Homework 2 solutions

STAT3530 Applied Linear Models

Solutions are shown in blue.

Throughout, $Y = X\beta + \epsilon$ with $\mathbb{E}[\epsilon] = 0$ and $\text{Var}[\epsilon] = \sigma^2 I$, and X has full column rank unless stated otherwise.

Recovering the simple linear regression estimates

On Homework 1 you derived the least squares estimates for the simple linear model one parameter at a time. Here you will verify that the matrix formula $\hat{\beta} = (X^T X)^{-1} X^T Y$ recovers the same estimates.

Take the design matrix with one predictor,

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

You may use the identities $\sum_{i=1}^n (x_i - \bar{x})^2 = (n-1)s_x^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$ from Homework 1 and the formula for the inverse of a 2×2 matrix,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

where $ad - bc$ is the matrix determinant.

1. Compute $X^T X$ and $X^T Y$, writing each entry as a sum. Then rewrite the two entries with $\sum_i x_i = n\bar{x}$ and $\sum_i Y_i = n\bar{Y}$.

Each entry of $X^T X$ is an inner product of two columns of X:

$$X^T X = \begin{bmatrix} \mathbf{1}^T\mathbf{1} & \mathbf{1}^T x \\ x^T\mathbf{1} & x^T x \end{bmatrix} = \begin{bmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_i x_i^2 \end{bmatrix}$$

$$X^T Y = \begin{bmatrix} \mathbf{1}^T Y \\ x^T Y \end{bmatrix} = \begin{bmatrix} \sum_i Y_i \\ \sum_i x_i Y_i \end{bmatrix} = \begin{bmatrix} n\bar{Y} \\ \sum_i x_i Y_i \end{bmatrix}$$

2. Show that $\det(X^TX) = n(n-1)s_x^2$, then write down $(X^TX)^{-1}$.

$$\det(X^TX) \;=\; n\sum_i x_i^2 - (n\bar{x})^2 \;=\; n\Big(\sum_i x_i^2 - n\bar{x}^2\Big) \;=\; n\sum_i (x_i-\bar{x})^2 \;=\; n(n-1)s_x^2$$

using the identity from Homework 1 in the third step. Then by the 2×2 formula,

$$(X^TX)^{-1} \;=\; \frac{1}{n(n-1)s_x^2}\begin{bmatrix} \sum_i x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

3. Multiply out $\hat{\beta} = (X^T X)^{-1} X^T Y$ and show that its second entry is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) Y_i}{(n-1) s_x^2}$$

which is what you found on Homework 1. *Hint: expand* $\sum_i (x_i - \bar{x}) Y_i$.

The second entry is the second row of $(X^T X)^{-1}$ times $X^T Y$:

$$\hat{\beta}_1 \;=\; \frac{1}{n(n-1)s_x^2}\Big(-n\bar{x} \cdot n\bar{Y} \;+\; n \sum_i x_i Y_i \Big) \;=\; \frac{\sum_i x_i Y_i - n\bar{x}\bar{Y}}{(n-1)s_x^2}$$

after cancelling one factor of n. Expanding the target numerator,

$$\sum_i (x_i - \bar{x}) Y_i \;=\; \sum_i x_i Y_i - \bar{x} \sum_i Y_i \;=\; \sum_i x_i Y_i - n\bar{x}\bar{Y}$$

so the two agree.

4. Show that the first entry simplifies to $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$. *Hint: put* $\bar{Y} - \hat{\beta}_1 \bar{x}$ *over the common denominator* $(n-1)s_x^2$*, then work backwards.*

The first entry is

$$\hat{\beta}_0 \;=\; \frac{1}{n(n-1)s_x^2}\Big(\sum_i x_i^2 \cdot n\bar{Y} \;-\; n\bar{x} \sum_i x_i Y_i \Big) \;=\; \frac{\bar{Y} \sum_i x_i^2 - \bar{x} \sum_i x_i Y_i}{(n-1)s_x^2}$$

Now start from the target and put it over the same denominator, using problem 3 for $\hat{\beta}_1$:

$$\bar{Y} - \hat{\beta}_1 \bar{x} \;=\; \frac{\bar{Y}(n-1)s_x^2 - \bar{x}\big(\sum_i x_i Y_i - n\bar{x}\bar{Y}\big)}{(n-1)s_x^2}$$

Replace $(n-1)s_x^2$ in the numerator by $\sum_i x_i^2 - n\bar{x}^2$:

$$= \frac{\bar{Y} \sum_i x_i^2 - n\bar{x}^2 \bar{Y} - \bar{x} \sum_i x_i Y_i + n\bar{x}^2 \bar{Y}}{(n-1)s_x^2} \;=\; \frac{\bar{Y} \sum_i x_i^2 - \bar{x} \sum_i x_i Y_i}{(n-1)s_x^2}$$

which matches.

Centering the predictors

In class we centered the predictor, giving the design matrix

$$X = \left[\begin{array}{cc} \mathbf{1} & \tilde{x} \end{array} \right], \qquad \tilde{x}_i = x_i - \bar{x}$$

This is the model you just worked through, with x replaced by $\tilde{x}$. Notice how much the algebra simplifies.

5. Show that $\mathbf{1}^T\tilde{x} = 0$.

$$\mathbf{1}^T\tilde{x} \;=\; \sum_i \tilde{x}_i \;=\; \sum_i (x_i - \bar{x}) \;=\; 0$$

by problem 1 on Homework 1.

6. Using problem 5, show that

$$X^TX = \left[\begin{array}{cc} n & 0 \\ 0 & (n-1)s_x^2 \end{array} \right] \qquad \text{and} \qquad (X^TX)^{-1} = \left[\begin{array}{cc} \frac{1}{n} & 0 \\ 0 & \frac{1}{(n-1)s_x^2} \end{array} \right]$$

Take the same inner products as in problem 1, now with $\tilde{x}$ in place of x. The off-diagonal entries are $\mathbf{1}^T\tilde{x} = 0$ by problem 5, and the lower right entry is

$$\tilde{x}^T\tilde{x} \;=\; \sum_i (x_i - \bar{x})^2 \;=\; (n-1)s_x^2$$

so X^TX is diagonal, as claimed. A diagonal matrix is inverted by taking reciprocals down the diagonal, which gives $(X^TX)^{-1}$.

7. What is $\mathrm{Cov}[\hat{\beta}_0, \hat{\beta}_1]$, and what does that say about the two estimators?

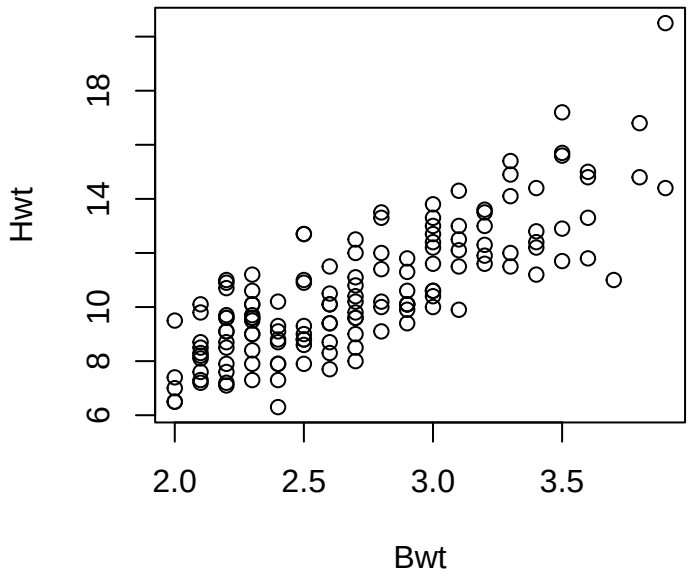
Since $\mathrm{Var}[\hat{\beta}] = \sigma^2(X^TX)^{-1}$ and $(X^TX)^{-1}$ is diagonal,

$$\mathrm{Cov}[\hat{\beta}_0, \hat{\beta}_1] \;=\; \sigma^2 \cdot 0 \;=\; 0$$

The two estimators are **uncorrelated**. Knowing that the slope estimate came out high tells you nothing about whether the intercept estimate came out high. Without centering they are correlated, because tilting the line about a point away from $\bar{x}$ moves the height at $x = 0$.

8. The `cats` data has heart weights in grams and body weights in kilograms for 144 adult cats. A plot of the data is shown below.

```
data(cats, package = 'MASS')
plot(Hwt ~ Bwt, data = cats)
```



   a. Fit the model `Hwt ~ Bwt` with and without centering the predictor. Report and interpret the intercept estimate in each case.

```
fit  <- lm(Hwt ~ Bwt, data = cats)
fit.c <- lm(Hwt ~ I(Bwt - mean(Bwt)), data = cats)
rbind(uncentered = coef(fit), centered = coef(fit.c))
```

```
           (Intercept)      Bwt
uncentered  -0.3566624 4.034063
centered    10.6305556 4.034063
```

```
mean(cats$Hwt)
```

```
[1] 10.63056
```

The slope is 4.034 in both. The uncentered intercept, −0.357, is the fitted heart weight of a cat weighing 0 kg — a negative heart weight for a cat that does not exist, and an extrapolation far outside the data. The centered intercept, 10.631, is the fitted heart weight of a cat of *average* body weight, and equals $\overline{\text{Hwt}}$ exactly.

   b. Estimate $\text{cov}(\hat{\beta}_0, \hat{\beta}_1)$ for each model.

```
c(uncentered = vcov(fit)[1, 2], centered = vcov(fit.c)[1, 2])
```

```
   uncentered      centered
-1.705820e-01 -3.434909e-17
```

Uncentered the covariance is −0.171; centered it is zero to machine precision, as problem 7 predicts.

c. Now refit the model after centering the *response*. What happens to the estimates? Explain.

```
fit.both <- lm(I(Hwt - mean(Hwt)) ~ I(Bwt - mean(Bwt)), data = cats)
coef(fit.both)
```

```
       (Intercept) I(Bwt - mean(Bwt))
     -3.169374e-15       4.034063e+00
```

```
c(sigma = sigma(fit.both), cov = vcov(fit.both)[1, 2])
```

```
        sigma           cov
 1.452373e+00 -3.434909e-17
```

The slope is unchanged at 4.034 and the intercept becomes 0. Centering the response subtracts $\overline{\text{Hwt}}$ from every fitted value, which lowers the whole line by that amount without tilting it, and the intercept was already $\overline{\text{Hwt}}$. Residuals, $\hat{\sigma}$, and the covariance are all unaffected: subtracting a constant from Y shifts the fit but changes nothing about its spread.

The hat matrix

Let $H = X(X^TX)^{-1}X^T$ and $M = I - H$.

9. Show that H is symmetric and idempotent, *i.e.*, that it is a projection.

Symmetric. Since X^TX is symmetric so is its inverse, so

$$H^T = \Big[X(X^TX)^{-1}X^T\Big]^T = X\Big[(X^TX)^{-1}\Big]^T X^T = X(X^TX)^{-1}X^T = H$$

Idempotent. The inner X^TX meets its own inverse and cancels:

$$HH = X(X^TX)^{-1}\underbrace{X^TX}(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H$$

10. Show that M is symmetric and idempotent, *i.e.*, that it is also a projection.

$$M^T = (I-H)^T = I - H^T = I - H = M$$
$$MM = (I-H)(I-H) = I - H - H + HH = I - 2H + H = I - H = M$$

using problem 9 for both $H^T = H$ and $HH = H$.

11. Show that $HX = X$, $MX = 0$, and $MH = 0$. Interpret each result briefly in words.

$$HX = X(X^TX)^{-1}X^TX = X, \qquad MX = (I-H)X = X - HX = 0$$
$$MH = (I-H)H = H - HH = H - H = 0$$

In words: H leaves alone any vector already in $\mathrm{col}(X)$ — there is nothing to project. M annihilates such vectors, since none of them lies in the orthogonal complement except 0. And $MH = 0$ says anything H produces is killed by M: the two projections land in orthogonal subspaces, which is why the fitted values and residuals are orthogonal.

Least squares for a population mean

Suppose we have a random sample $Y_1, \dots, Y_n$, each with mean μ and variance σ^2, and we want to estimate μ. Express this via the linear model

$$Y = X\beta + \epsilon \qquad \text{with} \qquad X = \mathbf{1}\,, \quad \beta = \mu$$

12. Compute X^TX, $(X^TX)^{-1}$, and X^TY for this model and show that the least squares estimator is the sample mean $\bar{Y}$.

With $X = \mathbf{1}$ every quantity is a scalar, so the inverse is just a reciprocal:

$$X^TX \;=\; \mathbf{1}^T\mathbf{1} \;=\; n, \qquad (X^TX)^{-1} \;=\; (\mathbf{1}^T\mathbf{1})^{-1}, \qquad X^TY \;=\; \mathbf{1}^TY \;=\; \sum_i Y_i$$

Therefore

$$\hat{\beta} \;=\; (X^TX)^{-1}X^TY \;=\; (\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^TY \;=\; \bar{Y}$$

which is exactly the expression for $\bar{Y}$ from problem 10 on Homework 1.

13. Use $\mathbb{E}[\hat{\beta}] = \beta$ and $\mathrm{Var}[\hat{\beta}] = \sigma^2(X^TX)^{-1}$ to show that the sample mean is unbiased with variance $\mathrm{Var}[\bar{Y}] = \frac{\sigma^2}{n}$.

Both facts come straight from the general results, with $\beta = \mu$ and $(X^TX)^{-1} = 1/n$:

$$\mathbb{E}[\bar{Y}] \;=\; \mathbb{E}[\hat{\beta}] \;=\; \beta \;=\; \mu$$

$$\mathrm{Var}[\bar{Y}] \;=\; \sigma^2(X^TX)^{-1} \;=\; \sigma^2(\mathbf{1}^T\mathbf{1})^{-1} \;=\; \frac{\sigma^2}{n}$$

No summation algebra was needed — the general theory already did that work.

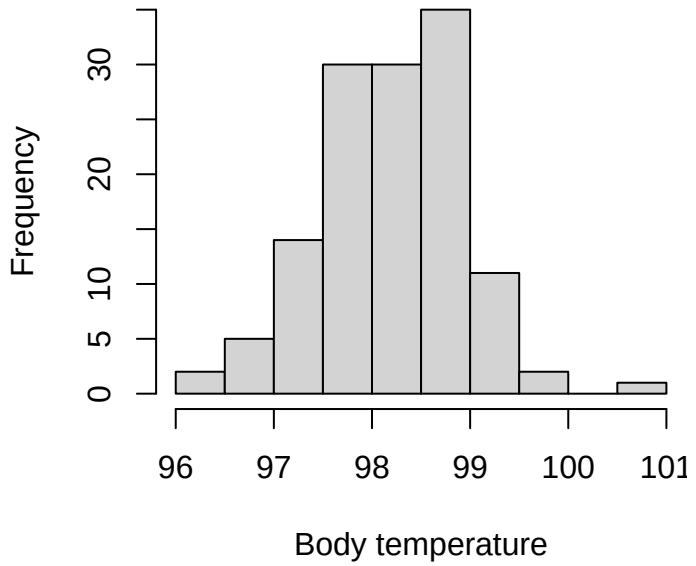
14. Compute the hat matrix for this model.

$$H \;=\; X(X^TX)^{-1}X^T \;=\; \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T \;=\; \frac{1}{n}\mathbf{1}\mathbf{1}^T \;=\; J$$

This is exactly the matrix J from Homework 1. The fitted values are $\hat{Y} = HY = \bar{Y}\mathbf{1}$: every observation is fitted by the same number, the sample mean. The residuals are $Y - \bar{Y}\mathbf{1}$, the centered data.

15. The `temps` dataset contains 130 measurements of body temperatures in Fahrenheit for a sample of adults. In R, you can fit an intercept-only model via the formula `lm(y ~ 1)`. Verify that the least squares estimate recovers the sample mean, and its estimated variance recovers the usual standard error for a sample mean. Then construct a 95% confidence interval for the mean body temperature. Is 98.6 a plausible value?

```
load(url("https://tdruiz-stat3530.share.connect.posit.cloud/_data/temps.RData"))
hist(temps$body.temp, xlab = 'Body temperature', main = NULL)
```




```
fit <- lm(body.temp ~ 1, data = temps)
coef(fit)                       # the least squares estimate
```

```
(Intercept)
   98.24923
```

```
mean(temps$body.temp)           # the sample mean -- identical
```

```
[1] 98.24923
```

```
sqrt(vcov(fit))                 # SE of the estimate
```

```
             (Intercept)
(Intercept)  0.06430442
```

```
sd(temps$body.temp) / sqrt(nrow(temps))    # s / sqrt(n) -- identical
```

```
[1] 0.06430442
```

Note this is *not* $\hat{\sigma}$: `sigma(fit)` is 0.733, the residual standard deviation, while the standard error of the estimate is $0.733/\sqrt{130} = 0.064$, smaller by a factor of $\sqrt{130}$.

```
confint(fit)
```

```
               2.5 %    97.5 %
(Intercept) 98.122 98.37646
```

The 95% interval is (98.12, 98.38). It does **not** contain 98.6, so 98.6 is not a plausible value for the population mean body temperature. The familiar figure appears to be about a third of a degree too

high.