Department of Industrial Engineering & Operations Research

IEOR 165 (Spring 2016)

Homework 5

Due: Friday, Apr 22

Question 1. The PCB concentration of a fish caught in Lake Michigan was measured by a technique that is known to result in an error of measurement that is normally distributed with a standard deviation of .08 ppm (parts per million). Suppose the results of 10 independent measurements of this fish are:

11.2, 12.4, 10.8, 11.6, 12.5, 10.1, 11.0, 12.2, 12.4, 10.6

(a) Give a 95 percent confidence interval for the PCB level of this fish.
(b) Give a 95 percent lower confidence bound.
(c) Give a 95 percent upper confidence bound.

Solution.
Note when $\alpha = 0.05$, $z(1-\alpha) = 1.645$ and $z(1-\alpha/2) = 1.96$.
(a) $n = 10$, $\overline{X} = 11.48$ and $\sigma = 0.08$. The 95% percent confidence interval is given by

$$\underline{\mu} = \overline{X} - \sigma \cdot z(1-\alpha/2)/\sqrt{n} = 11.48 - 0.08 \cdot 1.96/\sqrt{10} = 11.430$$

$$\overline{\mu} = \overline{X} + \sigma \cdot z(1-\alpha/2)/\sqrt{n} = 11.48 + 0.08 \cdot 1.96/\sqrt{10} = 11.530$$

(b) The 95% lower confidence bound is

$$\underline{\mu} = \overline{X} - \sigma \cdot z(1-\alpha)/\sqrt{n} = 11.48 - 0.08 \cdot 1.645/\sqrt{10} = 11.438$$

(c) The 95% upper confidence bound is

$$\overline{\mu} = \overline{X} - \sigma \cdot z(1-\alpha)/\sqrt{n} = 11.48 + 0.08 \cdot 1.645/\sqrt{10} = 11.522$$

Question 2. Let $X_1, \ldots, X_n, X_{n+1}$ be a sample from a normal population having an unknown mean μ and variance 1. Let $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ be the average of the first n of them.
(a) What is the distribution of $X_{n+1} - \bar{X}_n$?
(b) If $\bar{X}_n = 4$, give an interval that, within 90 percent confidence, will contain the value of X_{n+1}.

Solution.
(a) Since a linear combination of independent normal distributed random variables also follows normal distribution, $X_{n+1} - \overline{X}_n$ is normal for sure. Now we need to figure out its mean and variance.

$$E(X_{n+1} - \overline{X}_n) = E(X_{n+1}) - E(\overline{X}_n) = \mu - \mu = 0$$

$$Var(X_{n+1} - \overline{X}_n) = Var(X_{n+1}) + Var(\overline{X}_n) = 1 + \frac{1}{n} = \frac{n+1}{n}$$

So $X_{n+1} - \overline{X}_n \sim N(0, \frac{n+1}{n})$.

(b) Given $\overline{X}_n = 4$, we can use the following fact

$$\mathbb{P}\Big(\left| \frac{X_{n+1} - \overline{X}_n}{\sqrt{\frac{n+1}{n}}} \right| \leq z(1-\alpha/2) \Big) = 1 - \alpha \quad \big(\text{Note that } \frac{X_{n+1} - \overline{X}_n}{\sqrt{\frac{n+1}{n}}} \sim N(0,1)\ \big)$$

So the confidence interval for X_{n+1} is

$$\underline{X_{n+1}} = \overline{X}_n - z(1-\alpha/2)\sqrt{\frac{n+1}{n}} = 4 - 1.645\sqrt{\frac{n+1}{n}}$$

$$\overline{X_{n+1}} = \overline{X}_n + z(1-\alpha/2)\sqrt{\frac{n+1}{n}} = 4 + 1.645\sqrt{\frac{n+1}{n}}$$

Question 3. A sample of 10 fish were caught at lake A and their PCB concentrations were measured using a certain technique. The resulting data in parts per million were

$$\text{Lake A} : 11.5,\ 10.8,\ 11.6,\ 9.4,\ 12.4,\ 11.4,\ 12.2,\ 11,\ 10.6,\ 10.8$$

In addition, a sample of 8 fish were caught at lake B and their levels of PCB were measured by a different technique than that used at lake A. The resultant data were

$$\text{Lake B} : 11.8,\ 12.6,\ 12.2,\ 12.5,\ 11.7,\ 12.1,\ 10.4,\ 12.6$$

If it is known that the measuring technique used at lake A has a variance of .09 whereas the one used at lake B has a variance of .16, could you reject (at the 5 percent level of significance) a claim that the two lakes are equally contaminated? (assume the population distribution is normal)

Solution. The null hypothesis is that these two lakes are equally contaminated: $\mu_A = \mu_B$. Since σ_A^2 and σ_B^2 are known, we can perform a two-sample unpaired two-tailed Z-test, which gives

$$p = \mathbb{P}\Big(|Z| > \left| \frac{\overline{X}_A - \overline{X}_B}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} \right| \Big) = \mathbb{P}\Big(|Z| > \left| \frac{11.170 - 11.988}{\sqrt{0.09/10 + 0.16/8}} \right| \Big) = \mathbb{P}(|Z| > 4.803) = 1.56 \times 10^{-6}$$

So we reject the null hypothesis and we can conclude that these two lakes are not equally contaminated.

Question 4. A professor claims that the average starting salary of industrial engineering graduating seniors is greater than that of civil engineering graduates. To study this claim, samples of 16 industrial engineers and 16 civil engineers, all of whom graduated in 1993, were chosen and sample members were queried about their starting salaries. If the industrial engineers had a sample mean salary of \$47,700 and a sample standard deviation of \$2,400, and the civil engineers had a sample mean salary of \$46,400 and a sample standard deviation of\$2,200, has the professors claim been verified (at the 5 percent level of significance)? Find the appropriate p-value. (assume the population distribution is normal)

Solution. Let the average starting salary of industrial engineering seniors and civil graduates be μ_I and μ_C respectively. We may have two choices of null hypothesis: $\mu_I \leq \mu_C$ and $\mu_I \geq \mu_C$. You can choose either one and please state your decision clearly. Since σ_I^2 and σ_C^2 are unknown, we should perform a two-sample unpaired t-test. Under the null hypothesis $\mu_I \le \mu_C$, the p-value is given by

$$p = \mathbb{P}\Big(t_{n_I+n_C-2} > \frac{\overline{X}_I - \overline{X}_C}{s\sqrt{1/n_I + 1/n_C}}\Big)$$

where

$$s^2 = \frac{1}{n_I + n_C - 2}((n_I - 1)s_I^2 + (n_C - 1)s_C^2) = 5300000$$

So

$$p = \mathbb{P}\Big(t_{30} > \frac{47700 - 46400}{\sqrt{5300000}\sqrt{1/15 + 1/15}}\Big) = \mathbb{P}\Big(t_{30} > 1.597\Big) = 0.0604$$

So we accept the null hypothesis and we can not conclude that industrial engineering graduates earn more than civil engineers (if the null hypothesis is $\mu_I \ge \mu_C$, p-value is 0.9396 and we will conclude that industrial engineers earn more than civil engineers)

Question 5. A question of medical importance is whether jogging leads to a reduction in ones pulse rate. To test this hypothesis, 8 nonjogging volunteers agreed to begin a 1-month jogging program. After the month their pulse rates were determined and compared with their earlier values. If the data are as follows, can we conclude that jogging has had an effect on the pulse rates with the 5 percent level of significance (assume the population distribution is normal)?

Subject	1	2	3	4	5	6	7	8
Pulse Rate Before	74	86	98	102	78	84	79	70
Pulse Rate After	70	85	90	110	71	80	69	74

Solution. Note that we should perform a two-sample paired t-test. Let the average pulse rate before and after the jogging for subject i be $\mu_{X,i}$ and $\mu_{Y,i}$. If we want to test whether jogging leads to a reduction in one's pulse rate, the null hypothesis is $\mu_{X,i} \le \mu_{Y,i}$. The mean difference is

$$\overline{\Delta} = \overline{X - Y} = 2.75$$

And the sample variance for the difference is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\Delta_i - \overline{\Delta}) = 37.93$$

The p-value is given by

$$p = \mathbb{P}\Big(t_{n-1} > \frac{\overline{\Delta}}{s\sqrt{1/n}}\Big) = \mathbb{P}\Big(t_7 > \frac{2.75}{\sqrt{37.93}\sqrt{1/8}}\Big) = \mathbb{P}(t_7 > 1.263) = 0.1235$$

Thus we can not reject the null hypothesis and we can not conclude that jogging has had an effect on the pulse rate (if you use a different null hypothesis, e.g. two-tailed, it is OK if you state your hypothesis clearly).