1. In 'do students who study more score higher?', which is the response variable?

   ☐ The score

   ☐ The student

   ☐ The hours studied

   ☐ The subject

2. A table that cross-classifies two categorical variables is a ____ table (two words, first word).

   _____

3. Two categorical variables appear independent when their conditional distributions are...

   ☐ based on raw counts

   ☐ different across groups

   ☐ always 50/50

   ☐ the same across groups

4. A complete scatterplot description mentions which features?

   ☐ Form

   ☐ Strength

   ☐ Direction

   ☐ Unusual features / outliers

   Tick every correct option.

5. What is the largest possible value of a correlation coefficient r?

   _____

6. For y-hat = 55 + 4x, predict the score when x = 3 hours.

   _____

7. A house predicted at 300 sold for 330 (thousands). What is the residual?

   _____

8. With r = 0.8, s_x = 2, s_y = 10, find the slope b = r · (s_y/s_x).

   _____

9. A high-leverage point is one with an extreme value of...

   ☐ the correlation

   ☐ y only

   ☐ the residual

10. An observed association between two variables proves that one causes the other.

    ☐ True ☐ False

exercises.lol exercises.com.cn

1. **The score**

   The score is the outcome we want to explain – the response. Hours studied is explanatory.

2. **two-way**

   A two-way (contingency) table.

3. **the same across groups**

   Same conditional distributions → no association → independence.

4. **Form; Strength; Direction; Unusual features / outliers**

   All four – direction, form, strength, and any outliers, in context.

5. **1**

   r ranges from −1 to 1, so the maximum is 1.

6. **67**

   $55 + 4(3) = 55 + 12 = 67$.

7. **30**

   Residual = observed − predicted = $330 - 300 = 30$.

8. **4**

   $b = 0.8 \times (10/2) = 0.8 \times 5 = 4$.

9. **x (horizontal)**

   Leverage is about an extreme x-value; an outlier is a large residual.

10. **False**

    Association is not causation – other explanations remain possible.

exercises.lol exercises.com.cn