

Book Problems

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.*

TV The null hypothesis here is that spending on television advertising didn't affect sales. The p -value here is very small, so we can reject the null hypothesis at a significant level of $\alpha = 0.05$. This means that television advertising does have an affect on sales.

Radio The null hypothesis here is that spending on radio advertising didn't affect sales. The p -value here is very small, so we can reject the null hypothesis at a significant level of $\alpha = 0.05$. This means that radio advertising does have an affect on sales.

Newspaper The null hypothesis here is that spending on newspaper advertising doesn't affect sales. The p -value, at 0.8599 means that we fail to reject the null hypothesis, which in this circumstance implies that newspaper advertising doesn't affect sales.

4. *I collect a set of data ($n = 100$ observations) containing a single predictor and a quantitative response. I then fit a linear regression model to the data, as well as a separate cubic regression, i.e. $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$.*

- (a) *Suppose the true relationship between X and Y is linear, i.e. $Y = \beta_0 + \beta_1 X + \epsilon$. 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.*

Polynomial regression here should have a lower training RSS than a normal linear regression because of the possibility for a tighter fit of the data.

- (b) *Answer (a) using test rather than training RSS.*

This is the opposite as in part (a). Here the polynomial regression should have a larger/higher test RSS. The linear regression would have less error than the test RSS here, which could likely result in a model that overfits the "noise" that it finds in the data, leading to overfitting.

- (c) *Supposed that the true relationship between X and Y is not linear, but we don't know how far it is from linear. Consider the training 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.*

Sort of related to answer (a), again we will see that this polynomial regression will have a lower training RSS when compared to the linear fit option, again due to the tighter fit of the data (higher flexibility).

(d) *Answer (c) using test rather than training RSS.*

Not enough info. The training RSS could be both higher and lower. What if a linear model is the right model? Or a non-linear model is the right sort of model? There isn't enough here to say what which would be lower than the other/same.

5. *Consider the fitted values that result from performing linear regression without an intercept. In this setting, the i th fitted value takes the form*

$$\hat{y}_i = x_i \hat{\beta},$$

where

$$\hat{\beta} = \left(\sum_{i=1}^n x_i y_i \right) / \left(\sum_{i'=1}^n x_{i'}^2 \right).$$

Show that we can write

$$\hat{y}_i = \sum_{i'=1}^n a_{i'} y_{i'}.$$

What is $a_{i'}$?¹

$$\begin{aligned} \hat{y}_i &= x_i \hat{\beta} \\ \hat{\beta} &= \frac{\sum x_i y_i}{\sum x_i^2} \Rightarrow \\ \hat{y}_i &= x_i \left(\frac{\sum x_{i'} y_{i'}}{\sum x_{i'}^2} \right) \Rightarrow \\ \hat{y}_i &= \sum_{i'} \left(\frac{x_i x_{i'}}{\sum_j x_j^2} \right) y_{i'} \Rightarrow \\ a_{i'} &= \frac{x_i x_{i'}}{\sum_j x_j^2} \end{aligned}$$

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}) .*²

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{3.4}$$

That is equation 3.4 above, we will use it substitute it below to help us prove that the least squares line passes through the point (\bar{x}, \bar{y}) .

¹Note: We interpret this result by saying that the fitted values from linear regression are linear combinations of the response values.

²I talked to a friend from another college (an engineer who is much better at math than I am) and they helped me come to this elegant solution.

$$\begin{aligned}
y &= \hat{\beta}_0 + \hat{\beta}_1 x \\
0 &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} - \bar{y} \\
0 &= (\bar{y} - \hat{\beta}_1 \bar{x}) + \hat{\beta}_1 \bar{x} - \bar{y}
\end{aligned}$$

The above equation clearly simplifies to $0 = 0$, showing us that the least squares line does in fact go through a point that would be the means of both x and y .

Challenge Problem

Use the identities for expected value and variance to derive the bias-variance decomposition of

$$E \left[\left(y - \hat{f}(x) \right)^2 \right].$$

$E(c) = c$, where c is constant.

$$Var(X) = E(X^2) - [E(X)]^2$$

I'm sorry I didn't have time to take a crack at this. I'd still love to see it in the problem set key though! Thanks!