1.1 [1 point] ISLR 2e (Gareth James, et al.) Section 3.7 (Exercises), page 123: Exercise 6. (Hint: The least squares line is given by the equation below.)

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (i)$$

6. Using (3.4), argue that in the case of simple linear regression, the least squares line always passes through the point $(\bar{x}, \bar{y})$.

As we know from ISLR 2e (3.4), the definitions of $\hat{\beta}_0$ and $\hat{\beta}_1$ are:

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

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Substituting $\hat{\beta}_0$ in (i)

$$y_i = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i = \bar{y} + \hat{\beta}_1 (x_i - \bar{x})$$

Substituting $x_i = \bar{x}$, $y_i = \bar{y} + \hat{\beta}_1 (\bar{x} - \bar{x}) = \bar{y}$, that's how we prove that in the case of linear regression, the least square line (i) always passes through $(\bar{x}, \bar{y})$.

1.2 [1 point] Section 3.7 (Exercises), page 120: Exercises 1, 3, 4-a.

Exercise 1:

Describe the null hypotheses to which the p-values given in Table 3.4 correspond. Explain what conclusions you can draw based on these p-values. Your explanation should be phrased in terms of sales, TV, radio, and newspaper, rather than in terms of the coefficients of the linear model.

The null hypotheses to which the p-values correspond are:

- $H_0: \beta_0 = 0 (Null\ hypothesis\ for\ the\ intercept, there\ is\ no\ sales\ if\ TV\ and\ radio\ are\ 0)$
- $H_1: \beta_1 = 0\ (Null\ hypothesis\ for\ TV, TV\ doesn't\ affect\ sales)$
- $H_2: \beta_2 = 0\ (Null\ hypothesis\ for\ radio, radio\ doesn't\ affect\ sales)$
- $H_3: \beta_3 = 0\ (Null\ hypothesis\ for\ newspaper, newspaper\ doesn't\ affect\ sales)$

Based on these p-values, we can reject the null hypothesis for the first three cases but not for newspaper, affirming that if there isn't money spent in TV, radio and newspapers, sales will be different from 0, and that spending more money in TV and radio both result in an increase of sales. However, we can't affirm, seen its p-value and our level of confidence (either 90%, 95%,99%) money spent in newspaper advertising increase sales, but we can't also affirm the opposite, simply with what we know we can't affirm that newspaper advertising either affects or not sales.

Exercise 3:

3. Suppose we have a data set with five predictors, X1 = GPA, X2 = IQ, X3 = Level (1 for College and 0 for High School), X4 = Interaction between GPA and IQ, and X5 = Interaction between GPA and Level. The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model, and get ˆ β0 = 50, ˆ β1 = 20, ˆ β2 = 0.07, ˆ β3 = 35, ˆ β4 = 0.01, ˆ β5 = −10.

(a) Which answer is correct, and why?

i. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates.

False. Assuming that p-value for level is low enough for our level of confidence, β3 = 35, that meaning that, for a fixed value of IQ ang GPA, College Students earn 35k more, as Level codes 1 for college and 0 for high school.

ii. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates.

True. Assuming that p-value for level is low enough for our level of confidence, β3 = 35, that means that, for a fixed value of IQ ang GPA, College Students earn 35k more, as Level codes 1 for college and 0 for high school.

iii. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates provided that the GPA is high enough.

False, with this model, and given that GPA is high enough (near 3.5), the difference of level gives an advantage of +35k for college students. However, taking the interaction between level and GPA into account, they balance out, as this interaction increases 3.5*1*-10=-35 expected earnings for college graduates when GPA is 3.5, so the model cancels these two factors out, these two groups should have more or less the same income.

iv. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates provided that the GPA is high enough.

False, with this model, and given that GPA is high enough (near 3.5), the difference of level gives an advantage of +35k for college students. However, taking the interaction between level and GPA into account, they balance out, as this interaction increases 3.5*1*-10=-35 expected earnings for college graduates when GPA is 3.5, so the model cancels these two factors out, these two groups should have more or less the same income.

(b) Predict the salary of a college graduate with IQ of 110 and a GPA of 4.0.

Knowing that our model is:

$$salary = \hat{\beta}_0 + \hat{\beta}_1 \cdot GPA + \hat{\beta}_2 \cdot IQ + \hat{\beta}_3 \cdot Level + \hat{\beta}_4 \cdot GPA \cdot IQ + \hat{\beta}_5 \cdot GPA \cdot Level$$

we substitute with the values given in 3 b:

$$salary = 50 + 20 \cdot GPA + 0.07 \cdot IQ + 35 \cdot Level + 0.01 \cdot GPA \cdot IQ + -10 \cdot GPA \cdot Level$$

$$salary = 50 + 20 \cdot 4.0 + 0.07 \cdot 110 + 35 \cdot 1 + 0.01 \cdot 4.0 \cdot 110 + -10 \cdot 4.0 \cdot 1$$

$$salary = 50 + 20 \cdot 4.0 + 0.07 \cdot 110 + 35 \cdot 1 + 0.01 \cdot 4.0 \cdot 110 + -10 \cdot 4.0 \cdot 1$$

$$salary = \$137.100/year$$

(c) True or false: Since the coefficient for the GPA/IQ interaction term is very small, there is very little evidence of an interaction effect. Justify your answer.

False. Even if the effect may be low (for example, average IQ people, with IQ=100, would only expect 4k more because of this interaction depending on the GPA, between 0.0 and 4.0.), the existence of this interaction depends on the p-value and not the interaction coefficient, and this interaction may be existent but not very effective.

Exercise 4:

(a) Suppose that the true relationship between X and Y is linear, i.e. Y = β0 + β1X + ϵ. Consider the training residual sum of squares (RSS) for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.

As the cubic regression incorporates the linear case plus more terms, it will fit better the data, even if the data relation is linear in reality. With this little amount of observations, the cubic model, fitting a little bit better, will have a minor RSS than the linear case. However, the difference may be low and it may not be the wisest to choose this model instead of the linear one.