Statistics 250 Honors  
Monday, Feb. 18, 2002

Question 1 (6 points)
An experiment is performed in which a needle 1 inch long is tossed 1000 times onto a tabletop engraved with parallel lines exactly 1 inch apart. Of these 1000 tosses, 643 result in the needle crossing a line.

We wish to estimate the true proportion of tosses which result in a line crossing.

(a) Compute \( \hat{p} \) and its standard error (give numeric answers, not formulas). Tell whether each of these is an example of a statistic or a parameter. (1 point)

\[
\hat{p} = \frac{643}{1000} = 0.643 \\
\text{SE}_{\hat{p}} = \sqrt{\frac{0.643(0.357)}{1000}} = 0.015.
\]

Each of these quantities is a statistic.

(b) Give a 95\% confidence interval for the true population proportion. (1 point)

\[
\hat{p} \pm z^* (\text{SE}_{\hat{p}}) = 0.643 \pm 1.96(0.015) = 0.643 \pm 0.030
\]

(c) A friend performs a similar experiment, tossing a needle of unknown length onto the same lined tabletop 1000 times. She observes 625 line crossings. You wish to test

\[ H_0: \text{The true proportions in the two experiments are equal.} \]

Calculate the appropriate \( z \) statistic for the test (do not find a \( p \)-value). (1 point)

\[
z = \frac{0.643 - 0.625}{\sqrt{\frac{1268}{2000} \left(1 - \frac{1268}{2000}\right) \left(\frac{1}{1000} + \frac{1}{1000}\right)}} = 0.836
\]

It is possible to prove that the true proportion in the original 1000-toss experiment with the 1-inch needle is equal to \( 2/\pi \), which is 0.63662. Use this new information in parts (d) and (e) only.

(d) Give the true value of the mean and standard deviation of the \( \hat{p} \) statistic for repetitions of the experiment. (1 point)

The mean and standard deviation of the sampling distribution of \( \hat{p} \) are \( p \) and \( \sqrt{p(1-p)/n} \), respectively. In this case, we know that \( p = 0.63662 \). Thus, we get

\[
\mu_{\hat{p}} = 0.63662 \quad \text{and} \quad \sigma_{\hat{p}} = \sqrt{\frac{0.63662(0.36338)}{1000}} = 0.0152.
\]
(e) Name the theorem which tells us the approximate distribution of \( \hat{p} \). Using this approximate distribution, find the approximate probability that if you repeated the experiment, you would observe a value of \( \hat{p} \) less than 0.6215. (2 points)

The central limit theorem implies that \( \hat{p} \) is approximately normally distributed with mean \( \mu_p \) and standard deviation \( \sigma_p \) (see part (d)). Converting the number 0.6215 into a z-score, then, gives \((0.6215 - 0.63662)/(0.0152) = -0.995\). We then use Table III on page 520 to conclude that

\[
P(\hat{p} < 0.6215) \approx P(Z < -0.995) = 0.16.
\]

**Question 2 (7 points)**

Researchers studying vitamin C in a commodity called wheat soy blend (WSB) were concerned that some of the vitamin C content would be destroyed as a result of storage and shipment of the commodity to its final destination in Haiti. The researchers specially marked a collection of bags at the factory and took a sample from each of them to measure the vitamin C content. Five months later in Haiti they found each of the specially marked bags and took samples. The data consist of two vitamin C measures for each bag, one at the time of factory production and the other five months later in Haiti. The units are milligrams of vitamin C per 100 grams of WSB. There were 27 bags marked in the study.

The data are summarized below (you may not need all the information in the table):

<table>
<thead>
<tr>
<th>Factory</th>
<th>Haiti</th>
<th>Difference (Factory minus Haiti)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{x} = 42.85 )</td>
<td>( \bar{x} = 37.52 )</td>
<td>( \bar{x} = 5.33 )</td>
</tr>
<tr>
<td>( s = 4.79 )</td>
<td>( s = 2.44 )</td>
<td>( s = 5.59 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( n = 27 )</td>
</tr>
</tbody>
</table>

(a) Set up hypotheses to examine the question of interest to these researchers. Should the alternative hypothesis be 1-sided or 2-sided? (1 point)

\( H_0 : \) The mean difference (factory minus Haiti) equals 0 (\( \mu_D = 0 \)).

\( H_a : \) The mean difference is greater than 0 (\( \mu_D > 0 \)).

This is a one-sided alternative, since we’re only concerned with a loss of vitamin C.

(b) Perform the appropriate significance test at the 5% level and explain your results. In your answer, clearly mark your test statistic and p-value. (3 points)

We perform a paired (or matched) t-test, which is the same as a one-sample t-test on the differences. Using the third column in the table above, the t statistic is

\[
t = \frac{5.33 - 0}{5.59/\sqrt{27}} = 4.95
\]
on 26 degrees of freedom. From table VI on p. 527, we find that the corresponding p-value is less than 0.0005. This leads us to reject \( H_0 \) and conclude that there is a significant decrease in vitamin C content.

(c) Find a 90% confidence interval for the mean vitamin C content at the factory. (2 points)

We need a t-based confidence interval since we don’t know the exact standard deviation of the population of all possible factory measurements. On 26 df, the \( t^* \) value for the 90% interval is 1.706. Thus, the CI is

\[
\bar{x} \pm t^* \left( \frac{s}{\sqrt{n}} \right) = 42.85 \pm 1.706(4.79/\sqrt{27}) = 42.85 \pm 1.57.
\]

(d) State precisely what “90% confidence interval” means. (1 point)

If we resampled many times and constructed a new 90% CI each time, then 90% of these intervals would enclose the true factory mean \( \mu_F \).