# CENG/NANO 114

# Probability and Statistical Methods for Engineers

Problem Set 5

1. We would like to obtain a confidence interval for the gain in a circuit on a semiconductor device. Assume that gain is normally distributed with standard deviation = 20. We have calculated the mean gain of a sample of circuits to be 1000.
   i. Construct a 90% confidence interval if the sample size is 10.
   ii. Construct a 90% confidence interval if the sample size is 25.
   iii. Construct a 95% confidence interval if the sample size is 10.
   iv. Construct a 95% confidence interval if the sample size is 25.
   v. How does the width of the confidence intervals change with sample size and confidence level?

2. A student group concerned about the mean fat content of a new burger called low fat sandwich sold in their school restaurant. So they submit a random sample of 12 low fat sandwiches to an independent lab for analysis. The percentage of fat in each of the sandwich is:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 21 | 18 | 19 | 16 | 18 | 24 | 22 | 19 | 24 | 14 | 18 | 15 |

   The restaurant claims that the mean fat content of this new sandwich is less than 20%. Carry out an appropriate hypothesis test to advise the student group on the validity of the restaurant's claim. $\alpha = 0.05$

3. A textile fiber manufacturer is investigating a new drapery yarn, which the company claims has a mean thread elongation of 12 kilograms with a population standard deviation of 0.5 kilograms using a random sample of 16 specimens.
   i. What is the type I error probability if the critical region is defined as < 11.5 kilograms?
   ii. Calculate the probability of a type II error if the true mean elongation is normally distributed with a mean of 11.5 kilograms and the same standard deviation. $\alpha = 0.05$ (Hint: lecture 17 pg20)

4. In a chemical distillation process, we performed ten measurements of the purity of oxygen obtained (y) and the percentage of hydrocarbons present in the main condenser of the distillation unit (x), as given in the table below.

| Obs. Number | Hydrocarbon Level (x%) | Purity (y%) |
|---|---|---|
| **1** | 0.99 | 90.01 |
| **2** | 1.02 | 89.05 |
| **3** | 1.15 | 91.43 |
| **4** | 1.29 | 93.74 |
| **5** | 1.46 | 96.73 |
| **6** | 1.36 | 94.45 |
| **7** | 0.87 | 87.59 |
| **8** | 1.23 | 91.77 |
| **9** | 1.55 | 99.42 |
| **10** | 1.4 | 93.65 |

   i. Calculate the correlation coefficient between the purity of oxygen and the hydrocarbon level. Is this correlation coefficient statistically significant at the 5% level (show all steps clearly in your test, including the appropriate hypotheses)?
   ii. Perform a linear regression analysis to determine the relationship between the purity of oxygen and the hydrocarbon level.

5. A study compiled the following data on the compression strength (lb) for a sample of 12-oz aluminum cans filled with strawberry drink and another sample filled with cola. Perform a hypothesis test to determine if data suggest that the extra carbonation of cola results in a higher average compression strength at $\alpha$=0.01. Outline all steps clearly.

| Beverage | Sample Size | Sample Mean | Sample SD |
|---|---|---|---|
| Cola | 15 | 554 | 15 |
| Strawberry drink | 15 | 540 | 21 |

6. Two different formulations of an oxygenated motor fuel are being tested to study their road octane numbers. The variance of road octane number for formulation 1 is $\sigma_1^2 = 1.5$, and for formulation 2, it is $\sigma_2^2 = 1.2$. Two random samples of size $n_1 = 15$ and $n_2 = 20$ are tested, and the mean road octane numbers observed are $\overline{x}_1 = 89.6$ and $\overline{x}_2 = 92.5$. Assume normality.
   i. If formulation 2 produces a higher road octane number than formulation 1, the manufacturer would like to detect it. Formulate and test an appropriate hypothesis using $\alpha = 0.05$.

      (Hint: The population variance is given)

ii. Construct a 95% confidence interval on the difference in mean road octane number.

(Hint: We are concerned about the absolute value of difference.)

iii. What sample size would be required in each population if you wanted to be 95% confident that the error in estimating a difference in mean road octane number of 1?

(Hint: In this case, the size n1=n2, and it is related to the length of confidence interval.)

7. The life in hours of a 75-watt light bulb is known to be normally distributed with $\sigma = 25$ hours. A random sample of 20 bulbs has a mean life of $x = 1014$ hours.

i. Construct a 95% two-sided confidence interval on the mean life.

ii. Construct a 95% lower-confidence bound on the mean life. Compare the lower bound of this confidence interval with the one in part1.

8. Ten individuals have participated in a diet-modification program to stimulate weight loss. Their weight in pounds both before and after participation in the program is shown in the following list.

| **Subject** | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|
| **Before** | 195 | 213 | 247 | 201 | 210 |
| **After** | 187 | 195 | 221 | 190 | 197 |

| **Subject** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|
| **Before** | 187 | 294 | 215 | 310 | 246 |
| **After** | 175 | 278 | 199 | 285 | 221 |

i. Is there evidence to support the claim that this particular diet-modification program is effective in producing a mean weight reduction? Use $\alpha = 0.05$.

ii. Is there evidence to support the claim that this particular diet- modification program will result in a mean weight loss of at least 10 pounds? Use $\alpha = 0.05$.

iii. Suppose that, if the diet-modification program results in mean weight loss of at least 10 pounds, it is important to detect this with probability of at least 0.90. Was the use of 10 subjects an adequate sample size? If not, how many subjects should have been used?

9. In question 8, suppose that instead of measuring ten individuals before and after participation in the program, the experiment was conducted instead on twenty individuals

who were randomly assigned into a control group vs a treatment group. Assume that the results given in the Table in qn 8 still stands, with "Before" denoting the control group and "After" denoting the treatment group and all subjects are different individuals.

i. Perform a hypothesis test on whether the program is effective at stimulating weight loss at the 5% level.
ii. Construct a 95% confidence interval of the mean difference in weight between the treatment group and the control group.

10. A manufacturer of paper used for making grocery bags is interested in improving the product's tensile strength. Product engineering believes that tensile strength is a function of the hardwood concentration in the pulp and that the range of hardwood concentrations of practical interest is between 5 and 20%. A team of engineers responsible for the study decides to investigate four levels of hardwood concentration: 5%, 10%, 15%, and 20%. They decide to make up six test specimens at each concentration level by using a pilot plant. All 24 specimens are tested on a laboratory tensile tester in random order. The data from this experiment are shown below.

| **Hardwood** | **Observations** | | | | | |
|---|---|---|---|---|---|---|
| **Concentration (%)** | **1** | **2** | **3** | **4** | **5** | **6** |
| **5** | 7 | 8 | 15 | 11 | 9 | 10 |
| **10** | 12 | 17 | 13 | 18 | 19 | 15 |
| **15** | 14 | 18 | 19 | 17 | 16 | 18 |
| **20** | 19 | 25 | 22 | 23 | 18 | 20 |

Let us denote the mean tensile strength at each concentration as $\mu_{conc}$, e.g., $\mu_5$. We wish to use an ANOVA analysis to determine if there are any substantive differences in tensile strength for different concentrations of hardwood.

i. Construct the null hypothesis and alternative hypothesis for the test.
ii. Calculate the mean square variability within and between samples.
iii. Perform a hypothesis test at the 0.05 level and determine any significant differences in tensile strength with hardwood concentration. In particular, identify the specific pairs of concentrations with different tensile strengths using Tukey's HSD test.