| LAB 2 ASSIGNMENT |
|---|

# PROBABILITY  DISTRIBUTIONS

In this lab assignment you will use numerical and graphical tools available in Excel to examine the process of manufacturing of ultra-strong alloys used in fasteners, springs and instrument parts. You will use the process to explore the basic properties of binomial and normal distributions through computer experiment and simulation. In particular, you will verify the process target parameters and the process distribution assumptions.

We have used some functions available in Excel to develop three instructional templates (interactive worksheets) that allow you to review the properties of binomial and normal distributions. Some cells and charts in the interactive worksheets are protected and are therefore read only.

## Tensile Strength of Alloys

Carpenter MP35N alloy is a nonmagnetic, nickel-cobalt-chromium-molybdenum alloy possessing a unique combination of ultrahigh tensile strength and excellent corrosion resistance. In addition, this alloy displays exceptional resistance to high temperature oxidation. Because of its unique combination of properties, MP35N alloy has been used in a wide variety of applications including fasteners, springs, nonmagnetic electrical components and instrument parts in medical, seawater, and chemical and food processing environments. The alloys are delivered to customers in the form of rectangular in cross-section slabs.

According to the manufacturing specifications, the tensile strength of MP35N is supposed to follow a normal distribution with a mean of 285 ksi and a standard deviation of 5 ksi. The critical values for the tensile strength are 275 on the low side and 295 on the high side. The alloy slabs with the tensile strength values below 275 ksi or above 295 ksi are considered to be unacceptable. Too low tensile strength makes the alloys not strong enough for intended applications, whereas too high tensile strength makes them loose some desirable physical properties.

In this lab assignment you will examine the tensile strength distribution with Excel. The lab assignment is based on the Excel file *lab2.xls* located at http://www.stat.ualberta.ca/statslabs/index.htm (click *Stat 235* link, and *Data* for *Lab 2*). The file consists of four worksheets: *Normal Density*, *Normal Probabilities*, *Binomial Probabilities,* and *Data*. The first three are instructional templates; the fourth one contains data related to the above problem.

1.      Open the worksheet *Normal Density*. This worksheet contains a graph of the density function for the tensile strength (TS). The graph shares your screen with the values of μ and σ. This will enable you to change those values and immediately see the change in the density function without scrolling your screen or changing worksheets. Use the interactive graph to answer the following questions:

   (a)      Assume that the mean  of TS is set at 285.  Enter the value of σ as 4, then 5, and eventually 6. After each entry, carefully examine the shape of the corresponding density curve. You are not supposed to print the density curves. Describe briefly the change in the appearance of the density curve as σ increases from 4 to 6. What does the increase in the standard deviation value mean to the alloy strength in the  manufacturing process?

   (b)      Now assume that the standard deviation σ=5 and the mean TS changes. Enter the value of μ as 280, then  285, and finally 290.

            Describe briefly how the shape of the corresponding density curve changes. How does the change in the value of μ affects the strength of the alloy? Does the fraction of alloy slabs with unacceptable TS (below 275 or above 295 ksi) increase or decrease? Explain briefly. Do not print the density curves.

2.      In this part you will determine how the changes in the mean and standard deviation affect the fraction of alloy slabs that do not meet the TS specifications. Assume in this question that TS follows a normal distribution. Answer the questions below using either the template *Normal Probabilities* or the appropriate built-in functions in Excel.

   (a)      Suppose $\mu=285$ and $\sigma=5$. What fraction of alloy slabs will have unacceptable tensile strength (TS below 275 or above 295 ksi)? What fraction would have unacceptable TS when the value of $\mu$ were 283 with the same value of $\sigma$? You do not have to print out the values displayed on the worksheet, just copy the numerical values from the screen to your report.

   (b)      Now suppose that the mean TS is 285 with the value of $\sigma$ equal to 6. What fraction of alloy slabs will then be unacceptable?

   (c)      What is the fraction of alloy slabs with the TS value within 1 standard deviation of the mean of 285? What is the fraction of slabs with tensile strength within 2 standard deviations of the mean? Assume here that $\sigma = 5$ ksi.

   (d)      Assume $\mu=285$ and $\sigma=5$. What is the TS value exceeded by 95% of slabs? What would be the TS value exceeded by 99 % of slabs?

   (e)      One possible way to reduce the fraction of unacceptable alloy slabs is to reduce the process standard deviation. What must the standard deviation $\sigma$ be in order that only 1% of slabs have the TS values less than 275 or larger than 295? Assume $\mu = 285$. Notice that the template or appropriate statistical function in Excel is not able to provide the answer to the above question directly. You will have to vary the standard deviation until you reach the desired percentile. Report the approximate value of $\sigma$ with two digits after the point.

3.      Suppose that the manufacturing process is operating properly and tensile strength follows a normal distribution with a mean $\mu=285$ ksi and a standard deviation $\sigma=5$ ksi. You will simulate the TS measurements with a random number generator in Excel.

   You must use Excel in Windows environment to obtain the data (different data may be produced by *Random Number Generation* tool in Excel on MAC OS). Excel 2010 or higher version should be used in this part (earlier versions may produce different sequence of random numbers for the same seed).

   Use the *Random Number Generation* feature (seed 200) to generate 1 sample of size 200 from the normal distribution with the mean of 285 and standard deviation of 5 in a new worksheet. This corresponds to selecting randomly 200 slabs and obtaining a TS measurement for each of them. You may sort the data in ascending order to answer parts (a)-(c).

   (a)      Count the number of unacceptable slabs (TS below 275 or above 295) in the sample. Is the observed number of unacceptable slabs consistent with the theoretical prediction obtained in Question 2, part (a)?

   (b)      Determine the number of observations within 1, 2, and 3 standard deviations of the mean for the sample. Compare your results with the 68-95-99.7 Rule (see *Lab 2 Instructions*).

   (c)      Obtain the values of the standardized variable $Z=(TS-\mu)/\sigma$ for each observation in the sample of 200 CDS observations. Do not print the values in your report. What is the distribution of the standardized values? Refer to the theory to answer the question.

   (d)      Use the *Descriptive Statistics* to calculate the mean and standard deviation of the standardized observations obtained in (d). Report the two values with four digits after the decimal point. Are the values consistent with the distribution parameters provided by the theory? Explain.

4.    The manufacturer has recently implemented some changes in the manufacturing process that are expected to make the process more cost effective without compromising the process parameters. In order to evaluate the effect of the changes on the TS distribution, a random sample was taken consisting of 200 slabs, and the TS measurements were recorded in the worksheet *Data*. The data are not to be printed in your submission.

(a)    Use the *Descriptive Statistics* tool to calculate the mean and standard deviation of the 200 TS measurements. Are the values consistent with the target parameters (mean $\mu=285$ ksi and a standard deviation $\sigma=5$ ksi)? Comment briefly.

(b)    Obtain the relative frequency histogram for the 200 observations. Use the class intervals of width 5 with bins beginning at 270 and going to 300. The format of your histogram should be the same as the format of the sample histogram in the *Lab 1 Instructions* (labels, title, no gaps between bars). Paste the histogram into your report. Describe the pattern displayed in the plot. Is there any strong indication that the data do not follow a normal distribution? Explain briefly.

(c)    Assuming normality and the population parameters close to the values produced by the sample (large sample size), would the manufacturer be able to achieve the goal of not exceeding 5% limit on the proportion of unacceptable slabs (tensile strength less than 275 or more than 295 ksi)? Use the *Normal Probabilities* template to answer the question.

(d)    Use your findings in parts (a)-(b) to prepare a brief report to management that indicates whether or not the changes have changed the tensile strength. In particular, comment on the changes in the process parameters (mean and standard deviation) and normality of the TS distribution.

5.    In this question you will use the worksheet *Binomial* in *lab2.xls* file or the built-in functions in Excel for binomial and normal distributions. You do not have to print anything in this part. Just copy down the numerical values provided by the function or the template into your report.

(a)    The manufacturer's goal is not to exceed 5% limit on the proportion of unacceptable slabs (tensile strength below 275 or above 295 ksi). If the requirement is not met, the production process is discontinued while adjustments are made. One recognized way of monitoring production is to take a random sample from each hour's production and to determine the fraction of unacceptable slabs. The manufacturer developed a criterion according to which the manufacturing process is discontinued for adjustments when a random sample of 200 slabs results in more than 15 unacceptable slabs.

How likely is shutting down the manufacturing process erroneously based on the sample data? In other words, what is the probability of obtaining more than 15 unacceptable slabs in a random sample of 200 provided the true fraction of unacceptable slabs does not exceed 0.05?

(b)    The manufacturer claims that the target mean TS is 285 ksi with the value of $\sigma$ equal to 5 ksi. Assuming that this is true, calculate the probability that more than 15 unacceptable slabs (TS below 275 or above 295) will be obtained in a random sample of 200 slabs.

In both parts (a) and (b) give the parameters of the binomial distributions used and report the corresponding probabilities.

# LAB 2 ASSIGNMENT MARKING SCHEMA

Proper cover page (includes your name, with the surname capitalized, course section and lab assignment number) and appearance (i.a. lab report must be **typed**): 10 marks

**Question 1**

(a)     Description of the change in the density function as σ increases in the context of the process: 2 points
        Effect of the change in standard deviation on the strength of alloys: 2 points
(b)     Description of the change in the density function as μ increases: 2 points
        Effect of the change in the mean on the fraction of unacceptable alloys: 2 points.

**Question 2**

| PART | PARAMETERS | MARKS |
|------|------------|-------|
| (a) | μ=285 and σ=5 | 2 |
|     | μ=283 and σ=5 | 2 |
| (b) | μ=285 and σ=6 | 2 |
| (c) | μ=285 and σ=4 1, 2 St. deviation of μ | 2 |
| (d) | TS exceeded by 95% | 2 |
|     | TS exceeded by 99% | 2 |
| (e) | Value of σ | 3 |

**Question 3**

(a)     Number of unacceptable slabs: 2 points
        Consistency with the theoretical prediction: 2 points
(b)     Count of observations within 1-3 St. deviations: 2 points each (6 points total)
        Discussion of consistency with the 68-95-99.7 rule: 2 points
(c)     Distribution of the standardized values: 2 points
(d)     Mean and standard deviation of the standardized values: 2 points each (4 points total)
        Consistency with the theory: 2 points

**Question 4**

(a)     Mean and standard deviation of the 200 observations: 2 points each (4 points total)
        Consistency with the target parameters: 2 points
(b)     Histogram: 6 points
        Pattern and the normality assumption assessment: 2 points
(c)     Estimated proportion of unacceptable slabs: 2 points
(d)     Brief report: 3 points

**Question 5**

(a)     Parameters of the binomial distribution: 2 points
        Binomial probability: 3 points
(b)     Parameters of the binomial distribution: 3 points (the value of p calculated with the *Normal Probabilities* worksheet)
        Binomial probability: 3 points

# TOTAL SCORE = 83