In a study of the lizard *Sceloporis occidentalis*, researchers examined field-caught lizards for infection by the malarial parasite *Plasmodium*. To help assess the ecological impact of malarial infection, researchers tested 15 infected and 15 uninfected lizards for stamina, measured by the distance (in meters) each animal could run in two minutes. Summary statistics are tabulated below. You may also use the fact that the estimated degrees of freedom of the sampling distribution of the test statistic is 27.2.

Group	$\bar{y}$	s	n
Infected lizards	26.87	6.81	15
Uninfected lizards	32.23	8.07	15

(a) Find a 90% confidence interval for the difference in population means. Interpret this confidence interval in the context of the problem.

Solution: $-5.36 \pm 1.703 \times 2.73$, or -5.36 ± 4.64.

We are 90% confident that the difference in stamina (as measured by the distance in meters run in two minutes) between the mean of all infected lizards and the mean in all uninfected lizards in the species *Sceloporis occidentalis* is between -10 and -0.72.

(b) Test the null hypothesis of no difference in population means versus the alternative hypothesis that infection decreases stamina. Summarize the results in the context of the problem.

Solution: $H_0 : \mu_1 = \mu_2$
$H_A : \mu_1 < \mu_2$

$t = -5.36/2.7264 = -1.966$.

The p-value is the area to the left of -1.966 under a t distribution with 27.2 degrees of freedom, approximated here as 27 degrees of freedom. This area is between 0.025 and 0.05.

There is fairly strong evidence that infected lizards have less stamina (as measured by the distance in meters run in two minutes) than uninfected lizards (one-sided $p < 0.05$, independent sample t-test).

(c) From the list below, circle any statements that, if true, might invalidate the procedures used for constructing the confidence interval in part (a) or carrying out the hypothesis test in part (b).

1. Side-by-side boxplots of the data show that the distribution of each sample is fairly symmetric and without outliers.
2. A 90% confidence interval for $\mu_1 - \mu_2$ does not contain 0, but the one-sided (directional) p-value is less than 0.05. (Can this occur, or does it indicate a computation error?)
3. The distance measurements were taken for all of the infected lizards first and the uninfected lizards later in the day.
4. The investigators actually caught 49 uninfected lizards and randomly picked 15 of these to measure.
5. In the field, slower lizards are much easier to catch than quicker lizards, and are therefore more likely to be included in the sample.

Solution: Statements 3 and 5 might invalidate the inference procedures.

Statement 1 shows that the assumption that the sampling distribution of the test statistic is approximately t is probably okay.

Statement 2 is no problem. If the p-value were from a two-sided (non-directional) test, there would be reason for concern. But Statement 2 will occur whenever the one-sided (directional) p-value is between 0.025 and 0.05 as it is in this case.

Statement 3 indicates a possible bias in the measurement process if lizards tend to run different distances at different times in the day.

Statement 4 is no problem. A random sample of 15 from a random sample of 49 is a random sample of 15 from the population.

Statement 5 is a reason for concern. The mean of the sampled lizards in each group would tend to be biased too low if this were the case. The bias would be more strong in the group that was faster.