```r
students <- read.csv("students_train.csv")
```

A

```r
model <- lm(grade ~ sex + age + absences + internet, data = students)
coef(model)

(Intercept)         sex         age    absences    internet 
16.59305325  1.75035006 -0.34327942 -0.03659555  0.13974392 
```

We fitted a linear regression model using the variables `sex`, `age`, `absences`, and `internet` to predict the final math `grade`. The model was trained on the dataset `students_train.csv` using the quadratic loss function.

The estimated coefficients (θ) from the model are:

-intercept: 16.60;

-sex: 1.75;

-age: -0.34;

-absences: -0.03;

-internet:0.14.

B

```r
model <- lm(grade ~ sex + age + absences + internet, data = students)
summary(model)


Call:
lm(formula = grade ~ sex + age + absences + internet, data = students)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.8970  -1.8429  -0.1292   2.0093   7.1557 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) 16.59305    5.84572   2.838  0.00583 **
sex          1.75035    0.77481   2.259  0.02678 * 
age         -0.34328    0.33060  -1.038  0.30244   
absences    -0.03660    0.04267  -0.858  0.39382   
internet     0.13974    1.09181   0.128  0.89850   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3.415 on 75 degrees of freedom
Multiple R-squared:  0.1052,    Adjusted R-squared:  0.05743
F-statistic: 2.203 on 4 and 75 DF,  p-value: 0.07667
```

To confirm the estimate of θ from part a), we used the built-in R function `lm()` to fit the linear regression model. We then called `summary(model)` to display the estimated coefficients.

This confirms the same values of θ as obtained in part a).

C

```r
#model prediction
predictions <- predict(model, students)

#MSE
mse <- mean((students$grade - predictions)^2)
mse

[1] 10.93056
```

We calculated the training error as the Mean Squared Error (MSE) between the actual grades and the predicted grades from the model in part a).

The training error is 10.93

D

```r
students_test <- read.csv("students_test.csv")

# Predictions test set
test_predictions <- predict(model, students_test)

#  Mean Squared Error
test_mse <- mean((students_test$grade - test_predictions)^2)
test_mse

[1] 18.33673
```

We tested the model from part a) on a separate test dataset of 20 students. We predicted the final math grades and calculated the Mean Squared Error (MSE) on the test data.

The test error is 18.33

The test error is higher than the training error, which is expected, as the model performs better on the data it was trained on.

Exercise 2

$$L(y, \pi(x)) = -y\ln(\pi(x)) - (1-y)\ln(1-\pi(x))$$

For a constant model, $\pi(x) = p$:

$$L(y, p) = -y\ln(p) - (1-y)\ln(1-p)$$

Empirical risk $= \frac{1}{n}\sum_{i=1}^{n} L(y^{(i)}, \pi(x^{(i)})) = \frac{1}{n}\sum_{i=1}^{n}\left(-y^{(i)}\ln(p) - (1-y^{(i)})\ln(1-p)\right)$

$$\Rightarrow -\left(\frac{1}{n}\sum_{i=1}^{n} y^{(i)}\right)\ln(p) - \left(1 - \frac{1}{n}\sum_{i=1}^{n} y^{(i)}\right)\ln(1-p)$$

Now, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y^{(i)}$ (mean)

$$\Rightarrow -\bar{y}\ln(p) - (1-\bar{y})\ln(1-p)$$

To minimize loss, we take derivative with respect to p:

$$\frac{d}{dp}\left(-\bar{y}\ln(p) - (1-\bar{y})\ln(1-p)\right)$$

$$\Rightarrow -\frac{\bar{y}}{p} + \frac{(1-\bar{y})}{(1-p)} = 0$$

$$\Rightarrow \frac{1-\bar{y}}{1-p} = \frac{\bar{y}}{p}$$

$$\Rightarrow p(1-\bar{y}) = \bar{y}(1-p)$$

$$\Rightarrow p - \cancel{p\bar{y}} = \bar{y} - \cancel{p\bar{y}}$$

$$\Rightarrow \boxed{p = \bar{y}}$$

That means, $\boxed{p = \frac{1}{n}\sum_{i=1}^{n} y^{(i)} = \bar{y}}$