Statistics I, Fall 2012
Suggested Solution for Homework 11

Instructor: Ling-Chieh Kung
Department of Information Management
National Taiwan University

1. With the confidence level $1 - \alpha = 0.9$ and the sample size $n = 20$, we follow the typical three steps to do an interval estimation.

   - **Selection of the distribution**: Because the amounts spent are normally distributed and the population variance is unknown, we will use the t distribution to construct the confidence interval.
   - **Calculation**: The sample mean is $\bar{x} = 4.922$. The sample standard deviation is $s = 2.003$ and thus the multiplier is $\frac{s}{\sqrt{n}} = 0.448$. The critical t value is $t_{\frac{\alpha}{2},n-1} = t_{0.05,19} = 1.729$. Combining all the above, the confidence interval is
   $$\left[\bar{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}\right] = [4.148, 5.696].$$
   - **Conclusion**: With a 90% confidence level, the average amount a customer spends on a meal while this combination is purchased is between \$4.148 and \$5.696.

2. With the confidence level $1 - \alpha = 0.99$ and the sample size $n = 40$, we follow the typical three steps to do an interval estimation.

   - **Selection of the distribution**: Because the sample size $n = 40$ is larger than 30, we will use the z distribution to construct the confidence interval. Because the population variance is unknown, we will use the sample variance as a substitute.
   - **Calculation**: The sample mean is $\bar{x} = 11.3$. The sample standard deviation is $s = 8.209$ and thus the multiplier is $\frac{s}{\sqrt{n}} = 1.298$. The critical t value is $t_{\frac{\alpha}{2},n-1} = t_{0.005,39} = 2.708$. Combining all the above, the confidence interval is
   $$\left[\bar{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}\right] = [7.785, 14.815].$$
   - **Conclusion**: With a 99% confidence level, the average number of years of experience in supply chain among all supply chain transportation managers is between 7.785 years and 14.815 years.

3. With the confidence level $1 - \alpha = 0.95$ and the sample size $n = 1000$, we follow the typical three steps to do an interval estimation.

   - **Selection of the distribution**: Because the sample size $n = 1000$ is larger than 30, we will use the z distribution to construct the confidence interval. In calculating the standard error, because the population proportion is unknown, we will use the sample proportion as a substitute.
   - **Calculation**: The sample proportion is $\hat{p} = 0.64$. The approximated standard error is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.015$. The critical z value is $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$. Combining all the above, the confidence interval is
   $$\left[\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = [0.6103, 0.6698].$$
   - **Conclusion**: With a 95% confidence level, the proportion of registered votes supporting this candidate is between 61.03% and 66.98%.

4. (a) With the confidence level $1 - \alpha = 0.99$ and the sample size $n = 1003$, we follow the typical three steps to do an interval estimation.

   - **Selection of the distribution**: Because the sample size $n = 1003$ is larger than 30, we will use the z distribution to construct the confidence interval. In calculating the standard error, because the population proportion is unknown, we will use the sample proportion as a substitute.
   - **Calculation**: The sample proportion is $\hat{p} = 0.27$. The approximated standard error is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.014$. The critical z value is $z_{\frac{\alpha}{2}} = z_{0.005} = 2.576$. Combining all the above, the confidence interval is

$$\left[\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = [0.2339, 0.3061].$$

   - **Conclusion**: With a 99% confidence level, the Universal Music Group's market share is between 23.39% and 30.61%.

   (b) In this case, the approximated standard error is $\sqrt{\frac{\hat{p}(1-\hat{p})}{5000}} = 0.0063$. The confidence interval is

$$\left[\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{5000}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{5000}}\right] = [0.2538, 0.2862].$$

   With a 99% confidence level, the Universal Music Group's market share is between 25.38% and 28.62%. We can see that increasing the sample size reduces the size of the confidence interval.

5. (a) With the confidence level $1 - \alpha = 0.99$ and the sample size $n = 14$, we follow the typical three steps to do an interval estimation.

   - **Selection of the distribution**: Because the population is normal we will use the chi-square distribution to construct the confidence interval.
   - **Calculation**: The sample variance is $s^2 = 7.748$. The critical chi-square values are $\chi^2_{1-\frac{\alpha}{2},n-1} = \chi^2_{0.995,13} = 3.565$ and $\chi^2_{\frac{\alpha}{2},n-1} = \chi^2_{0.005,13} = 29.819$. Combining all the above, the confidence interval is

$$\left[\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)s^2}{\chi^2_{\frac{1-\alpha}{2},n-1}}\right] = [3.378, 28.252].$$

   - **Conclusion**: With a 99% confidence level, the population variance is between 3.378 and 28.252.

   (b) We first estimate the population variance. With the confidence level $1 - \alpha = 0.95$ and the sample size $n = 25$, we follow the typical three steps to do an interval estimation.

   - **Selection of the distribution**: Because the population is normal we will use the chi-square distribution to construct the confidence interval.
   - **Calculation**: The sample variance is $s^2 = 2100.682$. The critical chi-square values are $\chi^2_{1-\frac{\alpha}{2},n-1} = \chi^2_{0.975,24} = 12.401$ and $\chi^2_{\frac{\alpha}{2},n-1} = \chi^2_{0.025,24} = 39.364$. Combining all the above, the confidence interval is

$$\left[\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)s^2}{\chi^2_{\frac{1-\alpha}{2},n-1}}\right] = [1280.771, 4065.459].$$

   - **Conclusion**: With a 95% confidence level, the population variance is between 1280.771 and 4065.459.

   It then follows that, with a 95% confidence level, the population standard deviation is between 35.788 and 63.761.

6. **Note.** Because the problem was not stated clear when it was posted, all of you can get these 10 points for free.

   With the confidence level $1-\alpha = 0.95$ and the sample size $n = 14$, we follow the typical three steps to do an interval estimation.

   - **Selection of the distribution**: Because the population is normal we will use the chi-square distribution to construct the confidence interval.
   - **Calculation**: The sample variance is $s^2 = 71356553.85$. The critical chi-square values are $\chi^2_{1-\frac{\alpha}{2},n-1} = \chi^2_{0.975,13} = 5.009$ and $\chi^2_{\frac{\alpha}{2},n-1} = \chi^2_{0.025,13} = 24.736$. Combining all the above, the confidence interval is

   $$\left[\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)s^2}{\chi^2_{\frac{1-\alpha}{2},n-1}}\right] = [37502022.06, 185202914.95].$$

   - **Conclusion**: With a 95% confidence level, the population variance is between 37502022.06 and 185202914.95.

7. Suppose the sample size is n and $n \geq 30$. In this case, we will use the z distribution to estimate the population mean (if $n < 30$, because the population is not normal, we will not be able to do an interval estimation without Nonparametric Statistics). With a 90% confidence interval, the critical z value is $z_{\frac{\alpha}{2}} = z_{0.05} = 1.645$. It then follows that the error of the interval estimation is $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = \frac{822.5}{\sqrt{n}}$, where $\sigma = 500$ is the estimated population standard deviation. As the error should be no more than 100, we have

   $$\frac{822.5}{\sqrt{n}} \leq 100 \quad \Leftrightarrow \quad n \geq \left(\frac{822.5}{100}\right)^2 \approx 67.64.$$

   Therefore, the sample size should be 68.

8. Suppose the sample size is n and $n \geq 30$. In this case, we will use the z distribution to estimate the population mean (if $n < 30$, because the population is not normal, we will not be able to do an interval estimation without Nonparametric Statistics). With a 90% confidence interval, the critical z value is $z_{\frac{\alpha}{2}} = z_{0.05} = 1.645$. It then follows that the error of the interval estimation is $z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} = 1.645\sqrt{\frac{p(1-p)}{n}}$, where p is the population proportion. As p is unknown, we are unable to calculate $p(1-p)$. However, as $p(1-p)$ is maximized at $p = 0.5$, we have $p(1-p) \leq 0.25$ and thus $1.645\sqrt{\frac{p(1-p)}{n}} \leq \frac{0.8225}{\sqrt{n}}$. As the error should be no more than 0.02, we have

   $$\frac{0.8225}{\sqrt{n}} \leq 0.02 \quad \Leftrightarrow \quad n \geq \left(\frac{0.8225}{0.02}\right)^2 \approx 1691.27.$$

   Therefore, the maximum sample size we need is 1692.

9. (a) The population mean is 49.74 and the population variance is 101.63.
   (b) If we draw a histogram, we will see that the population is normal.
   (c) i. We should use the t distribution because the population variance is unknown and the population is normal. We cannot use the z distribution because the sample size is not large enough.
       ii. You should see around $200 \times 0.95 = 190$ intervals covering the population mean.
   (d) i. We should use the z distribution because the population variance is known and the population is normal. It does not matter whether the sample size is large or small.
       ii. You should see around $200 \times 0.9 = 180$ intervals covering the population mean.
   (e) i. We should use the chi-square distribution because the population is normal.
       ii. You should see around $200 \times 0.95 = 190$ intervals covering the population mean.