.162   0.2454
## regionWest      -12.09836   11.28447  -1.072   0.2840
## coupon            6.54916    4.89690   1.337   0.1815
## purchase         74.34426   13.08709   5.681 1.86e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61.92 on 819 degrees of freedom
## Multiple R-squared:  0.1565, Adjusted R-squared:  0.141
## F-statistic: 10.13 on 15 and 819 DF,  p-value: < 2.2e-16
```

```
#we should check for collinearity because #R-square is very low and standard error is very high for som
```

```
vif(mod.raw1)
```

```
##                  GVIF Df GVIF^(1/(2*Df))
## acctAge      1.030096  1        1.014937
## visitsMonth  1.214486  1        1.102037
## spendToDate  1.012564  1        1.006262
## satSite     11.784645  1        3.432877
## satQuality   2.903826  1        1.704062
## satPrice    13.838160  1        3.719968
## satOverall   5.975156  1        2.444413
## region       1.210931  6        1.016077
## coupon       1.169956  1        1.081645
## purchase     1.125148  1        1.060730
```

```r
# VIF>5 is a problem we refit the model removing satsite and satprice
```

```r
mod.raw2 <- lm(spendMonth ~ . - satSite - satPrice, sales.data.raw)
summary(mod.raw2)
```

```
##
## Call:
## lm(formula = spendMonth ~ . - satSite - satPrice, data = sales.data.raw)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -171.87  -19.74   -7.23    7.45  833.43
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      28.29500   12.65260   2.236 0.025600 *
## acctAge           0.21786    0.37603   0.579 0.562509
## visitsMonth       0.61471    1.01894   0.603 0.546487
## spendToDate       0.10154    0.01061   9.570  < 2e-16 ***
## satQuality        2.95012    1.62525   1.815 0.069861 .
## satOverall       -5.54966    1.66508  -3.333 0.000898 ***
## regionMidwest    -3.42900    6.79483  -0.505 0.613941
## regionMountains  -2.15716   18.54701  -0.116 0.907437
## regionNortheast -14.05713    9.75343  -1.441 0.149895
## regionSouth      -5.08406    5.76951  -0.881 0.378471
## regionSouthwest  -8.21644    6.89997  -1.191 0.234078
## regionWest      -12.62710   11.23608  -1.124 0.261427
## coupon            6.79339    4.85453   1.399 0.162073
## purchase         73.87912   13.05001   5.661 2.08e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61.86 on 821 degrees of freedom
## Multiple R-squared:  0.156,  Adjusted R-squared:  0.1427
## F-statistic: 11.67 on 13 and 821 DF,  p-value: < 2.2e-16
```

```r
vif(mod.raw2)
```

```
##                 GVIF Df GVIF^(1/(2*Df))
## acctAge     1.028891  1        1.014343
## visitsMonth 1.128422  1        1.062272
## spendToDate 1.012418  1        1.006190
## satQuality  1.488626  1        1.220093
## satOverall  1.640597  1        1.280858
## region      1.151821  6        1.011848
## coupon      1.151993  1        1.073309
## purchase    1.120915  1        1.058733
```

```r
#Corrected VIF with less than 5 value showing no collinearity
```

```r
#Answer2 for transforming the variables we use BoxCox from the forecast package
autoTransform <- function(x) {
  library(forecast)
  return(scale(BoxCox(x, BoxCox.lambda(x))))
}


sales.data <- sales.data.raw
sales.data[ , -9] <- lapply(sales.data.raw[ , -9], autoTransform)# Normalizing the numeric columns
```

```
## Warning: package 'forecast' was built under R version 3.6.3
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
mod.trns <- lm(spendMonth ~ . - satSite - satPrice, sales.data)
summary(mod.trns)
```

```
##
## Call:
## lm(formula = spendMonth ~ . - satSite - satPrice, data = sales.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.15330 -0.78249  0.07007  0.73324  2.02284
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.047437   0.065388   0.725  0.46837
## acctAge          0.001125   0.033781   0.033  0.97345
## visitsMonth     -0.011527   0.035151  -0.328  0.74305
## spendToDate      0.189670   0.033491   5.663 2.05e-08 ***
## satQuality       0.041778   0.040593   1.029  0.30370
## satOverall      -0.115826   0.042683  -2.714  0.00679 **
## regionMidwest   -0.070281   0.105716  -0.665  0.50636
## regionMountains -0.244129   0.288105  -0.847  0.39704
## regionNortheast -0.156577   0.151733  -1.032  0.30241
## regionSouth     -0.038861   0.089771  -0.433  0.66521
## regionSouthwest -0.088110   0.107412  -0.820  0.41228
## regionWest       0.057571   0.174525   0.330  0.74158
## coupon           0.036644   0.035728   1.026  0.30536
## purchase         0.206760   0.035030   5.902 5.24e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9615 on 821 degrees of freedom
## Multiple R-squared:  0.08996,    Adjusted R-squared:  0.07555
## F-statistic: 6.243 on 13 and 821 DF,  p-value: 2.724e-11
```

```r
vif(mod.trns)
```

```
##                 GVIF Df GVIF^(1/(2*Df))
## acctAge     1.029512  1        1.014649
## visitsMonth 1.114724  1        1.055805
## spendToDate 1.011893  1        1.005929
## satQuality  1.486594  1        1.219260
## satOverall  1.643603  1        1.282031
## region      1.153755  6        1.011990
## coupon      1.151608  1        1.073130
## purchase    1.107056  1        1.052167
```

```r
#Answer3
sat.pc <- prcomp(sales.data[ , 5:8])
summary(sat.pc)
```

```
## Importance of components:
##                           PC1    PC2    PC3     PC4
## Standard deviation     1.8404 0.6811 0.3143 0.22377
## Proportion of Variance 0.8468 0.1160 0.0247 0.01252
## Cumulative Proportion  0.8468 0.9628 0.9875 1.00000
```

```r
sales.data$satPC <- sat.pc$x[ , 1]
mod.trns2 <- lm(spendMonth ~ . , sales.data[ , c(-5, -6, -7, -8)])
summary(mod.trns2)
```

```
##
## Call:
## lm(formula = spendMonth ~ ., data = sales.data[, c(-5, -6, -7,
##     -8)])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.19027 -0.79163  0.06365  0.73625  2.03701
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.046513   0.065480   0.710    0.478
## acctAge         -0.001541   0.033806  -0.046    0.964
## visitsMonth     -0.007238   0.035510  -0.204    0.839
## spendToDate      0.191888   0.033522   5.724 1.46e-08 ***
## regionMidwest   -0.068493   0.105936  -0.647    0.518
## regionMountains -0.237397   0.288710  -0.822    0.411
## regionNortheast -0.145963   0.151971  -0.960    0.337
## regionSouth     -0.037440   0.089907  -0.416    0.677
## regionSouthwest -0.090496   0.107686  -0.840    0.401
## regionWest       0.054598   0.174614   0.313    0.755
## coupon           0.031186   0.035619   0.876    0.382
## purchase         0.207403   0.035134   5.903 5.21e-09 ***
## satPC            0.041262   0.020839   1.980    0.048 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.963 on 822 degrees of freedom
## Multiple R-squared:  0.08605,    Adjusted R-squared:  0.07271
## F-statistic:  6.45 on 12 and 822 DF,  p-value: 4.744e-11
```

```
#The coefficients are smaller now because the data have been standardized.
```

```
#Answer 4
sales.data$purchase <- sales.data.raw$purchase # Resetting the purchase variable from original dataset
purchase.lr1<- glm(purchase ~ coupon, data=sales.data, family=binomial)
summary(purchase.lr1)
```

```
##
## Call:
## glm(formula = purchase ~ coupon, family = binomial, data = sales.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.3728  -0.3728  -0.1598  -0.1598   2.9556
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.7708     0.2640 -14.281  < 2e-16 ***
## coupon        0.8162     0.2124   3.843 0.000122 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 231.59  on 834  degrees of freedom
## Residual deviance: 214.39  on 833  degrees of freedom
## AIC: 218.39
##
## Number of Fisher Scoring iterations: 7
```

```
# There is positive relation between purchase and coupon and the effect is statistically significant, c
```

```
#Answer 5
purchase.lr2 <- glm(purchase ~ coupon + spendToDate + region + satPC,data=sales.data, family=binomial)
summary(purchase.lr2)
```

```
##
## Call:
## glm(formula = purchase ~ coupon + spendToDate + region + satPC,
##     family = binomial, data = sales.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3084  -0.2330  -0.1437  -0.0870   3.5641
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)       -4.2390     0.4789  -8.852  < 2e-16 ***
## coupon             0.3943     0.2324   1.696   0.0898 .
## spendToDate        0.1658     0.2125   0.780   0.4351
## regionMidwest     -0.6461     0.6443  -1.003   0.3159
## regionMountains  -12.2614  1093.3447  -0.011   0.9911
## regionNortheast    0.1445     0.6964   0.207   0.8356
## regionSouth       -0.2645     0.5708  -0.463   0.6432
## regionSouthwest   -0.1294     0.8382  -0.154   0.8773
## regionWest         0.7853     1.1360   0.691   0.4894
## satPC             -0.6909     0.1399  -4.937 7.92e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 231.59  on 834  degrees of freedom
## Residual deviance: 181.18  on 825  degrees of freedom
## AIC: 201.18
##
## Number of Fisher Scoring iterations: 16
```

```
# The coefficient for coupon just became significant and also coefficient value decreased
```

```
#Answer 6
purchase.lr3 <- glm(purchase ~ coupon + satPC + coupon:satPC,data=sales.data, family=binomial)
summary(purchase.lr3)
```

```
##
## Call:
## glm(formula = purchase ~ coupon + satPC + coupon:satPC, family = binomial,
##     data = sales.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1570  -0.2241  -0.1524  -0.0967   3.4014
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.33999    0.35230 -12.319  < 2e-16 ***
## coupon        0.34263    0.31741   1.079 0.280384
## satPC        -0.63969    0.17263  -3.706 0.000211 ***
## coupon:satPC -0.04414    0.13958  -0.316 0.751842
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 231.59  on 834  degrees of freedom
## Residual deviance: 183.86  on 831  degrees of freedom
## AIC: 191.86
##
## Number of Fisher Scoring iterations: 8
```

```r
#There is no interaction between coupon and satPC as the coefficient is insignificant

#Answer 7
exp(coef(purchase.lr3))
```

```
##  (Intercept)       coupon        satPC coupon:satPC
##   0.01303662   1.40865205   0.52745405   0.95682462
```

```r
#Or
plogis(0.34263) / (1-plogis(0.34263))
```

```
## [1] 1.408647
```

```r
#This implies that customers are 1.40 times more likely to purchase the product when they have a coupon

#Answer 8
conjoint.df<- read.csv("bag.csv")
str(conjoint.df)
```

```
## 'data.frame':    4500 obs. of  6 variables:
##  $ resp.id: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ rating : int  3 10 9 6 2 6 10 5 2 7 ...
##  $ price  : int  17 15 15 19 19 15 15 17 19 17 ...
##  $ color  : Factor w/ 3 levels "black","gray",..: 2 3 1 3 3 2 1 3 3 3 ...
##  $ zipper : Factor w/ 2 levels "gold","silver": 2 2 1 1 2 2 2 2 2 1 ...
##  $ finish : Factor w/ 2 levels "matte","patent": 2 1 2 1 2 2 1 2 2 2 ...
```

```r
bag.lm <- lm(rating ~ price + color + zipper + finish, data=conjoint.df)
summary(bag.lm)
```

```
##
## Call:
## lm(formula = rating ~ price + color + zipper + finish, data = conjoint.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.2365 -1.5281  0.0228  1.4719  6.0228
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.24843    0.29332  45.167   <2e-16 ***
## price        -0.14309    0.01623  -8.817   <2e-16 ***
## colorgray    -2.02827    0.09064 -22.377   <2e-16 ***
## colornavy    -4.00164    0.08329 -48.044   <2e-16 ***
## zippersilver -0.86566    0.07604 -11.384   <2e-16 ***
## finishpatent -1.97148    0.06322 -31.183   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.015 on 4494 degrees of freedom
## Multiple R-squared:  0.4478, Adjusted R-squared:  0.4471
## F-statistic: 728.7 on 5 and 4494 DF,  p-value: < 2.2e-16
```

```r
#Answer 9
library(lme4)
```

```
## Loading required package: Matrix
```

```
## Registered S3 methods overwritten by 'lme4':
##   method                          from
##   cooks.distance.influence.merMod car
##   influence.merMod                car
##   dfbeta.influence.merMod         car
##   dfbetas.influence.merMod        car
```

```r
# model with random intercept & slope by respondent = (predictors | resp.id)
bag.hlm1 <- lmer(rating ~ price + color + zipper + finish +
                   (price + color + zipper + finish | resp.id),
                 data=conjoint.df,
                 control=lmerControl(optCtrl=list(maxfun=100000)))
```

```
## boundary (singular) fit: see ?isSingular
```

```r
#Ans 9b
# population estimate
fixef(bag.hlm1)
```

```
##  (Intercept)        price    colorgray    colornavy zippersilver finishpatent
##   13.2484297   -0.1430868   -2.0282684   -4.0016437   -0.8656622   -1.9714790
```

```r
#Ans 9c
head(ranef(bag.hlm1)$resp.id,4)
```

```
##   (Intercept)       price   colorgray   colornavy zippersilver finishpatent
## 1 -0.73588279  0.07137593  0.04778409  0.07592027  -0.02824101  -0.68788093
## 2 -1.96442837  0.10949705 -0.48502561 -1.20609445   1.01423206  -0.32917394
## 3 -0.83604993  0.02549676 -0.36595470 -0.17512117   0.37242325  -0.09238338
## 4 -0.02164348 -0.02440025 -0.19890568 -0.02295124   0.42991928  -0.01893050
```

```r
#Ans 9d
head(coef(bag.hlm1)$resp.id)
```

```
##   (Intercept)       price colorgray colornavy zippersilver finishpatent
## 1    12.51255 -0.07171087 -1.980484 -3.925723   -0.8939032   -2.6593599
## 2    11.28400 -0.03358975 -2.513294 -5.207738    0.1485699   -2.3006529
## 3    12.41238 -0.11759004 -2.394223 -4.176765   -0.4932389   -2.0638623
## 4    13.22679 -0.16748705 -2.227174 -4.024595   -0.4357429   -1.9904095
## 5    16.48041 -0.39005808 -1.735381 -2.338073   -1.9318254   -0.7084375
## 6    11.89763 -0.02912665 -2.069505 -4.205657   -0.9490550   -2.6179757
```

```r
#Ans 9e
coef(bag.hlm1)$resp.id[c(130,250),]
```

```
##     (Intercept)       price colorgray colornavy zippersilver finishpatent
## 130    11.79584  0.01481119 -1.805397 -3.488105    -1.512277    -3.358628
## 250    11.36265 -0.01017174 -2.283717 -2.875021    -1.603685    -3.236084
```