Math 610:Homework 2

Neal Kuperman

January 14, 2026

ISLP 3.10

This question should be answered using the Carseats data set.

(a) Fit a multiple regression model to predict Sales using Price, Urban, and US.

(b) Provide an interpretation of each coefficient in the model. Be careful—some of the variables in the model are qualitative!

(c) Write out the model in equation form, being careful to handle the qualitative variables properly.

(d) For which of the predictors can you reject the null hypothesis $H_0 : \beta_j = 0$?

(e) On the basis of your response to the previous question, fit a smaller model that only uses the predictors for which there is evidence of association with the outcome.

(f) How well do the models in (a) and (e) fit the data?

(g) Using the model from (e), obtain 95% confidence intervals for the coefficient(s).

(h) Is there evidence of outliers or high leverage observations in the model from (e)?

Note(s)

The data is a simulated data set containing sales of child car seats at 400 different stores. Information on the data set can be found on the ISLP documentation.

Variable	**Description**
Sales	Unit sales (in thousands) at each location
CompPrice	Price charged by competitor at each location
Income	Community income level (in thousands of dollars)
Advertising	Local advertising budget for company at each location (in thousands of dollars)
Population	Population size in region (in thousands)
Price	Price company charges for car seats at each site
ShelveLoc	Factor with levels Bad, Good, Medium — quality of shelving location
Age	Average age of the local population
Education	Education level at each location
Urban	Factor (No/Yes) — whether store is in urban or rural location
US	Factor (No/Yes) — whether store is in the US or not

Table 1: Carseats Dataset Variables

Solution

ISLP 3.10 (a)

Table 2 shows the results of a linear model to predict Sales using Price, Urban, and US. The table was generated using the `statsmodels` library in Python.

Dep. Variable:	Sales	**R-squared:**	0.239
Model:	OLS	**Adj. R-squared:**	0.234
Method:	Least Squares	**F-statistic:**	41.52
Date:	Tue, 13 Jan 2026	**Prob (F-statistic):**	2.39e-23
Time:	18:41:11	**Log-Likelihood:**	-927.66
No. Observations:	400	**AIC:**	1863.
Df Residuals:	396	**BIC:**	1879.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	13.0435	0.651	20.036	0.000	11.764	14.323
Price	-0.0545	0.005	-10.389	0.000	-0.065	-0.044
Urban_Yes	-0.0219	0.272	-0.081	0.936	-0.556	0.512
US_Yes	1.2006	0.259	4.635	0.000	0.691	1.710

Omnibus:	0.676	**Durbin-Watson:**	1.912
Prob(Omnibus):	0.713	**Jarque-Bera (JB):**	0.758
Skew:	0.093	**Prob(JB):**	0.684
Kurtosis:	2.897	**Cond. No.**	628.

Table 2: OLS Regression Results

ISLP 3.10 (b)

- **Price**: Sales is negatively related to price, with sales decreasing by 54 units per dollar increase in price
- **Urban_Yes**: There is not a significant statistical relationship between sales and whether or not a store is in an urban location
- **US_Yes**: There is a positive relationship between price and whether a store is located in the US or not. On average, a store located in the US will sell 1200 more units than if they were located outside the US.

ISLP 3.10 (c)

$$\text{Sales} = -0.0545 * \text{Price} + -0.0219 * \text{Urban\_Yes} + 1.2006 * \text{US\_Yes}$$

ISLP 3.10 (d)

The **Urban_Yes** predictor has a P value of 0.936, meaning we can reject the null hypothesis for it. We can not reject the null hypothesis for any of the other variables.

ISLP 3.10 (e)

Table 3 shows the results of a linear model to predict Sales using Price and US_Yes.

Dep. Variable:	Sales	**R-squared:**	0.239
Model:	OLS	**Adj. R-squared:**	0.235
Method:	Least Squares	**F-statistic:**	62.43
Date:	Tue, 13 Jan 2026	**Prob (F-statistic):**	2.66e-24
Time:	20:11:19	**Log-Likelihood:**	-927.66
No. Observations:	400	**AIC:**	1861.
Df Residuals:	397	**BIC:**	1873.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	13.0308	0.631	20.652	0.000	11.790	14.271
Price	-0.0545	0.005	-10.416	0.000	-0.065	-0.044
US_Yes	1.1996	0.258	4.641	0.000	0.692	1.708

Omnibus:	0.666	**Durbin-Watson:**	1.912
Prob(Omnibus):	0.717	**Jarque-Bera (JB):**	0.749
Skew:	0.092	**Prob(JB):**	0.688
Kurtosis:	2.895	**Cond. No.**	607.

Table 3: OLS Regression Results for Model with Price and US_Yes

ISLP 3.10 (f)

Models a and e fit the data equally poorly. Both have an R^2 value of 0.239

ISLP 3.10 (g)

Table 4 shows the 95% confidence intervals for the coefficients in the model from (e).

	0	1
Intercept	11.790320	14.271265
Price	-0.064760	-0.044195
US_Yes	0.691520	1.707766

Table 4: 95% Confidence Intervals for Coefficients in Model from (e)

ISLP 3.10 (h)

Table 5 shows the high leverage points in the model from (e). The Cook's Distances are plotted in Figure 1.

index	high_leverage_points
42	0.043338
125	0.025966
155	0.016106
156	0.015356
159	0.015707
165	0.028567
171	0.021014
174	0.029687
191	0.018039
203	0.015356
208	0.018235
269	0.019195
272	0.018687
313	0.023165
315	0.017049
356	0.018279
365	0.017399
367	0.023707
383	0.016514
386	0.016555

Table 5: High Leverage Points in Model from (e)

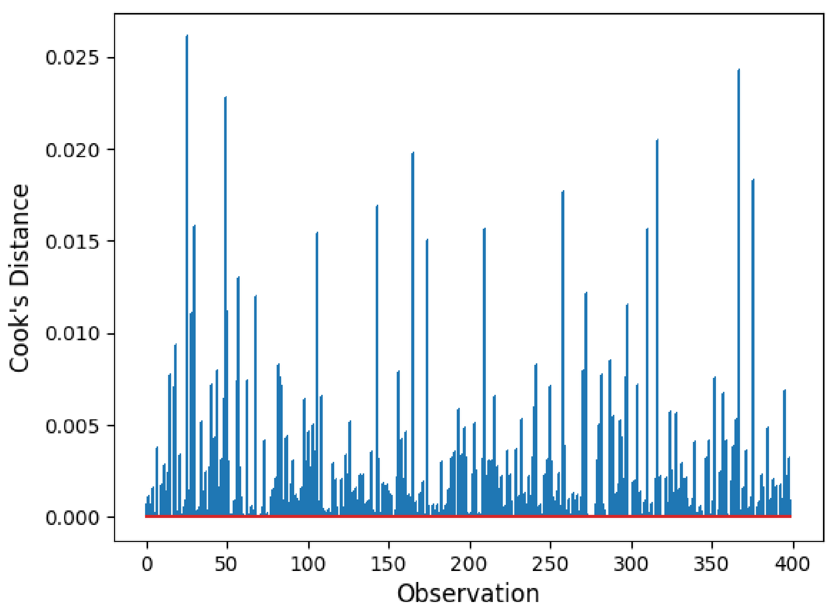


Figure 1: Cook's Distances for Model from (e)

ISLP 3.14

This problem focuses on the *collinearity* problem.

(a) Perform the following commands in Python:

```
rng = np.random.default_rng(10)
x1 = rng.uniform(0, 1, size=100)
x2 = 0.5 * x1 + rng.normal(size=100) / 10
y = 2 + 2 * x1 + 0.3 * x2 + rng.normal(size=100)
```

The last line corresponds to creating a linear model in which y is a function of x1 and x2. Write out the form of the linear model. What are the regression coefficients?

(b) What is the correlation between x1 and x2? Create a scatterplot displaying the relationship between the variables.

(c) Using this data, fit a least squares regression to predict y using x1 and x2. Describe the results obtained. What are $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$? How do these relate to the true β_0, β_1 and β_2? Can you reject the null hypothesis $H_0 : \beta_1 = 0$? How about the null hypothesis $H_0 : \beta_2 = 0$?

(d) Now fit a least squares regression to predict y using only x1. Comment on your results. Can you reject the null hypothesis $H_0 : \beta_1 = 0$?

(e) Now fit a least squares regression to predict y using only x2. Comment on your results. Can you reject the null hypothesis $H_0 : \beta_2 = 0$?

(f) Do the results obtained in (c)-(e) contradict each other? Explain your answer.

(g) Suppose we obtain one additional observation, which was unfortunately mismeasured. We use the function np.concatenate() to add this additional observation to each of x1, x2 and y.

```
x1 = np.concatenate([x1, [0.1]])
x2 = np.concatenate([x2, [0.8]])
y = np.concatenate([y, [6]])
```

Re-fit the linear models from (c) to (e) using this new data. What effect does this new observation have on the each of the models? In each model, is this observation an outlier? A high-leverage point? Both? Explain your answers.

Solution

ISLP 3.14 (a)

The linear model has the form

$$y = 2 + 2x_1 + 0.3x_2 + \varepsilon$$

$$\beta_0 = 2$$
$$\beta_1 = 2$$
$$\beta_2 = 0.3$$

ISLP 3.14 (b)

$$corr(x_1, x_2) = \frac{Cov(x_1 x_2)}{\sigma_{x_1} * \sigma_{x_2}} = \frac{E[(x_1 - \overline{x_1})(x_2 - \overline{x_2})]}{\sigma_{x_1} * \sigma_{x_2}} = \frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2 \cdot \sum_{i=1}^{n}(x_{2i} - \bar{x}_2)^2}}$$

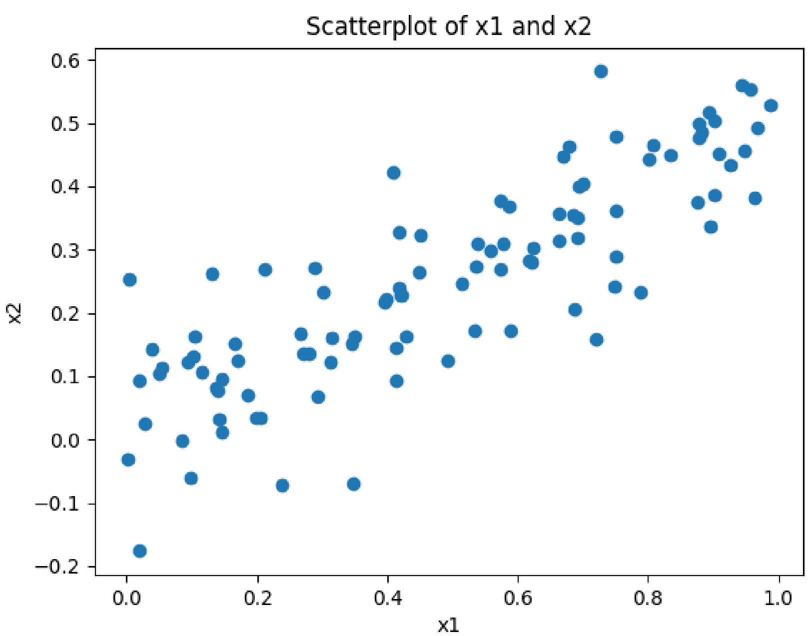


Figure 2: Scatterplot of x1 and x2

ISLP 3.14 (c)

Dep. Variable:	y	**R-squared:**	0.291
Model:	OLS	**Adj. R-squared:**	0.276
Method:	Least Squares	**F-statistic:**	19.89
Date:	Tue, 13 Jan 2026	**Prob (F-statistic):**	5.76e-08
Time:	21:24:01	**Log-Likelihood:**	-130.62
No. Observations:	100	**AIC:**	267.2
Df Residuals:	97	**BIC:**	275.1
Df Model:	2		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	1.9579	0.190	10.319	0.000	1.581	2.334
x1	1.6154	0.527	3.065	0.003	0.569	2.661
x2	0.9428	0.831	1.134	0.259	-0.707	2.592

Omnibus:	0.051	**Durbin-Watson:**	1.964
Prob(Omnibus):	0.975	**Jarque-Bera (JB):**	0.041
Skew:	-0.036	**Prob(JB):**	0.979
Kurtosis:	2.931	**Cond. No.**	11.9

Table 6: OLS Regression Results for Model with x1 and x2

The linear model obtained from the OLS is

$$y = 1.9579 + 1.6154x_1 + 0.9428x_2$$

and

Variable	**True**	**OLS Model**
β_0	2.0	1.9579
β_1	2.0	1.6154
β_2	0.3	0.9428

Table 7: Coefficients for Model with x1 and x2

We can reject the null hypothesis $H_0 : \beta_1 = 0$ as its t value is large ($t_{\beta_1} = 3.065$) and P value is small ($P_{\beta_1} = 0.003$). However, we can not reject the null hypothesis $H_0 : \beta_2 = 0$ because its t value is less than 2 ($t_{\beta_2} = 1.134$) along with a large P value ($P_{\beta_2} = 0.259$)

ISLP 3.14 (d)

Table 8 shows the results of the OLS regression to predict y using only x1. The goodness of fit of the least squares regression did not change meaningfully when we elmiited x2 as a predictor variable ($R^2_{x1,x2} = 0.291$, $R^2_{x1} = 0.281$). We can reject the null hypothesis $H_0 : \beta_1 = 0$ due to the large t value and small P value for x1

	coef	std err	t	P> |t|	[0.025	0.975]
const	1.9371	0.189	10.242	0.000	1.562	2.312
x1	2.0771	0.335	6.196	0.000	1.412	2.742

R-squared:	0.281	**Adj. R-squared:**	0.274
F-statistic:	38.39	**Prob (F-statistic):**	1.37e-08

Table 8: OLS Regression Results for Model with x1

ISLP 3.14 (e)

Table 9 shows the results of the OLS regression to predict y using only x2. The goodness of fit of the least squares regression with only x2 included did not change meaningfully when we elmiited x1 as a predictor variable ($R^2_{x1,x2} = 0.291$, $R^2_{x2} = 0.222$). We can reject the null hypothesis $H_0 : \beta_2 = 0$ due to the large t value and small P value for x2

	coef	std err	t	P> |t|	[0.025	0.975]
const	2.3239	0.154	15.124	0.000	2.019	2.629
x2	2.9103	0.550	5.291	0.000	1.819	4.002

R-squared:	0.222	**Adj. R-squared:**	0.214
F-statistic:	27.99	**Prob (F-statistic):**	7.43e-07

Table 9: OLS Regression Results for Model with x2

ISLP 3.14 (f)

They do not contradict each other, x1 and x2 explain the variance in y equally well (poorly) with $R^2_{x1} = 0.281$ and $R^2_{x2} = 0.222$. Additionally the null hypothesis can not be rejected when only including 1 predictor, either x1 or x2. These outcomes can be explained by x1 and x2 being colinear (they are highly coorelated).

ISLP 3.14 (g)

Table 10 shows the results of the OLS regression to predict y using x1, x2, x1 only, and x2 only. Figure 3 shows some diagnostic plots for the full model.

Full Model

	coef	**t value**	**p value**	**std err**
intercept	2.0618	10.7201	0.0000	0.1923
x_1	0.8575	1.8383	0.0690	0.4665
x_2	2.2663	3.2160	0.0018	0.7047
R-squared	0.2916	F-statistic		20.1727

x_1 **Only Model**

	coef	**t value**	**p value**	**std err**
intercept	2.0739	10.3098	0.0000	0.2012
x_1	1.8760	5.2360	0.0000	0.3583
R-squared	0.2169	F-statistic		27.4159

x_2 **Only Model**

	coef	**t value**	**p value**	**std err**
intercept	2.2840	15.0879	0.0000	0.1514
x_2	3.1458	6.0082	0.0000	0.5236
R-squared	0.2672	F-statistic		36.0984

Table 10: OLS Regression Results for Models with x1, x2, x1 only, and x2 only

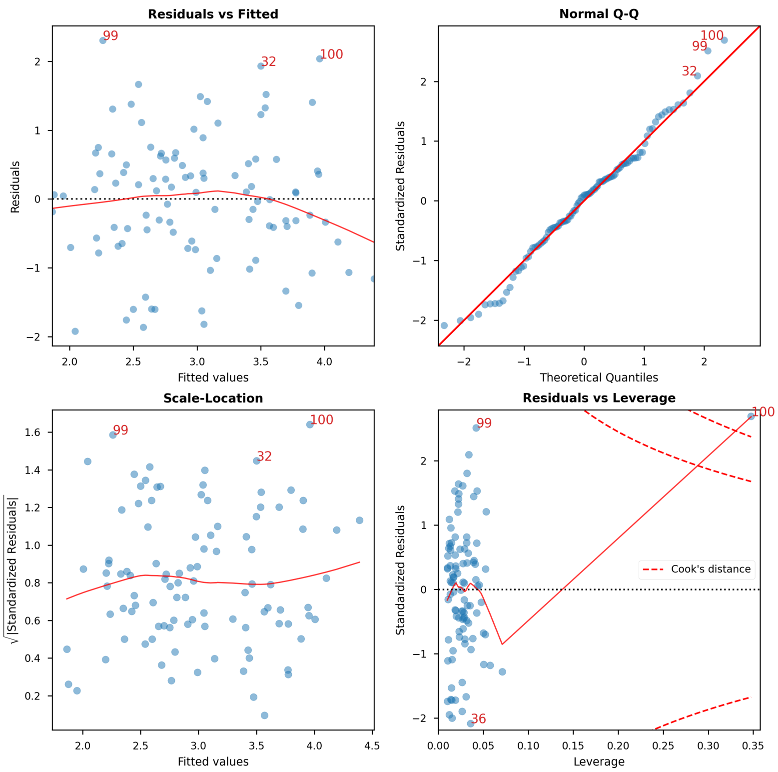


Figure 3: Diagnostic Plots for Full Model

With the added point, we can no longer reject the null hypothesis for β_1, as the t value is now 1.8383 and the P value is now 0.0690. We can now, however, reject the null hypothesis for β_2, as the t value is now 3.2160 and the P value is now 0.0018. The added point is a high-leverage point and an outlier in the full model. All models fit the data equally well (poorly), with R^2 values all less than 0.3.

The added point does have a high leverage (0.347) but is not an outlier is an outlier because its Cook's Distance is not greater than 1 (Cook's Distance = 0.316).

ISLP 3.15 (a, b, d)

This problem involves the Boston data set, which we saw in the lab for this chapter. We will now try to predict per capita crime rate using the other variables in this data set. In other words, per capita crime rate is the response, and the other variables are the predictors.

(a) For each predictor, fit a simple linear regression model to predict the response. Describe your results. In which of the models is there a statistically significant association between the predictor and the response? Create some plots to back up your assertions.

(b) Fit a multiple regression model to predict the response using all of the predictors. Describe your results. For which predictors can we reject the null hypothesis $H_0 : \beta_j = 0$?

(d) Is there evidence of non-linear association between any of the predictors and the response? To answer this question, for each predictor X, fit a model of the form

$$y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon \tag{1}$$

Solution

ISLP 3.15 (a)

Figure 4 shows the scatterplots for each predictor. Figure 5 shows the residuals for each predictor. Table 11 shows the R-squared for each predictor. All R-squared values are less than 0.4, indicating that the univariatemodels are not very good at predicting the response.

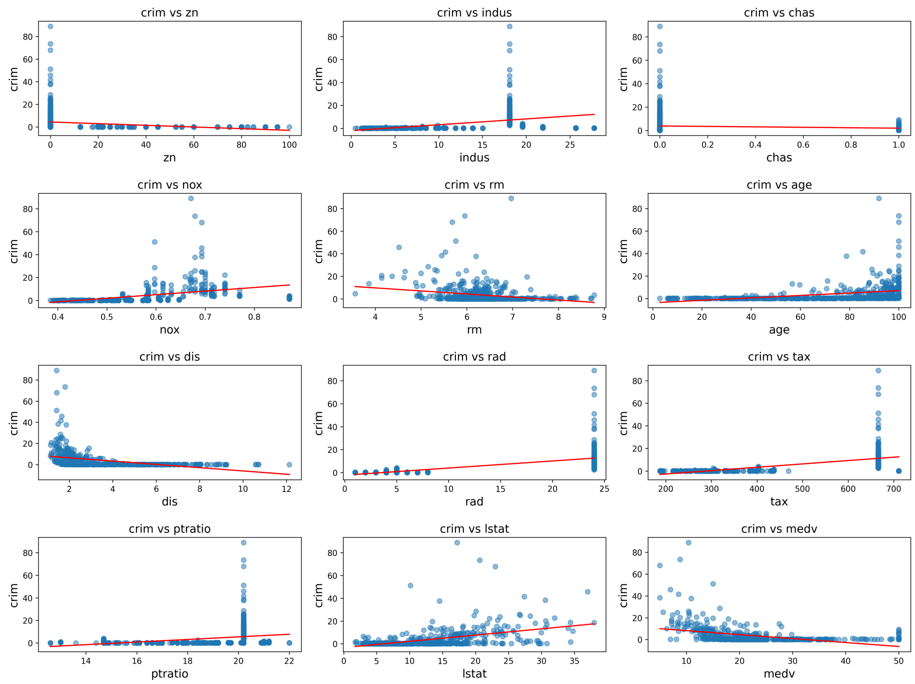


Figure 4: Scatterplots for Each Predictor

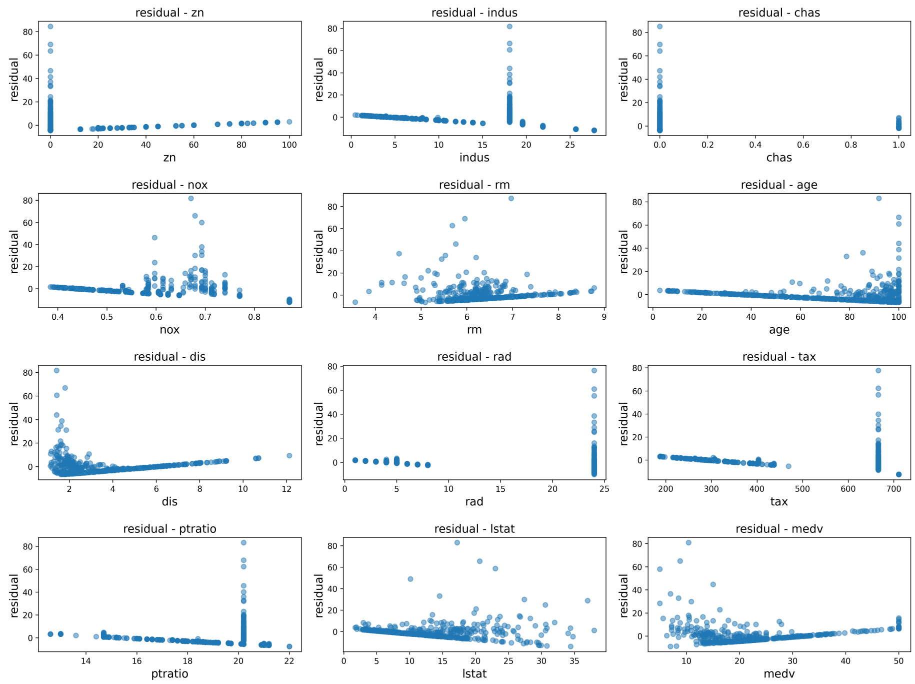


Figure 5: Residuals for Each Predictor

predictor	r_squared
rad	0.391257
tax	0.339614
lstat	0.207591
nox	0.177217
indus	0.165310
medv	0.150780
dis	0.144149
age	0.124421
ptratio	0.084068
rm	0.048069
zn	0.040188
chas	0.003124

Table 11: R-squared for Each Predictor

ISLP 3.15 (b)

Table 12 shows the results of the OLS regression to predict crim using all of the predictors. We can reject the null hypothesis, with a 95% confidence interval $\alpha = 0.05$, for zn, dis, rad, and medv.

We can not reject the null hypothesis for all other predictors.

Dep. Variable:	crim	**R-squared:**	0.449
Model:	OLS	**Adj. R-squared:**	0.436
Method:	Least Squares	**F-statistic:**	33.52
Date:	Tue, 13 Jan 2026	**Prob (F-statistic):**	2.03e-56
Time:	23:53:03	**Log-Likelihood:**	-1655.4
No. Observations:	506	**AIC:**	3337.
Df Residuals:	493	**BIC:**	3392.
Df Model:	12		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	13.7784	7.082	1.946	0.052	-0.136	27.693
zn	0.0457	0.019	2.433	0.015	0.009	0.083
indus	-0.0584	0.084	-0.698	0.486	-0.223	0.106
chas	-0.8254	1.183	-0.697	0.486	-3.150	1.500
nox	-9.9576	5.290	-1.882	0.060	-20.351	0.436
rm	0.6289	0.607	1.036	0.301	-0.564	1.822
age	-0.0008	0.018	-0.047	0.962	-0.036	0.034
dis	-1.0122	0.282	-3.584	0.000	-1.567	-0.457
rad	0.6125	0.088	6.997	0.000	0.440	0.784
tax	-0.0038	0.005	-0.730	0.466	-0.014	0.006
ptratio	-0.3041	0.186	-1.632	0.103	-0.670	0.062
lstat	0.1388	0.076	1.833	0.067	-0.010	0.288
medv	-0.2201	0.060	-3.678	0.000	-0.338	-0.103

Omnibus:	663.436	**Durbin-Watson:**	1.516
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	80856.852
Skew:	6.579	**Prob(JB):**	0.00
Kurtosis:	63.514	**Cond. No.**	1.24e+04

Table 12: OLS Regression Results for Multiple Regression Model

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.24e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

ISLP 3.15 (d)

Tables 13 - 15 show the results of the polynomial regressions for each predictor. Figure 6 and figure 7 show the scatter and residuals plots for the polynomial regression models respectively.

Looking at the scatter plots, we can see that none of the variables have great non-linear fits to the response. Looking exclusively at the P values for the different polynomial models, we can not reject a non-linear relationship for indius, nox, age, dis, ptratio, and medv.

Model for zn:

	coef	t value	p value	std err
intercept	4.8461	11.1922	0.0000	0.4330
zn	-0.3322	-3.0252	0.0026	0.1098
zn^2	0.0065	1.6791	0.0938	0.0039
zn^3	-0.0000	-1.2030	0.2295	0.0000
R-squared	0.0582	F-statistic		10.3485

Model for indus:

	coef	t value	p value	std err
intercept	3.6626	2.3269	0.0204	1.5740
indus	-1.9652	-4.0773	0.0001	0.4820
$indus^2$	0.2519	6.4070	0.0000	0.0393
$indus^3$	-0.0070	-7.2920	0.0000	0.0010
R-squared	0.2597	F-statistic		58.6883

Model for chas:

	coef	t value	p value	std err
intercept	3.7444	9.4530	0.0000	0.3961
chas	-0.6309	-1.2567	0.2094	0.5020
$chas^2$	-0.6309	-1.2567	0.2094	0.5020
$chas^3$	-0.6309	-1.2567	0.2094	0.5020
R-squared	0.0031	F-statistic		1.5794

Model for nox:

	coef	t value	p value	std err
intercept	233.0866	6.9282	0.0000	33.6431
nox	-1279.3713	-7.5082	0.0000	170.3975
nox^2	2248.5441	8.0334	0.0000	279.8993
nox^3	-1245.7029	-8.3446	0.0000	149.2816
R-squared	0.2970	F-statistic		70.6867

Table 13: Polynomial Regression Results for Each Predictor

Model for rm:

	coef	**t value**	**p value**	**std err**
intercept	112.6246	1.7457	0.0815	64.5172
rm	-39.1501	-1.2503	0.2118	31.3115
rm^2	4.5509	0.9084	0.3641	5.0099
rm^3	-0.1745	-0.6615	0.5086	0.2637
R-squared	0.0678	F-statistic		12.1677

Model for age:

	coef	**t value**	**p value**	**std err**
intercept	-2.5488	-0.9204	0.3578	2.7691
age	0.2737	1.4683	0.1427	0.1864
age^2	-0.0072	-1.9878	0.0474	0.0036
age^3	0.0001	2.7237	0.0067	0.0000
R-squared	0.1742	F-statistic		35.3061

Model for dis:

	coef	**t value**	**p value**	**std err**
intercept	30.0476	12.2850	0.0000	2.4459
dis	-15.5544	-8.9600	0.0000	1.7360
dis^2	2.4521	7.0783	0.0000	0.3464
dis^3	-0.1186	-5.8135	0.0000	0.0204
R-squared	0.2778	F-statistic		64.3741

Model for rad:

	coef	**t value**	**p value**	**std err**
intercept	-0.6055	-0.2954	0.7678	2.0501
rad	0.5127	0.4913	0.6234	1.0436
rad^2	-0.0752	-0.5061	0.6130	0.1485
rad^3	0.0032	0.7031	0.4823	0.0046
R-squared	0.4000	F-statistic		111.5727

Table 14: Polynomial Regression Results for Each Predictor

Model for tax:

	coef	**t value**	**p value**	**std err**
intercept	19.1836	1.6263	0.1045	11.7955
tax	-0.1533	-1.6023	0.1097	0.0957
tax^2	0.0004	1.4877	0.1375	0.0002
tax^3	-0.0000	-1.1668	0.2439	0.0000
R-squared	0.3689	F-statistic		97.8047

Model for ptratio:

	coef	**t value**	**p value**	**std err**
intercept	477.1840	3.0434	0.0025	156.7950
ptratio	-82.3605	-2.9793	0.0030	27.6439
$ptratio^2$	4.6353	2.8821	0.0041	1.6083
$ptratio^3$	-0.0848	-2.7433	0.0063	0.0309
R-squared	0.1138	F-statistic		21.4839

Model for lstat:

	coef	**t value**	**p value**	**std err**
intercept	1.2010	0.5920	0.5541	2.0286
lstat	-0.4491	-0.9660	0.3345	0.4649
$lstat^2$	0.0558	1.8522	0.0646	0.0301
$lstat^3$	-0.0009	-1.5170	0.1299	0.0006
R-squared	0.2179	F-statistic		46.6294

Model for medv:

	coef	**t value**	**p value**	**std err**
intercept	53.1655	15.8405	0.0000	3.3563
medv	-5.0948	-11.7438	0.0000	0.4338
$medv^2$	0.1555	9.0455	0.0000	0.0172
$medv^3$	-0.0015	-7.3120	0.0000	0.0002
R-squared	0.4202	F-statistic		121.2721

Table 15: Polynomial Regression Results for Each Predictor

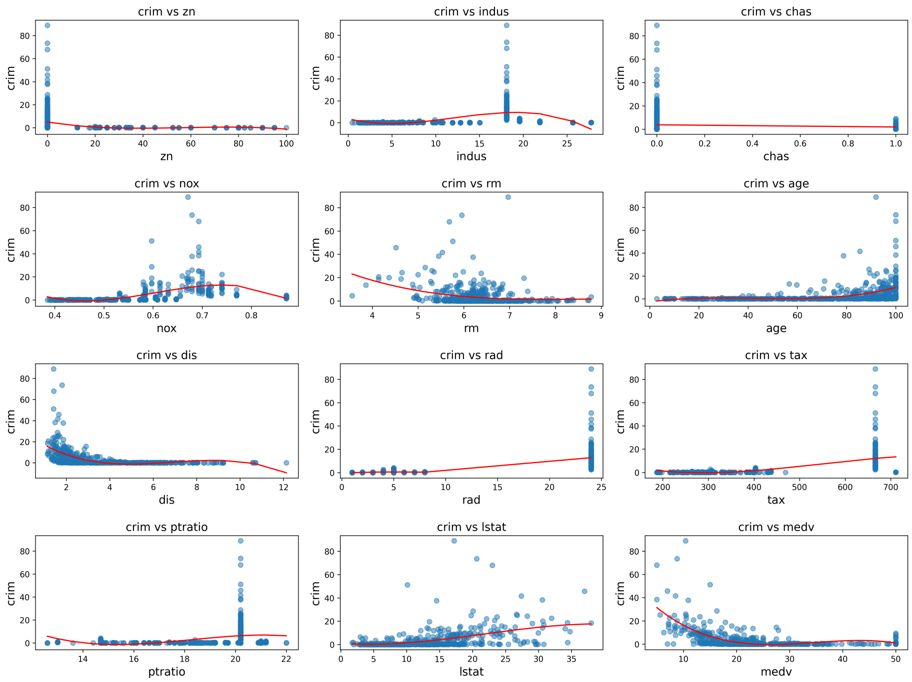
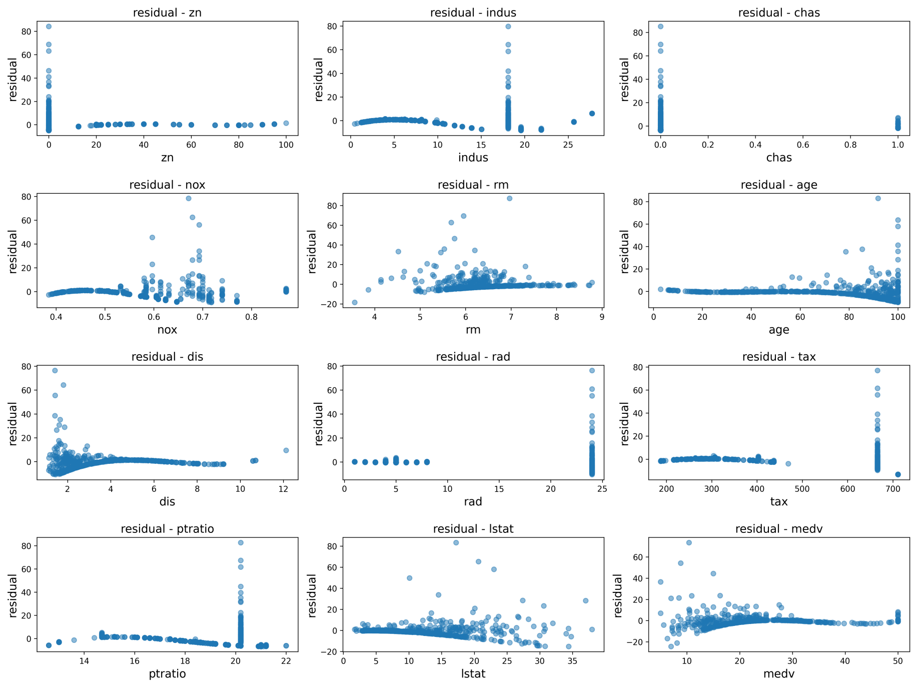


Figure 6: Scatterplots for Polynomial Regression Models



0

Figure 7: Residuals for Polynomial Regression Models

ESL 3.17

Repeat the analysis of Table 3.3 on the spam data discussed in Chapter 1. Include LS, Best Subset, Ridge regression, and Lasso. (skip PCR and PLS)

Note(s)

- **Best Subset**: Best-subset selection drops all variables with coefficients smaller than the Mth largest; this is a form of “hard-thresholding.”
- **Ridge regression**:

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\text{argmin}} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\} \qquad \text{with } \lambda \geq 1$$

  which can be written in matrix form as

$$\text{RSS}(\lambda) = (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \lambda\beta^2\beta$$

  with

$$\hat{\beta}^{ridge} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$$

- **Lasso**: Similar to ridge but with an L_1 penelty instead of L_2:

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\text{argmin}} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \right\}$$

$$\text{subject to } \sum_{j=1}^{p} |\beta_j| \leq t$$

  or equivalently in *Lagrangian form*

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\text{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \qquad \text{with } \lambda \geq 1$$

  NOTE: t should be adaptively chosen to minimize an estimate of expected prediction error.

  NOTE: Ridge regression does a proportional shrinkage. Lasso translates each coefficient by a constant factor λ, truncating at zero (“soft thresholding)

Solution