

Solution to Statistics 571 Midterm 2
Hanlon/Larget, Fall 2011

1.

Solution:

(a) According to the normal table, the 30th percentile is about 0.525 standard deviations below the mean, which is $11.2 - (0.525)(2.5) = 9.9$.

(b)

$$\begin{aligned} P(10.0 < Y < 12.4) &= P\left(\frac{10.0 - 11.2}{2.5} < \frac{Y - 11.2}{2.5} < \frac{12.4 - 11.2}{2.5}\right) \\ &= P(-0.48 < Z < 0.48) \end{aligned}$$

This is twice the area to the right of 0.48 under a standard normal density, $2(0.3156) = 0.6312$.

(c) The mean of the sampling distribution of \bar{Y} is the population mean of 11.2. The standard deviation of the sampling distribution of \bar{Y} is σ/\sqrt{n} which is $2.5/\sqrt{4} = 1.25$.

(d)

$$\begin{aligned} P(10.0 < \bar{Y} < 12.4) &= P\left(\frac{10.0 - 11.2}{1.25} < \frac{\bar{Y} - 11.2}{1.25} < \frac{12.4 - 11.2}{1.25}\right) \\ &= P(-0.96 < Z < 0.96) \end{aligned}$$

This is twice the area to the right of 0.96 under a standard normal density, $2(0.1685) = 0.337$.

(e) The probability will be greater. The reason is that with a larger sample size, the standard deviation is smaller so it is more likely that the sample mean will be close to the population mean of 11.2. Thus, the probability in an interval around 11.2 will go up.

2.

Solution:

(a) There are 14 degrees of freedom, so the associated critical t^* for the middle 95% is 2.14.

$$23.0 \pm 2.14 \times \frac{22.1}{\sqrt{15}}, \quad \text{or} \quad 23.0 \pm 12.2$$

The 95% confidence interval for the population mean time for the venom to travel from foot to groin is $10.8 < \mu < 35.2$ minutes.

(b) The 10,000 bootstrap means were sorted from smallest to largest, and the 0.025 and 0.975 sample quantiles were found. As 2.5% of 10,000 is 250, the 250th and 9751st values from the sorted list were determined (counting 250 from each end).

(c) The bootstrap method may be more reliable than the t -method for this data because of the apparent skewness in the population as revealed by skewness in the sample shown in the dot plot. The t -methods are susceptible to problems with nonnormal populations, especially skewness, when sample sizes are relatively small.

3.

Solution:

- (a) The null hypothesis is $H_0: \mu_D = 0$, the mean difference in number of aphids per plant between top and bottom is zero.

The alternative hypothesis is $H_A: \mu_D \neq 0$, the mean difference in number of aphids per plant between top and bottom is not zero.

- (b) This is a paired test because data is collected in pairs from each of the ten plants.

$$T = \frac{10.6 - 8.1}{2.92/\sqrt{10}} \doteq 2.71$$

- (c) There are $10 - 1 = 9$ degrees of freedom.

- (d) The two-sided p-value is between 0.02 and 0.05.

- (e) There is strong evidence that there is a difference between the mean number of aphids found between the top and bottom of this type of plant under the experimental conditions ($p < 0.05$, $T = 2.71$, two-sided paired t -test, 9 df).

4.

Solution:

- (a) This is a two-independent sample setting because the individual observations (time for spoonfuls of ice cream to melt) are not paired.

Select the pooled estimate of the common population standard deviation.

$$s_p = \sqrt{\frac{8(20.7)^2 + 8(78.2)^2}{8 + 8}} \doteq 57.2$$

The standard error is

$$SE = 57.2 \times \sqrt{\frac{1}{9} + \frac{1}{9}} \doteq 27$$

There are $9 + 9 - 2 = 16$ degrees of freedom, and the corresponding critical t^* for a 95% confidence interval is 2.12 from the table. Thus, the margin of error is $2.12 \times 27 \doteq 57.2$.

The 95 confidence interval is 67.1 ± 57.2 or $9.9 < \mu_A - \mu_B < 124.3$.

- (b) We are 95% confident that the mean time for a teaspoon of ice cream A to melt is between 9.9 and 124.3 seconds longer than that for ice cream B under the experimental conditions.
- (c) Welch's method may be more reliable because it does not assume equal population variances and the data summary suggests that melting times for Flavor B are much more variable (with the SD about four times as large).

5.

Solution:

- (a) Let μ_F and μ_M be the mean populations weights of female and male mosquitos, measured in milligrams. The hypotheses are:

$$H_0: \mu_F = \mu_M$$

$$H_A: \mu_F > \mu_M$$

- (b) The pooled estimate of the common population standard deviation is

$$s_p = \sqrt{\frac{10(0.0785)^2 + 8(0.0342)^2}{18}} \doteq 0.0628$$

and the test statistic is

$$T = \frac{0.2648 - 0.1743}{0.0628 \sqrt{\frac{1}{11} + \frac{1}{9}}} \doteq 3.21$$

- (c) The p-value for the hypothesis test is 1 times the area to the RIGHT of the correct test statistic calculated in part (b) beneath a t density with 18 degrees of freedom.
- (d) From the table, the one-sided p-value is between 0.001 and 0.01, so 0.002448 must be correct.
- (e) One questionable assumption is normality, as the graph of the sample of female weights exhibits some skewness. A second questionable assumption is equal population variances as the plot and numerical summaries indicate that female weights may be more variable than male weights. Either answer is acceptable.
- (f) To obtain this p-value, the 20 observations were placed in a random order and the difference between the means of the first 11 and last 9 values was calculated. This process was repeated a total of 10,000 times, resulting in 10,000 differences in sample means. Only 18 of these 10,000 differences were at least as large as the observed difference $0.2648 - 0.1743 = 0.0905$ between the female and male mean sample weights. This indicates that the actual difference in sample means is quite unusual as compared to the distribution of differences in sample means from random groupings.