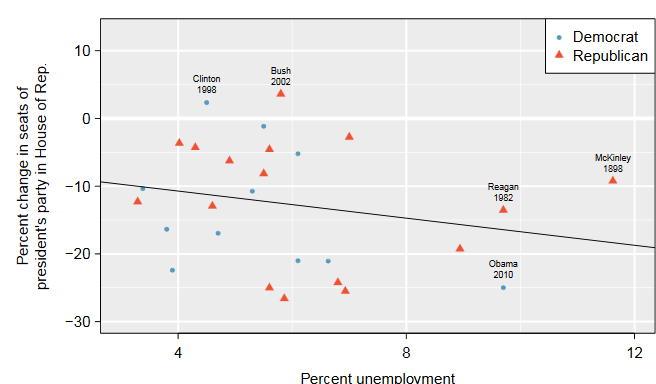
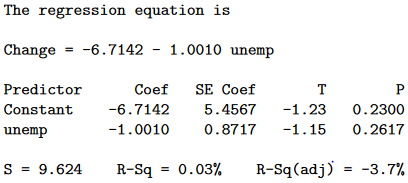
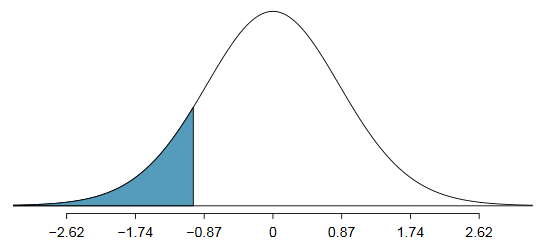
**Linear Regression**

* Elections for the House of Representatives occur every 2 years, coinciding every 4 years w/ the Presidential election, w/ the elections in the middle of a term called **midterm elections**.
* In America’s 2-party system, 1 political theory suggests the higher the unemployment rate, the worse the President’s party will do in the midterms.
* To assess the validity of this claim, we can compile historical data and look for a connection.
* We consider every midterm election from 1898 to 2010, with the exception of those elections during the Great Depression.
* 
* Figure above shows these data and the **least-squares regression line**:
* *% change in House seats for President’s party = -6.71 – 1.00\* (unemployment rate)*
* Consider the % change in number of seats of the President’s party (e.g. % change in the number of seats for Democrats in 2010) against the unemployment rate
* Examining the data, there are no clear deviations from linearity, the constant variance condition, or in the normality of residuals (though we don’t examine a normal probability plot here).
* While the data are collected sequentially, a separate analysis was used to check for any apparent correlation between successive observations; no such correlation was found.
* The data for the Great Depression (1934 and 1938) were removed b/c the unemployment rate was 21% and 18%, respectively. Do you agree they should be removed for this investigation?
* We will provide 2 considerations.
* Each of these points would have *very high* leverage on *any* least-squares regression line, and years w/ such high unemployment may not help us understand what would happen in other years where unemployment is only modestly high.
* On the other hand, these are exceptional cases, and we would be discarding important information if we exclude them from a final analysis.
* We have a negative slope, however, this slope (+ the y-intercept) are only *estimates* of the parameter values.
* We might wonder, *is this convincing evidence that the “true” linear model has a negative slope*?
* That is, *do the data provide strong evidence that the political theory is accurate*?
* We can frame this investigation into a 1-sided statistical hypothesis test:
* H(0): B1 = 0 🡪 The true linear model has slope zero.
* H(a): B1 < 0 🡪 The true linear model has a slope less than zero. The higher the unemployment, the greater the loss for the President’s party in the House of Representatives.
* We would reject H(0) in favor of H(a) if the data provide strong evidence that the true slope parameter is less than zero.
* To assess the hypotheses, we ID a **standard error for the estimate**, compute an appropriate test statistic, + identify the p-value.

Just like other point estimates seen before, we can compute a standard error + test statistic for B1, + will generally label the test statistic w/ “t”, since it follows a t-distribution (don’t know parameters)

* ***TIP:*** *Use a t-test with n−2 dF when performing a hypothesis test on the slope of a regression line*
* We will rely on statistical software to compute the standard error
* 
* The row labeled **unemp** represents the info for the slope, which is the coefficient of the unemployment variable.
* The entries in the 1st column **Coef** represent the least squares estimates, B0 and B1
* The values in the 2nd column **SE Coef** correspond to the standard errors of each estimate.



* This distribution shown here is the sampling distribution for B1 if the null was true.
* The shaded tail represents the p-value for the hypothesis test evaluating whether there is convincing evidence that higher unemployment corresponds to a greater loss of House seats for the President’s party during a midterm election
* We previously used a T-test statistic for hypothesis testing in the context of numerical data.
* Regression is very similar.
* In the hypotheses we are considering, the null value for the slope is 0, so we can compute the test statistic using the t-score formula:



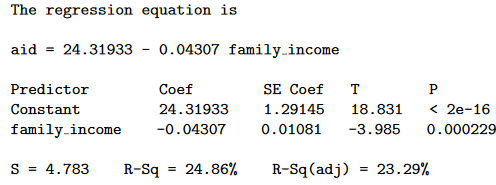
* We can look for the 1-sided p-value using the probability table, + in this example, the sample size n = 27, so df = 27 − 2 = 25
* We’d then see that the absolute value of our test statistic is smaller than any value listed, which means the tail area + therefore also the p-value is larger than 0.100 (one tail!).
* B/c the p-value is so large, we FAIL to reject the null hypothesis = the data do NOT provide convincing evidence that a higher unemployment rate has any correspondence w/ smaller or larger losses for the President’s party in the House of Representatives in midterm elections.
* We could have ID’ed the T-test statistic from the software output of the model (unemp row, t-value column), while the last column = the p-value for the 2-sided hypothesis test where the null = zero.
* *The corresponding 1-sided test would have a p-value half of the listed value.*

Inference for regression

* We usually rely on statistical software to ID point estimates + standard errors for parameters of a regression line.
* After verifying conditions hold for fitting a line, we can use to create confidence intervals for regression parameters or to evaluate hypothesis tests.

*Caution: Don’t carelessly use the p-value from regression output*

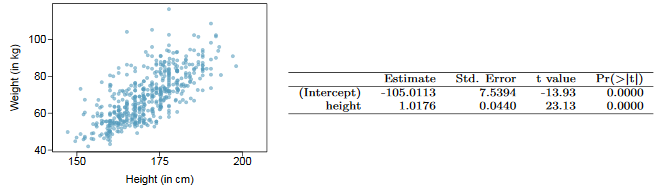
* The last column in regression output often lists p-values for 1 particular hypothesis: a *TWO-sided* test where the null value is zero.
* If your test is ONE-sided *and the point estimate is in the direction of H(A),* you can halve the software’s p-value to get the 1-tail area.
* If neither of these scenarios match your hypothesis test, be cautious about using the software output to obtain the p-value.
* Recall that **B1 = R\*(s(y)/s(x))**
* If the slope of the *true* regression line = 0, then the population correlation coefficient must also be 0
* The linear regression test for β1 = 0 is equivalent, then, to a test for the population correlation coefficient ρ = 0.
* The regression summary below shows statistical software output from fitting a least squares regression line. Use this output to formally evaluate the following hypotheses;
* H(0): The true slope of the regression line is zero
* H(a): The true slope of the regression line is not zero.



* Look in the 2nd row corresponding to the family income variable + see the point estimate of the slope of the line = -0.0431, standard error of this estimate = 0.0108, + the T-test statistic = -3.98.
* The p-value 0.0002 corresponds *exactly* to the *two*-sided test we are interested in, which is so small that we are able to reject the null + conclude that family income + financial aid at Elmhurst College for freshman entering in the year 2011 are negatively correlated + the true slope parameter is indeed less than 0
* **TIP**: Always check assumptions
* If conditions for fitting the regression line do not hold, then the methods presented here should not be applied.
* The standard error or distribution assumption of the point estimate (assumed to be normal when applying the T-test statistic) may not be valid
* Hypothesis test for the slope of regression line
* 1. State the name of the test being used.
* Linear regression t-test
* 2. Verify conditions.
* The residual plot has no pattern.
* 3. Write the hypotheses in plain language. No mathematical notation is needed for this test.
* H(0): β1 = 0 🡪 There is no significant linear relationship between [x] and [y].
* H(A): β1 <>, <, > 0 🡪 There is a significant linear relationship between [x] and [y].
* 4. Identify the significance level α.
* 5. Calculate the test statistic and dF
* T = (point estimate − null value) / SE of estimate
* The point estimate = B1
* SE can be located on regression summary table next to value of B1
* dF = n - 2
* 6. Find the p-value, compare it to α, + state whether to reject or not reject the null hypothesis.
* 7. Write the conclusion in the context of the question.
* Constructing a confidence interval for the slope of regression line
* 1. State the name of the CI being used.
* t-interval for slope of regression line
* 2. Verify conditions.
* The residual plot has no pattern.
* 3. Plug in the numbers and write the interval in the form
* **point estimate ± t\* x SE of estimate**
* The point estimate = B1
* dF = n - 2
* The critical value t∗ can be found on the t-table at row dF = n − 2
* SE can be located on regression summary table next to value of B1
* 4. Evaluate the CI and write in the form ( \_, \_).
* 5. Interpret the interval: “We are [XX]% confident that this interval contains the true average increase in [y] for each additional [unit] of [x].
* 6. State the conclusion to the original question.

Inference for the slope of a regression line

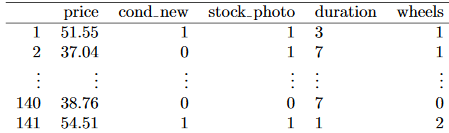
* In the following exercises, visually check the conditions for fitting a least squares regression line, but you do not need to report these conditions in your solutions.



* Describe the relationship between height and weight.
* Positive linear relationship (moderate-to-strong)
* Write the equation of the regression line. Interpret the slope and intercept in context.
* y = 1.0176\*x – 105.0113 height = 1.0176\*weight – 105.0113
* Do the data provide strong evidence that an increase in height is associated w/ an increase in weight? State the null and alternative hypotheses, report the p-value, and state your conclusion
* H(0): The true slope coefficient of height is 0 🡪 There is no relationship between height and weight
* H(A): The true slope coefficient of height > 0 🡪 There is a relationship between height and weight
* The p-value for this is incredibly small for a two-sided t-test, so it would be even smaller than a one-sided t-test
* We therefore can reject the Null 🡪 **The data provides convincing evidence that height and weight are positively correlated**
* The correlation coefficient for height and weight is 0.72. Calculate R2 and interpret it in context.
* R2 = 0.72^2 = 0.5184 = **Approximately 52% of the variability in weight can be attributed to the variability in height**
* 8.6. EXERCISES
* 419
* 8.36 Beer and blood alcohol content.
* Many people believe that gender, weight, drinking
* habits, and many other factors are much more important in predicting blood alcohol content (BAC)
* than simply considering the number of drinks a person consumed. Here we examine data from
* sixteen student volunteers at Ohio State University who each drank a randomly assigned number
* of cans of beer. These students were evenly divided between men and women, and they differed
* in weight and drinking habits. Thirty minutes later, a police officer measured their blood alcohol
* content (BAC) in grams of alcohol per deciliter of blood.
* 27
* The scatterplot and regression table
* summarize the findings.
* l
* l
* l
* l
* l
* l
* l
* l
* l
* l
* l
* l
* l
* l
* l
* l
* 2
* 4
* 6
* 8
* 0.05
* 0.10
* 0.15
* Cans of beer
* BAC (grams per deciliter)
* Estimate
* Std. Error
* t value
* Pr(
* >
* |
* t
* |
* )
* (Intercept)
* -0.0127
* 0.0126
* -1.00
* 0.3320
* beers
* 0.0180
* 0.0024
* 7.48
* 0.0000
* (a) Describe the relationship between the number of cans of beer and BAC.
* (b) Write the equation of the regression line. Interpret the slope and intercept in context.
* (c) Do the data provide strong evidence that drinking more cans of beer is associated with an
* increase in blood alcohol? State the null and alternative hypotheses, report the p-value, and
* state your conclusion.
* (d) The correlation coefficient for number of cans of beer and BAC is 0.89. Calculate
* R
* 2
* and
* interpret it in context.
* (e) Suppose we visit a bar, ask people how many drinks they have had, and also take their BAC.
* Do you think the relationship between number of drinks and BAC would be as strong as the
* relationship found in the Ohio State study?
* 8.37 Husbands and wives, Part II.
* The scatterplot below summarizes husbands’ and wives’
* heights in a random sample of 170 married couples in Britain, where both partners’ ages are below
* 65 years. Summary output of the least squares fit for predicting wife’s height from husband’s
* height is also provided in the table.
* Husband's height (in inches)
* Wife's height (in inches)
* 60
* 65
* 70
* 75
* 55
* 60
* 65
* 70
* Estimate
* Std. Error
* t value
* Pr(
* >
* |
* t
* |
* )
* (Intercept)
* 43.5755
* 4.6842
* 9.30
* 0.0000
* height
* husband
* 0.2863
* 0.0686
* 4.17
* 0.0000
* (a) Is there strong evidence that taller men marry taller women? State the hypotheses and include
* any information used to conduct the test.
* (b) Write the equation of the regression line for predicting wife’s height from husband’s height.
* (c) Interpret the slope and intercept in the context of the application.
* (d) Given that
* R
* 2
* = 0
* .
* 09, what is the correlation of heights in this data set?
* (e) You meet a married man from Britain who is 5’9” (69 inches). What would you predict his
* wife’s height to be? How reliable is this prediction?
* (f) You meet another married man from Britain who is 6’7” (79 inches). Would it be wise to use
* the same linear model to predict his wife’s height? Why or why not?

**Multiple And Logistic Regression**

* The principles of simple linear regression lay the foundation for more sophisticated regression methods used in a wide range of challenging settings.
* **Multiple linear regression** introduces the possibility of more than 1 predictor + **logistic regression** is a technique for predicting *categorical* outcomes with 2 possible categories.
* **Multiple regression** extends simple 2-variable regression to the case that still has *one response* but many predictors (denoted x1, x2, x3…)
* Multiple regression also allows for categorical variables with *many levels*
* The method is motivated by scenarios where many variables may be simultaneously connected to an output.
* Consider eBay auctions of a video game called Mario Kart for the Nintendo Wii.
* Outcome variable of interest = the total price of an auction = the highest bid + the shipping cost.
* We will try to determine how total price is related to each characteristic in an auction while simultaneously controlling for other variables.
* For instance, all other characteristics held constant, *are longer auctions associated w/ higher or lower prices*?
* And, *on average, how much more do buyers tend to pay for additional Wii wheels (plastic steering wheels that attach to the Wii controller) in auctions?*
* Multiple regression will help us answer these and other questions.
* The data set mario\_kart includes results from 141 auctions.



* Notice that the condition and stock photo variables are indicator variables.
* **Price =** final auction price + shipping costs, in US dollars
* **cond\_new =** a coded 2-level categorical variable, which takes value 1 when the game is new and 0 if the game is used
* **stock\_photo =** a coded 2-level categorical variable, which takes value 1 if the primary photo used in the auction was a stock photo and 0 if the photo was unique to that auction
* **duration =** the length of the auction, in days, taking values from 1 to 10
* **wheels =** the number of Wii wheels included with the auction
* Using indicator variables in place of category names allows for these variables to be directly used in regression.
* Let’s fit a linear regression model with the game’s condition as a predictor of auction price.
* The model may be written as
* ̂
* price
* = 42
* .
* 87 + 10
* .
* 90
* ×
* cond
* new
* Results of this model are shown in Table 8.3 and a scatterplot for price versus game con-
* dition is shown in Figure 8.4.
* Estimate
* Std. Error
* t value
* Pr(
* >
* |
* t
* |
* )
* (Intercept)
* 42.8711
* 0.8140
* 52.67
* 0.0000
* cond
* new
* 10.8996
* 1.2583
* 8.66
* 0.0000
* df
* = 139
* Table 8.3: Summary of a linear model for predicting auction price based
* on game condition.
* ⊙
* Guided Practice 8.1
* Examine Figure 8.4. Does the linear model seem reason-
* able?
* 2
* 2
* Yes. Constant variability, nearly normal residuals, and linearity all appear reasonable.
* 374
* CHAPTER 8. MULTIPLE AND LOGISTIC REGRESSION
* Price
* 30
* 40
* 50
* 60
* 70
* 0
* (used)
* 1
* (new)
* Condition
* Figure 8.4: Scatterplot of the total auction price against the game’s condi-
* tion. The least squares line is also shown.
* Example 8.2
* Interpret the coefficient for the game’s condition in the model. Is this
* coefficient significantly different from 0?
* Note that
* cond
* new
* is a two-level categorical variable that takes value 1 when the
* game is new and value 0 when the game is used. So 10.90 means that the model
* predicts an extra
* $
* 10.90 for those games that are new versus those that are used.
* (See Section 7.2.7 for a review of the interpretation for two-level categorical predictor
* variables.) Examining the regression output in Table 8.3, we can see that the p-
* value for
* cond
* new
* is very close to zero, indicating there is strong evidence that the
* coefficient is different from zero when using this simple one-variable model.
* 8.1.2 Including and assessing many variables in a model
* Sometimes there are underlying structures or relationships between predictor variables.
* For instance, new games sold on Ebay tend to come with more Wii wheels, which may
* have led to higher prices for those auctions. We would like to fit a model that includes all
* potentially important variables simultaneously. This would help us evaluate the relationship
* between a predictor variable and the outcome while controlling for the potential influence
* of other variables. This is the strategy used in
* multiple regression
* . While we remain
* cautious about making any causal interpretations using multiple regression, such models
* are a common first step in providing evidence of a causal connection.
* 8.1. INTRODUCTION TO MULTIPLE REGRESSION
* 375
* We want to construct a model that accounts for not only the game condition, as in Sec-
* tion 8.1.1, but simultaneously accounts for three other variables:
* stock
* photo
* ,
* duration
* ,
* and
* wheels
* .
* ̂
* price
* =
* β
* 0
* +
* β
* 1
* ×
* cond
* new
* +
* β
* 2
* ×
* stock
* photo
* +
* β
* 3
* ×
* duration
* +
* β
* 4
* ×
* wheels
* ˆ
* y
* =
* β
* 0
* +
* β
* 1
* x
* 1
* +
* β
* 2
* x
* 2
* +
* β
* 3
* x
* 3
* +
* β
* 4
* x
* 4
* (8.3)
* In this equation,
* y
* represents the total price,
* x
* 1
* indicates whether the game is new,
* x
* 2
* indicates whether a stock photo was used,
* x
* 3
* is the duration of the auction, and
* x
* 4
* is the
* number of Wii wheels included with the game. Just as with the single predictor case, a
* multiple regression model may be missing important components or it might not precisely
* represent the relationship between the outcome and the available explanatory variables.
* While no model is perfect, we wish to explore the possibility that this one may fit the data
* reasonably well.
* We estimate the parameters
* β
* 0
* ,
* β
* 1
* , ...,
* β
* 4
* in the same way as we did in the case of a
* single predictor. We select
* b
* 0
* ,
* b
* 1
* , ...,
* b
* 4
* that minimize the sum of the squared residuals:
* SSE
* =
* e
* 2
* 1
* +
* e
* 2
* 2
* +
* ···
* +
* e
* 2
* 141
* =
* 141
* ∑
* i
* =1
* e
* 2
* i
* =
* 141
* ∑
* i
* =1
* (
* y
* i
* −
* ˆ
* y
* i
* )
* 2
* (8.4)
* Here there are 141 residuals, one for each observation. We typically use a computer to
* minimize the sum in Equation (8.4) and compute point estimates, as shown in the sample
* output in Table 8.5. Using this output, we identify the point estimates
* b
* i
* of each
* β
* i
* , just
* as we did in the one-predictor case.
* Estimate
* Std. Error
* t value
* Pr(
* >
* |
* t
* |
* )
* (Intercept)
* 36.2110
* 1.5140
* 23.92
* 0.0000
* cond
* new
* 5.1306
* 1.0511
* 4.88
* 0.0000
* stock
* photo
* 1.0803
* 1.0568
* 1.02
* 0.3085
* duration
* -0.0268
* 0.1904
* -0.14
* 0.8882
* wheels
* 7.2852
* 0.5547
* 13.13
* 0.0000
* df
* = 136
* Table 8.5: Output for the regression model where
* price
* is the outcome
* and
* cond
* new
* ,
* stock
* photo
* ,
* duration
* , and
* wheels
* are the predictors.
* Multiple regression model
* A multiple regression model is a linear model with many predictors. In general,
* we write the model as
* ˆ
* y
* =
* β
* 0
* +
* β
* 1
* x
* 1
* +
* β
* 2
* x
* 2
* +
* ···
* +
* β
* k
* x
* k
* when there are
* k
* predictors. We often estimate the
* β
* i
* parameters using a computer.
* 376
* CHAPTER 8. MULTIPLE AND LOGISTIC REGRESSION
* ⊙
* Guided Practice 8.5
* Write out the model in Equation (8.3) using the point
* estimates from Table 8.5. How many predictors are there in this model?
* 3
* ⊙
* Guided Practice 8.6
* What does
* β
* 4
* , the coefficient of variable
* x
* 4
* (Wii wheels),
* represent? What is the point estimate of
* β
* 4
* ?
* 4
* ⊙
* Guided Practice 8.7
* Compute the residual of the first observation in Table 8.1
* on page 373 using the equation identified in Guided Practice 8.5.
* 5
* Example 8.8
* We estimated a coefficient for
* cond
* new
* in Section 8.1.1 of
* b
* 1
* = 10
* .
* 90
* with a standard error of
* SE
* b
* 1
* = 1
* .
* 26 when using simple linear regression. Why might
* there be a difference between that estimate and the one in the multiple regression
* setting?
* If we examined the data carefully, we would see that some predictors are correlated.
* For instance, when we estimated the connection of the outcome
* price
* and predictor
* cond
* new
* using simple linear regression, we were unable to control for other variables
* like the number of Wii wheels included in the auction. That model was biased by the
* confounding variable
* wheels
* . When we use both variables, this particular underlying
* and unintentional bias is reduced or eliminated (though bias from other confounding
* variables may still remain).
* Example 8.8 describes a common issue in multiple regression: correlation among pre-
* dictor variables. We say the two predictor variables are
* collinear
* (pronounced as
* co-linear
* )
* when they are correlated, and this collinearity complicates model estimation. While it is
* impossible to prevent collinearity from arising in observational data, experiments are usu-
* ally designed to prevent predictors from being collinear.
* ⊙
* Guided Practice 8.9
* The estimated value of the intercept is 36.21, and one might
* be tempted to make some interpretation of this coefficient, such as, it is the model’s
* predicted price when each of the variables take value zero: the game is used, the
* primary image is not a stock photo, the auction duration is zero days, and there are
* no wheels included. Is there any value gained by making this interpretation?
* 6
* 8.1.3 Adjusted
* R
* 2
* as a better estimate of explained variance
* We first used
* R
* 2
* in Section 7.2 to determine the amount of variability in the response that
* was explained by the model:
* R
* 2
* = 1
* −
* variability in residuals
* variability in the outcome
* = 1
* −
* V ar
* (
* e
* i
* )
* V ar
* (
* y
* i
* )
* where
* e
* i
* represents the residuals of the model and
* y
* i
* the outcomes. This equation remains
* valid in the multiple regression framework, but a small enhancement can often be even
* more informative.
* 3
* ˆ
* y
* = 36
* .
* 21 + 5
* .
* 13
* x
* 1
* + 1
* .
* 08
* x
* 2
* −
* 0
* .
* 03
* x
* 3
* + 7
* .
* 29
* x
* 4
* , and there are
* k
* = 4 predictor variables.
* 4
* It is the average difference in auction price for each additional Wii wheel included when holding the
* other variables constant. The point estimate is
* b
* 4
* = 7
* .
* 29.
* 5
* e
* i
* =
* y
* i
* −
* ˆ
* y
* i
* = 51
* .
* 55
* −
* 49
* .
* 62 = 1
* .
* 93, where 49.62 was computed using the variables values from the
* observation and the equation identified in Guided Practice 8.5.
* 6
* Three of the variables (
* cond
* new
* ,
* stock
* photo
* , and
* wheels
* ) do take value 0, but the auction duration
* is always one or more days. If the auction is not up for any days, then no one can bid on it! That means
* the total auction price would always be zero for such an auction; the interpretation of the intercept in this
* setting is not insightful.
* 8.1. INTRODUCTION TO MULTIPLE REGRESSION
* 377
* ⊙
* Guided Practice 8.10
* The variance of the residuals for the model given in Guided
* Practice 8.7 is 23.34, and the variance of the total price in all the auctions is 83.06.
* Calculate
* R
* 2
* for this model.
* 7
* This strategy for estimating
* R
* 2
* is acceptable when there is just a single variable.
* However, it becomes less helpful when there are many variables. The regular
* R
* 2
* is a less
* estimate of the amount of variability explained by the model. To get a better estimate, we
* use the adjusted
* R
* 2
* .
* Adjusted R
* 2
* as a tool for model assessment
* The
* adjusted R
* 2
* is computed as
* R
* 2
* adj
* = 1
* −
* V ar
* (
* e
* i
* )
* /
* (
* n
* −
* k
* −
* 1)
* V ar
* (
* y
* i
* )
* /
* (
* n
* −
* 1)
* = 1
* −
* V ar
* (
* e
* i
* )
* V ar
* (
* y
* i
* )
* ×
* n
* −
* 1
* n
* −
* k
* −
* 1
* where
* n
* is the number of cases used to fit the model and
* k
* is the number of
* predictor variables in the model.
* Because
* k
* is never negative, the adjusted
* R
* 2
* will be smaller – often times just a
* little smaller – than the unadjusted
* R
* 2
* . The reasoning behind the adjusted
* R
* 2
* lies in the
* degrees of freedom
* associated with each variance.
* 8
* ⊙
* Guided Practice 8.11
* There were
* n
* = 141 auctions in the
* mario
* kart
* data set
* and
* k
* = 4 predictor variables in the model. Use
* n
* ,
* k
* , and the variances from Guided
* Practice 8.10 to calculate
* R
* 2
* adj
* for the Mario Kart model.
* 9
* ⊙
* Guided Practice 8.12
* Suppose you added another predictor to the model, but
* the variance of the errors
* V ar
* (
* e
* i
* ) didn’t go down. What would happen to the
* R
* 2
* ?
* What would happen to the adjusted
* R
* 2
* ?
* 10
* Adjusted
* R
* 2
* could have been used in Chapter 7. However, when there is only
* k
* = 1
* predictors, adjusted
* R
* 2
* is very close to regular
* R
* 2
* , so this nuance isn’t typically important
* when considering only one predictor.
* 7
* R
* 2
* = 1
* −
* 23
* .
* 34
* 83
* .
* 06
* = 0
* .
* 719.
* 8
* In multiple regression, the degrees of freedom associated with the variance of the estimate of the
* residuals is
* n
* −
* k
* −
* 1, not
* n
* −
* 1. For instance, if we were to make predictions for new data using our current
* model, we would find that the unadjusted
* R
* 2
* is an overly optimistic estimate of the reduction in variance
* in the response, and using the degrees of freedom in the adjusted
* R
* 2
* formula helps correct this bias.
* 9
* R
* 2
* adj
* = 1
* −
* 23
* .
* 34
* 83
* .
* 06
* ×
* 141
* −
* 1
* 141
* −
* 4
* −
* 1
* = 0
* .
* 711.
* 10
* The unadjusted
* R
* 2
* would stay the same and the adjusted
* R
* 2
* would go down.