Week 2 Quiz

Quiz, 8 questions

✖ **Try again once you are ready.**

Required to pass: 80% or higher

You can retake this quiz up to 3 times every 8 hours.

Back to Week 2

Retake

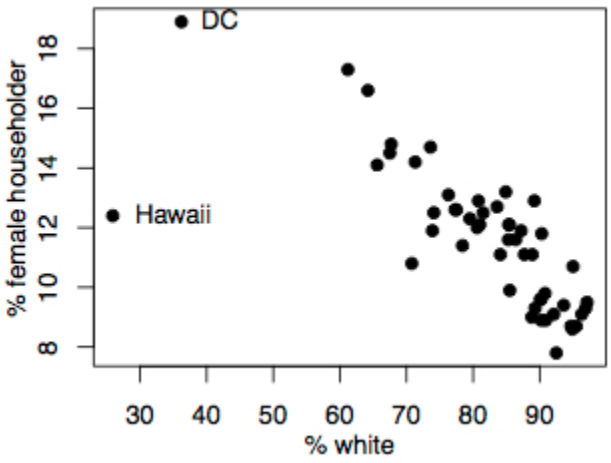
Week 2 Quiz

Quiz, 8 questions

1 / 1 point

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?



- DC is more influential than Hawaii, but it has lower leverage than Hawaii.
- Hawaii has higher leverage and is more influential than DC.

Correct

Hawaii has higher leverage than DC because it is farther away from the bulk of the data in the x direction.

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

- Define a leverage point as a point that lies away from the center of the data in the horizontal direction.
- Define an influential point as a point that influences (changes) the slope of the regression line.

1. This is usually a leverage point that is away from the trajectory of the rest of the data

Week 2 Quiz

Quiz, 8 questions

1 / 1 point

2.

The model below is for predicting the heart weight (in g) of cats from their body weight (in kg). The coefficients are estimated using a dataset of 144 domestic cats. The correlation between the heart and body weight is 0.8. Which of the following is **false**?

	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	-0.36	0.69	-0.52	0.61
body_wt	4.03	0.25	16.12	0.00

- The slope estimate would not change if body weights were measured in pounds.

Correct

The correlation coefficient is unitless but the slope is not therefore this statement is false.

- The intercept is meaningless in context of the data and only serves to adjust the height of the regression line.
- The correlation coefficient would not change if body weights were measured in pounds.
- The explanatory variable is body weight, and the response variable is heart weight.

Week 2 Quiz

Quiz, 8 questions

1 / 1 point

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

- Because $Pr(>|t|) = 0$ for the gender variable, we can say that gender is not a significant predictor of heart weight in cats.

Correct

- The expected heart weight for female cats is 9.2 grams, on average.
- On average, male cats are expected to have hearts that weigh 2.12 grams more than the hearts of female cats.
- Gender is a significant predictor of heart weight in cats.

Week 2 Quiz

Quiz, 8 questions

1 / 1 point

4.

We fit a linear regression model for predicting the best used price of 23 GMC pickup trucks from their list price, both measured in thousands. Which of the following is **false** based on this model output?

	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	0.43	0.18	2.5	0.02
list_price	0.85	0.01	84.7	<2e-16

- For each additional $1,000 in the list price of a GMC pickup truck we would expect the best used price to be higher on average by $850.
- The linear model is $\widehat{best\_used\_price} = 0.43 + 0.85\ list\_price$.
- The 95% confidence interval for the slope can be calculated as $0.85 \pm 84.7 \times 0.01$.

Correct

False. We need the critical t score, not the observed t score, in calculation of the margin of error.

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.

Week 2 Quiz

Quiz, 8 questions

1 / 1 point

5.

Answer Question 5, 6 and 7 based on the information below:

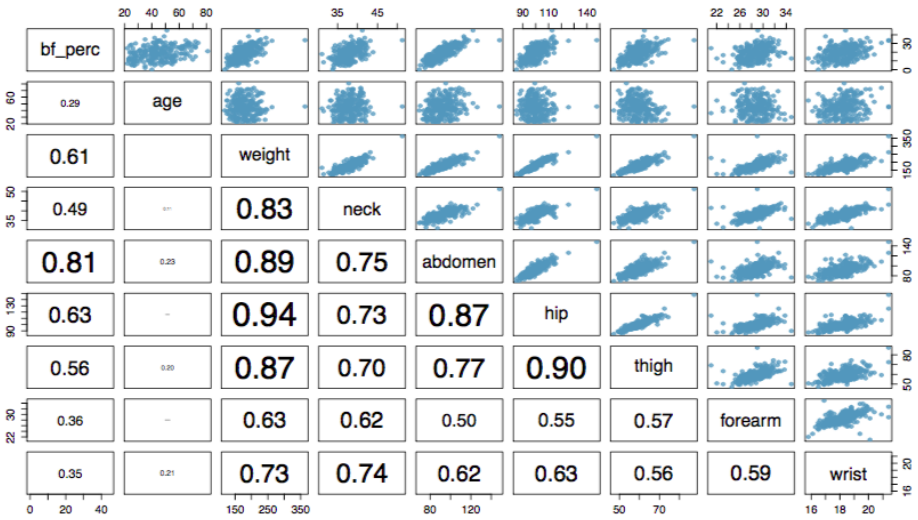
Body fat percentage can be complicated to estimate, while variables such age, height, weight, and measurements of various body parts are easy to measure. Based on data on body fat percentage and other various easy to obtain measurements, we develop a model to predict body fat percentage based on the following variables:

-age (years) - abdomen circumference (cm) - forearm circumference (cm)

-wight (pounds) - hip circumference (cm) - wrist circumference (cm)

-neck circumference (cm) - thigh circumference (cm)

The plot below shows the relationship between each of these variables and body fat percentage (the response variable) as well as the correlation coefficients between these variables:



And the following are the model outputs associated with this analysis:

Regression Summary	Estimate	Std. Error	t value	Pr(> | t |)
(Intercept)	-20.062	10.847	-1.850	0.066
age	0.059	0.028	2.078	0.039
weight	-0.084	0.037	-2.277	0.024
neck	-0.432	0.208	-2.077	0.039
abdomen	0.877	0.067	13.170	0.000
hip	-0.186	0.128	-1.454	0.147
thigh	0.286	0.119	2.397	0.017
forearm	0.483	0.173	2.797	0.006

ANOVA	Df	Sum Sq	Mean Sq	F value	Pr(>F)
age	1	1260.93	1260.93	80.21	0.0000
weight	1	5738.41	5738.41	365.04	0.0000
neck	1	153.37	153.37	9.76	0.0020
abdomen	1	3758.51	3758.51	239.09	0.0000
hip	1	6.42	6.42	0.41	0.5234
thigh	1	122.04	122.04	7.76	0.0058
forearm	1	79.91	79.91	5.08	0.0251
wrist	1	139.46	139.46	8.87	0.0032

Week 2 Quiz

Quiz, 8 questions

0 / 1 point

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 0.000. Since the p-value is small we reject the null hypothesis of no relationship.

This should not be selected

Recall that for a predictor i, the p-value given in the Regression Summary output refers to the following t-test:

- $H_0: \beta_i = 0$; with all other predictors in the model, predictor i does not explain a significant portion of the variance in body fat once all other predictors are included in the model, so the coefficient is not significantly different from 0.
- $H_A: \beta_i \neq 0$; predictor i does explain a significant portion of the variance in body fat even when all other predictors are included in the model, so the coefficient is significantly different from 0.

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.

- 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 0.039. Since the p-value is small we reject the null hypothesis of no relationship

Week 2 Quiz

Quiz, 8 questions

1 / 1 point

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 percentage point increase in body fat, abdomen circumference is expected to be higher by 0.745 to 1.009 cm.
- (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 cm in abdomen circumference, body fat percentage is expected to be higher by 0.745 to 1.009 percentage points.

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.

- (-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.

Week 2 Quiz

Quiz, 8 questions

0 / 1 point

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%.

This should not be selected

False, future random samples do not need to yield confidence intervals that fall within the first confidence interval.

- None of these.