# Attendance Workbook For Statistics 513

# Statistical Control Theory

# Spring 2004

by

Jonathan Kuhn, Ph.D.
Associate Professor of Statistics,
Mathematics, Statistics and Physics,
Purdue University North Central

# Preface

This is a *second* course in statistics. The aim of this course is acquaint a student with statistical control theory. Matrix algebra, calculus and numerical computation are used. The statistical software package called SAS is used.

This workbook is a necessary component for a student to successfully complete this course. Without the workbook, a student will not be able to participate in the course.

- This attendance workbook is *based* on the text.
- Although the material covered in each is very similar, the *presentation* of the material in the workbook is quite different from the presentation given in the text. The text consists essentially of definitions, formulas, worked out examples and exercises; this workbook, on the other hand, consists *solely* of exercises to be worked out by the student.
- The overheads presented during each lecture are based *exclusively* on the workbook. A student is to use this workbook to follow along with during a lecture.
- There are different kinds of exercises, including multiple choice, true/false, matching and fill–in–the–blank.
- Each week, I recommend you read the text, answer the questions given here in the attendance workbook and then do either the quiz or homework assignment, in that order.

On the one hand, the workbook is, as you will see, quite a bit more elaborate than typical lecture notes, which are usually a summary of what the instructor finds important in a recommended course text. On the other hand, this workbook is not quite a text, because although it has many exercises, it does not have quite enough exercises to qualify it as a complete text. I should also point out that this workbook, unfortunately, possesses a number of typographical errors. In short, this workbook aspires to be text and, in the next few years, when enough exercises have been collected, and when most of the typographical errors have been weeded out, it will become a text.

Dr. Jonathan Kuhn,
Associate Professor of Statistics,
Mathematics, Statistics and Physics,
Purdue University North Central,
May 2004

# Chapter 1

# Quality Improvement in the Modern Business Environment

We look at what "quality" means.

## 1.1 The Meaning of Quality and Quality Improvement

We look at the definitions associated with quality and quality improvement. We find out quality is about reducing variability. This means reduced waste and, thus, reduced costs for a company.

**Exercise 1.1 (The Meaning of Quality and Quality Improvement)**

1. *Quality.*
   **True / False**
   There is a traditional definition and a modern definition of *quality*. The traditional definition is "fitness of use". The modern definition is

   Quality is inversely proportional to variability.

   The modern definition is more specific than the traditional definition; the traditional definition tended to be associated with "quality of conformance", whether or not the product was actually fit–for–use by the customer.

2. *Quality Improvement.*
   **True / False**
   Quality improvement is the reduction of variability in processes and products.

3. *Dimensions of Quality.*
   Match the columns.

| | |
|---|---|
| A performance | a Is the product made exactly as the designed intended? |
| B reliability | b What is the reputation of the company or its product? |
| C durability | c What does the product do? |
| D serviceability | d What does the product look like? |
| E aesthetics | e How easy is it to repair the product? |
| F features | f How long does the product last? |
| G perceived quality | g How often does the product fail? |
| H conformance to standards | h Will the product do the intended job? |

| A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|
| h | | | | | | | |

4. *Quality Characteristics.*
   Match the columns.

| | |
|---|---|
| A physical | a length, weight, voltage, viscosity |
| B sensory | b reliability, durability, serviceability |
| C time orientation | c taste, appearance, color |

| A | B | C |
|---|---|---|
| | | |

5. *Data types, statistical methods.*
   Match the columns.

| | |
|---|---|
| A attributes | a counts, discrete data |
| B variables | b continuous data |

| A | B |
|---|---|
| | |

6. *Specifications.*
   **True / False**
   The *specifications* are the desired (nominal, target) measurements for quality characteristics. The largest allowable value of a quality characteristic is the *upper specification limit (USL)* and smallest allowable value is the *lower specification limit (LSL)*.

7. *Noncomforming Versus Defective.*
   **True / False**
   A nonconforming product is one which fails to mean specifications. A defective product is a product that has nonconformities so serious as to affect the sale or effective use of the product.

## 1.2 A Brief History of Quality Control and Improvement

This is an interesting section that describes the development of quality control, from Taylor in 1875 through to the present day.

## 1.3 Statistical Methods for Quality Control and Improvement

In this course we will look at three statistical methods used in quality control: statistical process control, design of experiments and (to a lesser extent) acceptance sampling.

**Exercise 1.2 (Statistical Methods for Quality Control and Improvement)**

1. *Production process.*
   **True / False**
   An *on–line* production process can be described by the following flow chart.

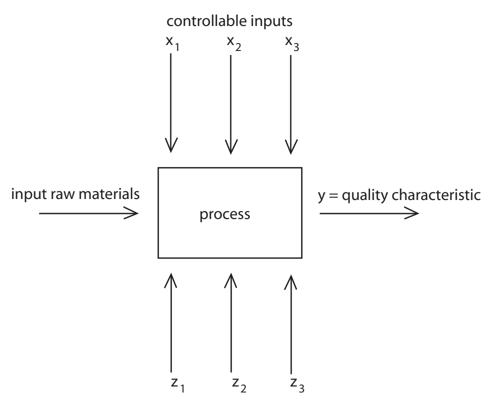



Figure 1.1 (Production process inputs and outputs)

2. *Statistical Process Control*
   **True / False**
   An example of a control chart is given below.

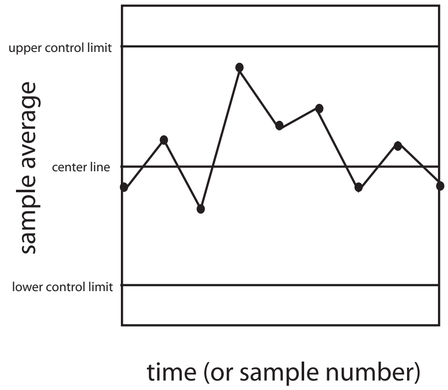



Figure 1.2 (Control chart)

Control charts are used to monitor a production process to make sure the quality characteristics fall within specifications, to make sure the variability in the process is acceptable.

3. *Design of experiments.*
   A designed experiment is used to discover which of a number of key variables/attributes influences the quality characteristics. Since a designed experiment is done before the production process is set up, it is called an (choose one) **on–line** / **in–process** / **off–line** quality control tool.

4. *Acceptance sampling.*
   Acceptance sampling is defined as the inspection of a sample of units selected at random from a larger batch (or lot) and the ultimate decision as to either accept or reject the lot. If the lot is rejected, the items may be (choose none, one or more) **scrapped** / **recycled** / **reworked** / **replaced**. Acceptance sampling usually takes place at the end of the production or just before arriving to a customer or both.

## 1.4 Other Aspects of Quality Control and Improvement

This section emphasizes that quality control is only effective if management in addition to the production people accepts and implements quality control.

**Exercise 1.3 (Other Aspects of Quality Control and Improvement)**

1. *Six–Sigma: Part diameters.*

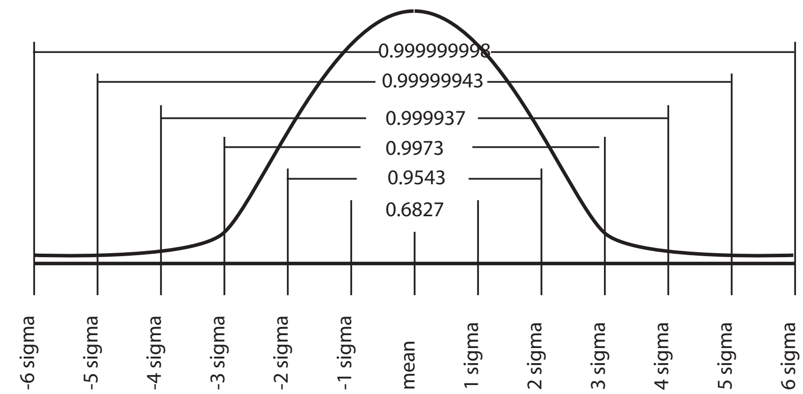



Figure 1.3 (Standard deviation)

(a) *One Interpretation of six–sigma*
A part has a specified average diameter of 50 mm and a standard deviation in diameter of 1 mm. If the diameter follows a normal distribution, then, as shown in the figure, about 68% of the part diameters fall in the range $(50-1, 50+1) = (49, 51)$ mm. In other words, 68% fall within 1 standard deviation (or one "sigma") of the average diameter of the part.
Also, 99.9999998% fall within (choose one) **five** / **six** / **seven** sigma of the average diameter of the part. That is, 99.9999998% fall within the range $(44, 56)$ mm. In this case, as the sigma increases, the range in the variability of the part diameter increases.

(b) *A statistical quality control interpretation of six–sigma.*
A part has a specified average diameter of 50 mm and it is *required that the range in the variability in the part diameter remain fixed at 1 (one) mm.* One–sigma implies that 68% of the part diameters fall in the range $(50-1, 50+1) = (49, 51)$ mm and six–sigma implies that 99.9999998% of the part diameters fall in the range (choose one) $\mathbf{49, 51}$ / $\mathbf{44, 56}$ / $\mathbf{40, 60}$. In this case, as sigma increases, the range in variability of the part diameter remains constant; that is, as sigma increases, a greater percentage of the variability of the diameters fall in the same range (the product is less variable).

(c) *Implication of six–sigma.*
On the one hand, if 100 (independent) parts of a product have three–sigma quality performance, where each part of this product must be conforming for the product to work, then the part will conform with probability

$$0.9973^{100} = 0.763$$

On the other hand, if there is six–sigma quality performance, then the part will conform with probability (choose one) **0.899** / **0.998** / **0.9999998**

2. *Pareto Principle: Part diameters.*
   **True** / **False**
   The Pareto principle involves solving the problems with the largest costs first.

# Part I

# Statistical Methods Useful in Quality Improvement

# Chapter 2

# Modeling Process Quality

We quickly look at some basic results in descriptive statistics and probability, that will be used throughout the course.

## 2.1 Describing Variation

We look at stem and leaf plots, frequency distributions, the box plot and the average, variance and standard deviation.

**Exercise 2.1 (Stem and Leaf Plot)**

1. *Age Data.*
   The ages of the twenty patients in a high blood pressure study are given by,

   32, 37, 39, 40, 41, 41, 41, 42, 42, 43,
   44, 45, 45, 45, 46, 47, 47, 49, 50, 51

   A stem and leaf plot of this data is given as follows,

   | | | | | | | | | | | | | | | | | |
   |---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
   | 3 | 2 | 7 | $9^*$ | | | | | | | | | | | | | |
   | 4 | 0 | 1 | 1 | $1^{**}$ | 2 | 2 | 3 | 4 | 5 | 5 | 5 | 6 | 7 | 7 | 9 | stem: 10s |
   | 5 | 0 | | | | | | | | | | | | | | | leaf: 1s |

   Table 2.1 (Stem–and–Leaf For Age Data)

   (a) The starred number, $\mathbf{9^*}$, represents the age (circle one) **39** / **93** / **9**. The double–starred number, $1^{**}$, represents the age (circle one) **41** / **14** / **1**.

   (b) All of the numbers to the left of double line (in the first column) are called (circle one) **stems** / **leaves**; all of the numbers to the right of this double line are called (circle one) **stems** / **leaves**.

(c) The starred number 9* is a leaf of the stem (circle one) **3** / **4** / **5**.

(d) **True** / **False** The note to the right of the stem–and–leaf plot specifies the numbers used as stems are "tens" (or "10s") and the numbers used as leaves are "ones" (or "1s"). So, for instance, the stem "3" represents 3x10 = 30 and the leaf "2" represents 1x2 = 2.

(e) The stem–and–leaf plot is ordered where, in the first stem, for example, 32 is followed by (circle one) **37** / **39** / **40**.

(f) A stem–and–leaf plot is useful in identifying the center of the data set, or, where "most" data values are located. In this case, this is in the (circle one) **30**s / **40**s / **50**s stem.

2. *Split Stem and Leaf Plots: Age Data.*
The ages of the twenty patients in a high blood pressure study are given by,

32, 37, 39, 40, 41, 41, 41, 42, 42, 43,
44, 45, 45, 45, 46, 47, 47, 49, 50, 51

Sometimes, to spread the data out, the stems are *split* as, for example, in the following table.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 3L | 2 | | | | | | | | | |
| 3H | 7 | 9 | | | | | | | | |
| 4L | 0 | 1 | 1 | 1 | 1 | 2 | 2 | 3 | 4 | |
| 4H | 5 | 5 | 5 | 6 | 7 | 7 | 7 | 9 | | |
| 5L | 0 | | | | | | | | | stem: 10s |
| 5H | | | | | | | | | | leaf: 1s |

Table 2.2 (Stem and Leaf For Age Data With Split Stems)

(a) Notice how all stems have each been split into two.
Stem 3 is split into low stem 3, or (circle one) **3L** / **3H**,
and high stem 3, or (circle one) **3L** / **3H**.

(b) **True** / **False** Low stem 3L contains one half of the leaves, 0, 1, 2, 3 or 4; high stem 3H contains the other half of the leaves, 5, 6, 7, 8 and 9.

(c) **True** / **False** Stem and leaf plots can have stems split not only twice, but also three or more times. A stem–and-leaf plot with each stem split three times might, say, consist of a low stem which contains leaves 0, 1 and 2; a middle stem with leaves 3, 4, 5 and 6 and a high stem with leaves 7, 8 and 9.

(d) A stem and leaf plot with 10s as stems can be split at most
(circle one) **5** / **7** / **10** / **100** times.

(e) **True** / **False** Although there is no one "best" way of constructing a stem–and–leaf plot, most stem–and–leaf plots consist of 5 to 20 stems.

**Exercise 2.2 (Frequency Distribution Table and Histogram)**

1. *A First Look: Age.*
   Twenty patients in a high blood pressure study have the following ages.

   32, 37, 39, 40, 41, 41, 41, 42, 42, 43,
   44, 45, 45, 45, 46, 47, 47, 49, 50, 51

   (a) The frequency (number) of patients between the ages of 35 and up but not including 40 (class interval 35 to 40) is (circle one) **2** / **3** / **4** / **5**.

   (b) The *width* of class interval 35 to 40 is (circle one) **2** / **3** / **4** / **5** years.

   (c) The *width* of class interval 40 to 45 is $45 - 40 =$ (circle one) **2** / **3** / **4** / **5** years.

   (d) The *width* of class interval 41 to 44 is $44 - 41 =$ (circle one) **2** / **3** / **4** / **5** years.

   (e) The *proportion of patients in the five years* in the class interval 40 to 45 is (circle one) $\mathbf{\frac{6}{20} = 0.30}$ / $\mathbf{\frac{7}{20} = 0.35}$ / $\mathbf{\frac{8}{20} = 0.40}$ / $\mathbf{\frac{9}{20} = 0.45}$.

   (f) The *proportion of patients in the five years* in the class interval 50 to 55 is (circle one) $\mathbf{\frac{1}{20} = 0.05}$ / $\mathbf{\frac{2}{20} = 0.10}$ / $\mathbf{\frac{3}{20} = 0.15}$ / $\mathbf{\frac{4}{20} = 0.20}$.

   (g) **True** / **False** If the proportion of patients in a class interval of width *five* years is 0.35, then the proportion of patients per *one* year in this class interval is $\frac{0.35}{5} = 0.07$.

   (h) **True** / **False** If the proportion of patients in a class interval of width *ten* years is 0.35, then the proportion of patients per one year in this class interval is $\frac{0.35}{10} = 0.035$.

   (i) If the proportion of patients in a class interval of width *seven* years is 0.35, then the proportion of patients per one year in this class interval is (circle one) $\mathbf{0.35 \times 7 = 2.45}$ / $\mathbf{\frac{7}{0.35} = 20}$ / $\mathbf{\frac{0.35}{7} = 0.05}$ / $\mathbf{\frac{0.35}{10} = 0.035}$.

   (j) **True** / **False** If the *proportion* of patients per one year in some class interval is 0.07, then the *percentage* of patients per one year $0.07 \times 100 = 7$ percent.

2. *Frequency Distribution Table.*
   Consider the following incomplete distribution table for the age data,

   32, 37, 39, 40, 41, 41, 41, 42, 42, 43,
   44, 45, 45, 45, 46, 47, 47, 49, 50, 51

| class interval | frequency | relative frequency | proportion per 5 years | proportion per 1 year | % per 1 year |
|---|---|---|---|---|---|
| 30 to 35 | 1 | $\frac{1}{20} = 0.05$ | $\frac{1}{20} = 0.05$ | $\frac{0.05}{5} = 0.01$ | 1 |
| 35 to 40 | 2 | $\frac{2}{20} = 0.10$ | (c) | $\frac{0.10}{5} = 0.02$ | 2 |
| 40 to 45 | 8 | (a) | $\frac{8}{20} = 0.40$ | $\frac{0.40}{5} = 0.08$ | 8 |
| 45 to 50 | 7 | $\frac{7}{20} = 0.35$ | $\frac{7}{20} = 0.35$ | $\frac{0.35}{5} = 0.07$ | 7 |
| 50 to 55 | 2 | (b) | (d) | (e) | 2 |
| total | 20 | 1.0 | 1.0 | | |

(a) Complete the distribution table by filling in the following table.

| (a) | (b) | (c) | (d) | (e) |
|---|---|---|---|---|
| | | | | |

(b) The first class interval is (circle one) **30 to 35** / **30 to 40** / **40 to 45**.

(c) The number of class intervals is (circle one) **3** / **4** / **5** / **6**.

(d) The width of each class interval is (circle one) **3** / **4** / **5** / **6** years.

(e) **True** / **False** The class intervals given here are the *only* possible class intervals that could have been used for this data. For example, it would not be possible to have, instead, class intervals of *un*equal width, such as "30 to 40", "40 to 45" and "45 to 55", say.

3. *Histogram.*
Use your calculators to help draw the two possible graphs given below for the age data. Notice that both the "relative frequency" graph and the "proportion per 1 year" graph have the same *shape*.

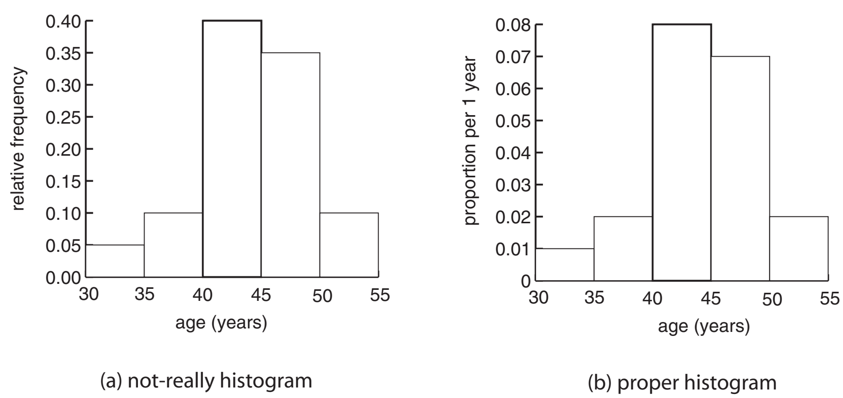



Figure 2.1 (Two Possible Graphs For The Age Data)

(STAT EDIT 32.5, 37.5, 42.5, 47.5, 52.5 in $L_1$ and 0.01, 0.02, 0.08, 0.07, 0.0-2 into $L_2$ and then 2nd Y= ON, choose histogram, then WINDOW 30, 55, 5, -0.05, 0.1, 0.1, 1 GRAPH to get the "final" histogram. Type TRACE to see the frequency in each class.)

(a) The *total* area of all the vertical bars in the "not–really" histogram (circle one) **is less than one** / **equal to one** / **greater than one**.
Hint: $(35 - 30) \times 0.05 + \cdots + (55 - 50) \times 0.10 = 5$

(b) The *total* area of all the vertical bars in the "proper" histogram is (circle one) **less than one** / **equal to one** / **greater than one**.
Hint: $(35 - 30) \times 0.01 + \cdots + (55 - 50) \times 0.02 = 1$

(c) The proportion of ages in the 30 to 40 class interval is the proportion in the 30 to 35 class interval plus the proportion in the 35 to 40 class interval, or, in other words, equal to the total area in the two vertical bars in these two classes, or
$(5 \times 0.01) + (5 \times 0.02) =$ (circle one) **0.05** / **0.10** / **0.15**.

(d) The proportion of ages in the 35 to 50 class interval is equal to the total area in the vertical bars for the three class intervals, 35 to 40, 40 to 45 and 45 to 50, or
$(5 \times 0.02) + (5 \times 0.08) + (5 \times 0.07) =$ (circle one) **0.80** / **0.85** / **0.90**.

(e) The proportion of ages in the 35 to 37 portion of the 35 to 40 class interval is equal to the area of the portion of the vertical bar associated with 35 to 37. Since the width of the 35 to 37 portion is $37 - 35 = 2$, and the height of the 35 to 40 class interval is 0.02, the area must be
$(2 \times 0.02) =$ (circle one) **0.03** / **0.04** / **0.05**.

4. *How Do We Use Histograms?*
We use histograms to (choose none, one or more)

(a) identify the shape of the data.

(b) determine the location of the center of data.

(c) determine the scatter in the data.

**Exercise 3.3 (Sample Mean, Median, Standard Deviation, Five Number Summary and Box–Plots)**

1. *Sample Mean and Median: Temperatures.*
Consider the following ordered sample of temperatures taken from 10 different locations in Westville during the second day in January of a recent year:

$$0, 0, 0, 1, 1, 2, 2, 3, 3, 4.$$

(a) The average is, since there are $n = 10$ temperatures and

$$x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 1, x_5 = 1, x_6 = 2, x_7 = 2, x_8 = 3, x_9 = 3, x_{10} = 4$$

Then:

$$\begin{aligned}\bar{x} &= \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10}}{n} \\ &= \frac{\sum_{i=1}^{10} x_i}{n}\end{aligned}$$

and so $\bar{x} = \frac{0+0+0+1+1+2+2+3+3+4}{10} =$ (circle one) **1.0** / **1.6** / **1.7**.
(Use your calculator: STAT ENTER; then type the ten temperatures, 0,0,...,4 into $L_1$; then calculate average ($\bar{x}$) by typing STAT CALC ENTER 2nd $L_1$ ENTER; read "$\bar{x} = 1.6$".)

(b) **True** / **False** The mean (or average) calculated for a *sample* is called a *statistic*; the mean for a *population* is called a *parameter*. The ten temperatures here could be considered a sample of all the possible locations in Westville and so, in this case, the average, 1.6 degrees, is the value of a statistic and not a parameter.

(c) The *median*, $\tilde{x}$, is the center value of the 10 ordered set of temperatures. Since there are an *even* number of observations, $n = 10$, in this case, there is *no* center value, or, in other words, no observation where there is the same number of observations both above and below this value. Consequently, the sample median is set equal to the average of the center *two* observations (circle one) **1.5** / **1.6** / **1.7**.
(Use your calculator: as above, but, now, arrow down to "Med = 1.5".)

(d) If there had been an *odd* number of observations, then the center value would have been used as the sample median. For example, if the data set had consisted of the *nine* values,

$$0, 0, 0, 1, 1, 2, 2, 3, 3,$$

then the sample median would have been
(circle one) **1.0** / **1.6** / **1.7**.

(e) **True** / **False** The *location* of the median can be calculated using the $\frac{n+1}{2}$ formula. For example, if there are 9 observations, then the $\frac{9+1}{2} =$ 5th ordered observation is the median. If there are 10 observations, then the $\frac{10+1}{2} = 5.5$th ordered observation (average of 5th and 6th ordered observations) is used as the median value.

2. *Average Sensitive to Outliers; Median Robust to Outliers: Temperatures.* Consider the following ordered sample of temperatures taken from 10 different locations in Minneapolis during a cold day in January where, because of a typing mistake, the last temperature, 4 degrees, is mistakenly recorded as 40 degrees:

0, 0, 0, 1, 1, 2, 2, 3, 3, 40.

(a) The average is (circle one) **1.5** / **1.6** / **5.2**.

(b) The median is (circle one) **1.5** / **1.6** / **5.2**.

(c) The present average is
(circle one) **much bigger** / **about the same** / **much smaller** than the average calculated for the ten temperatures, 0, 0, 0, 1, 1, 2, 2, 3, 3, 4.

(d) The present *median* is
(circle one) **much bigger** / **about the same** / **much smaller** than the median calculated for the ten temperatures, 0, 0, 0, 1, 1, 2, 2, 3, 3, 4.

(e) The average is said to be (circle one) **sensitive** / **robust** to outliers, whereas the median is said to be (circle one) **sensitive** / **robust** to outliers.

3. *Standard Deviation.*
Consider the algae weights (in ounces) given in the following three shipments of six containers each:

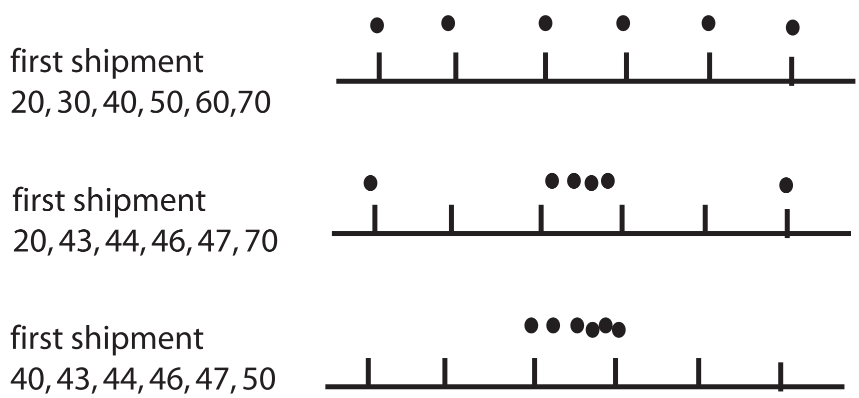



Figure 2.2 (Three Shipments of Six Containers of Algae Data)

It is often useful to have a *single* number which summarizes the fact that, for example, the weights for the first shipment are more variable (more "spread out") than the weights for the third shipment.

(a) It is fairly clear from the three diagrams above that the tire weights for the first shipment are (circle one) **more** / **less** spread out (or variable) than the tire weights for the third shipment.

(b) If a single number, called the *standard deviation*, is large when a data set was spread out and small when a data set is jammed together, then the standard deviation is largest for shipment (circle one) **one** / **two** / **three**.

(c) *Standard Deviation Using TI–83.* The *standard deviation*, denoted $S$, of the first shipment is given by:

$$
\begin{aligned}
S &= \sqrt{\frac{\sum_{i=1}^{6}(x_i - \bar{x})^2}{n-1}} \\
&= \sqrt{\frac{(20-45)^2 + (30-45)^2 + \cdots + (70-45)^2}{6-1}} \\
&=
\end{aligned}
$$

(circle one) **3.46** / **15.87** / **18.71**.
(Type STAT ENTER; then type the six weights into $L_1$; then STAT CALC ENTER 2nd $L_1$ ENTER; then read $s_x = 3.46...$)
The sample standard deviation for the second shipment of weights is (circle one) **3.46** / **15.87** / **18.71**.
and for the third shipment is
(circle one) **3.46** / **15.87** / **18.71**.

(d) Since $S$ for the first shipment is
(circle one) **larger than** / **about the same as** / **smaller than** the $S$ for the third shipment, this means the variability in weights for the first shipment is
(circle one) **larger than** / **about the same as** / **smaller than** the variability in tire weights for the third shipment.

(e) The formula for the *variance* is given by

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

and so the standard deviation is the (circle one) **square** / **square root** of the variance.

(f) *Review.* The three SDs in the algae weights for the three shipments are probably all examples of (circle one) **populations** / **samples** / **statistics** / **parameters**. The three true (or actual) SDs in weights for all of the algae in the three shipments are all examples of (circle one) **populations** / **samples** / **statistics** / **parameters**.

4. *Percentiles: Temperatures.*
Reconsider the following sample of 10 temperatures,

$$0, 0, 0, 1, 1, 2, 2, 3, 3, 4.$$

(a) The 50th percentile is that temperature such that 50% of the data is *below* this temperature. If 50% of the data is below this temperature, then 50%

must be above this temperature. In other words, the 50th percentile must be the
(circle one) **median** / **average**.

(b) Assume the 50th percentile is the median. The *location* of the median can be calculated by adding one to the total number of data points, $n$, and dividing by 2, the $\frac{n+1}{2}$ formula. The *location* of the median for the $n = 10$ temperatures is
(circle one) **5** / **5.5** / **6**.
And so the median is equal to
(circle one) **0** / **1** / **2**.

(c) The 25th percentile is that temperature such that 25% of the data is below this temperature and 75% is above this temperature. The 75th percentile must be
(circle one) **above** / **equal to** / **below** the 50th percentile.

(d) The 65th percentile is that temperature such that what percentage of all the temperatures is below this temperature?
(circle one) **55%** / **65%** / **70%**.

(e) The location formula for the 50th percentile, $\frac{n+1}{2}$, could be rewritten as $\frac{1}{2}(n+1)$ or interpreted as "50% of $(n+1)$". Not surprisingly, then, the location formula for the 25th percentile could be interpreted as "25% of $(n+1)$" or $\frac{1}{4}(n+1)$. Consequently, the *location* of the 25th percentile for the $n = 10$ temperatures is
(circle one) **1.25** / **2.75** / **3.50**.
And so the 25th percentile is equal to
(circle one) **0** / **1** / **2**.

(f) The location formula for the 65th percentile could be interpreted as "65% of $(n+1)$" or $0.65 \times (n+1)$. Consequently, the *location* of the 65th percentile for the $n = 10$ temperatures is
(circle one) **5.25** / **5.55** / **7.15**.
And so the 65th percentile is equal to
(circle one) **2.5** / **2.75** / **3.25**.

(g) The 25th percentile has a special name called the lower (or first) quartile and denoted $Q1$. Similarly, the 50th percentile is often called
(circle one or more) **middle** / **second** / **upper** / **third** quartile.
Finally, the 75th percentile is often called the
(circle one or more) **middle** / **second** / **upper** / **third** quartile and denoted $Q3$.

5. *Five Number Summary: Temperatures.*
Reconsider the following sample of 10 temperatures,

$$0, 0, 0, 1, 1, 2, 2, 3, 3, 4.$$

(a) The *five–number summary* is given by,

minimum, lower quartile, median, upper quartile, maximum,

and so, for the 10 temperatures, the five number summary (circle one) **is** / **is not** $\{0, 0, 1.5, 3, 4\}$. (Use your calculator: STAT ENTER; type 0, 0, ..., 4, into $L_1$; then STAT CALC ENTER 2nd $L_1$ ENTER; then arrow down to read minX = 0, $Q_1 = 0$, Med = 1.5, $Q_3 = 3$, maxX = 4.)

(b) The *interquartile range* is equal to the upper quartile minus the lower quartile. Consequently, the interquartile range for the set of 10 temperatures is given by, $3 - 0 =$ (circle one) **1** / **2** / **3**.

(c) The interquartile range is robust to outliers whereas the variance and standard deviation are both sensitive to outliers. This means, for example, if one of the 10 temperatures above was mistyped as 40, instead of 4, the interquartile range would (circle one) **not change much** / **would change a lot**, but both the variance and standard deviation would (circle one) **not change much** / **change a lot**.

6. *Box–Plot: Ph Levels Of Soil.*
Consider the ordered set of the Ph levels of soil data given below.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4.3 | 5 | 5.9 | 6.5 | 7.6 | 7.7 | 7.7 | 8.2 | 8.3 | 9.5 |
| 10.4 | 10.4 | 10.5 | 10.8 | 11.5 | 12 | 12 | 12.3 | 12.6 | 12.6 |
| 13 | 13.1 | 13.2 | 13.5 | 13.6 | 14.1 | 14.1 | 15.1 | | |

(a) Six pieces of information are required to draw a box and whisker plot: median, upper and lower quartiles, the interquartile range, upper and lower fences. The lower quartile, median and upper quartile are 7.95, 11.15 and 13.05, respectively and, so, the interquartile range is $13.05 - 7.95 =$(circle one) **4.95** / **5.10** / **6.15**.

(b) The "upper fence" is determined by adding $1.5\times$(interquartile range) to the upper quartile,
$13.05 + 1.5(5.1) =$(circle one) **20.95** / **21.15** / **6.15**.

(c) The "lower fence" is determined by subtracting $1.5\times$(interquartile range) from the lower quartile,
$7.95 - 1.5(5.1) =$(circle one) **0.1** / **0.2** / **0.3**.

(d) Use your calculator to show the box and whiskers plot is given below.

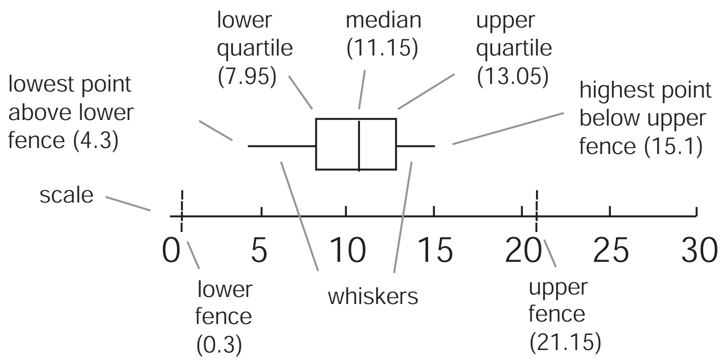



Figure 2.3 (Box and Whisker Plot For Ph Levels Of Soil Data)

The purpose of the "fences" is to determine if any data points are outliers or not. Data points outside the fences, are considered outliers. Any outliers are indicated as "dots" that are in line with, but out and away from, the whiskers.
(After STAT ENTER, type Ph levels into $L_1$; then 2nd STAT PLOT, choose second box plot; then ZOOM 9:ZoomStat ENTER; TRACE to see five number summary[1].)

(e) *How is a box–plot used?*
The box–plot is to identify the (choose none, one or more)

i. location of the data.
ii. spread in the data.
iii. departure from symmetry of the data.
iv. outliers in the data.

## 2.2 Important Discrete Distributions

In this section, we will look at various important discrete distributions.

[1] The calculator *can* give a slightly different box and whisker plot which uses hinges, rather than quartiles; in this case, the calculator gives the same box and whiskers as given above. Either box and whiskers plot is acceptable.

| DISCRETE | $p(x)$ | $E[X] = \mu$ | $\text{Var}(X) = \sigma^2$ |
|---|---|---|---|
| General | $P(X = x)$ | $\sum_{x:p(x)>0} xp(x)$ $= \sum_{i=1}^{\infty} x_i p(x_i)$ | $E[(X-\mu)^2]$ $= E(X^2) - [E(X)]^2$ |
| Hypergeometric | $\dfrac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}$ | $np$, $p = \frac{D}{N}$ | $\frac{N-n}{N-1} np(1-p)$, $p = \frac{m}{N}$ |
| Binomial | $\binom{n}{x} p^x (1-p)^{n-x}$ | $np$ | $np(1-p)$ |
| Bernoulli | $\binom{n}{x} p^x (1-p)^{1-x}$, $x = 0, 1$ | $p$ | $p(1-p)$ |
| Poisson | $e^{-\lambda}\frac{\lambda^x}{x!}$ | $\lambda$ | $\lambda$ |
| Pascal | $\binom{x-1}{r-1} p^r (1-p)^{x-r}$ | $r/p$ | $r(1-p)/p^2$ |
| Geometric | $p(1-p)^{x-1}$ | $1/p$ | $(1-p)/p^2$ |

**Exercise 2.4 (Important Discrete Distributions)**

1. *Hypergeometric: Televisions.*
   Seven television tubes are chosen at random from a shipment of 240 television tubes of which 15 are defective.

   (a) The probability that four of the seven televisions chosen are defective is (circle one)

   $$\frac{\binom{15}{4}\times\binom{225}{3}}{\binom{240}{4}} \;/\; \frac{\binom{15}{3}\times\binom{225}{3}}{\binom{240}{7}} \;/\; \frac{\binom{15}{4}\times\binom{225}{3}}{\binom{240}{7}}$$

   (b) The probability that *five* of the seven televisions chosen are defective is (circle none, one or more)

   i. $\boldsymbol{p(5) = \dfrac{\binom{D}{x}\times\binom{N-D}{n-x}}{\binom{N}{n}}, \; D = 15, N = 240, n = 7, x = 5}$

   ii. $\boldsymbol{p(5) = \dfrac{\binom{15}{5}\times\binom{225}{2}}{\binom{240}{7}}}$

   iii. $\boldsymbol{p(5) = 9.08 \times 10^{-6}}$

   (c) The probability that *two* of the seven televisions chosen are defective is (circle one)

   i. $\boldsymbol{p(2) = \dfrac{\binom{D}{x}\times\binom{N-D}{n-x}}{\binom{N}{n}}, \; D = 15, N = 240, n = 7, x = 2}$

ii. $\boldsymbol{p(2)} = \dfrac{\begin{pmatrix} \mathbf{15} \\ \mathbf{2} \end{pmatrix} \times \begin{pmatrix} \mathbf{225} \\ \mathbf{5} \end{pmatrix}}{\begin{pmatrix} \mathbf{240} \\ \mathbf{7} \end{pmatrix}}$

iii. $\boldsymbol{p(2)} = \mathbf{0.0579}$

(d) *Expectation.* The expected number of defective TVs chosen from the seven is given by $E[X] = \frac{Dn}{N} = \frac{(15)(7)}{240} \approx$ (circle one) **0.4375** / **4.53** / **6.8**.

(e) *Variance.* The variance in the number of defective TVs chosen from the seven is given by (remember $p = \frac{D}{N} = \frac{15}{240}$)
$\text{Var}(X) = \frac{N-n}{N-1} np(1-p) = \frac{240-7}{240-1}(7)(15/240)(1 - 15/240) \approx$ (circle one) **0.39986** / **5.7** / **6.2**.

(f) **True** / **False** We sample with*out* replacement; that is, every time a TV is chosen, we do *not* replace it to be potentially chosen again. In other words, the chance of choosing a defective TV, every time a TV is chosen, *changes* or *depends* on the number of defective TVs that were chosen before it.

(g) If the number of defective TVs, $X$, and initial number of TVs, $N$, are large relative to the sample size, $n$, the hypergeometric can be approximated by a binomial. Let $p = \frac{D}{N} = \frac{15}{240} = 0.0625$ and so
$E[X] \approx np = (7)(0.0625) =$ (circle one) **0.39986** / **0.4375** / **6.2**.
$\text{Var}(X) \approx (7)(0.0625)(1 - 0.0625) =$ (circle one) **0.39986** / **0.4102** / **6.2**.

2. *Binomial Random Variable: Airplane Engines.*
One engine of a four ($n = 4$) engine airplane fails 11% ($p = 0.11$) of the time. Assume this problem obeys the conditions of a binomial experiment.

(a) The chance three engines fail is $\frac{n!}{x!(n-x)!} \times p^x \times (1-p)^{n-x}$ where $n = 4$, $x = 3$ and $p = 0.11$, in other words (circle one) **0.005** / **0.011** / **0.040**.

(b) The chance *at most* three engines fail is (circle one) **0.995** / **0.997** / **0.999**.

(c) The *expected* number of failures is $np = 4(0.11) =$ **0.44** / **0.51** / **0.62**.

(d) The *variance* in the number of failures is $np(1-p) =$
(circle one) **0.15** / **0.40** / **0.51**.

3. *Bernoulli, Special Case of Binomial.*
The Bernoulli distribution is a special case of the binomial distribution where $n = 1$. In other words, the Bernoulli distribution is given by (circle none, one or more)

(a) $\begin{pmatrix} 1 \\ x \end{pmatrix} p^x(1-p)^{1-x}; \; x = 0, 1$

(b) $\begin{pmatrix} 1 \\ 0 \end{pmatrix} p^0(1-p)^{1-0}$ when $x = 0$ and $\begin{pmatrix} 1 \\ 1 \end{pmatrix} p^1(1-p)^{1-1}$ when $x = 1$

(c) $(1-p)$ when $x=0$ and $p$ when $x=1$

4. *Poisson: Photons.*
A piece of iron is bombarded with electrons and, as a consequence, releases a number of photons. A number, $x$, of the photon particles released hit a magnetic detection field that surround the piece of iron being tested. It is found that an *average* (or expectation) of $\lambda = 5$ particles hit the magnetic detection field *per microsecond*.

(a) The chance that 2 particles hit the field per microsecond is
$\frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-5}5^2}{2!} \approx$ (circle one) **0.06** / **0.07** / **0.08**.
(Use your calculator: 2nd DISTR B:poissonpdf(5,2) ENTER.)

(b) The chance that 0 particles hit the field per microsecond is
$\frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-5}5^0}{0!} \approx$ (circle one) **0.007** / **0.008** / **0.009**.

(c) If an average of 5 particles hit the field every one microsecond time interval, then, in a *two* microsecond time interval, an average of $2 \times 5 = 10$ particles will hit the field. In a similar way, in a *six* microsecond time interval, an average of (circle one) **25** / **30** / **35** particles will hit the field.

(d) Since an average of $\lambda = 2 \times 5 = 10$ particles hit the field in a two microsecond time interval, the chance that 3 particles hit the field in a two microsecond time interval, is
$\frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-10}10^3}{3!} \approx$ (circle one) **0.007** / **0.008** / **0.009**.

(e) The chance that $x = 21$ particles hit the field in a four microsecond time interval, is $\frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-20}20^{21}}{21!} \approx$ (circle one) **0.073** / **0.085** / **0.091**.

(f) *Expected Value and Standard Deviation.* The variance in the number of particles hitting the field is equal to the average number (expected number) of particles hitting the field, $\lambda$. Consequently, the standard deviation must be (circle one) $\boldsymbol{\sqrt{\lambda}}$ / $\boldsymbol{\lambda^2}$ / $\boldsymbol{\frac{1}{2}\lambda^2}$.

(g) *What Is Poisson Used For In Statistical Quality Control?*
**True** / **False** The Poisson distribution is often used as a model of the number of defects (nonconformities) that occur in a "unit of a product".

5. *Pascal: Shooting Hoops.*
There is a 35% ($p = 0.35$) chance of making a basket on a free throw.

(a) The chance that the third basket will occur on the *second* try is
$p(2) \approx$ (circle one) **0** / **0.148** / **0.198**.

(b) The chance that the 15th basket will occur on the 50th try is (circle one)

i. $p(x) = \binom{x-1}{r-1} p^r(1-p)^{x-r},\ x = 50, r = 15, p = 0.65$

ii. $p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \; x = 25, r = 19, p = 0.15$

iii. $p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \; x = 50, r = 15, p = 0.35$

iv. $p(x) = \binom{x-1}{r-1} p^r (1-p)^{i-r}, \; x = 31, r = 19, p = 0.15$

(c) The chance that the 7th basket will occur on the 50th try is $\binom{50-1}{7-1} 0.35^7 (1-0.35)^{50-7} =$ (circle one) **5.8** / **0.058** / **0.00008117**.

(d) *Expectation.* The expected number of attempts until a second basket is given by
$E[X] = \frac{r}{p} = \frac{2}{0.35} \approx$ (circle one) **5.7** / **6.2** / **6.8**.

(e) *Variance.* The variance in the number of attempts until a second basket is given by
$\text{Var}(X) = \frac{r(1-p)}{p^2} = \frac{2(1-0.35)}{0.35^2} \approx$ (circle one) **0.4** / **1.8** / **2.2** / **10.6**.

(f) **True** / **False** We have assumed that each attempt is independent of every other throw and chance of a basket remains constant, $p = 0.35$, on every throw.

## 2.3 Important Continuous Distributions

We now review important continuous distributions[2].

| CONTINUOUS | $f(x)$ | $F(x)$ | $\mu$ | $\sigma^2$ |
|---|---|---|---|---|
| General | | | $\int_{-\infty}^{\infty} x f(x)\, dx$ | $\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$ |
| Normal | $\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ | $\Phi(x)$ | $\mu$ | $\sigma^2$ |
| Exponential | $\lambda e^{-\lambda x}$ | $1 - e^{-\lambda a}$ | $1/\lambda$ | $1/\lambda^2$ |
| Gamma | $\lambda e^{-\lambda x} (\lambda x)^{r-1}/\Gamma(r)$ | $1 - \int_a^{\infty} \lambda e^{-\lambda t} (\lambda t)^{r-1}/\Gamma(r)\, dt$ | $r/\lambda$ | $r/\lambda^2$ |
| Weibull | $\frac{\beta}{\theta} \left(\frac{x-\nu}{\theta}\right)^{\beta-1} \exp\left\{-\left(\frac{x-\nu}{\theta}\right)\right\}$ | $1 - \exp\left\{-\left(\frac{a-\nu}{\theta}\right)^{\beta}\right\}$ | $\theta\Gamma\left(1+\frac{1}{\beta}\right)$ | $\theta^2 \left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma\left(1+\frac{1}{\beta}\right)^2\right]$ |

**Exercise 2.5 (Important Continuous Distributions)**

1. *Normal Distribution: IQ Scores.*

[2]The Weibull given here is a little bit more general than the one given in the text; the text assumes $\nu = 0$.

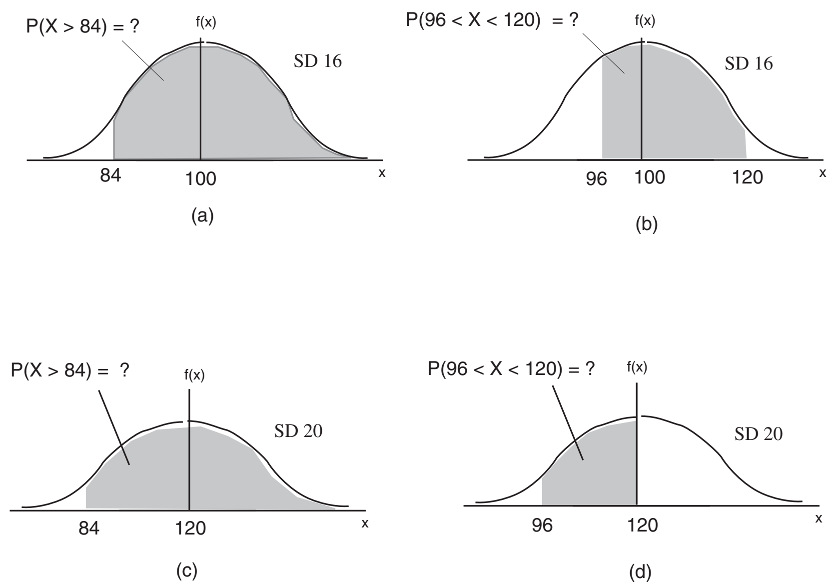



Figure 2.4 (Probabilities For Nonstandard Normal Distributions of IQ Scores)

(a) The *upper two* (of the four) normal curves above represent the IQ scores for *sixteen* year olds. Both are nonstandard normal curves because the (circle none, one or more)

  i. the average is 100 and the SD is 16.
  ii. neither the average is 0, nor is the SD equal to 1.
  iii. the average is 16 and the SD is 100.
  iv. the average is 0 and the SD is 1.

  The *lower* two normal curves above represent the IQ scores for *twenty* year olds ($\mu = 120, \sigma = 20$).

(b) Since the sixteen year old distribution is symmetric, (circle one) **25%** / **50%** / **75%** of the IQ scores are above (to the right) of 100.

(c) The probability of the IQ scores being less than 84, $P\{X < 84\}$, for the sixteen year old distribution is (circle one) **greater than** / **about the same as** / **smaller than** 0.50.

(d) $P\{X > 84\} =$

$$\begin{aligned} P\{X > 84\} &= 1 - \Phi\left(\frac{84-\mu}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{84-100}{16}\right) \\ &= 1 - \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{84} e^{-(1/2)[(84-100)/16]^2}\, dy \ = \end{aligned}$$

(circle one) **0.8413** / **0.1587** / **−0.1587**
(Use 2nd DISTR 2:normalcdf(84, 2nd EE 99, 100, 16).)

(e) Match the columns.

| Column I | Column II |
|---|---|
| **(a)** $P\{X > 84\}$, "sixteen year old" normal | **(a)** 0.4931 |
| **(b)** $P\{96 < X < 120\}$, "sixteen year old" normal | **(b)** 0.9641 |
| **(c)** $P\{X > 84\}$, "twenty year old" normal | **(c)** 0.8413 |
| **(d)** $P\{96 < X < 120\}$, "twenty year old" normal | **(d)** 0.3849 |

| Column I | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| Column II | | | | |

(f) **True** / **False** $P\{Z < 84\}$ for standard normal equals $P\{X < 84\}$ for the *non*standard normal

2. *Standardizing Nonstandard Normal Random Variables.*
*Non*standard random variable $X$, with mean $\mu$ and standard deviation $\sigma$, can be "standardized" into a standard random variable $Z$ using the following formula:

$$Z = \frac{X - \mu}{\sigma}$$

(a) The IQ scores for the 16 year olds are normal with $\mu = 100$ and $\sigma = 16$. The standardized value of the nonstandard IQ score of 110 for the 16 year olds, then, is
$Z = \frac{X-\mu}{\sigma} = \frac{110-100}{16} =$ (circle one) **0.625** / **1.255** / **3.455**
and so $P\{X > 110\} = P\{Z > 0.625\}$.
(Compare 2nd DISTR 2:normalcdf(110, 2nd EE 99, 100, 16) with 2nd DISTR 2:normalcdf(0.625, 2nd EE 99, 0, 1).)

(b) The IQ scores for the 20 year olds are normal with $\mu = 120$ and $\sigma = 20$. The standardized value of the nonstandard IQ score of 110 for the 20 year olds, then, is
$Z = \frac{X-\mu}{\sigma} = \frac{110-120}{20} =$ (circle one) **0.5** / **-0.5** / **0.25**.
and so $P\{X > 110\} = P\{Z > -0.5\}$.
(Compare 2nd DISTR 2:normalcdf(110, 2nd EE 99, 120, 20) with 2nd DISTR 2:normalcdf($-0.5$, 2nd EE 99, 0, 1).)

(c) If both a 16 year old and 20 year old score 110 on an IQ test, (check none, one or more)

i. the 16 year old is brighter relative to his age group than the 20 year old is relative to his age group
ii. the z–score is higher for the 16 year old than it is for the 20 year old
iii. the z–score allows us to compare the IQ score for a 16 year old with the IQ score for a 20 year old

(d) If $\mu = 100$ and $\sigma = 16$, then
$P\{X > 130\} = P\left\{Z > \frac{130-100}{16}\right\} =$ (circle one) **0.03** / **0.31**

(e) If $\mu = 120$ and $\sigma = 20$, then
$P\{X > 130\} = P\left\{Z > \frac{130-120}{20}\right\} =$ (circle one) **0.03** / **0.31**

(f) If $\mu = 25$ and $\sigma = 5$, then
$P\{27 < X < 32\} = P\left\{\frac{27-25}{5} < Z < \frac{32-25}{5}\right\} =$
(circle one) **0.03** / **0.26** / **0.31**

3. *Central Limit Theorem.*

(a) *Sum.*
Suppose $X$ has a (*any!*) distribution where $\mu_X = 2.7$ and $\sigma_X = 0.64$. If $n = 35$, then determine $P\left(\sum_{i=1}^{35} X_i > 99\right)$.

i. $\mu_{\sum X_i} = n\mu = 35(2.7) =$ (circle one) **93.5** / **94.5** / **95.5**.

ii. $\sigma_{\sum X_i} = \sigma\sqrt{n} = 0.64\sqrt{35} =$ (circle one) **3.5** / **3.8** / **4.1**.

iii. $P\left(\sum_{i=1}^{35} X_i > 99\right) \approx$ (circle one) **0.09** / **0.11** / **0.15**.
(2nd DISTR normalcdf(99,E99,94.5,3.8))

(b) *Another Sum.*
Suppose $X$ has a (*any!*) distribution where $\mu_X = -1.7$ and $\sigma_X = 1.6$. If $n = 43$, then determine $P\left(-76 < \sum_{i=1}^{43} X_i < -71\right)$.

i. $\mu_{\sum X_i} = n\mu = 43(-1.7) =$ (circle one) **−73.5** / **−73.1** / **−72.9**.

ii. $\sigma_{\sum X_i} = \sigma\sqrt{n} = 1.6\sqrt{43} =$ (circle one) **9.5** / **9.8** / **10.5**.

iii. $P\left(-76 < \sum_{i=1}^{43} X_i < -71\right) \approx$ (circle one) **0.09** / **0.11** / **0.19**.
(2nd DISTR normalcdf(-76,-71,-73.1,10.5))

(c) *Average.*
Suppose $X$ has a distribution where $\mu = 2.7$ and $\sigma = 0.64$. If $n = 35$, determine the chance the *average* (not sum!) is larger than 2.75, $P(\bar{X} > 2.75)$.

i. $\mu_{\bar{X}} = \frac{n\mu}{n} = \mu =$ (circle one) **2.7** / **2.8** / **2.9**.

ii. $\sigma_{\bar{X}} = \frac{\sigma\sqrt{n}}{n} = \frac{\sigma}{\sqrt{n}} = \frac{0.64}{\sqrt{35}} =$ (circle one) **0.11** / **0.12** / **0.13**.

iii. $P\left(\bar{X} > 2.75\right) \approx$ (circle one) **0.30** / **0.32** / **0.35**.
(2nd DISTR normalcdf(2.75,E99,2.7,0.11))

(d) *Another Average.*
Suppose $X$ has a distribution where $\mu_X = -1.7$ and $\sigma_X = 1.5$. If $n = 49$, then
$P(-2 < \bar{X} < 2.75) \approx$ (circle one) **0.58** / **0.58** / **0.92**.

4. *Exponential Distribution.*

(a) *Some Calculations.*

i. Probability By Distribution Function[3]. For $\lambda = \frac{1}{2}$,

$$\begin{aligned} P\{X \leq 1.1\} &= F(1.1) \\ &= 1 - e^{-\lambda(1.1)} \\ &= 1 - e^{-\frac{1}{2}(1.1)} = \end{aligned}$$

(circle one) **0.32** / **0.42** / **0.52**.

ii. For $\lambda = 3$, $P\{X < 1.1\} = F(1.1) = 1 - e^{-3(1.1)} =$ (circle one) **0.32** / **0.42** / **0.96**.

iii. For $\lambda = 5$, $P\{X < 1.1\} = F(1.1) = 1 - e^{-5(1.1)} =$ (circle one) **0.32** / **0.42** / **0.996**.

iv. The probability of waiting less than 1.1 minutes for an email when $\lambda = \frac{1}{2}$ is (circle one) **greater than** / **about the same as** / **smaller than** probability of waiting less than 1.1 minute for an email when $\lambda = 5$.

(b) For $\lambda = 3$,
$P\{X > 0.54\} = 1 - F(0.54) = 1 - \left(1 - e^{-(3)(0.54)}\right) = e^{-(3)(0.54)} =$
(circle one) **0.20** / **0.22** / **0.29**.

(c) For $\lambda = 3$,
$P\{1.13 < X < 1.62\} = F(1.62) - F(1.13) = e^{-(3)(1.13)} - e^{-(3)(1.62)} =$
(circle one) **0.014** / **0.026** / **0.29**.

(d) *Expectation and Variance.*
For $\lambda = \frac{1}{2}$, $E[X] = \frac{1}{\lambda} = \frac{1}{1/2} =$ (circle one) **2** / **3** / **4**.
For $\lambda = 3$, $E[X] = \frac{1}{\lambda} = \frac{1}{3} =$ (circle one) $\mathbf{\frac{1}{2}}$ / $\mathbf{\frac{1}{3}}$ / $\mathbf{\frac{1}{4}}$.
For $\lambda = \frac{1}{2}$, $\text{Var}(X) = \frac{1}{\lambda^2} = \frac{1}{(1/2)^2}$ (circle one) **2** / **3** / **4**.
For $\lambda = 3$, $\text{Var}(X) = \frac{1}{\lambda^2} = \frac{1}{3^2}$ (circle one) $\mathbf{\frac{1}{2}}$ / $\mathbf{\frac{1}{5}}$ / $\mathbf{\frac{1}{9}}$.

(e) *Exponential and Statistical Quality Control.*

i. **True** / **False**
In $\mu = \frac{1}{\lambda}$, $\lambda$ could be considered the failure rate and $\mu$ is the mean time to failure and, for this reason, this distribution is used in *reliability engineering*.

ii. **True** / **False**
$P\left\{x \leq \frac{1}{\lambda}\right\} = 0.6312$ *always*!

iii. **True** / **False**
If the number of occurrences of an event has a Poisson distribution with parameter $\lambda$, the distribution of the interval *between* occurrences is an exponential with parameter $\lambda$.

[3] The distribution function, $F(x)$, is often easier to calculate than integrating the distribution, $\int f(x)\,dx$.

5. *Gamma Distribution.*

   (a) *Getting Familiar With The Gamma Function,* $\Gamma(r) = \int_0^\infty e^{-y} y^{r-1}\, dy$.

      i. $\Gamma(1.2) = \int_0^\infty e^{-y} y^{1.2-1}\, dy =$ (circle closest one) **0.92** / **1.12** / **2.34**. (First let Y= MATH 9:fnInt($e^{-Y} Y^{X-1}$,Y,0,100), then 2nd QUIT and then VARS Y-VARS 1:Function ENTER $\text{Y}_1$ (1.2) ENTER)

      ii. $\Gamma(2.4) = \int_0^\infty e^{-y} y^{2.4-1}\, dy =$ (circle closest one) **0.92** / **1.24** / **2.34**. (2nd ENTER and type 2.4 over 1.2 ENTER)

      iii. $\Gamma(1) =$ (circle *none, one or more*) $\mathbf{(1-1)!}$ / **0!** / **1**.

      iv. $\Gamma(2) =$ (circle *none, one or more*) $\mathbf{(2-1)!}$ / **1!** / **1**.

      v. $\Gamma(3) =$ (circle *none, one or more*) $\mathbf{(3-1)!}$ / **2!** / **2**.

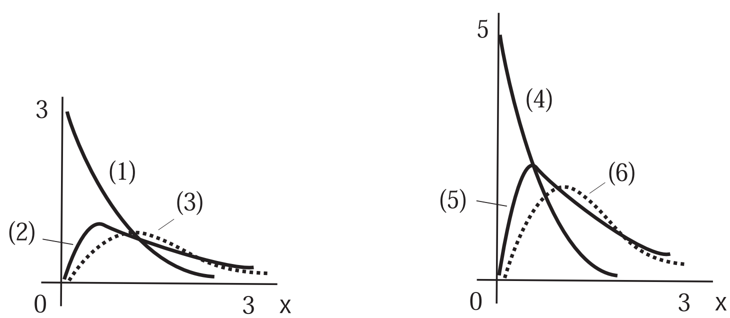
      vi. $\Gamma(4) =$ (circle *none, one or more*) $\mathbf{(4-1)!}$ / **3!** / **6**.

      vii. $\Gamma(5) =$ (circle *none, one or more*) $\mathbf{(5-1)!}$ / **4!** / **24**.

      viii. **True** / **False**. In general,

$$\Gamma(n) = (n-1)!, \quad \text{if } n \text{ is a positive integer}$$

   (b) *Calculating* $\frac{\lambda e^{-\lambda x}(\lambda x)^{r-1}}{\Gamma(r)}$. The graphs of different (different $(r, \lambda)$) gamma density functions are given the figure below.




(a) changing r, k= 3 (b) changing r, k= 5

Figure 2.5 (Gamma Probability Density Functions)

Match each of the gamma density functions ((1) to (6)) given below to each of the graphs given above. Notice that $r$ are all positive integers and so $\Gamma(r) = (r-1)!$

| $(r, \lambda) =$ | $(1,3)$ | $(2,3)$ | $(3,3)$ | $(1,5)$ | $(2,5)$ | $(3,5)$ |
|---|---|---|---|---|---|---|
| graph | | | | | | |

So $r$ is the *shape* parameter and $\lambda$ is the scale parameter. (Hint: Use your calculators; use WINDOW 0 3 1 0 5 1;
for $(r, \lambda) = (1, 3)$, use $\text{Y}_1 = 3e^{-3X}$,
for $(r, \lambda) = (2, 3)$, use $\text{Y}_2 = 3e^{-3X}(3X)$,

for $(r, \lambda) = (3, 3)$, use $Y_3 = 3e^{-3X}(3X)^2/2$
and for $(r, \lambda) = (1, 5)$, use $Y_4 = 5e^{-5X}$,
for $(r, \lambda) = (2, 5)$, use $Y_5 = 5e^{-5X}(5X)$,
for $(r, \lambda) = (3, 5)$, use $Y_6 = 5e^{-5X}(5X)^2/2$

(c) For $(r, \lambda) = (1, 5)$, $P\{X < 0.2\} =$ (circle one) **0.35** / **0.41** / **0.63**.
(MATH 9:fnInt($Y_4$,X,0,0.2) ENTER)

(d) For $(r, \lambda) = (2, 5)$, $P\{X > 0.4\} =$ (circle one) **0.35** / **0.41** / **0.63**.
(First MATH 9:fnInt($Y_5$,X,0,0.4) ENTER, then subtract the result from 1.)

(e) For $(r, \lambda) = (2, 5)$, $P\{0.1 < X < 0.3\} =$ (circle one) **0.35** / **0.41** / **0.63**.

(f) *Expectation and Variance.*
For $(r, \lambda) = (1, 5)$, $E[X] = \frac{r}{\lambda} =$ (circle one) **1/5** / **3** / **4**.
For $(r, \lambda) = (3, 3)$, $E[X] =$ (circle one) **1/5** / **1** / **3**.
For $(r, \lambda) = (1, 5)$, $\text{Var}(X) = \frac{r}{\lambda^2} =$ (circle one) **1/25** / **3** / **8**.
For $(r, \lambda) = (3, 3)$, $\text{Var}(X) =$ (circle one) $\mathbf{\frac{1}{2}}$ / $\mathbf{\frac{1}{3}}$ / $\mathbf{\frac{1}{4}}$.

6. *Weibull Distribution.*

(a) *Sketching* $\frac{\beta}{\theta}\left(\frac{x-\nu}{\theta}\right)^{\beta-1} exp\left\{-\left(\frac{x-\nu}{\theta}\right)^{\beta}\right\}$. The graphs of different (different $(\nu, \theta, \beta)$) Weibull density functions are given the figure below.

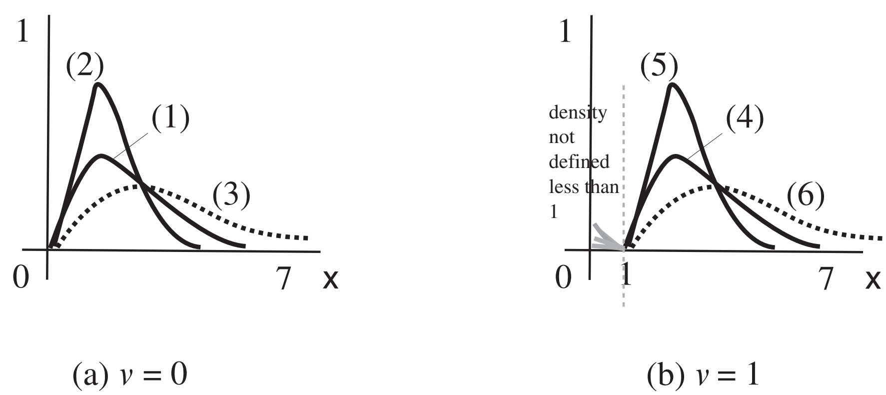



Figure 2.6 (Weibull Probability Density Functions)

Match each of the gamma density functions ((1) to (6)) given below to each of the graphs in (a) and (b) given above.

| $(\nu, \theta, \beta) =$ | $(0, 2, 3)$ | $(0, 2, 5)$ | $(0, 4, 3)$ | $(1, 2, 3)$ | $(1, 2, 5)$ | $(1, 4, 3)$ |
|---|---|---|---|---|---|---|
| graph | | | | | | |

(Hint: Use your calculators; use WINDOW 0 7 1 0 1 0.1;
for $(0, 2, 3)$, use $Y_1 = (3/2)((X - 0)/2)^{(3-1)}e^{-((X-0)/2)^3}$,
for $(0, 2, 5)$, use $Y_2 = (5/2)((X - 0)/2)^{(5-1)}e^{-((X-0)/2)^5}$,
for $(0, 4, 3)$, use $Y_3 = (3/4)((X - 0)/4)^{(3-1)}e^{-((X-0)/4)^3}$

and for $(1,2,3)$, use $\mathrm{Y}_4 = (3/2)((X-1)/2)^{(3-1)}e^{-((X-1)/2)^3}$,
for $(1,2,5)$, use $\mathrm{Y}_5 = (5/2)((X-1)/2)^{(5-1)}e^{-((X-1)/2)^5}$,
for $(1,4,3)$, use $\mathrm{Y}_6 = (3/4)((X-1)/4)^{(3-1)}e^{-((X-1)/4)^3}$)

(b) *Probability By Integration.* For $(\nu, \theta, \beta) = (1,2,3)$,

$$\begin{aligned} P\{1 < X < 1.7\} &= \int_1^{1.7} \frac{\beta}{\theta}\left(\frac{x-\nu}{\theta}\right)^{\beta-1} \exp\left\{-\left(\frac{x-\nu}{\theta}\right)^{\beta}\right\} dx \\ &= \int_1^{1.7} \frac{3}{2}\left(\frac{x-1}{2}\right)^{2-1} \exp\left\{-\left(\frac{x-1}{2}\right)^{3}\right\} dx \ = \end{aligned}$$

**0.04** / **0.41** / **0.98**.
(MATH fnInt($\mathrm{Y}_4$,X,1,1.7) ENTER)

(c) *Probability By Distribution Function.* For $(\nu, \theta, \beta) = (1,2,3)$,

$$\begin{aligned} P\{1 < X < 1.7\} &= F(1.7) - F(1) \\ &= \left(1 - \exp\left\{-\left(\frac{1.7-\nu}{\theta}\right)^{\beta}\right\}\right) - \left(1 - \exp\left\{-\left(\frac{1-\nu}{\theta}\right)^{\beta}\right\}\right) \\ &= \exp\left\{-\left(\frac{1-\nu}{\theta}\right)^{\beta}\right\} - \exp\left\{-\left(\frac{1.7-\nu}{\theta}\right)^{\beta}\right\} \\ &= \exp\left\{-\left(\frac{1-1}{2}\right)^{3}\right\} - \exp\left\{-\left(\frac{1.7-1}{2}\right)^{3}\right\} \ = \end{aligned}$$

**0.04** / **0.41** / **0.98**.

(d) For $(\nu, \theta, \beta) = (1,2,3)$, $P\{1 < X < 1.7\} =$
(circle one) **0.04** / **0.41** / **0.63**.
(MATH 9:fnInt($\mathrm{Y}_4$,X,1,1.7) ENTER)

(e) For $(\nu, \theta, \beta) = (1,2,3)$, $P\{X < 1.7\} =$
(circle one) **0.04** / **0.41** / **0.63**.
(MATH 9:fnInt($\mathrm{Y}_4$,X,1,1.7) ENTER)

(f) For $(\nu, \theta, \beta) = (0,2,3)$, $P\{0 < X < 0.7\} =$
(circle one) **0.04** / **0.41** / **0.63**.
(MATH 9:fnInt($\mathrm{Y}_1$,X,0,0.7) ENTER)

(g) For $(\nu, \theta, \beta) = (0,2,3)$, $P\{X < 1.7\} =$
(circle one) **0.04** / **0.46** / **0.63**.
(MATH 9:fnInt($\mathrm{Y}_4$,X,0,1.7) ENTER)

(h) For $(\nu, \theta, \beta) = (1,2,3)$, $P\{X > 1.4\} =$
(circle one) **0.35** / **0.41** / **0.992**.
(Subtract MATH 9:fnInt($\mathrm{Y}_4$,X,1,0.4) ENTER from 1.)

(i) *Weibull and Statistical Quality Control.*
**True** / **False** The Weibull random variable is used in engineering contexts when the distribution of the lifetime of some item, which consists of many parts, fails if any one of the parts fails.

## 2.4 Some Useful Approximations

Sometimes it is useful to approximate one distribution with another; in particular,

- Binomial approximation to Hypergeometric
- Poisson approximation to Binomial
- Normal approximation to Binomial