University of Texas at Austin

STA 235 Mastery Exam A 2025-03-03

Personal Data

Family Name:

Given Name:

Signature:

checked

Registration Number

0 1 2 3 4 5 6 7 8 9

In this section **no** changes or modifications must be made!

Scrambling 0 0

Type: 031

Exam ID: 25030300001

Please mark the boxes carefully: ☒ Not marked: ☐ or ■

This document is scanned automatically. Please keep clean and do not bend or fold. For filling in the document please use a **blue or black pen**.

Only clearly marked and positionally accurate crosses will be processed!

Answers 1 - 15

	a	b	c	d	e
1	☐	☐			
2	☐	☐	☐	☐	☐
3	☐	☐	☐	☐	☐
4	☐	☐	☐	☐	
5	☐	☐	☐	☐	☐
6	☐	☐	☐	☐	☐
7	☐	☐	☐	☐	☐
8	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐
10	☐	☐	☐	☐	
11	☐	☐	☐	☐	☐
12	☐	☐	☐	☐	☐
13	☐	☐	☐	☐	☐
14	☐	☐	☐	☐	☐
15	☐	☐	☐	☐	☐
	a	b	c	d	e

Answers 16 - 30

	a	b	c	d	e
16	☐	☐	☐	☐	☐
17	☐	☐	☐	☐	
18	☐	☐	☐	☐	☐
19	☐	☐	☐	☐	☐
20	☐	☐	☐	☐	☐
21	☐	☐	☐	☐	☐
22	☐	☐	☐	☐	
23	☐	☐	☐	☐	☐
24	☐	☐	☐	☐	☐
25	☐	☐	☐	☐	☐
26	☐	☐	☐	☐	☐
27	☐	☐	☐	☐	
28	☐	☐	☐	☐	☐
29	☐	☐	☐	☐	☐
30	☐	☐	☐	☐	☐
	a	b	c	d	e

Answers 31 - 31

	a	b	c	d	e
31	☐	☐	☐	☐	☐
	a	b	c	d	e

For each question, select the best answer. (For numeric questions, this is the answer that is closest.)

1. *(0 points)* On this exam you may use one 8.5 $\times$ 11 sheet of notes (that you have personally prepared) and a graphing or scientific calculator (without a keyboard or internet access)—but no other resources. You can use the last page of the exam as scratch paper.

   Mark the box for this question to indicate that you understand this and will complete this exam on your own, without assistance from anyone else.

   a. I agree

   b. I do not agree (your exam will not be graded)

2. *(5 points)* A call center employs salespeople who make unsolicited phone calls to prospects to sell a product. For each salesperson, `calls` represents the number of calls they made, and `sales` represents their total sales. You have access to a simple random sample of 100 salespeople and perform the following regression (Model C1):

```
C1 <- lm(sales ~ calls, data = callcenter)
summary(C1)
```

```
|
| Call:
| lm(formula = sales ~ calls, data = callcenter)
|
| Residuals:
|     Min      1Q  Median      3Q     Max
| -643.26 -159.50  -22.49  173.70  561.27
|
| Coefficients:
|             Estimate Std. Error t value            Pr(>|t|)
| (Intercept)   167.57     136.60   1.227               0.223
| calls         135.03      13.18  10.243 <0.0000000000000002 ***
| ---
| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
|
| Residual standard error: 261.7 on 98 degrees of freedom
| Multiple R-squared:  0.517,   Adjusted R-squared:  0.5121
| F-statistic: 104.9 on 1 and 98 DF,  p-value: < 0.00000000000000022
```

```
confint(C1)
```

```
|                 2.5 %   97.5 %
| (Intercept) -103.5125 438.6489
| calls        108.8698 161.1921
```

```
predict(C1, list(calls = 10), interval = "confidence")
```

```
|        fit      lwr      upr
| 1 1517.878 1465.752 1570.003
```

```
predict(C1, list(calls = 10), interval = "prediction")
```

```
|        fit      lwr      upr
| 1 1517.878 995.9114 2039.844
```

What do you predict the total sales will be for a salesperson that makes 18 calls?

a. $168
b. $303
c. $2598
d. $2431
e. $135

3. *(5 points)* A salesperson comes to you at the end of the day; they are thinking of staying late and making 5 more calls. According to Model C1, how much more can they expect to make in sales?

   a. $675
   b. $303
   c. $843
   d. $135
   e. $168

4. *(5 points)* The company is interested in generalizing the results of Model C1 to the thousands of other salespeople in their workforce. What can you conclude?

   a. In this sample, there is no relationship between sales and the number of calls made.
   b. In this sample, there is a relationship between sales and the number of calls made, but we can't be confident that it will generalize to the rest of the workforce.
   c. We can be confident that in the entire workforce there is a relationship between sales and the number of calls made.
   d. This model doesn't provide any information about whether there is a relationship betewen sales and the number of calls made.

5. *(5 points)* A new salesperson has joined the company who is planning to make 10 calls per day and is asking your advice. Based on Model C1, you can advise them they can be 95% confident that they'll generate sales of *at least*

   a. $1465.75
   b. $995.91
   c. $1517.88
   d. $1570.00
   e. $2039.84

6. *(5 points)* The CFO is planning to use Model C1 to make sales predictions for each employee in their entire sales force. She would like to know how accurate her predictions will be in general—how can you advise her?

   a. About 95% of her predictions will be accurate to within $\pm$ $52.13.
   b. About 95% of her predictions will be accurate to within $\pm$ $261.71.
   c. About 95% of her predictions will be accurate to within $\pm$ $523.42.
   d. About 95% of her predictions will be accurate to within $\pm$ $785.13.
   e. You can't provide any evidence about the quality of this model.

7. *(5 points)* A startup company began operations in November 2024. They have recently implemented the use of website ads and decide to invest some money in marketing. They conducted a regression (Model R) based on revenue data they had gathered over the first few weeks of their operations, to understand how their `revenue` can be predicted from the number of `ads` they run and the amount they spend on `marketing`:

```
R <- lm(revenue ~ ads + marketing, data = webmktg)
summary(R)
```

```
|
| Call:
| lm(formula = revenue ~ ads + marketing, data = webmktg)
|
| Residuals:
|    Min     1Q Median     3Q    Max
| -34514  -4893  -1459   7971  20843
|
| Coefficients:
|              Estimate Std. Error t value Pr(>|t|)
| (Intercept) 19085.517  24523.172   0.778  0.44711
| ads          8817.764   2242.429   3.932  0.00107 **
| marketing       6.241      1.671   3.734  0.00165 **
| ---
| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
|
| Residual standard error: 15500 on 17 degrees of freedom
| Multiple R-squared:  0.5527,  Adjusted R-squared:  0.5001
| F-statistic:  10.5 on 2 and 17 DF,  p-value: 0.001071
```

```
confint(R)
```

```
|                     2.5 %       97.5 %
| (Intercept) -32653.852895 70824.886195
| ads           4086.653247 13548.875646
| marketing        2.714837     9.766186
```

How much revenue do you expect them to generate when they display 6 ads and $90360 in marketing?

a. $71992
b. $591840
c. $635929
d. $27910
e. $19086

8. *(5 points)* The intercept 19085.52 in Model R represents

   a. The predicted revenue for a startup that spends no money on marketing and runs no ads.
   b. The predicted revenue for a startup that spends the average amount on marketing and runs the average number of ads.
   c. The predicted revenue for a startup that runs no ads, regardless of how much they spend on marketing.
   d. The predicted amount spent on marketing for a startup that has no revenue and runs no ads.
   e. The predicted revenue for a startup that spends no money on marketing, regardless of how many ads they run.

9. *(5 points)* Which of these statements about Model R is correct?

   a. About 50.01% of the variation in revenue is explained by the number of ads run and the amount spent on marketing.
   b. About 55.27% of the variation in revenue is explained by the number of ads run and the amount spent on marketing.

c. The probability that we will make an accurate prediction of revenue is 0.55.

d. If we add a third predictor variable to this model, we expect that R^2 will decrease.

e. None of these are true.

10. *(5 points)* A venture capital firm is hoping to generalize the results to advise other startups in their portfolio. Assuming all regression assumptions for Model R are satisfied, which of the following would be correct inferences about the broader population of all startups, based on Model R?

    a. We can be confident that there is no relationship between revenue and number of ads run.

    b. We can be confident that there is a relationship between revenue and number of ads run.

    c. Considering two startups with the same level of marketing spend, we can be confident that there is a relationship between revenue and number of ads run.

    d. Regression cannot be used to generalize from a sample to the broader population it was drawn from.

11. *(5 points)* The startup is considering running 7 more ads. Which statement about Model R accurately describes the predicted effect on their evenue?

    a. If they run an additional 7 ads next week, they should expect to generate an additional $8817.76 in revenue.

    b. If they run an additional 7 ads next week and keep their marketing budget fixed, they should expect to generate an additional $8817.76 in revenue.

    c. If they run an additional 7 ads next week and keep their marketing budget fixed, they should expect to generate an additional $80809.87 in revenue.

    d. If they run an additional 7 ads next week, they should expect to generate an additional $61724.35 in revenue.

    e. If they run an additional 7 ads next week and keep their marketing budget fixed, they should expect to generate an additional $61724.35 in revenue.

12. *(5 points)* A booking agent for touring bands has collected data on the most recent tour of a simple random sample of bands. For each band, the variable `attendance` indicates the number of attendees at various types of venues: either `concerts` (concert halls), `arenas`, or `festivals`. The variable `ticket` indicates the ticket price on that tour, measured in dollars.

    The booker builds Model T1 predicting attendance from ticket price and the type of venue:

```
T1 <- lm(attendance ~ ticket + venue, data = music)
summary(T1)
```

```
|
| Call:
| lm(formula = attendance ~ ticket + venue, data = music)
|
| Residuals:
|      Min       1Q   Median       3Q      Max
| -17143.1  -3855.9    199.6   3877.4  14116.1
|
| Coefficients:
|               Estimate Std. Error t value             Pr(>|t|)
| (Intercept)   47894.98    2187.97  21.890 < 0.0000000000000002 ***
| ticket          -32.80       7.72  -4.249     0.00003332423238 ***
| venueConcert  -6063.79    1183.24  -5.125     0.00000071723006 ***
| venueFestival  9499.24    1286.94   7.381     0.00000000000449 ***
```

```
| ---
| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
|
| Residual standard error: 6207 on 194 degrees of freedom
| Multiple R-squared:  0.5709,  Adjusted R-squared:  0.5642
| F-statistic: 86.02 on 3 and 194 DF,  p-value: < 0.00000000000000022
```

```
confint(T1)
```

```
|                     2.5 %      97.5 %
| (Intercept)   43579.70636 52210.24835
| ticket          -48.02572   -17.57519
| venueConcert  -8397.45698 -3730.12836
| venueFestival  6961.05376 12037.43415
```

```
predict(T1, list(ticket = 100, venue = "Concert"), interval = "confidence")
```

```
|        fit      lwr      upr
| 1 38551.14 35890.29 41211.99
```

```
predict(T1, list(ticket = 100, venue = "Concert"), interval = "prediction")
```

```
|        fit      lwr      upr
| 1 38551.14 26022.69 51079.59
```

Using Model T1, what is the predicted attendance if the show is held in a concert hall and the ticket price is $178?

a. 51556
b. 35993
c. 42056
d. 36218
e. 45492

13. *(5 points)* The `ticket` coefficient of 32.8 in Model T1 indicates that

    a. Two bands that have ticket prices that differ by $1 and that both played in arenas are predicted to have attendance that differs by 32.8.
    b. Two bands that have ticket prices that differ by $1 and that both played in concert halls are predicted to have attendance that differs by 32.8.
    c. Two bands that have ticket prices that differ by $1 and that both played in the same kind of venue are predicted to have attendance that differs by 32.8.
    d. Two bands that have ticket prices that differ by $1 and that both played in different kinds of venues are predicted to have attendance that differs by 32.8.
    e. Two bands that have ticket prices that differ by $1 are predicted to have attendance that differs by 32.8.

14. *(5 points)* According to Model T1, among bands with the same ticket prices,

    a. bands that played at festivals are predicted to have the highest attendance, and bands that played in concert halls the lowest.
    b. bands that played at festivals are predicted to have the highest attendance, and bands that played in arenas the lowest.
    c. bands that played at concert halls are predicted to have the highest attendance, and bands that played in festivals the lowest.

d. bands that played at arenas are predicted to have the highest attendance, and bands that played in festivals the lowest.

e. bands that played at arenas are predicted to have the highest attendance, and bands that played in concert halls the lowest.

15. *(5 points)* The band *Mary and the Models* previously played at concert halls but is considering festivals for their next tour. According to Model T1, by how much do you predict their attendance will change, assuming they don't also change their ticket price?

    a. 3435

    b. -6064

    c. 47895

    d. 15563

    e. 9499

16. *(5 points)* The band *Rey and the RStudios*, which played (and will continue to play) in concert halls, is considering increasing their ticket price by $100. They would like to understand the uncertainty in your prediction of the effect on attendance of such a change. Based on Model T1, we be 95% confident that making a change like that to their ticket price would lower the attendance of *Rey and the RStudios* by no more than

    a. 41212

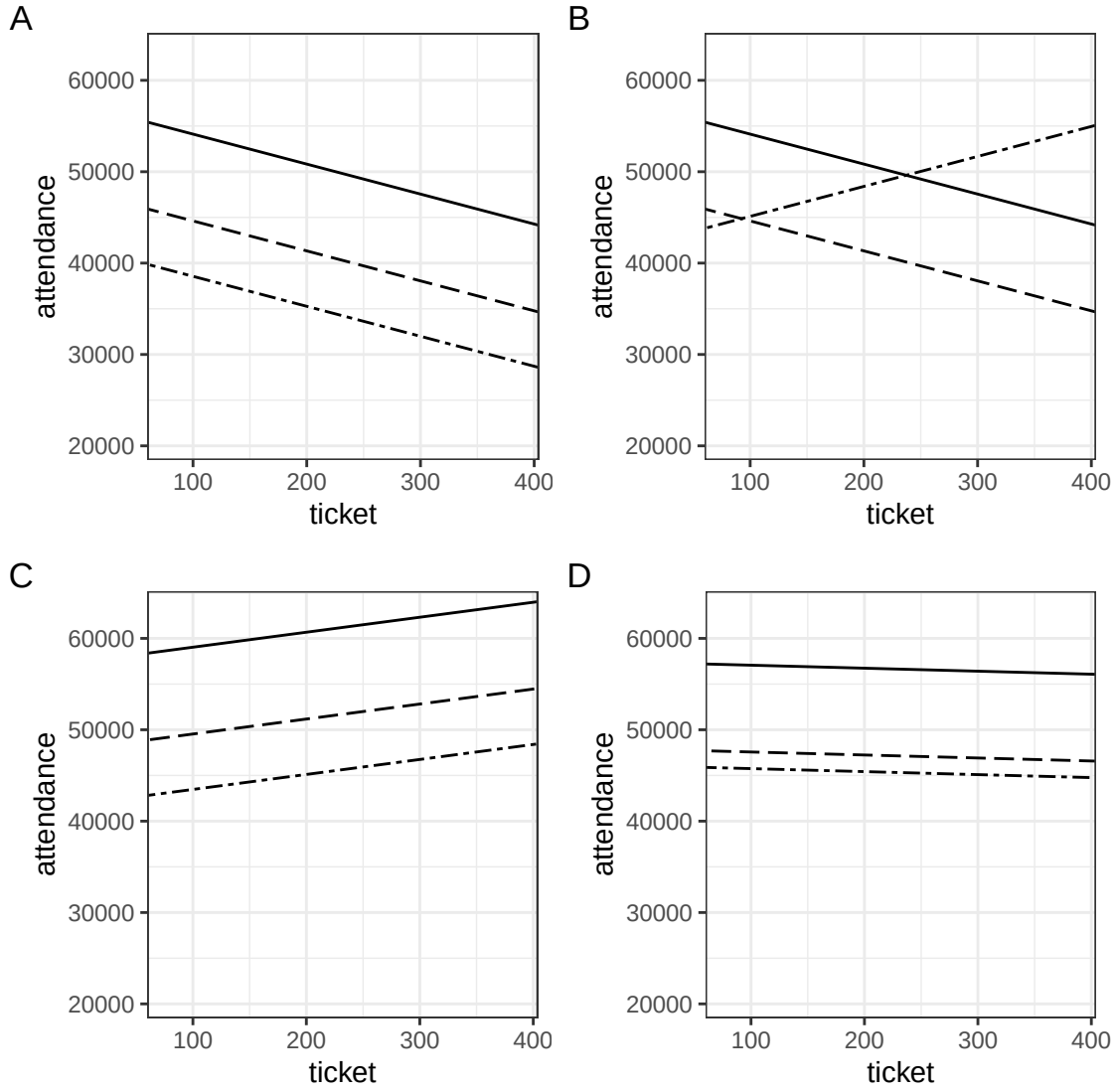
    b. 4803

    c. 51080

    d. 1758

    e. 38551

17. *(5 points)* Which graph corresponds to Model T1?



a. Graph A
b. Graph B
c. Graph C
d. Graph D

18. *(5 points)* The booker has also collected data on the most recent tour of each band they represent. For each band, `shows` indicates the number of shows they performed at. `sales` indicates the total sales from that tour, comprised of both ticket sales and merchandise sales (e.g., t-shirts, stickers, etc). They represent both rock and country bands; `rock` is 1 for rock bands and 0 for country bands.

The booker builds a regression predicting sales from the number of shows, the type of band, an an interaction between the two (Model T2):

```
T2 <- lm(sales ~ shows * rock, data = music)
summary(T2)
```

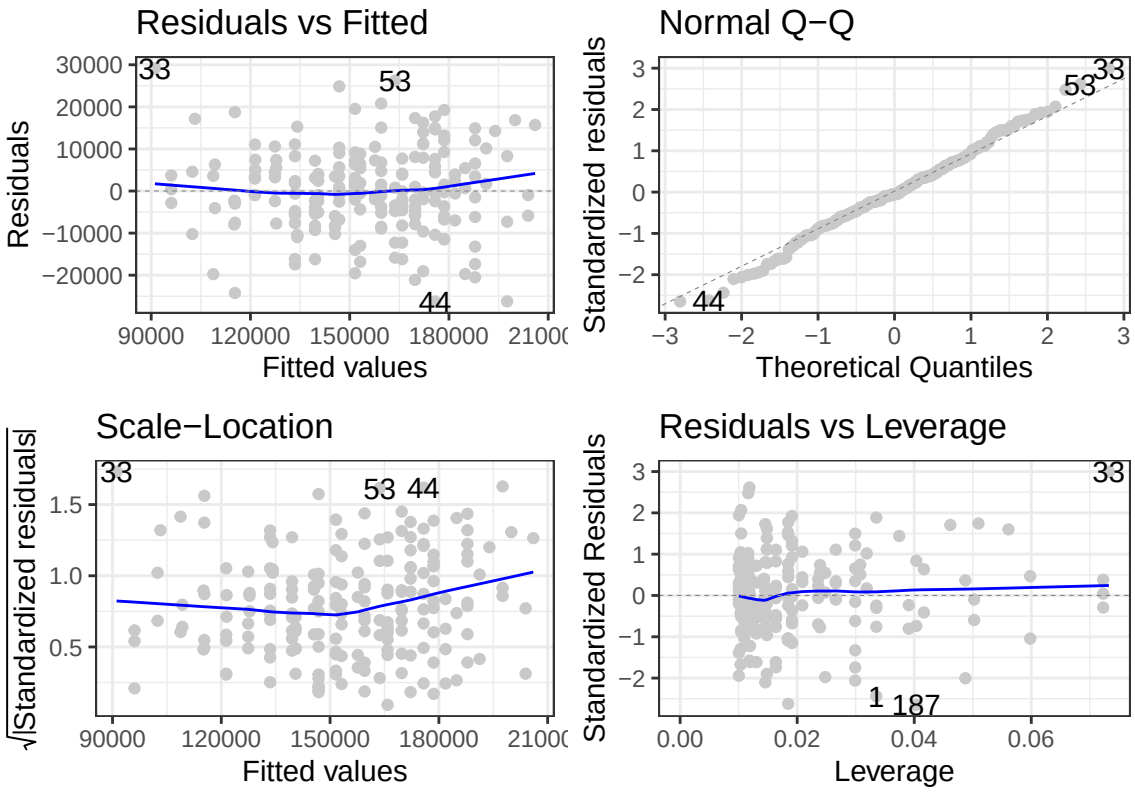
```
|
| Call:
| lm(formula = sales ~ shows * rock, data = music)
```

```
|
| Residuals:
|      Min       1Q   Median       3Q      Max
| -26258.3  -5878.1   -544.5   6337.6  29151.2
|
| Coefficients:
|             Estimate Std. Error t value            Pr(>|t|)
| (Intercept)    910.7     6605.5   0.138               0.890
| shows         6344.5      267.2  23.749 <0.0000000000000002 ***
| rock          -515.6     9159.2  -0.056               0.955
| shows:rock    -296.2      364.6  -0.812               0.418
| ---
| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
|
| Residual standard error: 10110 on 194 degrees of freedom
| Multiple R-squared:  0.857,   Adjusted R-squared:  0.8547
| F-statistic: 387.4 on 3 and 194 DF,  p-value: < 0.00000000000000022
```

What do you predict the total sales will be for a rock band that played 27 shows on their last tour?

a. $171400
b. $171696
c. $172211
d. $163700
e. $162789

19. *(5 points)* Below are the diagnostic plots from Model T2:



Based on these plots alone, which assumption(s) is/are satisfied by Model T2?

a. No assumptions are satisfied

b. Linearity only
c. Linearity and Equal variance
d. Linearity, Normality, and Equal variance
e. All assumptions are satisfied

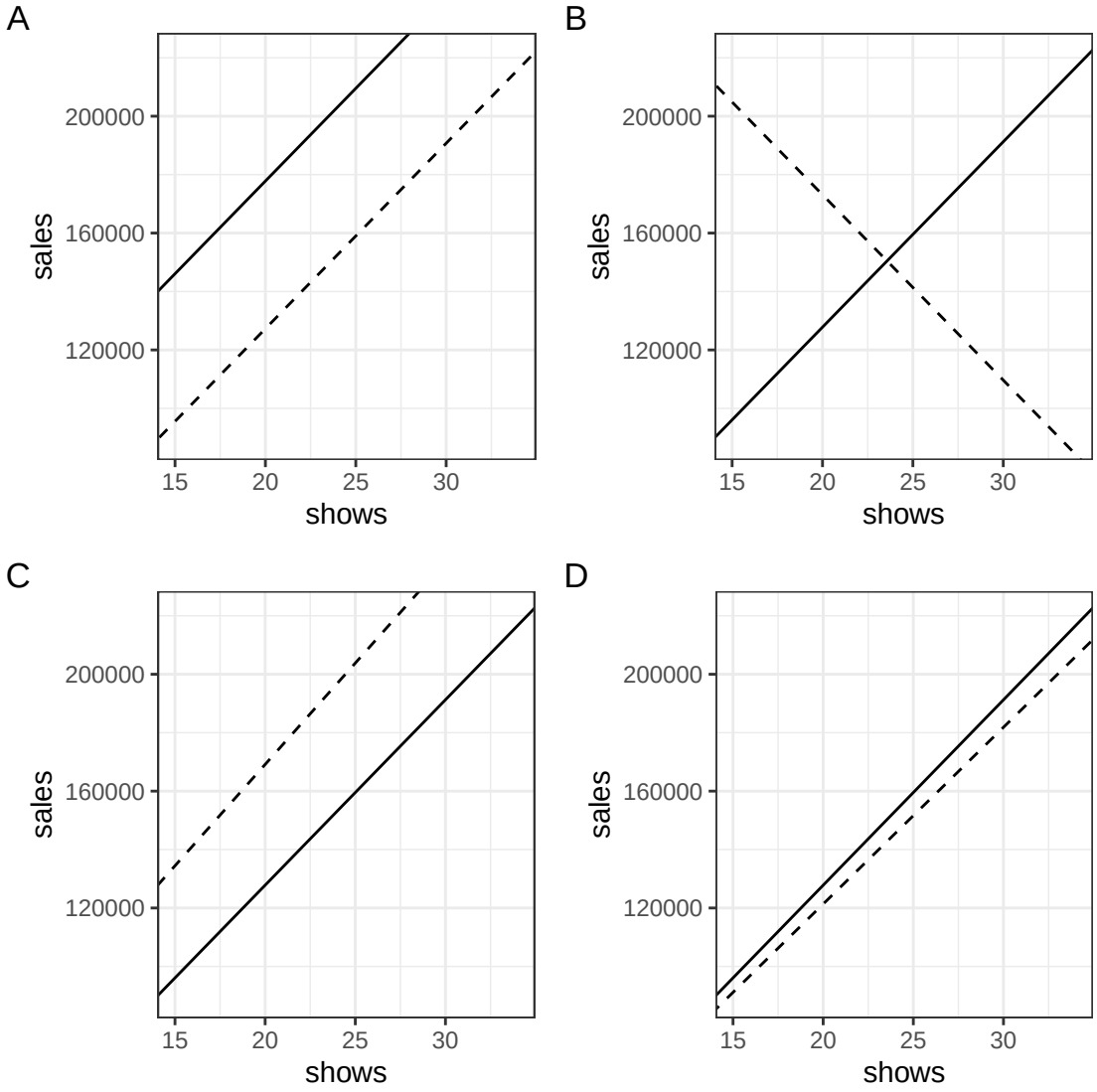
20. *(5 points)* Removing which observation would we expect to change the coefficients of Model T2 the most?

a. 33
b. 53
c. 1
d. 44
e. 187

21. *(5 points)* The booking agent would like to use Model T2 to create a 95% prediction interval for predicting a particular band's total sales. Should they believe that the interval they get from the `predict()` function is correct? Assume that independence is satisfied.

a. No, because the model has an interaction term.
b. Yes, because all regression assumptions are satisfied.
c. No, because some regression assumptions are violated.
d. Yes, because a 95% prediction interval can be created for any model in R.
e. There is not enough information to determine whether `predict()` will give an accurate interval.

22. *(5 points)* Which graph corresponds to Model T2?



a. Graph A
b. Graph B
c. Graph C
d. Graph D

23. *(5 points)* Based on Model T2, which of the following can the booker conclude about the earning power of rock versus country bands?

a. Rock bands are always predicted to earn the same amount as country bands.
b. Rock bands are always predicted to earn less than country bands.
c. Rock bands are always predicted to earn more than country bands.
d. Rock bands who play a small number of shows are predicted to earn less than country bands, but rock bands who play a large number of shows are predicted to earn more than country bands.
e. Rock bands who play a small number of shows are predicted to earn more than country bands, but rock bands who play a large number of shows are predicted to earn less than country bands.

24. *(5 points)* The rock band *Indigo and the Interactions* is considering extending their tour and play one additional show. Based on Model T2, how much more are they expected to earn?

    a. $6048
    b. $-296
    c. $911
    d. $6344
    e. $-516

25. *(5 points)* The booker is also trying to determine how to advise bands on building their brands on Instagram, so they collected data about how many Instagram posts each band made in the last year (`posts`) and how many followers they had at the end of the year (`followers`). They built Model T3:

```
T3 <- lm(followers ~ shows * posts, data = music)
summary(T3)
```

```
|
| Call:
| lm(formula = followers ~ shows * posts, data = music)
|
| Residuals:
|    Min     1Q Median     3Q    Max
| -34565 -10595   -733   8947  40292
|
| Coefficients:
|             Estimate Std. Error t value Pr(>|t|)
| (Intercept) 42456.47   48244.97   0.880  0.37994
| shows        5959.56    1923.00   3.099  0.00223 **
| posts        2651.39    1542.01   1.719  0.08713 .
| shows:posts   213.15      61.49   3.466  0.00065 ***
| ---
| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
|
| Residual standard error: 15230 on 194 degrees of freedom
| Multiple R-squared:  0.9449,  Adjusted R-squared:  0.944
| F-statistic:  1108 on 3 and 194 DF,  p-value: < 0.00000000000000022
```

    What is the meaning of the coefficient for `posts` in Model T3?

    a. Among all bands, a band with 1 additional Instagram post is expected to have about 2651 more followers.
    b. If we only compare bands to other bands that played the same number of shows, a band with 1 additional Instagram post is expected to have about 2651 more followers.
    c. Among all country bands that played no shows, a band with 1 additional Instagram post is expected to have about 2651 more followers.
    d. Among all bands that played no shows, a band with 1 additional Instagram post is expected to have about 2651 more followers.
    e. Among all country bands, a band with 1 additional Instagram post is expected to have about 2651 more followers.

26. *(5 points)* The band *Rick and the Regressions* is considering extending their tour to play one more show. Which of the following would Model T3 predict about their Instagram following?

    a. Their new follower count will be about 48416.

b. The number of followers they should expect to gain from each additional post will increase by 213.

c. They will gain about 213 followers.

d. The number of followers they should expect to gain from each additional post will increase by 5960.

e. They will gain about 5960 followers.

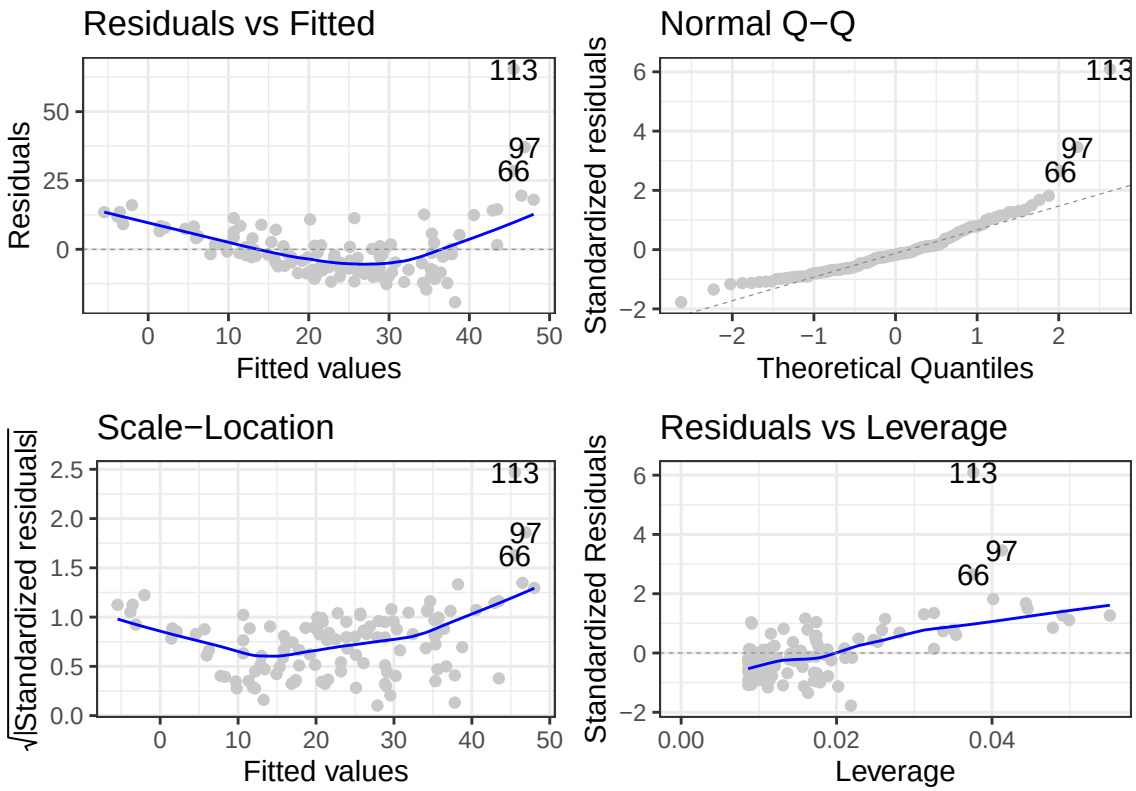
27. *(5 points)* A grocery store chain plans to open at other locations within the city. One aspect that their marketing team is analyzing is the purchasing decisions made by shoppers at the dairy section, particularly the milk. They are looking at two models that explore the relationship between the number of bottles that are likely to be bought at a certain retail price.

Below are the two models (B1 and B2) that they are considering:

```
B1 <- lm(bottles ~ retail, data = milk)
summary(B1)
```

```
|
| Call:
| lm(formula = bottles ~ retail, data = milk)
|
| Residuals:
|     Min      1Q  Median      3Q     Max
| -19.224  -7.253  -1.871   4.444  65.471
|
| Coefficients:
|             Estimate Std. Error t value            Pr(>|t|)
| (Intercept)   65.790      3.688   17.84 <0.0000000000000002 ***
| retail       -13.783      1.143  -12.06 <0.0000000000000002 ***
| ---
| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
|
| Residual standard error: 10.95 on 114 degrees of freedom
| Multiple R-squared:  0.5606,  Adjusted R-squared:  0.5567
| F-statistic: 145.4 on 1 and 114 DF,  p-value: < 0.00000000000000022
```

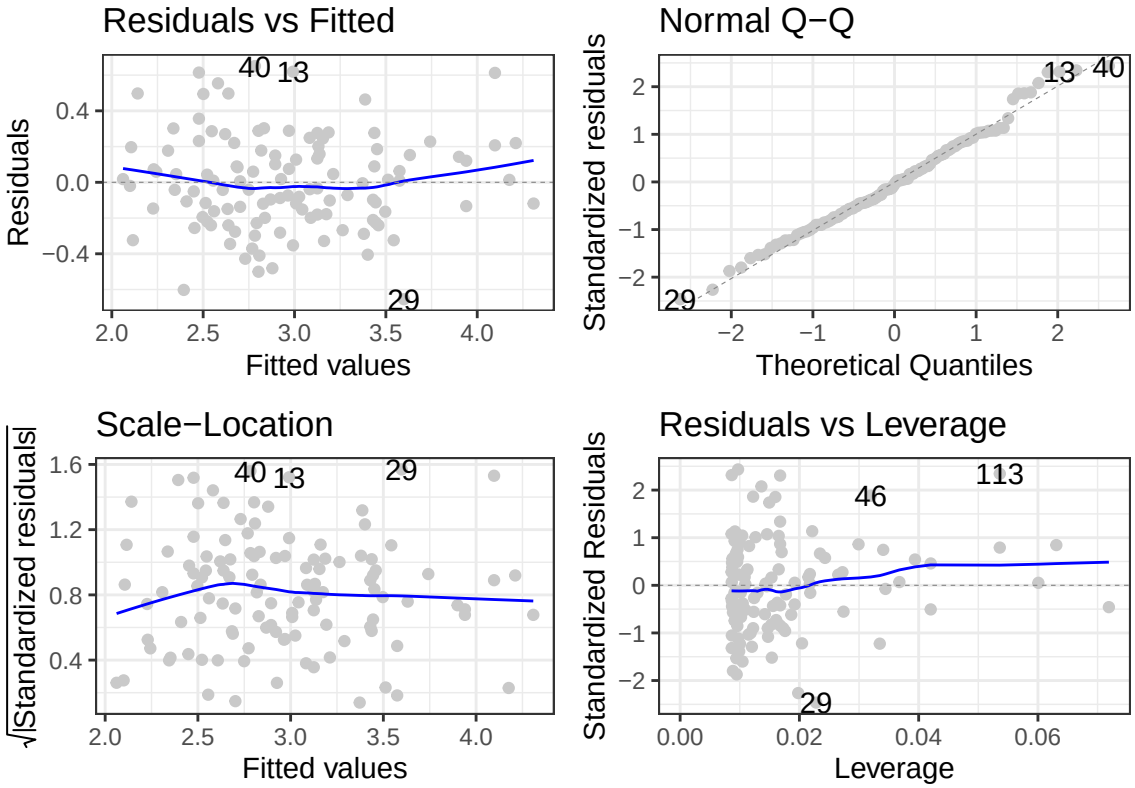
Model B1 has the following diagnostic plots:




```
B2 <- lm(log(bottles) ~ log(retail), data = milk)
summary(B2)
```

```
|
| Call:
| lm(formula = log(bottles) ~ log(retail), data = milk)
|
| Residuals:
|      Min       1Q   Median       3Q      Max
| -0.65425 -0.18405 -0.01262  0.17986  0.65074
|
| Coefficients:
|             Estimate Std. Error t value            Pr(>|t|)
| (Intercept)  4.72060    0.09172   51.47 <0.0000000000000002 ***
| log(retail) -1.61858    0.08116  -19.94 <0.0000000000000002 ***
| ---
| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
|
| Residual standard error: 0.2687 on 114 degrees of freedom
| Multiple R-squared:  0.7772,  Adjusted R-squared:  0.7753
| F-statistic: 397.7 on 1 and 114 DF,  p-value: < 0.00000000000000022
```

Model B2 has the following diagnostic plots:



Which models would be appropriate to use for prediction or inference?

a. Only Model B2 can be used for prediction or inference.
b. Only Model B1 can be used for prediction or inference.
c. Either model can be used for either prediction or inference.
d. Neither model can be used for either prediction or inference.

28. *(5 points)* Model B2 indicates that

a. a $1 increase in the retail price is associated with an expected decrease of 1.62% of the number of bottles purchased.
b. a $1 increase in the retail price is associated with an expected decrease of 1.62 bottles purchased.
c. a 1% increase in the retail price is associated with an expected decrease of 1.62% of the number of bottles purchased.
d. a 1% increase in the retail price is associated with an expected decrease of 1.62 bottles purchased.
e. a 1% increase in the retail price is associated with an expected decrease of 161.86% of the number of bottles purchased.

29. *(5 points)* Model B2 predicts that when the retail price is 5 dollars, the number of bottles of milk purchased will be

a. 6
b. 24
c. 8
d. 13
e. 5

30. *(5 points)* Consider Model B3, which also predicts the number of bottles sold from the retail price:

```
B3 <- lm(bottles ~ retail + I(retail^2), data = milk)
summary(B3)
```

```
|
| Call:
| lm(formula = bottles ~ retail + I(retail^2), data = milk)
|
| Residuals:
|     Min      1Q  Median      3Q     Max
| -23.461  -4.126  -0.192   3.282  50.113
|
| Coefficients:
|             Estimate Std. Error t value             Pr(>|t|)
| (Intercept) 133.4321     7.8950  16.901 < 0.0000000000000002 ***
| retail      -60.0686     5.1240 -11.723 < 0.0000000000000002 ***
| I(retail^2)   7.2914     0.7955   9.166  0.00000000000000259 ***
| ---
| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
|
| Residual standard error: 8.333 on 113 degrees of freedom
| Multiple R-squared:  0.748,   Adjusted R-squared:  0.7435
| F-statistic: 167.7 on 2 and 113 DF,  p-value: < 0.00000000000000022
```

Model B3 predicts that an increase in the retail price by $1 will

a. increase the number of bottles sold by 7.29.
b. change the number of bottles sold, but by an amount that depends on the original retail price.
c. decrease the number of bottles sold by 60.07.
d. not change the number of bottles sold.
e. increase the number of bottles sold by 133.43.

31. *(5 points)* Model B3 predicts that when the retail price is 2.5 dollars, the number of bottles of milk purchased will be

a. 46
b. 1
c. -17
d. -105
e. 29