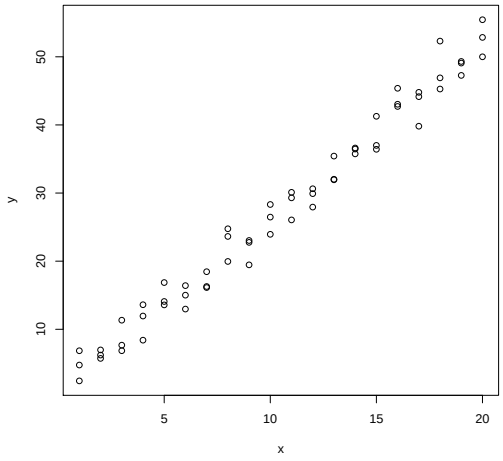
Exam 1 Practice Problems

In addition to the exercises below, please see the textbook exercises that correspond to the relevant sections of Chapters 2 and 3 covered on this exam.

1. Which of the following best describes the least-squares line fit to the data shown in the plot.



(a) $\widehat{\beta}_0 = 2.9$, $\widehat{\beta}_1 = -1.0$

(b) $\widehat{\beta}_0 = -1.2$, $\widehat{\beta}_1 = -2.5$

(c) $\widehat{\beta}_0 = 2.1$, $\widehat{\beta}_1 = 11.0$

(d) $\widehat{\beta}_0 = 1.2$, $\widehat{\beta}_1 = 2.5$

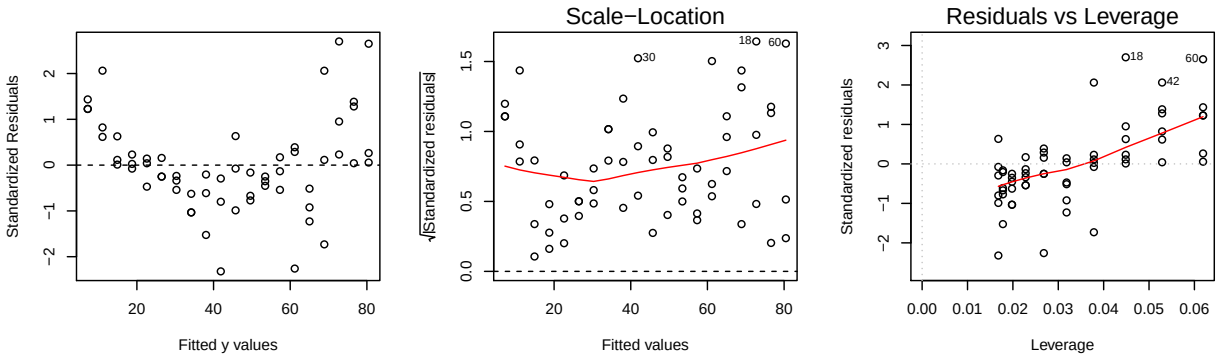
(e) $\widehat{\beta}_0 = -2.9$, $\widehat{\beta}_1 = 1.0$

2. If $n = 25$, $\bar{x} = 4$, $S_{XX} = 16$, $\bar{y} = 6$, and $S_{XY} = 8$, then what are the least squares estimates of $\widehat{\beta}_0$ and $\widehat{\beta}_1$?

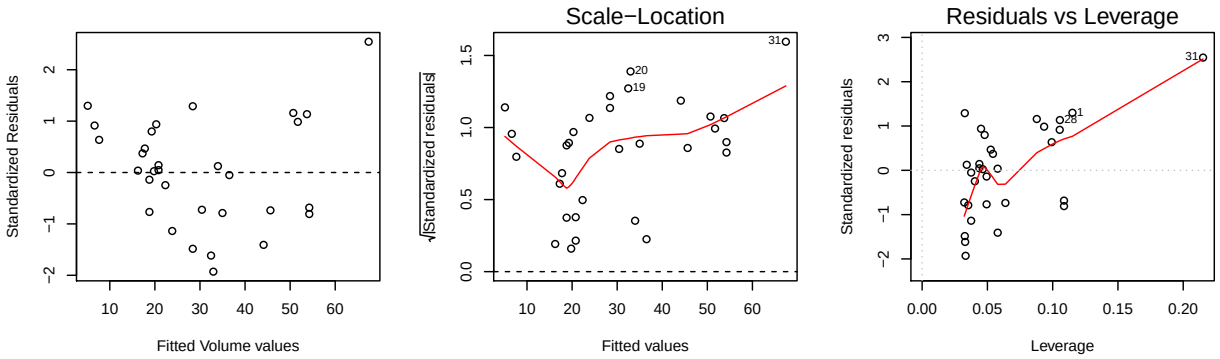
3. Which of the following would (always) results in a larger prediction interval for y_i?

(a) a larger sample size (n);
(b) a larger value of $\hat{y}_i$;
(c) a larger confidence level (smaller α);
(d) an x_i with lower leverage;
(e) a smaller estimated residual standard deviation (S);
(f) none of these

4. List all Simple Linear Regression assumptions that might not be satisfied for the following data.



5. Consider the Volume and Girth data in R's `trees` dataset:




```
##
## Call:
## lm(formula = Volume ~ Girth, data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.065 -3.107  0.152  3.495  9.587
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -36.9435     3.3651  -10.98 7.62e-12 ***
## Girth         5.0659     0.2474   20.48  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.252 on 29 degrees of freedom
## Multiple R-squared:  0.9353,Adjusted R-squared:  0.9331
## F-statistic: 419.4 on 1 and 29 DF,  p-value: < 2.2e-16
```

Answer the following:

(a) What is the 95% confidence interval for the regression intercept?

(b) Based on this interval, is it reasonable to conclude that the *true* intercept is zero, i.e., $\beta_0 = 0$?

(c) Write out the null and alternate hypotheses, and explain what the test means in terms of tree structure.

(d) Do the diagnostic plots above make you trust these model results, or not? Explain.

True or False

For each question, circle either T (true) or F (false). Answering "true" implies that the given statement is *always* true. Statements are made in the context of this class, and the usual SLR assumptions.

1. T F Assuming our simple linear regression model, each least squares coefficient $\widehat{\beta}_j$ has expected value β_j/n (j is either 0 or 1).

2. T F The observations y_i (aka $Y|X = x_i$), for $i = 1$ to n, are all independent and identically distributed.

3. T F Uncertainty about the regression coefficients depends upon the variance of the residuals.

4. T F If x_i has high leverage, then $E[e_i^2]$ is large relative to the true residual variance (σ^2).

5. T F The true variance (σ^2) of the residuals will decrease as the sample size increases.

6. T F Least squares estimates of the coefficients $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are chosen to minimize S, the residual standard error.

7. T F In an analysis of variance, the F-statistic follows a Student's t distribution.
8. T F An observation with a residual of more than 1000 is an outlier.

9. T F A Normal Q-Q plot shows standardized residuals versus the expected order statistics for a Normal distribution with mean $\bar{y}$ and standard deviation S.
10. T F In simple linear regression, the slope of the regression line is equal to R^2.