Homework 5

Due: Thursday, May 4

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 .8 ppm (parts per million). Suppose the results of 10 independent measurements of this fish are:

$$12.8, \ 11.0, \ 11.1, \ 10.6, \ 12.2, \ 9.7, \ 12.9, \ 10.9, \ 11.8, \ 11.3$$

(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) $n = 10$, $\overline{X} = 11.43$ and $\sigma = 0.8$. The 95% percent confidence interval is given by

$$\underline{\mu} = \overline{X} - \sigma \cdot z(1 - \alpha/2)/\sqrt{n} = 11.43 - 0.8 \cdot 1.96/\sqrt{10} = 10.93$$

$$\overline{\mu} = \overline{X} + \sigma \cdot z(1 - \alpha/2)/\sqrt{n} = 11.43 + 0.8 \cdot 1.96/\sqrt{10} = 11.93$$

(b) The 95% lower confidence bound is

$$\underline{\mu} = \overline{X} - \sigma \cdot z(1 - \alpha)/\sqrt{n} = 11.43 - 0.8 \cdot 1.645/\sqrt{10} = 11.01$$

(c) The 95% upper confidence bound is

$$\overline{\mu} = \overline{X} - \sigma \cdot z(1 - \alpha)/\sqrt{n} = 11.43 + 0.8 \cdot 1.645/\sqrt{10} = 11.84$$

Question 2. 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

$$\text{Lake A}: 10.4, \ 10.7, \ 10.6, \ 10.2, \ 9.8, \ 10.8, \ 10.8, \ 10.0, \ 11.0, \ 10.4$$

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

$$\text{Lake B}: 12.3, \ 11.8, \ 11.6, \ 10.9, \ 11.5, \ 11.5, \ 11.6, \ 11.3$$

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

$$p = \mathbb{P}\left(|Z| > \left|\frac{\overline{X}_A - \overline{X}_B}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}}\right|\right) = \mathbb{P}\left(|Z| > \left|\frac{10.47 - 11.56}{\sqrt{0.09/10 + 0.16/8}}\right|\right) = \mathbb{P}(|Z| > 6.41) = 7.02 \times 10^{-11}$$

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

Question 3. Suppose we have conducted 6 null hypothesis tests, with p-values as

Test #	p-value
1	0.007
2	0.025
3	0.012
4	0.051
5	0.003
6	0.068

Use both Bonferroni correction and Holm-Bonferroni method to determine which tests should be rejected when the family-wise error rate is $\alpha = 0.05$.

Solution.
Bonferroni correction: we should compare each p-value with corrected $\alpha' = \alpha/6 = 0.0083$. Thus we should reject tests 1, 5 and accept tests 2, 3, 4, 6.

Holm-Bonferroni method: first we sort p-values from the smallest to the largest: 0.003, 0.007, 0.012, 0.025, 0.051, 0.068. Then we calculated adjusted $\alpha(k) = \alpha/(n - k + 1)$ for $k = 1, 2, \ldots, 6$: 0.0083, 0.01, 0.0125, 0.0167, 0.025, 0.05. We now need to find the smallest k such that $p_{(k)} > \alpha(k)$. For this case, the smallest $k = 4$, hence we should reject tests 1, 3, 5 and accept the remaining tests.

Question 4. A machine shop contains 3 ovens that are used to heat metal specimens. Subject to random fluctuations, they are all supposed to heat to the same temperature. To test this hypothesis, temperatures were noted on 15 separate heatings. The following data resulted.

Oven	Temperature
1	488.1, 494.7, 490.2, 489.6, 487.9
2	483.7, 480.5, 475.5, 482.7, 479.6
3	500.3, 492.2, 513.3, 490.1, 495.5

Do the ovens appear to operate at the same temperature? Test at the 5 percent level of significance. What is the p-value? (assume the population distribution is normal)

Solution. We need to perform ANOVA in this case. The null hypothesis is $\mu_1 = \mu_2 = \mu_3$. First, compute the sample averages as

$$\overline{X}^1 = 490.1 \quad \overline{X}^2 = 480.4 \quad \overline{X}^3 = 498.3$$

Then MSG and MSE are

$$MSG = 400.58$$

$$MSE = 34.37$$

So

$$p = \mathbb{P}(F_{2,12} > \frac{MSG}{MSE}) = \mathbb{P}(F_{2,12} > 11.66) = 0.0015 < 0.05$$

Thus we should reject the null hypothesis and conclude that these ovens do not operate at the same temperature.

Question 5. Suppose that a process is in control with $\mu = 10$ and $\sigma = 1.5$. An $\overline{X}$-control chart based on subgroups of size 6 with significance level of 0.05 is employed. If a shift in the mean of 3.1 units occurs, what is the probability that the next subgroup average will fall outside the control limits? On average, how many subgroups will have to be looked at in order to detect this shift?

Solution. First we calculate the LCL and UCL:

$$LCL = \mu - z(1 - \alpha/2) \cdot \sigma/\sqrt{n} = 8.80$$

$$UCL = \mu + z(1 - \alpha/2) \cdot \sigma/\sqrt{n} = 11.20$$

If a shift occurs, the distribution is now $\mathcal{N}(10 \pm 3.1, (1.5)^2)$. Since the normal distribution is symmetric, we can just compute the probability under the case $\mathcal{N}(10 + 3.1, (1.5)^2)$. With this shift, the subgroup average $\overline{X}$ will follow a normal distribution $\mathcal{N}(13.1, (1.5)^2/6)$. The probability that the next subgroup average will fall outside the control limits is

$$\mathbb{P}(\overline{X} < LCL) + \mathbb{P}(\overline{X} > UCL)$$
$$= \mathbb{P}(Z < \frac{LCL - 13.1}{1.5/\sqrt{6}}) + \mathbb{P}(Z > \frac{UCL - 13.1}{1.5/\sqrt{6}})$$
$$= P(Z < -7.02) + P(Z > -3.10)$$
$$= P(Z < -7.02) + 1 - P(Z < -3.10) = 0.999$$

On average, we need to look at $\frac{1}{0.999} = 1.01$ subgroups to detect this shift.