

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

$$\text{Test } H_0 : \mu_1 = \mu_2 \text{ vs. } H_a : \mu_1 \neq \mu_2 \text{ with } t = \frac{\bar{Y}_2 - \bar{Y}_1 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{176.47 - 59.40 - 0}{121.86 \sqrt{\frac{1}{5} + \frac{1}{15}}} = 1.86 \quad df = n_2 + n_1 - 2 = 18$$

Two sided p -value is between 0.05 and .10. Some weak evidence that 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.

σ_2 probably $>$ σ_1 and $n_2 > n_1 \Rightarrow$ pooled estimate s_p is probably too high.

Data indicate right skewed distributions. Perhaps even some outliers

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.

μ_2^z = mean log-amount of mRNA produced by population of plants grown in cold

μ_1^z = mean log-amount of mRNA produced by population of plants grown at normal temp

$$\text{Test } H_0 : \mu_1 = \mu_2 \text{ vs. } H_a : \mu_1 \neq \mu_2 \text{ with } t = \frac{\bar{Y}_2 - \bar{Y}_1 - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{4.87 - 3.74 - 0}{0.88 \sqrt{\frac{1}{5} + \frac{1}{15}}} = 2.48 \quad df = n_2 + n_1 - 2 = 18$$

Two sided p -value is between 0.02 and .04. Some moderate evidence that the mean amount of log mRNA differs between plants grown at cold and normal temperatures (cold growing temp seemingly increases log mRNA production).

5. What does $\bar{Z}_2 = 4.87$ estimate?

The pop mean μ_2^z of the log mRNA amounts for the population of plants grown in cold or the log of the median of mRNA amounts for the population of plants grown in cold $\log(\text{median}(Y_2))$

That is, $\bar{Z}_2 = 4.87$ estimates $\mu_2^z = \log(\text{median}(Y_2))$

6. What does $e^{\bar{Z}_2} = e^{4.87} = 130.32$ estimate?

median mRNA amount for population of plants grown in cold

That is, $e^{\bar{Z}_2} = e^{4.87} = 130.32$ estimates $e^{\log(\text{median}(Y_2))} = \text{median}(Y_2)$

7. What does $\bar{Z}_2 - \bar{Z}_1 = 4.87 - 3.74 = 1.13$ estimate?

$\log(\text{median mRNA amount for population of plants grown in cold}) - \log(\text{median mRNA amount for population of plants grown at normal temp})$

That is, $\bar{Z}_2 - \bar{Z}_1$ estimates $\mu_2^z - \mu_1^z = \log(\text{median}(Y_2)) - \log(\text{median}(Y_1))$

8. What does $e^{\bar{Z}_2 - \bar{Z}_1} = e^{4.87 - 3.74} = e^{1.13} \approx 3.10$ estimate? (Hint: $e^{\log(a) - \log(b)} = \frac{a}{b}$)

the ratio given by $\frac{\text{median mRNA amount for population of plants grown in cold}}{\text{median mRNA amount for population of plants grown at normal temp}}$

That is, $e^{\bar{Z}_2 - \bar{Z}_1}$ estimates $e^{\log(\text{median}(Y_2)) - \log(\text{median}(Y_1))} = \frac{\text{median}(Y_2)}{\text{median}(Y_1)}$

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.

95% CI for $\mu_2^z - \mu_1^z = \log(\text{median}(Y_2)) - \log(\text{median}(Y_1))$ is given by

$$\bar{Z}_2 - \bar{Z}_1 \pm t_{n_1+n_2-2}^{(0.975)} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 4.87 - 3.74 \pm 2.101(0.88) \sqrt{\frac{1}{5} + \frac{1}{15}} = 0.18 \text{ to } 2.08$$

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

There is moderate 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 0.02 and 0.04. We estimate

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

That is, the median amount of mRNA produced by the gene under cold temp is estimated to be about 3.10 times larger/greater than the median amount of mRNA produced under normal growing temperature

In general, we estimate median of population 2 is $e^{\bar{Z}_2 - \bar{Z}_1}$ times as large as the median of population 1

To get a 95% CI for $\frac{\text{median}(Y_2)}{\text{median}(Y_1)} \Rightarrow$ back-transform the 95% CI for $\mu_2^z - \mu_1^z = \log(\text{median}(Y_2)) - \log(\text{median}(Y_1))$ by e

Here, 95% CI for $\frac{\text{median}(Y_2)}{\text{median}(Y_1)}$ is $e^{0.18} = 1.20$ to $e^{2.08} = 8.00$; **that is, estimate the median amount of mRNA produced by plants grown in cold is between 1.20 and 8.00 times greater than the median amount of mRNA produced by plants grown at normal temperature.**