STT 351

1. Four pumps are pulled from a process under statistical control. The four pumps average 3.2 gallons per minute with sample s.d. = \( s_x = 0.7 \).

   a. Write the formula for a 99\% confidence interval for \( \mu_x \).

      \[
      \bar{x} \pm 3.182 \frac{s_x}{\sqrt{n}}
      \]

   b. Numerically evaluate (a) for the information given but do not reduce.

      \[
      3.2 \pm 3.182 \frac{0.7}{\sqrt{4}}
      \]

   c. What is your numerical estimate of the population sd \( \sigma_x \) from the information given? Do not reduce.

      \[
      \text{est } \sigma_x = 0.7
      \]

   d. What is your numerical estimate of the sd of the sample mean (i.e. your estimate of \( \sigma_x \)) based on the information given? Don't reduce.

      \[
      \text{est } \sigma_x = \frac{s_x}{\sqrt{4}}
      \]

   e. What performance claim is made for a 99\% confidence interval?

      - 99\% of samples of \( n=4 \) yield a 99\%
      - \( t \) is "exact"
      - \( \pm \) confidence interval \( (1) \) containing \( \mu_x \)

   f. For large \( n \), what happens to the width of a CI if \( n \) is replaced by \( 4n \)?

      \[
      \frac{1}{\sqrt{4n}} = \frac{1}{2n} \text{ so width of CI is half as large.}
      \]

   g. What happens to \( \frac{\bar{x} - \mu_x}{s_x / \sqrt{n}} \) if scores \( x_i \) are replaced by \( (3 x_i - 6) \)?

      - \( \frac{3x_i - \mu_x}{s_x / \sqrt{n}} \) is unchanged!

      - \( \frac{x + c - \mu_x}{s_x / \sqrt{n}} \) is unchanged as for \( x \)

      - \( 3 \) is same ratio
2. Sixty with-replacement samples are selected from parts plated with process x. Independently of these, 100 samples are selected from parts plated with process y. The following data apply to a measurement of corrosion resistance

<table>
<thead>
<tr>
<th>sample mean</th>
<th>sample sd</th>
<th>sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>24.6</td>
<td>60</td>
</tr>
<tr>
<td>y</td>
<td>23.8</td>
<td>100</td>
</tr>
</tbody>
</table>

a. Numerically determine, but do not reduce, your large-n estimate of the margin of error for $\bar{x} - \bar{y}$.

$$1.96 \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}} = 1.96 \sqrt{\frac{3.7^2}{60} + \frac{3.9^2}{100}}$$

b. Same as (a) except the samples are without replacement and the population sizes are 800 and 600 for x and y respectively.

$$1.96 \sqrt{\frac{3.7^2}{800} + \frac{3.9^2}{600-1}}$$

3. It is desired to estimate the mean of $y = \text{burst pressure}$ for a type of plastic inflatable already in use by the public. A random sample of 50 is selected. The sample is effectively with replacement since the population size N is so large relative to 50. It is anticipated that another measurement $x = \text{time in use}$ may have a bearing on this, so both x and y scores are measured for each of the 50. We find sample s.d. are

$s_x = 17.8 \text{ months}$ \hspace{1cm} $s_y = 20 \text{ psi}$

If the sample correlation of x with y is $r = 0.8$ determine the numerical value of the margin of error of the regression based estimator of $\mu_y$. Do not reduce.

$$1.96 \sqrt{1-r^2} \frac{s_y}{\sqrt{n}}$$

$$1.96 \sqrt{1-0.8^2} \frac{20}{\sqrt{50}}$$
4. A random sample of 100 lead seals is selected from production. Of these are 29 found to be defective.
   a. Give the formula for a 95% \( z \)-based CI for the population fraction of defective seals.

\[
\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}
\]

\[
\frac{29}{100} \pm 1.96 \sqrt{\frac{\frac{29}{100}(\frac{71}{100})}{100}}
\]

b. Numerically evaluate (a) but do not reduce.

5. A 95% bootstrap ci for the population mean, based on a dataset named "prices" is given by a call to the function
   \text{bootci}[\text{mean, prices, 10000, 0.95}]
   a. Give the call required to generate a \[99\%\] bootstrap ci for the population \[\text{median}\] but with \[\text{twice the number of bootstrap replications}\].

   \text{bootci}[\text{median, prices, 20000, 0.99}]

b. Is it correct to say that the bootstrap method effectively increases the sample size of the dataset? \text{NO, OUR BOOTSTRAP REPLICATE SAMPLES ARE NOT NEW DATA, THEY MERELY SUBSTITUTE FOR ANALYTICAL CALCULATION.}

6. Machine processes are scored for \( x \) = efficiency. It is desired to obtain a 95% ci for \( \mu_x \) by the method of \text{proportional stratified sampling} with strata corresponding to three levels of cycling rate.

   \[
   \begin{array}{ccc}
   \text{stratum} & 1 & 2 & 3 \\
   \text{stratum size} & 1000 & 1500 & 800 \\
   N = 3300
   \end{array}
   \]

   a. If a total sample size of 33 is used, how is this allocated among the three strata?

\[
\frac{1000}{3300} \cdot 33 = 10
\]

\[
\frac{1500}{3300} = 15
\]

\[
\frac{800}{3300} = 8.
\]

\[
\sum n = 33
\]
b. The stratum by stratum sample means are

\[
\bar{x} = \sum_i^3 w_i \bar{x}_i = \frac{10}{33}(2.6) + \frac{15}{33}(3.1) + \frac{8}{33}(2.8)
\]

Determine the overall sample mean of all 66. Don't reduce.

\[
\bar{x} = \sum_i^3 w_i \bar{x}_i = \frac{10}{33}(2.6) + \frac{15}{33}(3.1) + \frac{8}{33}(2.8)
\]

c. The estimated s.d. of \(\bar{x}\) (from this stratified sample) is given by

\[
\sqrt{\sum_i^3 W_i^2 \frac{s_i^2}{n_i}}
\]

Give the numerical values of the weights \(W_i\).

\[
\omega_1 = \frac{10}{33}, \quad \omega_2 = \frac{15}{33}, \quad \omega_3 = \frac{8}{33}
\]

7. A maximum likelihood estimator selects the model giving the most probability to what has been seen (the data). By this way of thinking, which model best explains the event R1 G2?

- model 1: select two with replacement from [ R R G G G ]
- model 2: select two without replacement from [ R G G ]

Show your calculations.

1. \(P(R1 G2) = \frac{2}{5} \cdot \frac{3}{5} = \frac{6}{25}\)

2. \(P(R1 G2) = \frac{1}{3} \cdot 1 = \frac{1}{3}\)

\(\frac{1}{3} > \frac{6}{25}\) (i.e. 25 > 18) so

model 2 is the more likely choice among the two models.
<table>
<thead>
<tr>
<th>Degrees of freedom</th>
<th>Central area = confidence/prediction level for two-sided interval:</th>
<th>Cumulative area = confidence/prediction level for one-sided interval:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80%</td>
<td>90%</td>
</tr>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
</tr>
<tr>
<td>2</td>
<td>1.886</td>
<td>2.920</td>
</tr>
<tr>
<td>3</td>
<td>1.638</td>
<td>2.353</td>
</tr>
<tr>
<td>4</td>
<td>1.533</td>
<td>2.132</td>
</tr>
<tr>
<td>5</td>
<td>1.476</td>
<td>2.015</td>
</tr>
<tr>
<td>6</td>
<td>1.440</td>
<td>1.943</td>
</tr>
<tr>
<td>7</td>
<td>1.415</td>
<td>1.895</td>
</tr>
<tr>
<td>8</td>
<td>1.397</td>
<td>1.860</td>
</tr>
<tr>
<td>9</td>
<td>1.383</td>
<td>1.833</td>
</tr>
<tr>
<td>10</td>
<td>1.372</td>
<td>1.812</td>
</tr>
<tr>
<td>11</td>
<td>1.363</td>
<td>1.796</td>
</tr>
<tr>
<td>12</td>
<td>1.356</td>
<td>1.782</td>
</tr>
<tr>
<td>13</td>
<td>1.350</td>
<td>1.771</td>
</tr>
<tr>
<td>14</td>
<td>1.345</td>
<td>1.761</td>
</tr>
<tr>
<td>15</td>
<td>1.341</td>
<td>1.753</td>
</tr>
<tr>
<td>16</td>
<td>1.337</td>
<td>1.746</td>
</tr>
<tr>
<td>17</td>
<td>1.333</td>
<td>1.740</td>
</tr>
<tr>
<td>18</td>
<td>1.330</td>
<td>1.734</td>
</tr>
<tr>
<td>19</td>
<td>1.328</td>
<td>1.729</td>
</tr>
<tr>
<td>20</td>
<td>1.325</td>
<td>1.725</td>
</tr>
<tr>
<td>21</td>
<td>1.323</td>
<td>1.721</td>
</tr>
<tr>
<td>22</td>
<td>1.321</td>
<td>1.717</td>
</tr>
</tbody>
</table>