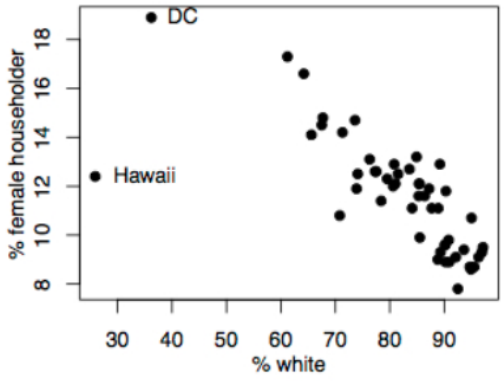
1. The scatterplot on the right shows the relationship between percentage of white residents and percentage of households with a female head in all 50 US States and the District of Columbia (DC). Which of the below **best** describes the two points marked as DC and Hawaii?

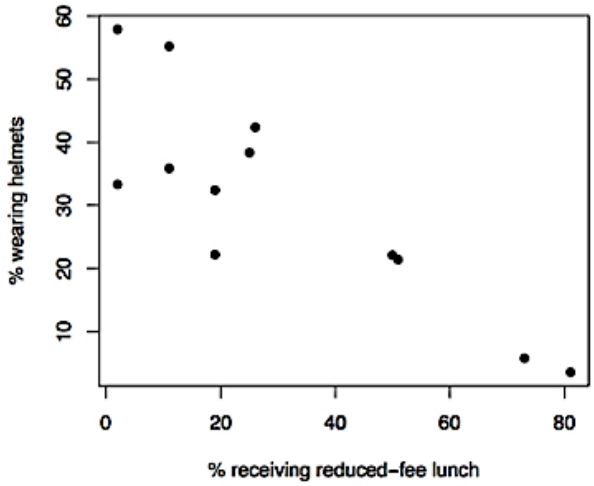


- ○ Hawaii is not an outlier, and DC is not a leverage point.
- ○ DC and Hawaii should both be excluded from a simple linear regression analysis.
- ○ Neither DC nor Hawaii appear to be leverage points.
- ◉ Hawaii has higher leverage and is more influential than DC.

2. The scatterplot below shows the relationship between socioeconomic status measured as the percentage of children in a neighborhood receiving reduced-fee lunches at school (lunch) and bike helmet use measured as the percentage of bike riders in the neighborhood wearing helmets (helmet). The equation of the regression line is

$$helmet = 47.49 - 0.54 \quad lunch$$

and the R^2 is 72%. Which of the following is **true**?



- ○ The correlation coefficient is 0.85.
- ○ 72% of the percentage of children receiving reduced-fee lunches at school can be accurately predicted by the model.
- ○ Decreasing the percentage of children receiving reduced-fee lunches at school by 5% will increase the percentage of bike riders wearing helmets in that neighborhood by 2.7%.
- ◉ Neighborhoods where no students receive reduced-fee lunches are expected on average to have 47.49% of bike riders wearing helmets.

3. The model below is for predicting the heart weight (in g) of cats from their gender (female and male). The coefficients are estimated using a dataset of 144 domestic cats. Which of the following is **false**?

	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	9.20	0.33	28.31	0.00
sex:male	2.12	0.40	5.35	0.00

- Female cats on average are expected to have hearts that weigh 2.12 grams less than those of male cats.
- If the regression equation is written $\hat{y} = b_0 + b_1 x$, then plugging in $x = 0$ would give you the predicted heart weight for a female cat.
- The expected heart weight for male cats is, on average, 11.32 grams.
- The intercept is meaningless.

✓ **Correct**

For a categorical explanatory variable like we have here (gender), a value of 0 for the explanatory variable corresponds to the baseline level.

4. Determine if I or II is higher, or if they are equal:

The uncertainty associated with the slope estimate when

I. there is a lot of scatter around the regression line

II. there is very little scatter around the regression line

- ○ I is higher
- ◉ II is higher
- ○ I and II are equal

Incorrect

This question refers to the following learning objective(s):

- Calculate a confidence interval for the slope as

$$b_1 \pm t^\star_{df} SE_{b_1},$$

where $df = n - 2$ and $t^\star_{df}$ is the critical score associated with the given confidence level at the desired degrees of freedom.

- Note that the standard error of the slope estimate SE_{b_1} can be found on the regression output.

Question 5

Which of the following is **not** supported by information provided in the model outputs above?

- The F-test for the significance of the model overall suggests that at least one of the slope coefficients is significantly different than 0.
- The sample size is 252.
- Wrist circumference is the most significant predictor of body fat percentage since the slope associated with this variable has the highest magnitude.
- All else held constant, people with wider hips tend to have lower body fat percentages.
- All else held constant, for each additional cm the forearm circumference is higher, body fat percentage is expected to be higher by 0.483 percentage points.

✓ **Correct**

This question refers to the following learning objective:

Determine whether an explanatory variable is a significant predictor for the response variable using the t-test and the associated p-value in the regression output.

6. Do these data provide convincing evidence that age and body fat percentage are significantly **positively** associated? Why or why not? Use quantitative information based on the model output to support your answer, and make sure to note the p-value you use to make this decision.

- (●) Yes, the p-value for testing for a positive correlation between age and body fat percentage is 0.039 / 2 = 0.0195. Since the p-value is small we reject the null hypothesis of no relationship.
- ( ) Yes, the p-value for testing for a positive correlation between age and body fat percentage is 0.000. Since the p-value is small we reject the null hypothesis of no relationship.
- ( ) Yes, the p-value for testing for a positive correlation between age and body fat percentage is 2e^{-16}. Since the p-value is small we reject the null hypothesis of no relationship.
- ( ) Yes, the p-value for testing for a positive correlation between age and body fat percentage is 0.039. Since the p-value is small we reject the null hypothesis of no relationship.

✓ **Correct**

This question refers to the following learning objective:

Determine whether an explanatory variable is a significant predictor for the response variable using the t-test and the associated p-value in the regression output.

7. Construct a 95% confidence interval for the slope of abdomen circumference and interpret it in context of the data.

- (●) (0.745, 1.009); All else held constant, for each additional cm in abdomen circumference, body fat percentage is expected to be higher by 0.745 to 1.009 percentage points.
- ( ) (-0.00539, 1.75); All else held constant, for each additional cm in abdomen circumference, body fat percentage is expected to change by -0.00539 to 1.75 percentage points.
- ( ) (0.00539, 0.88239); All else held constant, for each additional cm in abdomen circumference, body fat percentage is expected to be higher by 0.00539 to 0.88239 percentage points.
- ( ) (0.745, 1.009); All else held constant, for each additional percentage point increase in body fat, abdomen circumference is expected to be higher by 0.745 to 1.009 cm.

✓ **Correct**

We recall that this confidence interval is supposed to capture $\beta_{abdomen}$, ie the impact of increasing abdomen circumference by 1 cm on the response of body fat percentage.

This question refers to the following learning objective:

Calculate a confidence interval for the slope as $b_1 \pm t^*_{df} SE_{b_1}$ where $df = n - 2$ and t^*_{df} is the critical score associated with the given confidence level at the desired degrees of freedom. Note that the standard error of the slope estimate SE_{b_1} can be found on the regression output.

8. *Physical activity in the US.* The CDC monitors the physical activity level of Americans. A recent survey on a random sample of $23,129$ Americans yielded a 95% confidence interval of 61.1% to 62.9% for the proportion of Americans who walk for at least 10 minutes per day. Which of the following is the **correct** interpretation of this confidence interval?

- ○ 61.1% to 62.9% of the time Americans walk for at least 10 minutes per day.
- ○ 95% of the time the true proportion of Americans who walk for at least 10 minutes per day is between 61.1% to 62.9%.
- ○ Between 61.1% and 62.9% of random samples of $23,129$ Americans are expected to yield confidence intervals that contain the true proportion of Americans who walk for at least 10 minutes per day.
- ○ 95% of random samples of $23,129$ Americans will yield confidence intervals between 61.1% and 62.9 %.
- ◉ None of these.

✓ **Correct**