

Introduction to Linear Regression
STAT-UB.0003: Regression and Forecasting Models

Hypothesis tests (review)

1. We collect a simple random sample of size $n = 100$ from a population. The sample mean is $\bar{x} = 12.4$ and the sample standard deviation is $s = 8.0$. Use this data to test the null hypothesis $H_0 : \mu = 10.0$ against the alternative $H_a : \mu \neq 10.0$, where μ denotes the population mean:

(a) Compute the test statistic.

(b) If the null hypothesis were true and we were to repeat the experiment, we would get a new test statistic. In this hypothetical setting, approximately what is the probability of getting a new test statistic at least as extreme as the observed test statistic we computed in part (a)?

(c) What is the p-value for performing this hypothesis test? Give a one-sentence explanation.

(d) Using a significance level (α) of 5%, what is the result of the hypothesis test?

Linear regression

2. In the following scenarios, which would you consider to be predictor (x) and which would you consider to be response (y)?
- (a) Sales revenue; Advertising expenditures
 - (b) Starting salary after college; Undergraduate GPA
 - (c) The current month's sales; the previous month's sales
 - (d) The size of an apartment; the sale price of an apartment.
 - (e) A restaurant's Zagat Price rating; a restaurant's Zagat Food rating.

3. Let y be the payment (in dollars) for a repair which takes x hours. Suppose that

$$y = 25 + 30x.$$

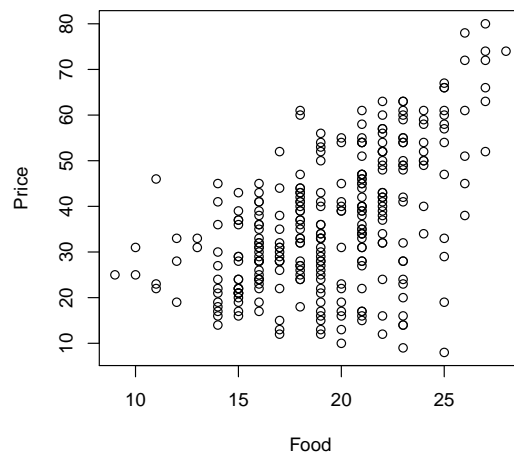
What is the interpretation of this model?

4. Consider two variables measured on 294 restaurants in the 2003 Zagat guide:

y = typical dinner price, including one drink and tip (\$)

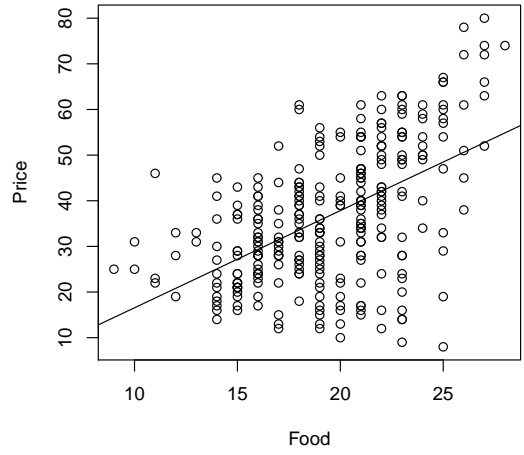
x = Zagat quality rating (0–30).

Here is a scatterplot of y on x :



Why is an exact linear relationship inappropriate to describe the relationship between y and x ?

5. Here is the least squares regression fit to the Zagat restaurant data:



Here is the Minitab output from the fit:

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
12.5559	27.93%	27.68%	26.86%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-4.74	3.95	-1.20	0.232	
Food	2.129	0.200	10.64	0.000	1.00

Regression Equation

Price = -4.74 + 2.129 Food

- (a) What are the estimated intercept and slope?

- (b) Use the estimated regression model to estimate the average dinner price of all restaurants with a quality rating of 20.

- (c) In the estimated regression model, what is the interpretation of the slope?

- (d) In the estimated regression model, why doesn't the intercept have a direct interpretation?

6. Refer to the Minitab output from the previous problem, the regression analysis of the Zagat data.

(a) What is the estimated standard deviation of the error (the “standard error of the regression”)? What is the interpretation of this value?

(b) According to the estimated regression model, what is the range of typical prices for restaurants with quality ratings of 20?

(c) According to the estimated regression model, what is the range of typical prices for restaurants with quality ratings of 10?