4. Among the factors that can influence the price of a house stand out the area and age of the house. To understand the relationship between price, area and age, 20 American households selected randomly were studied. The houses observed areas are between 0.8 and 3.6 and ages are between 1 and 33 (the area is measured in thousands of square feet, age is measured in years, and the price in thousands of US dollars). Assuming the multiple linear regression model is appropriate to this dataset, R was used to estimate the parameters, and part of the output is presented below.

```
Call:
lm(formula = Price ~ Area + Age, data = x)

Residuals:

     Min       1Q   Median       3Q      Max
-3.27322 -1.63194  0.05641  1.59341  3.90155
```

$\hat{\beta}_i$ $\quad s_{\hat{\beta}_i}$ $\quad H_0: \beta_i = 0$ vs $H_1: \beta_i \neq 0$ $\quad$ P-value

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.39367    1.80957   5.744 2.39e-05 ***
Area        12.75077    0.60261  21.159 1.19e-13 ***
Age         -0.07187    0.05312  -1.353    0.194
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$i=0$ (Intercept), $i=1$ Area, $i=2$ Age

$\hat{\sigma}$

```
Residual standard error: 2.208 on 17 degrees of freedom
Multiple R-squared: 0.9643,     Adjusted R-squared: 0.9602
F-statistic: 229.9 on 2 and 17 DF,  p-value: 4.932e-13
```

$H_0: \beta_1 = \beta_2 = 0$ vs $H_1: \exists \beta_i \neq 0$

```
Analysis of Variance Table

Response: Price
          Df  Sum Sq Mean Sq  F value    Pr(>F)
Area       1 2232.74 2232.74 457.9580 9.878e-14 ***
Age        1    8.92    8.92   1.8305    0.1938
Residuals 17   82.88    4.88
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
solve(t(X)%*%X)
             [,1]          [,2]          [,3]
[1,]  0.67164255 -0.1630071739 -0.0137501489
[2,] -0.16300717  0.0744831743  0.0006530616
[3,] -0.01375015  0.0006530616  0.0005787919
```

$= (\underset{\sim}{X}^T \underset{\sim}{X})^{-1} = \underset{\sim}{C}$

(a) Explain the meaning of the estimates 12.75077 and -0.07187.

(b) Calculate the 95% confidence interval on the mean price of houses with an area of 2200 square feet and 26 years of age. Calculate the 95% prediction interval on a house with the same characteristics. Compare and discuss the results obtained from the two intervals.

(c) From the ANOVA table we can easily conclude that the regression model is significant. But it is suspected that between the two predictors involved only the area is meaningful.

i. Confirm the assumption "*the regression model is significant*", by making the appropriate analysis of the ANOVA table.

ii. Confirm that only area is a meaningful predictor using a t-test.

(d) The aim is to predict the price of a house with 50 years of age. What do you think about the usefulness of this model to make this prediction?

a) $\hat{E}[Y|\underset{\sim}{x}] = \hat{\mu}|_{\underset{\sim}{x}} = 10.394 + 12.751\,\text{Area} - 0.072\,\text{Age}$

b) $T = \dfrac{\hat{\mu}|_{\underset{\sim}{x}_0} - \mu|_{x_0}}{\sqrt{\hat{\sigma}^2\, \underset{\sim}{x}_0^T \underset{\sim}{C}\, \underset{\sim}{x}}} \sim t(n-p) = t(17)$ $\quad \underset{\sim}{C} = (\underset{\sim}{X}^T \underset{\sim}{X})^{-1}$

$n = 20$

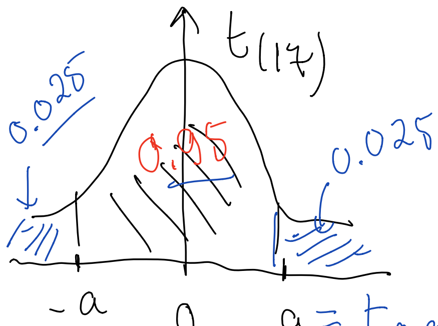
$p = 3$

$\hat{E}[Y|\underset{\sim}{x}_0] = \hat{\mu}|_{\underset{\sim}{x}_0} = 10.393 \times 1 + 12.75 \times 2.2 - 0.072 \times 26 = 36.57674$

$\underset{\sim}{x}_0^T = (1, 2.2, 26)$

Area age

$CI_{95\%}(\mu|_{\underset{\sim}{x}_0}) = \left[\hat{\mu}|_{\underset{\sim}{x}_0} \pm t_{0.975}(17) \times \sqrt{\hat{\sigma}^2 \underset{\sim}{x}_0^T \underset{\sim}{C}\, \underset{\sim}{x}_0}\right]$



$$\underset{\sim}{C} = (\underset{\sim}{X}^T \underset{\sim}{X})^{-1}$$

$$a = t_{0.975}(17)$$
$$= 2.1001$$

$$\sqrt{\hat{\sigma}^2 \, \underset{\sim}{x}_0^T \underset{\sim}{C} \, \underset{\sim}{x}_0} = ?$$

$$\overset{\underset{\sim}{x}_0^T}{(1 \quad 2.2 \quad 26)} \; \overset{\underset{\sim}{C}}{\left[\quad\quad\quad \right]} \; \overset{\underset{\sim}{x}_0}{\begin{pmatrix} 1 \\ 2.2 \\ 26 \end{pmatrix}}$$

$$= 0.06587536$$

$$\hat{\sigma}^2 = MSE = 4.88$$

$$\sqrt{4.88 \times 0.0658736} = 0.5667096 = \sqrt{\hat{\sigma}^2 \underset{\sim}{x}_0^T \underset{\sim}{C} \underset{\sim}{x}_0}$$

$$CI_{95\%}(\mu \mid \underset{\sim}{x}_0) = [36.58 \pm 2.1001 \times 0.5667096]$$
$$= [35.386; 37.773]$$

$$E[Y \mid \underset{\sim}{x}_0] = \beta_0 + \beta_1 \text{Area} + \beta_2 \text{Age}$$

$PI(Y_0) = ?$

$\hat{Y}_0 = \hat{E}[Y | \underset{\sim}{x}_0] = \hat{\mu} | \underset{\sim}{x}_0$

$Y_0 = Y | \underset{\sim}{x}_0 = \beta_0 + \beta_1 x_{01} + \beta_2 x_{02} + \varepsilon$

$$T = \frac{\hat{Y}_0 - Y_0}{\sqrt{\hat{\sigma}^2 (1 + \underset{\sim}{x}_0^T \underset{\sim}{C} \, \underset{\sim}{x}_0)}} \sim t_{(17)}$$

$P.I_{95\%}(Y_0) = [31.78 ; 41.36]$