DATA 1220 Standard Test

Name:

"**LungCapData**" data describe the lung capacity (**LungCap**) of a sample of 725 US citizens. The data also include the following variables: **Age**, **Height**, **Smoke** (smoking status), **Gender**, and whether or not a person was born by **Caesarean**. The dataset can be downloaded from Canvas. Import the data into RStudio and complete the following questions. Use the level of significance $\alpha = 0.05$.

1. What are the values of quartiles of lung capacity in this sample?

2. Construct a 95% confidence interval for smoking rate of all US citizens (keep 4 decimal places). And write one sentence to interpret the interval. To get the partial credits, you need to show your work.

3. Suppose you want to investigate whether smokers have larger lung capacity than non-smokers, on average.
   (a) Which statistical methodology is appropriate?

   (b) Conduct the statistical methodology in part (a). To get the partial credits, you need to show your work.

   (c) A researcher concludes that smoking causes larger lung capacity. Do you agree with this conclusion or not? Please briefly explain your rationale.