

An Analysis Using the Log Transformation

Researchers wish to compare the amount of messenger RNA produced by a certain gene when corn is grown under normal temperature to the amount of messenger RNA produced by the gene when corn is grown under cold growing temperature. If the amount of messenger RNA produced appears to change in response to the cold, there is some evidence that this gene is involved in a corn plant's response to cold temperature. The researchers obtain a measurement of the amount of messenger RNA produced in each of 15 independent pools of plants grown in cold temperature and 5 independent pools of plants grown at normal temperature. The data along with some summary statistics are provided below.

Treatment	Amount of messenger RNA														
2. Cold	29	32	43	82	89	107	119	139	175	202	211	261	281	359	518
1. Normal	14	20	42	81	140										

$$\bar{Y}_2 = 176.47 \quad s_2 = 135.34 \quad \bar{Y}_1 = 59.40 \quad s_1 = 52.15$$

1. Find s_p , the pooled-data estimate of population standard deviation.

$$\text{Verify } s_p = \sqrt{\frac{(5-1)52.15^2 + (15-1)135.34^2}{15+5-2}} = 121.86$$

2. Conduct the usual two-sample t -test using this data. Determine the test statistic, a p -value, and provide a conclusion.

μ_2 = mean amount of mRNA produced by population of plants grown in cold

μ_1 = mean amount of mRNA produced by population of plants grown at normal temp

$$t = \frac{\bar{Y}_2 - \bar{Y}_1 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{176.47 - 59.40 - 0}{121.86 \sqrt{\frac{1}{5} + \frac{1}{15}}} = 1.86$$

Two sided p -value is between _____. Some weak evidence the mean amount of mRNA differs between plants grown at cold and normal temperatures.

3. List some reasons why the analysis above may not be valid.

4. When each of the data values is replaced by its natural logarithm, the summary statistics become

$$\bar{Z}_2 = 4.87 \quad s_2 = 0.86 \quad \bar{Z}_1 = 3.74 \quad s_1 = 0.95 \quad s_p = 0.88$$

Conduct a the two-sample t -test using the log-transformed data. Determine the test statistic and a p -value.

5. What does $\bar{Z}_2 = 4.87$ estimate?
6. What does $e^{\bar{Z}_2} = e^{4.87} = 130.32$ estimate?
7. What does $\bar{Z}_2 - \bar{Z}_1 = 4.87 - 3.74 = 1.13$ estimate?
8. What does $e^{\bar{Z}_2 - \bar{Z}_1} = e^{4.87 - 3.74} = e^{1.13} \approx 3.10$ estimate? (Hint: $e^{\log(a) - \log(b)} = \frac{a}{b}$)
9. Find a 95% confidence interval for the difference between the mean of the log messenger RNA amount produced by the corn gene under cold growing temperature and the mean of the log messenger RNA amount produced by the corn gene under normal growing temperature.

10. Provide a clear statement that explains the results of the these analyses.

There is _____ evidence that the mean amount of mRNA produced differs between plants grown at cold and normal temperatures (cold growing temp may increase the gene's mRNA production) as two sided p-value is between _____. We estimate

$$\frac{\text{(pop.) median mRNA amount for plants grown in cold}}{\text{(pop.) median mRNA amount for plants grown at normal temp}} = 3.10.$$

That is, the median amount of mRNA produced by the gene under cold temp is estimated to be about _____ times greater than the median amount of mRNA produced under normal growing temperature