Week 2 Quiz

Back to Week 2

8/8 points earned (100%)

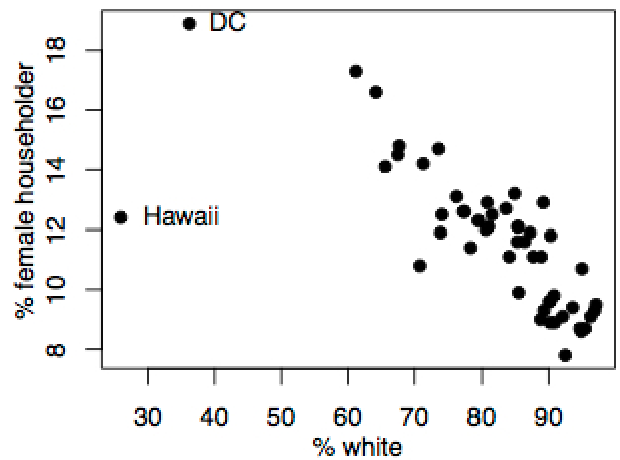
Quiz passed!

---

1 / 1 points

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?



- [ ] Neither DC nor Hawaii appear to be leverage points.
- [x] 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.

- [ ] 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.

1 / 1
points

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 explanatory variable is body weight, and the response variable is heart weight.

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.

1 / 1
points

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.
- 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.

- The expected heart weight for male cats is, on average, 11.32 grams.
- 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.

1 / 1

points

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

- [x] 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.

- [ ] List price is a significant predictor of the best used price.
- [ ] 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 intercept is meaningless in this context.

The linear model is $\widehat{best\_used\_price} = 0.43 + 0.85\ list\_price.$

1 / 1
points

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

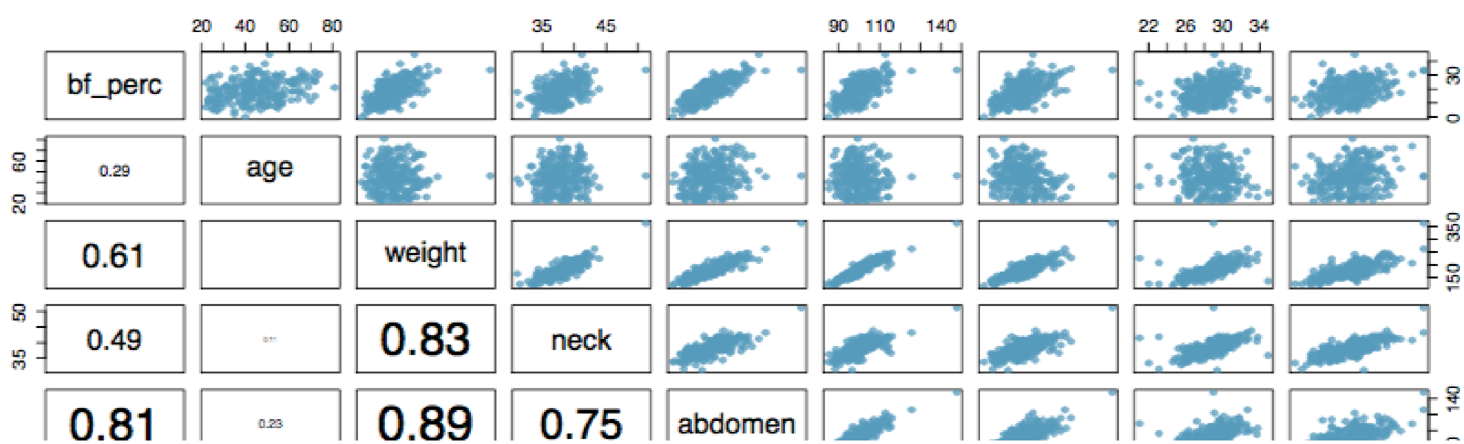
Body fat percentage can be complicated to estimate, while variables such age, height, weight, and measurements of various body parts are easy to measure. Based on data on body fat percentage and other various easy to obtain measurements, we develop a model to predict body fat percentage based on the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

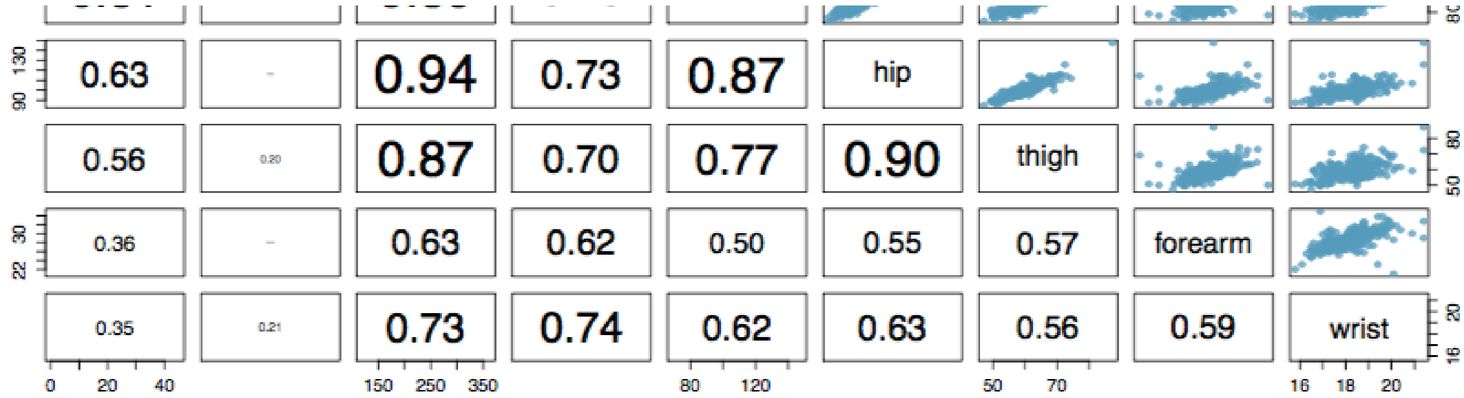
-age (years) - abdomen circumference (cm) - forearm circumference (cm)

-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
wrist	-1.405	0.472	-2.978	0.003

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
Residuals	243	3819.99	15.72		
Total	251	15079.02			

Residual standard error: 3.965 on 243 degrees of freedom

F-statistic: 89.53 on 8 and 243 DF, p-value: $< 2.2e-16$

Source: Penrose, K., Nelson, A., and Fisher, A. (1985), *Generalized Body Composition Prediction Equation for Men Using Simple Measurement Techniques*, Medicine and Science in Sports and Exercise, 7(2), 189.

Question 5

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

○

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.

- [x] Wrist circumference is the most significant predictor of body fat percentage since the slope associated with this variable has the highest magnitude.

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.

- [ ] The sample size is 252.
- [ ] All else held constant, people with wider hips tend to have lower body fat percentages.
- [ ] The F-test for the significance of the model overall suggests that at least one of the slope coefficients is significantly different than 0.

---

1 / 1
points

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. 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 / 2 = 0.0195. 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.

---

1 / 1
points

7.

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

- (-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.

- (0.745, 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.

- (0.745, 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, 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.

---

1 / 1
points

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%.