You may use a calculator and one page of (two-sided) formula sheet.

1. (20 points) Bottles of a popular cola drink are supposed to contain 300 milliliters (ml) of cola. There is some variation from bottle to bottle because the filling machinery is not perfectly precise. The distribution of the contents is normal with standard deviation \( \sigma = 3 \) ml. A student who suspects that the bottle is under-filling measures the contents of six bottles. The results are

\[
299.4 \quad 297.7 \quad 301.0 \quad 298.9 \quad 300.2 \quad 297.0.
\]

(a). Calculate the sample mean \( \bar{x} \) and sample variance \( s^2 \).

Sample mean = 299.0333,
Sample variance = 2.258667.

(b). What is the variance of \( \bar{x} \), \( \text{Var}(\bar{x}) \)?

\[
\frac{s^2}{6} = 1.5
\]

(c). Specify the hypotheses the student wants to test.

\( H_0: \ \mu = 300 \quad \text{vs} \quad H_1: \ \mu < 300 \)

(d). What is the statistic you suggest to the test? Give the value.

\[
z = \frac{\sqrt{n} (\bar{x} - 300)}{\sigma} = -0.7893
\]

(e). Is this convincing evidence that the mean contents of cola bottle is less than the advertised 300 ml? Draw your conclusion at significance level \( \alpha = 0.05 \).

One-sided test. Critical value \( z(0.05) = -1.645 \)

\( z = -0.7893 > -1.645. \)

Cannot reject \( H_0 \). At level 0.05 we can conclude that the bottles contain 300 milliliters (ml) of cola.

2. (20 points) You measure the weights of 16 male runners. These runners are not a random sample from a population, but you are willing to assume that their weights represent the weights of similar runners. Here are their weights in kilograms:

\[
67.8, 61.9, 63.0, 53.1, 62.3, 59.7, 55.4, 58.9,
\]


60.9, 69.2, 63.7, 68.3, 64.7, 65.6, 56.0, 57.8
\( \bar{x} = 61.77, s^2 = 22.74 \).
(a). Find a 95\% confidence interval for the mean weight of the population of all such runners. You may leave your answer as \( \bar{xxx} \pm xxx \times \sqrt{xxxx} \).
\[ \bar{x} \pm t(n-1, \alpha/2) \frac{s}{\sqrt{n}} = 61.77 \pm (2.1314) \sqrt{22.78/16} \]

(b). For a future observation, provide a 95\% confidence interval for its value. You may leave your answer as \( xxx \pm xxx \times \sqrt{xxxx} \).
\[ \bar{x} \pm t(n-1, \alpha/2) \frac{s}{\sqrt{n}} \sqrt{1 + \frac{1}{n}} = 61.77 \pm (2.1314) \sqrt{22.78 \times \sqrt{1 + \frac{1}{16}}} \]

(c). What assumptions did you make in order to make your conclusions in (a) and (b) to be valid?

For (a), the sample is taken at random from a normal population;
For (b), the observation is from the same population

3. (30 points) Researchers studying the leaning of speech often compare measurements made on the recorded speech of adults and children. One variable of interest is called the voice onset time (VOT). Here are the results for six-year old children and adults asked to pronounce the words “bees.” The VOT is measured in milliseconds and can be either positive or negative.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>( \bar{x} )</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children</td>
<td>10</td>
<td>-3.67</td>
<td>33.89</td>
</tr>
<tr>
<td>Adults</td>
<td>20</td>
<td>-23.17</td>
<td>50.74</td>
</tr>
</tbody>
</table>

(a). What is the estimate of the difference of the means for two groups?

\[ \mu_c - \mu_a: \ \text{estimate} \ -3.67 \times (-23.17) = 19.5 \]

(b). The researchers were investigating whether VOT distinguishes adults from children. State \( H_0 \) and \( H_1 \) (specify the meanings of your notation)

\[ H_0: \ \mu_c - \mu_a = 0; \ \ \text{vs} \ \ H_1: \ \mu_c - \mu_a \neq 0 \]

(c). Test the hypotheses in (b) under the assumption that the variances from two
populations are equal. Draw your conclusion.

\[ s^2_p = \frac{(n_1 - 1)s^2_1 + (n_2 - 1)s^2_2}{N - 2} = \frac{(9)(33.89^2) + (19)(50.74^2)}{10 + 20 - 2} = 2116.185 \]

\[ t = \frac{19.5}{s_p \sqrt{\frac{1}{10} + \frac{1}{20}}} = 1.095 \]

Two-sided test, critical value \( t(28, 0.025) = 2.0484 \)

\( |1.095| < 2.0484 \)

Cannot reject \( H_0 \) and conclude that there is no significant difference between groups.

(d). Test the hypotheses in (b) without assuming that the variances from two populations are equal. You only need to give the statistic for the test and specify its distributions. Determine the degrees of freedom by

\[ k = \frac{1}{n_1 - 1} \left[ \frac{s^2_1 / n_1 + s^2_2 / n_2}{s^2_1 / n_1 + s^2_2 / n_2} \right]^2 + \frac{1}{n_2 - 1} \left[ \frac{s^2_2 / n_1 + s^2_2 / n_2}{s^2_1 / n_1 + s^2_2 / n_2} \right]^2 \]

\[ = 0.0394 \]

\[ k = 25.38, \quad v = 25. \]

\[ t = \frac{19.5}{\sqrt{\frac{s^2_1}{10} + \frac{s^2_2}{20}}} = 1.249 \sim t(25), \quad t(25, 0.025) = 2.0595. \quad \text{Accept } H_0. \]

(e). Let \( S_c \) and \( S_a \) denote the sample standard deviations for the children and adults groups. In order to test the equality of the variances of the two populations, we use the statistic

\[ F = \frac{S^2_c}{S^2_a}. \]

Specify the distribution of the statistic under the null hypothesis and give the rejection region. You don’t have to give the exact critical values here.

F-distribution with dfs 9 and 19.

Reject \( H_0 \) if \( F > F(9, 19, 0.025) \) or \( F < F(9, 19, 0.975) \).

4. (8 points) For a set of observations: 3.5, 4.6, 2.7, 5.6, 7.0, 2.0, 9.0, 1.0, 8.5, 6.0. Find the first, second and third quartiles.

Sort the data: 1.0, 2.0, 2.7, 3.5, 4.6, 5.6, 6.0, 7.0, 8.5, 9.0

\( n = 10. \)

\( (10)(0.25) = 2.5 \) \quad \text{The first quartile} = 2.7;

\( (10)(0.50) = 5.0 \) \quad \text{The second quartile} = (4.6 + 5.6) / 2 = 5.1 \text{ (median)}
(10)(0.75)=7.5 \quad \text{The third quartile}=7.0;

5. (22 points) Let X be a normal distribution with mean 80 and standard deviation 10.
(a) Find the 90% percentile.
\[
\frac{x - 80}{10} = 1.28, \quad x=92.8
\]
(b) Calculate the probability \( P(X>90) \).
\[
\frac{X - 80}{10} > 1 \quad \text{prob}=0.1587
\]
(c) A sample of size 101 is taken from the population, and the sample mean and variance are denoted by \( \bar{X} \) and \( S^2 \).
(c1) What is the distribution of \( \bar{X} \)?
Normal with mean 80 and variance 100/101.
(c2) What is the distribution of \( S^2 \)?
Chi-square with df= 100.
(c3) What is the distribution of \( \frac{\bar{X} - 80}{S/\sqrt{101}} \)?
\( t(100) \)