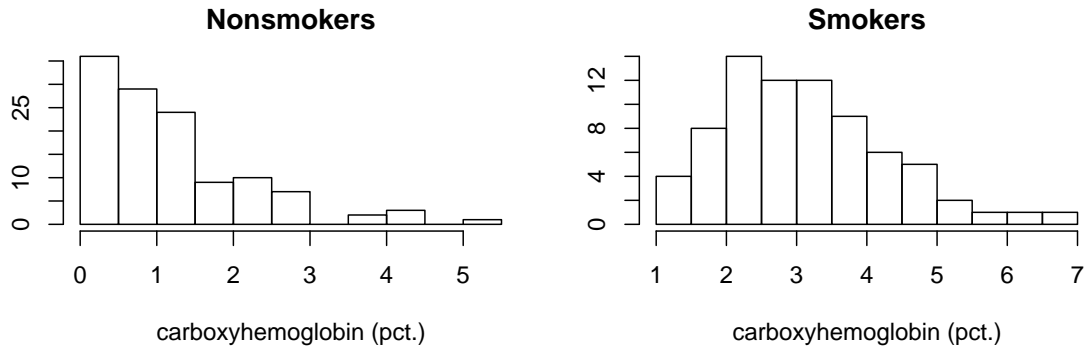


An investigator is interested in testing if the population mean levels of carboxyhemoglobin (as a percentage of total hemoglobin) differ between smokers and nonsmokers. The data may be considered to be random samples of 121 nonsmokers and 75 smokers from a particular population. Here are histograms of the two carboxyhemoglobin level samples.



Nonsmokers have mean 1.23 and standard deviation 1.03. Smokers have mean 3.18 and standard deviation 1.19. Answer parts (a) and (b) on the back side of the paper.

- (a) Use a t -test to test the hypothesis that the population mean carboxyhemoglobin levels are equal versus the alternative that they are higher in smokers. You may use the fact there are 139.8 degrees of freedom. State hypotheses, compute a test statistic, find a p -value (or a range for the p -value), and interpret the results of the test in the context of the problem.

Solution: Let μ_s and μ_{ns} be the population mean carboxyhemoglobin levels for smokers and nonsmokers. The hypotheses are as follows.

$$H_0: \mu_s = \mu_{ns} \quad \text{and} \quad H_A: \mu_s > \mu_{ns}$$

The test statistic is

$$\frac{3.18 - 1.23}{\sqrt{\frac{(1.19)^2}{75} + \frac{(1.03)^2}{121}}} = 11.73$$

From the table, the p -value is (substantially) smaller than 0.0005.

There is very strong evidence that the mean carboxyhemoglobin level is higher in smokers than in nonsmokers from this population.

- (b) List several assumptions that the previous test makes.

Solution: The t test assumes that we have two independent samples, that individuals are independently sampled within samples, and that the populations are normally distributed. The last of these is not critical because the sample sizes are relatively high.

For the following questions, *circle* TRUE or FALSE. If you answer FALSE, modify the statement to make it true or briefly justify your response.

- (c) TRUE or FALSE:

The skewness in the histograms indicates that the t test in part (a) is invalid.

Solution: False. The samples are skewed to the right which indicates that the populations are likely skewed as well. But the sample sizes are fairly large, and the Central Limit Theorem may be used as a justification for inferences based on the t distribution. The skewness in the populations would need to be substantially stronger for the sampling distribution to be so skewed as to make the

t distribution method invalid.

(d) TRUE or FALSE:

In question (a), if the population of smokers was older than the population of nonsmokers the effects of smoking and age on carboxyhemoglobin would be confounded.

Solution: True. The statistical method can only indicate strong evidence that the group mean for smokers is larger than that for nonsmokers. To attribute smoking as the *cause* of this difference requires further assumptions. If there are other variables that are different between the groups (such as age), the possible effects of these other variables would be confounded with the effect due to smoking.

(e) TRUE or FALSE:

If the p -value of a hypothesis test is 0.44, this indicates strong evidence that the null hypothesis is true.

Solution: False. The p -value indicates that the data is consistent with the null hypothesis. This is a much weaker interpretation than that there is strong evidence in favor of the null hypothesis.

There might be a large range of parameter values that are consistent with the observed data and we can't have strong evidence that each one is correct. A hypothesis test never indicates strong evidence in favor of the null, only consistency with it. A high confidence level confidence interval that says that we can be highly confident that the parameter values are close to what is stated in the null hypothesis is the right way to indicate support for the null hypothesis.

(f) TRUE or FALSE:

If a 95% confidence interval for $\mu_1 - \mu_2$ is strictly less than 0, the p -value for a one-sided test with alternative hypothesis $\mu_1 < \mu_2$ would be greater than 0.05.

Solution: False. The two-sided p -value would be smaller than 0.05. The one-sided p -value would then be smaller than 0.025.