

## Review Exercises

1. For each of the claims that follow, choose which of the following tests would be most appropriate.
  - (a) One sample z/t test
  - (b) Two sample z/t test
  - (c) Paired difference test
  - (d) One sample proportion test
  - (e) Two sample proportion test
  - (f) Analysis of variance
  - (g) Chi-square goodness-of-fit test
  - (h) Regression slope test
    - i. More than half of all babies born are boys.
    - ii. The probability a student will pass Engineering 5 is the same as the probability a student will pass Engineering 7.
    - iii. The average cost of a pack of cigarettes is higher in Canada than in the U. S.
    - iv. Smokers have a shorter life expectancy than non-smokers.
    - v. Increased spending on schools corresponds to better student performance.
    - vi. Gas prices in Santa Cruz are the same as in San Jose.
    - vii. The average GPA is the same for students in each of the ten colleges.
    - viii. A coin is fair.
    - ix. The average rent for a one-bedroom apartment in Santa Cruz is \$1000.

2. In a class I taught earlier, regressing final exam scores on midterm scores gives the following output:

```
Call:  lm(formula = Final ~ Midterm)
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	43.16489	3.74433	11.528	1.56e-08
Midterm	0.42349	0.05005	8.462	7.09e-07

```
Multiple R-Squared: 0.8364
```

- (a) Test if there is an association between midterm and final exam scores in this class.
  - (b) If a student got 75 out of 100 on the midterm, what is their predicted final exam score?
  - (c) Interpret the R-Squared. Is this a good fit or a poor one?
  - (d) If the  $t$ -test for regression has 14 degrees of freedom, how many students were in the class?
3. Regression models are often used to help determine the appraised value of single-family houses. This dataset involves houses from a suburban New York community. Available variables for prediction are lot size, number of bedrooms, number of bathrooms, age of the house, and whether it has central air conditioning. Use the regression output and plots (next page) that follow to answer these questions:
- (a) Is there a relationship between appraised value and the five explanatory variables?
  - (b) Interpret the  $R^2$  in the context of the problem.
  - (c) Interpret the slope coefficient for the number of bathrooms.
  - (d) If you had access to a computer with a statistical package, what would be your next step? Using the available information, what model do you think will be the best model to use?

Call: `lm(formula = Value ~ Lotsize + Bed + Bath + Age + CAC)`

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	153.0023	21.0526	7.268	1.51e-09
Lotsize	-0.8226	0.9590	-0.858	0.394849
Bed	-4.6830	4.4999	-1.041	0.302666
Bath	31.8970	7.6369	4.177	0.000109
Age	0.2055	0.2660	0.773	0.443135
CAC	50.4346	14.1393	3.567	0.000766

Residual standard error: 26.06 on 54 degrees of freedom

Multiple R-Squared: 0.415, Adjusted R-squared: 0.3608

F-statistic: 7.662 on 5 and 54 DF, p-value: 1.68e-05