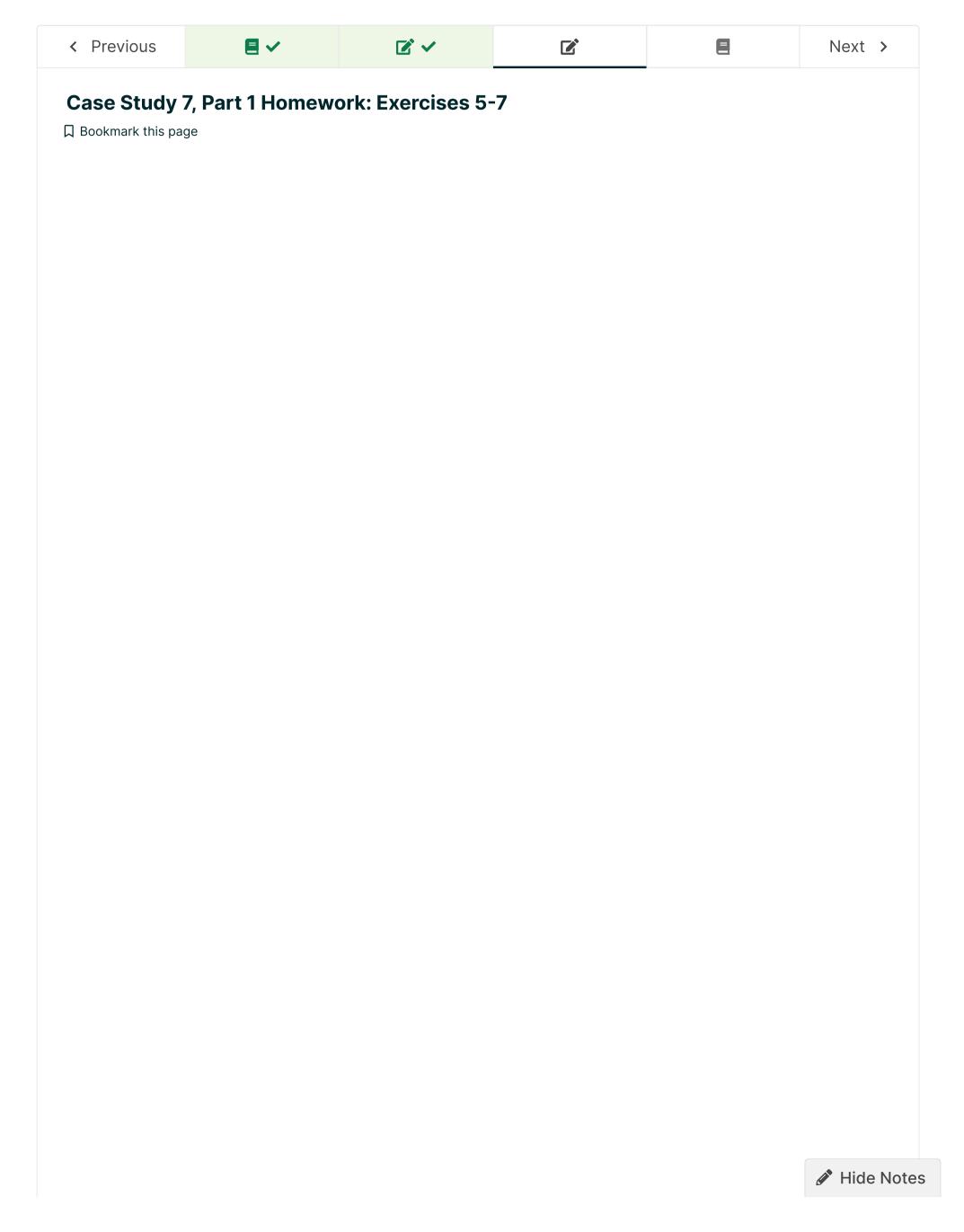
<u>Course</u> <u>Progress</u>

<u>Dates</u>

<u>Discussion</u> <u>Syllabus and FAQs</u>

<u>Notes</u>

☆ Course / Week 5: Statistical Learning / Homework: Case Study 7, Part 1



Homework due Jul 14, 2021 05:59 +06

Exercise 5

1/1 point (graded)

Some variables in the dataset are already numeric and perhaps useful for regression and classification. In Exercise 5, we will store the names of these variables for future use. We will also take a look at some of the continuous variables and outcomes by plotting each pair in a scatter plot. Finally, we will evaluate the skew of each variable.

Instructions

- Call plt.show() to observe the plot generated by the code given below. Which of the covariates and/or outcomes are correlated with each other?
- Call skew() on the columns outcomes_and_continuous_covariates in df. Is the skew above 1 for any of these variables?

Here is the code to get you started:

Which continuous covariate appears to be the most skewed?

willian continuous sevariate appears to be the most skewed.
budget
popularity
<pre>runtime</pre>
O vote_count
<pre>vote_average</pre>
revenue
<pre>profitable</pre>
Submit You have used 1 of 2 attempts
✓ Correct (1/1 point)

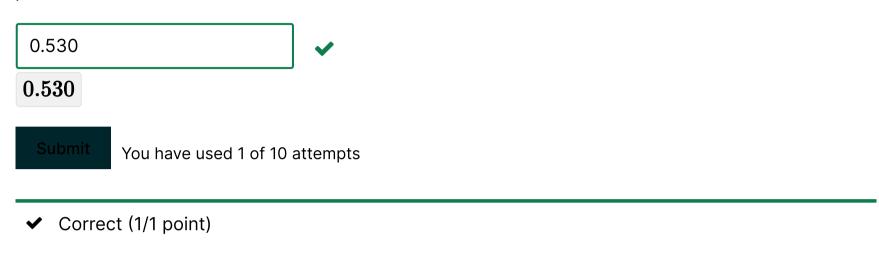
1/1 point (graded)

It appears that the variables <code>budget</code>, <code>popularity</code>, <code>runtime</code>, <code>vote_count</code>, and <code>revenue</code> are all right-skewed. In Exercise 6, we will transform these variables to eliminate this skewness. Specifically, we will use the <code>np.log10()</code> method. Because some of these variable values are exactly 0, we will add a small positive value to each to ensure it is defined; this is necessary because log(0) is negative infinity.

Instructions

For each above-mentioned variable in df, transform value x into np.log10(1+x).

What is the new value of <code>skew()</code> for the covariate <code>runtime</code>? Please provide the answer to 3 decimal points.



Exercise 7

1/1 point (graded)

Now we're going to save our dataset to use in Part 2 of this case study.

Instructions

Use to_csv() to save the df object as movies_clean.csv.

What is the correct way to save the df object?





edX

<u>About</u>

Affiliates

edX for Business

Open edX

<u>Careers</u>

News

Legal

Terms of Service & Honor Code

Privacy Policy

Accessibility Policy

<u>Trademark Policy</u>

<u>Sitemap</u>

Connect

<u>Blog</u>

Contact Us

Help Center

Media Kit

Donate















© 2021 edX Inc. All rights reserved.

深圳市恒宇博科技有限公司 <u>粤ICP备17044299号-2</u>