STATS 500 HW7

Minxuan Chen

2023-11-08

Table of contents

Github repo: https://github.com/PKUniiiiice/STATS_500

Problem 1

(a)

```
library(faraway)
data(teengamb)

m1 <- lm(gamble ~ ., data=teengamb)
summary(m1)
```

```
Call:
lm(formula = gamble ~ ., data = teengamb)

Residuals:
    Min      1Q  Median      3Q     Max
-51.082 -11.320  -1.451   9.452  94.252

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  22.55565   17.19680   1.312   0.1968
sex         -22.11833    8.21111  -2.694   0.0101 *
status        0.05223    0.28111   0.186   0.8535
income        4.96198    1.02539   4.839 1.79e-05 ***
verbal       -2.95949    2.17215  -1.362   0.1803
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22.69 on 42 degrees of freedom
Multiple R-squared:  0.5267,    Adjusted R-squared:  0.4816
F-statistic: 11.69 on 4 and 42 DF,  p-value: 1.815e-06
```

```
m1 <- update(m1, . ~ . - status)
summary(m1)
```

```
Call:
lm(formula = gamble ~ sex + income + verbal, data = teengamb)

Residuals:
    Min      1Q  Median      3Q     Max
-50.639 -11.765  -1.594   9.305  93.867

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  24.1390    14.7686   1.634   0.1095
```

```
sex         -22.9602     6.7706  -3.391   0.0015 **
income        4.8981     0.9551   5.128 6.64e-06 ***
verbal       -2.7468     1.8253  -1.505   0.1397
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22.43 on 43 degrees of freedom
Multiple R-squared:  0.5263,    Adjusted R-squared:  0.4933
F-statistic: 15.93 on 3 and 43 DF,  p-value: 4.148e-07
```

```r
m1 <- update(m1, . ~ . - verbal)
summary(m1)
```

```
Call:
lm(formula = gamble ~ sex + income, data = teengamb)

Residuals:
    Min      1Q  Median      3Q     Max
-49.757 -11.649   0.844   8.659 100.243

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.041      6.394   0.632  0.53070
sex          -21.634      6.809  -3.177  0.00272 **
income         5.172      0.951   5.438 2.24e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22.75 on 44 degrees of freedom
Multiple R-squared:  0.5014,    Adjusted R-squared:  0.4787
F-statistic: 22.12 on 2 and 44 DF,  p-value: 2.243e-07
```

The best model is the last summary output shown above, in which `sex` and `income` are used as predictors.

(b)

```r
#AIC
m2 <- lm(gamble ~ ., data=teengamb)
step(m2)
```

```
Start:  AIC=298.18
gamble ~ sex + status + income + verbal
```

```
         Df Sum of Sq   RSS    AIC
- status  1      17.8 21642 296.21
<none>                21624 298.18
- verbal  1     955.7 22580 298.21
- sex     1    3735.8 25360 303.67
- income  1   12056.2 33680 317.00

Step:  AIC=296.21
gamble ~ sex + income + verbal

         Df Sum of Sq   RSS    AIC
<none>                21642 296.21
- verbal  1    1139.8 22781 296.63
- sex     1    5787.9 27429 305.35
- income  1   13236.1 34878 316.64


Call:
lm(formula = gamble ~ sex + income + verbal, data = teengamb)

Coefficients:
(Intercept)          sex       income       verbal
     24.139      -22.960        4.898       -2.747
```

The best model selected by AIC is that uses `sex`, `income` and `verbal` as predictors.

(c)

```r
#Adjusted R^2
library(leaps)
m3 <- regsubsets(gamble ~ ., data=teengamb)
res <- summary(m3); res
```

```
Subset selection object
Call: regsubsets.formula(gamble ~ ., data = teengamb)
4 Variables  (and intercept)
       Forced in Forced out
sex        FALSE      FALSE
status     FALSE      FALSE
income     FALSE      FALSE
verbal     FALSE      FALSE
1 subsets of each size up to 4
Selection Algorithm: exhaustive
         sex status income verbal
```

```
1  ( 1 ) " " " "    "*"    " "
2  ( 1 ) "*" " "    "*"    " "
3  ( 1 ) "*" " "    "*"    "*"
4  ( 1 ) "*" "*"    "*"    "*"
```

```r
#select model with largest adjusted r^2
which.max(res$adjr2)
```

```
[1] 3
```

```r
res$adjr2[3]
```

```
[1] 0.4932879
```

```r
summary(lm(gamble ~ .-status, data=teengamb))
```

```
Call:
lm(formula = gamble ~ . - status, data = teengamb)

Residuals:
    Min      1Q  Median      3Q     Max
-50.639 -11.765  -1.594   9.305  93.867

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  24.1390    14.7686   1.634   0.1095
sex         -22.9602     6.7706  -3.391   0.0015 **
income        4.8981     0.9551   5.128 6.64e-06 ***
verbal       -2.7468     1.8253  -1.505   0.1397
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22.43 on 43 degrees of freedom
Multiple R-squared:  0.5263,    Adjusted R-squared:  0.4933
F-statistic: 15.93 on 3 and 43 DF,  p-value: 4.148e-07
```

The best model selected by Adjusted R^2 is that uses `sex`, `income` and `verbal` as predictors, which is the same as the choice made by AIC.

(d)

```r
# Mallows Cp
which.min(res$cp)
```

```
[1] 3
```

```r
res$cp[3]
```

```
[1] 3.034526
```

```r
summary(lm(gamble ~ .-status, data=teengamb))
```

```
Call:
lm(formula = gamble ~ . - status, data = teengamb)

Residuals:
    Min      1Q  Median      3Q     Max
-50.639 -11.765  -1.594   9.305  93.867

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  24.1390    14.7686   1.634   0.1095
sex         -22.9602     6.7706  -3.391   0.0015 **
income        4.8981     0.9551   5.128 6.64e-06 ***
verbal       -2.7468     1.8253  -1.505   0.1397
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22.43 on 43 degrees of freedom
Multiple R-squared:  0.5263,    Adjusted R-squared:  0.4933
F-statistic: 15.93 on 3 and 43 DF,  p-value: 4.148e-07
```

The best model selected by Mallows C_p is that uses `sex`, `income` and `verbal` as predictors, which is the same as the choice made by AIC and Adjusted R^2.

Problem 2

(a)

```r
data("seatpos")
m.sea <- lm(hipcenter ~ ., data=seatpos)
summary(m.sea)
```

```
Call:
lm(formula = hipcenter ~ ., data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-73.827 -22.833  -3.678  25.017  62.337

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 436.43213  166.57162   2.620   0.0138 *
Age           0.77572    0.57033   1.360   0.1843
Weight        0.02631    0.33097   0.080   0.9372
HtShoes      -2.69241    9.75304  -0.276   0.7845
Ht            0.60134   10.12987   0.059   0.9531
Seated        0.53375    3.76189   0.142   0.8882
Arm          -1.32807    3.90020  -0.341   0.7359
Thigh        -1.14312    2.66002  -0.430   0.6706
Leg          -6.43905    4.71386  -1.366   0.1824
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.72 on 29 degrees of freedom
Multiple R-squared:  0.6866,    Adjusted R-squared:  0.6001
F-statistic:  7.94 on 8 and 29 DF,  p-value: 1.306e-05
```

Note that the coefficient of `Leg` is -6.43905, so it means `hipcenter` will decrease 6.43905 if we increase `Leg` by 1 unit, when all other predictors are held constant. Moreover, the p-value of `Leg` is 0.1824, greater than 0.05. So, in this model, this effect in `hipcenter` may be not such significant.

(b)

```r
newx <- data.frame(as.list(colMeans(seatpos)[-9]))
predict(m.sea, newdata=newx, interval="prediction")
```

```
        fit     lwr       upr
1 -164.8849 -243.04 -86.72972
```

The prediction interval is $[-243.04 - 86.72972]$.

(c)

```r
g <- lm(hipcenter ~ ., data=seatpos)
summary(g)
```

```
Call:
lm(formula = hipcenter ~ ., data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-73.827 -22.833  -3.678  25.017  62.337

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 436.43213  166.57162   2.620   0.0138 *
Age           0.77572    0.57033   1.360   0.1843
Weight        0.02631    0.33097   0.080   0.9372
HtShoes      -2.69241    9.75304  -0.276   0.7845
Ht            0.60134   10.12987   0.059   0.9531
Seated        0.53375    3.76189   0.142   0.8882
Arm          -1.32807    3.90020  -0.341   0.7359
Thigh        -1.14312    2.66002  -0.430   0.6706
Leg          -6.43905    4.71386  -1.366   0.1824
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.72 on 29 degrees of freedom
Multiple R-squared:  0.6866,    Adjusted R-squared:  0.6001
F-statistic:  7.94 on 8 and 29 DF,  p-value: 1.306e-05
```

```r
g <- update(g, . ~ . - Ht)
summary(g)
```

```
Call:
lm(formula = hipcenter ~ Age + Weight + HtShoes + Seated + Arm +
    Thigh + Leg, data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-74.107 -22.467  -4.207  25.106  62.225

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 436.84207  163.64104   2.670   0.0121 *
```

```
Age           0.76574    0.53590   1.429   0.1634
Weight        0.02897    0.32244   0.090   0.9290
HtShoes      -2.13409    2.53896  -0.841   0.4073
Seated        0.54959    3.68958   0.149   0.8826
Arm          -1.30087    3.80833  -0.342   0.7350
Thigh        -1.09039    2.46534  -0.442   0.6615
Leg          -6.40612    4.60272  -1.392   0.1742
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.09 on 30 degrees of freedom
Multiple R-squared:  0.6865,    Adjusted R-squared:  0.6134
F-statistic: 9.385 on 7 and 30 DF,  p-value: 4.014e-06
```

```
g <- update(g, . ~ . - Weight)
summary(g)
```

```
Call:
lm(formula = hipcenter ~ Age + HtShoes + Seated + Arm + Thigh +
    Leg, data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-74.263 -22.571  -4.842  24.647  61.926

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 427.5073   124.3877   3.437   0.0017 **
Age           0.7757     0.5158   1.504   0.1427
HtShoes      -2.0823     2.4329  -0.856   0.3986
Seated        0.5858     3.6083   0.162   0.8721
Arm          -1.2826     3.7415  -0.343   0.7341
Thigh        -1.1153     2.4101  -0.463   0.6468
Leg          -6.3572     4.4966  -1.414   0.1674
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36.49 on 31 degrees of freedom
Multiple R-squared:  0.6864,    Adjusted R-squared:  0.6257
F-statistic: 11.31 on 6 and 31 DF,  p-value: 1.122e-06
```

```
g <- update(g, . ~ . - Seated)
summary(g)
```

```
Call:
lm(formula = hipcenter ~ Age + HtShoes + Arm + Thigh + Leg, data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-73.966 -22.403  -4.725  24.989  60.834

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 436.5463   109.5266   3.986 0.000365 ***
Age           0.7667     0.5049   1.518 0.138717
HtShoes      -1.7716     1.4786  -1.198 0.239648
Arm          -1.3390     3.6683  -0.365 0.717498
Thigh        -1.1983     2.3193  -0.517 0.608955
Leg          -6.4910     4.3527  -1.491 0.145686
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 35.93 on 32 degrees of freedom
Multiple R-squared:  0.6862,    Adjusted R-squared:  0.6371
F-statistic: 13.99 on 5 and 32 DF,  p-value: 2.823e-07
```

```r
g <- update(g, . ~ . - Arm)
summary(g)
```

```
Call:
lm(formula = hipcenter ~ Age + HtShoes + Thigh + Leg, data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-77.069 -24.643  -3.584  26.092  59.182

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 445.7977   105.1452   4.240  0.00017 ***
Age           0.6525     0.3910   1.669  0.10462
HtShoes      -1.9171     1.4050  -1.365  0.18164
Thigh        -1.3732     2.2392  -0.613  0.54391
Leg          -6.9502     4.1118  -1.690  0.10040
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 35.46 on 33 degrees of freedom
Multiple R-squared:  0.6849,    Adjusted R-squared:  0.6467
F-statistic: 17.93 on 4 and 33 DF,  p-value: 6.535e-08
```

```r
g <- update(g, . ~ . - Thigh)
summary(g)
```

```
Call:
lm(formula = hipcenter ~ Age + HtShoes + Leg, data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-79.269 -22.770  -4.342  21.853  60.907

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 456.2137   102.8078   4.438 9.09e-05 ***
Age           0.5998     0.3779   1.587   0.1217
HtShoes      -2.3023     1.2452  -1.849   0.0732 .
Leg          -6.8297     4.0693  -1.678   0.1024
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 35.13 on 34 degrees of freedom
Multiple R-squared:  0.6813,    Adjusted R-squared:  0.6531
F-statistic: 24.22 on 3 and 34 DF,  p-value: 1.437e-08
```

In the last model, we use three predictors, `Age`, `HtShoes` and `Leg`. Although in this model, not all p-values are less than 0.05, we don't need to strictly obey this criteria. And if we eliminate `Age`, we will find there is a large decrease in R^2 comparing to previous eliminations.

So we conclude this model to be the best.

```r
#AIC
g2 <- lm(hipcenter ~ ., data=seatpos)
step(g2)
```

```
Start:  AIC=283.62
hipcenter ~ Age + Weight + HtShoes + Ht + Seated + Arm + Thigh +
    Leg

          Df Sum of Sq   RSS    AIC
- Ht       1      5.01 41267 281.63
- Weight   1      8.99 41271 281.63
- Seated   1     28.64 41290 281.65
- HtShoes  1    108.43 41370 281.72
- Arm      1    164.97 41427 281.78
- Thigh    1    262.76 41525 281.87
<none>                41262 283.62
```

```
- Age      1   2632.12 43894 283.97
- Leg      1   2654.85 43917 283.99

Step:  AIC=281.63
hipcenter ~ Age + Weight + HtShoes + Seated + Arm + Thigh + Leg

          Df Sum of Sq   RSS    AIC
- Weight   1     11.10 41278 279.64
- Seated   1     30.52 41297 279.66
- Arm      1    160.50 41427 279.78
- Thigh    1    269.08 41536 279.88
- HtShoes  1    971.84 42239 280.51
<none>                 41267 281.63
- Leg      1   2664.65 43931 282.01
- Age      1   2808.52 44075 282.13

Step:  AIC=279.64
hipcenter ~ Age + HtShoes + Seated + Arm + Thigh + Leg

          Df Sum of Sq   RSS    AIC
- Seated   1     35.10 41313 277.67
- Arm      1    156.47 41434 277.78
- Thigh    1    285.16 41563 277.90
- HtShoes  1    975.48 42253 278.53
<none>                 41278 279.64
- Leg      1   2661.39 43939 280.01
- Age      1   3011.86 44290 280.31

Step:  AIC=277.67
hipcenter ~ Age + HtShoes + Arm + Thigh + Leg

          Df Sum of Sq   RSS    AIC
- Arm      1    172.02 41485 275.83
- Thigh    1    344.61 41658 275.99
- HtShoes  1   1853.43 43166 277.34
<none>                 41313 277.67
- Leg      1   2871.07 44184 278.22
- Age      1   2976.77 44290 278.31

Step:  AIC=275.83
hipcenter ~ Age + HtShoes + Thigh + Leg

          Df Sum of Sq   RSS    AIC
- Thigh    1     472.8 41958 274.26
<none>                 41485 275.83
- HtShoes  1    2340.7 43826 275.92
- Age      1    3501.0 44986 276.91
```

```
- Leg      1    3591.7 45077 276.98

Step:  AIC=274.26
hipcenter ~ Age + HtShoes + Leg

          Df Sum of Sq   RSS    AIC
<none>                  41958 274.26
- Age      1    3108.8 45067 274.98
- Leg      1    3476.3 45434 275.28
- HtShoes  1    4218.6 46176 275.90

Call:
lm(formula = hipcenter ~ Age + HtShoes + Leg, data = seatpos)

Coefficients:
(Intercept)          Age      HtShoes          Leg
   456.2137       0.5998      -2.3023      -6.8297
```

The best model selected by AIC is that uses `Age`, `Leg` and `HtShoes` as predictors, which is the same as the choice made by Backward Elimination.

```
# Mallows Cp
g3 <- regsubsets(hipcenter ~ ., data=seatpos)
res.g3 <- summary(g3); res
```

```
Subset selection object
Call: regsubsets.formula(gamble ~ ., data = teengamb)
4 Variables  (and intercept)
       Forced in Forced out
sex        FALSE      FALSE
status     FALSE      FALSE
income     FALSE      FALSE
verbal     FALSE      FALSE
1 subsets of each size up to 4
Selection Algorithm: exhaustive
         sex status income verbal
1  ( 1 ) " " " "    "*"    " "
2  ( 1 ) "*" " "    "*"    " "
3  ( 1 ) "*" " "    "*"    "*"
4  ( 1 ) "*" "*"    "*"    "*"
```

```
#select model with largest adjusted r^2
which.min(res.g3$cp)
```

```
[1] 1
```

```r
res.g3$cp[1]
```

```
[1] -0.5342143
```

```r
summary(lm(hipcenter ~ Ht, data=seatpos))
```

```
Call:
lm(formula = hipcenter ~ Ht, data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-99.956 -27.850   5.656  20.883  72.066

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 556.2553    90.6704   6.135 4.59e-07 ***
Ht           -4.2650     0.5351  -7.970 1.83e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36.37 on 36 degrees of freedom
Multiple R-squared:  0.6383,    Adjusted R-squared:  0.6282
F-statistic: 63.53 on 1 and 36 DF,  p-value: 1.831e-09
```

The best model selected by Mallows C_p is that only uses `Ht` as predictors, which is different from the choices made by Backward Elimination and AIC.

(d)

The model is

```r
summary(g.aic <- lm(formula = hipcenter ~ Age + HtShoes + Leg, data = seatpos))
```

```
Call:
lm(formula = hipcenter ~ Age + HtShoes + Leg, data = seatpos)

Residuals:
    Min      1Q  Median      3Q     Max
-79.269 -22.770  -4.342  21.853  60.907

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) 456.2137   102.8078   4.438 9.09e-05 ***
Age           0.5998     0.3779   1.587   0.1217
HtShoes      -2.3023     1.2452  -1.849   0.0732 .
Leg          -6.8297     4.0693  -1.678   0.1024
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 35.13 on 34 degrees of freedom
Multiple R-squared:  0.6813,    Adjusted R-squared:  0.6531
F-statistic: 24.22 on 3 and 34 DF,  p-value: 1.437e-08
```

The coefficient of `Leg` is -6.8297, so it means `hipcenter` will decrease 6.8297 if we increase `Leg` by 1 unit, when all other predictors are held constant. Moreover, the p-value of `Leg` is 0.1024, which is smaller than the previous 0.1824. So, in this model, this effect in `hipcenter` becomes more significant.

For the prediction interval

```r
predict(g.aic, newdata=newx[ ,c(1,3,8)], interval="prediction")
```

```
        fit      lwr       upr
1 -164.8849 -237.209 -92.56072
```

The prediction interval is $[-237.209, -92.56072]$.

Comparison: The estimated mean value of `hipcenter` in full model is -164.8849, and the CI is $[-243.04, -86.72972]$, which has length 156.3103.

The estimated mean value of `hipcenter` in AIC-selected model is -164.8849, and the CI is $[-237.209, -92.56072]$, which has length 144.6483.

We observe these two models give very close (even the same) prediction of the mean value of response and they have close R^2. So the fitting ability of them are close.

However, the AIC-selected model gives a narrower CI, which shows it is more accurate when performing predictions. Moreover, the AIC-selected model is easier to interpret since it has less variables.

Problem 3

We can rewrite the design matrix

$$X = \begin{bmatrix} \mathbf{1} & X_p \end{bmatrix}$$

in which

$$X_p = \begin{bmatrix} | & | & | \\ x_1 - \bar{x}_1 & \cdots & x_p - \bar{x}_p \\ | & | & | \end{bmatrix}$$

The standard formula of $\hat{\beta}$ is

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

We have

$$X^T X = \begin{bmatrix} \mathbf{1}^T \mathbf{1} & \mathbf{1}^T X_p \\ X_p^T \mathbf{1} & X_p^T X_p \end{bmatrix}$$

Note that

$$\begin{aligned} \mathbf{1}^T X_p &= \begin{bmatrix} \mathbf{1}^T (x_1 - \bar{x}_1) & \cdots & \mathbf{1}^T (x_p - \bar{x}_p) \end{bmatrix} \\ &= \begin{bmatrix} \sum_i x_{1i} - n\bar{x}_1 & \cdots & \sum_i x_{pi} - n\bar{x}_p \end{bmatrix} \\ &= \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix} \quad \text{(by definiton of mean)} \end{aligned}$$

Therefore, by the formula of block matrix inversion (ref:click)

$$\begin{aligned} (X^T X)^{-1} &= \begin{bmatrix} n & 0 \\ 0 & X_p^T X_p \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1/n & 0 \\ 0 & (X_p^T X_p)^{-1} \end{bmatrix} \end{aligned}$$

so

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T y \\ &= \begin{bmatrix} 1/n & 0 \\ 0 & (X_p^T X_p)^{-1} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{1}^T \\ X_p^T \end{bmatrix} \cdot y \\ &= \begin{bmatrix} 1/n & 0 \\ 0 & (X_p^T X_p)^{-1} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{1}^T y \\ X_p^T y \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n} \mathbf{1}^T y \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \vdots \\ \vdots \end{bmatrix} \end{aligned}$$

Therefore, the resultant estimate of the intercept is

$$\hat{\beta}_o = \bar{y}$$