***Note: In addition to Chapters 1 and 2, the Midterm exam on October 16th will cover Sections 3.1-3.3 inclusive as well as the Applications on pp. 106-110. Material to study for Midterm: pp. 1 – 126.***

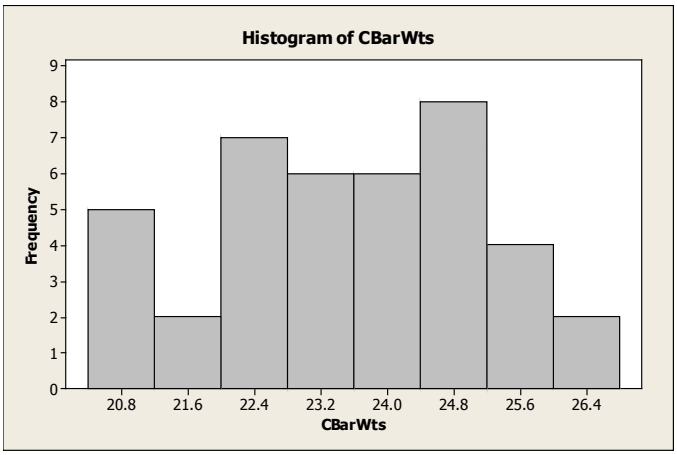
# Chapter 3 – Continuous Distributions

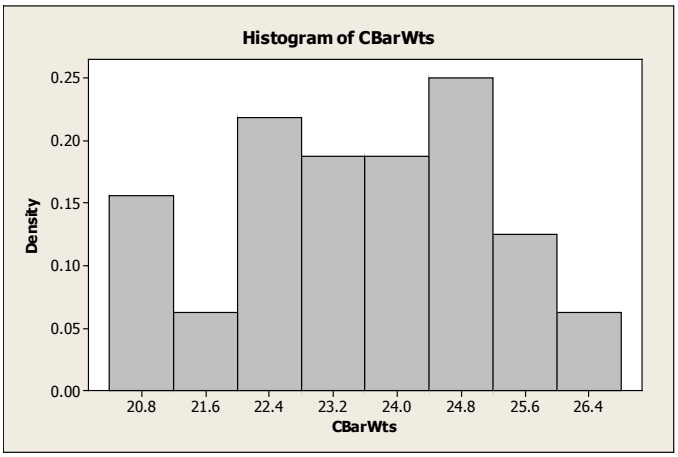
**We'll spend 6 classes on this 59-page chapter.**

## Descriptive Statistics and EDA (Section 3.1, pp. 97-105)

**Even though they may be rounded off, if measurements of a RV X at least conceptually could come from an *interval* of possible outcomes, then it is a continuous RV. Continuous data can be summarized in a number of ways, both graphically and numerically.**

**On pp.97-98, steps are given to group continuous data; in the center of p.98, a formula is given to find relative frequencies and to then construct a relative frequency histogram. The key here is that *areas represent probabilities*, so the area under the curve sums to one. Consider the Baby Ruth Candy Bar Weights dataset on p.98.**




**Minitab has been used in the above graphs. In contrast with our text, it has chosen the intervals [20.4 – 21.2), [21.2 – 22.0), [22.0 – 22.8], etc. The graph above at left gives the Frequencies ($f_m$) on the y-axis; the**

**graph on the right plots the Densities, h(x), which are the Relative Frequencies divided by the class widths:**

$$h(x) = \frac{f_m}{(n)(c_m - c_{m-1})} \quad \text{for } c_{m-1} < x \leq c_m, \text{ and } m = 1, 2 \ldots k$$

**Usually, we calculate the actual sample mean and variance from a set of data, but once it has been grouped and discarded, we could find the grouped sample mean and variance using the formulas on p.100; interestingly, all points in an interval are then assumed to occur at the midpoint of the respective interval (called the class marks).**

**The relevance of graphing *densities* instead of *relative frequencies* is illustrated in Ex. 3.1-3 on pp.100-1 since *intervals of differing widths* are chosen.**

**Another plot of data includes Stem and Leaf Plots; these are very quickly obtained and give a good idea of the distribution of the data.**

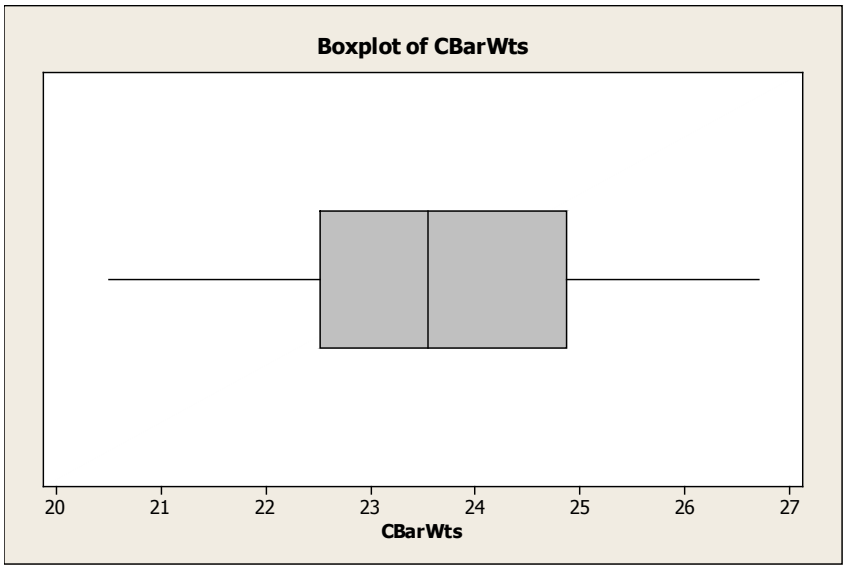
**Stem-and-Leaf Display: CBarWts**

```
Stem-and-leaf of CBarWts  N  = 40
Leaf Unit = 0.10

 3    20  578
 6    21  004
 7    21  5
 9    22  01
 16   22  5667799
 19   23  134
(5)   23  56669
 16   24  13
 14   24  558899
 8    25  112
 5    25  689
 2    26  1
 1    26  7
```

**Additional examples of Stem and Leaf plots are given on pp.102-3; as on p.102, usually the Stem and Leaf plot is first given for the unordered data, but then must be given for the *ordered data*.**

**The following plot – the Box Plot – graphs the 5-number summary (discussed below).**




**In the above plot: at the left is the minimum (20.5), at the right is the maximum (26.7), in the center is the median (23.55 – also called the 50% percentile and the second quartile $Q_2$), on the left of the box is the first quartile ($Q_1$) or $25^{th}$ percentile (22.525), and on the right of the box is the third quartile ($Q_3$) or $75^{th}$ percentile (24.875).**

**Descriptive Statistics: CBarWts**

| Variable | N | Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
|---|---|---|---|---|---|---|---|---|
| CBarWts | 40 | 23.505 | 1.641 | 20.500 | 22.525 | 23.550 | 24.875 | 26.700 |

**Once the data are ordered – as in the Stem and Leaf plot – it's easy to obtain the sample order statistics including those listed above; see the discussion at the top of p.104. Further, the position of the $100p^{th}$ sample percentile is (n+1)p. For example, the positions above of $Q_1$, $Q_2$, and $Q_3$ are $41\times0.25 = 10.25$, $41\times0.50 = 20.5$, and $41\times0.75 = 30.75$.**

***Don't confuse the position with the number itself.*** **It follows that $Q_1 = 0.75\times22.5 + 0.25\times22.6 = 22.525$, $Q_2 = \frac{1}{2}\times23.5 + \frac{1}{2}\times23.6 = 23.55$, and $Q_1 = 0.25\times24.8 + 0.75\times24.9 = 24.875$. These values are given above in the Minitab output and graphed in the Box Plot. Finally, the IQR (interquartile range) is $Q_3 - Q_1$; here, it's equal to 2.35.**

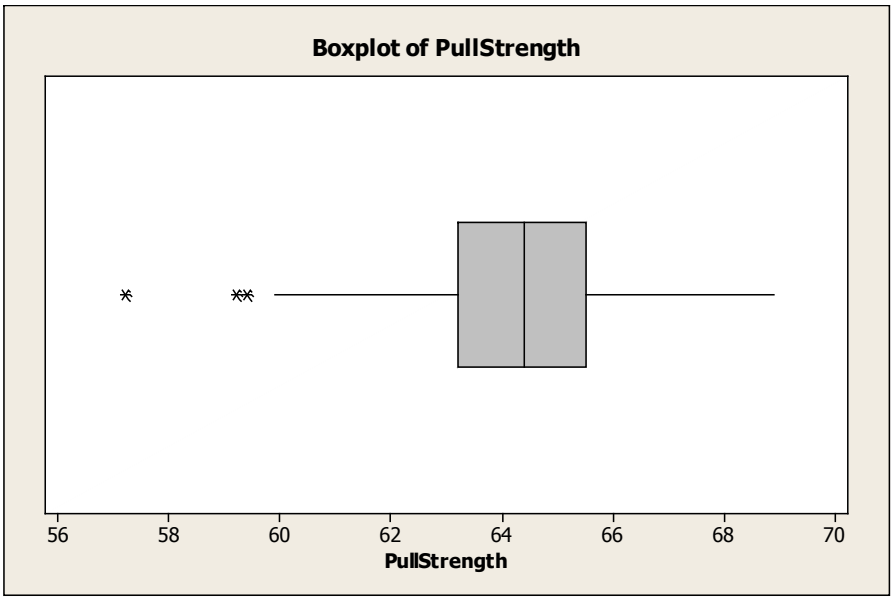
**The Box Plot on the bottom of p.105 shows the Left skew in the data.**

## Applications (pp. 106-110)

## Example I. Solder Joint Strengths.

**Descriptive Statistics: PullStrength**

| Variable | N | Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
|---|---|---|---|---|---|---|---|---|
| PullStrength | 98 | 64.202 | 2.055 | 57.200 | 63.200 | 64.400 | 65.500 | 68.900 |



## Example II. Edwin Moses Race Times.

**Descriptive Statistics: RaceTime**

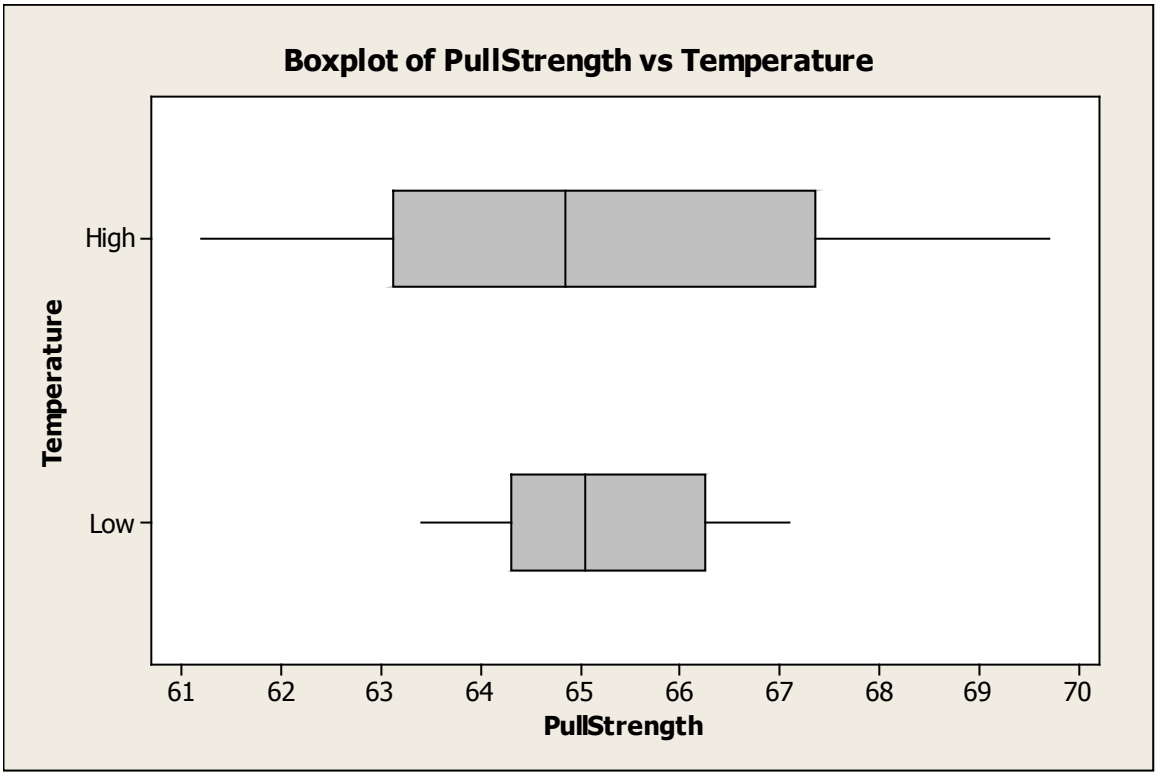
| Variable | N | Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
|---|---|---|---|---|---|---|---|---|
| RaceTime | 122 | 48.487 | 0.736 | 47.000 | 47.900 | 48.500 | 49.000 | 50.100 |

**The Olympic runner is shown below.**

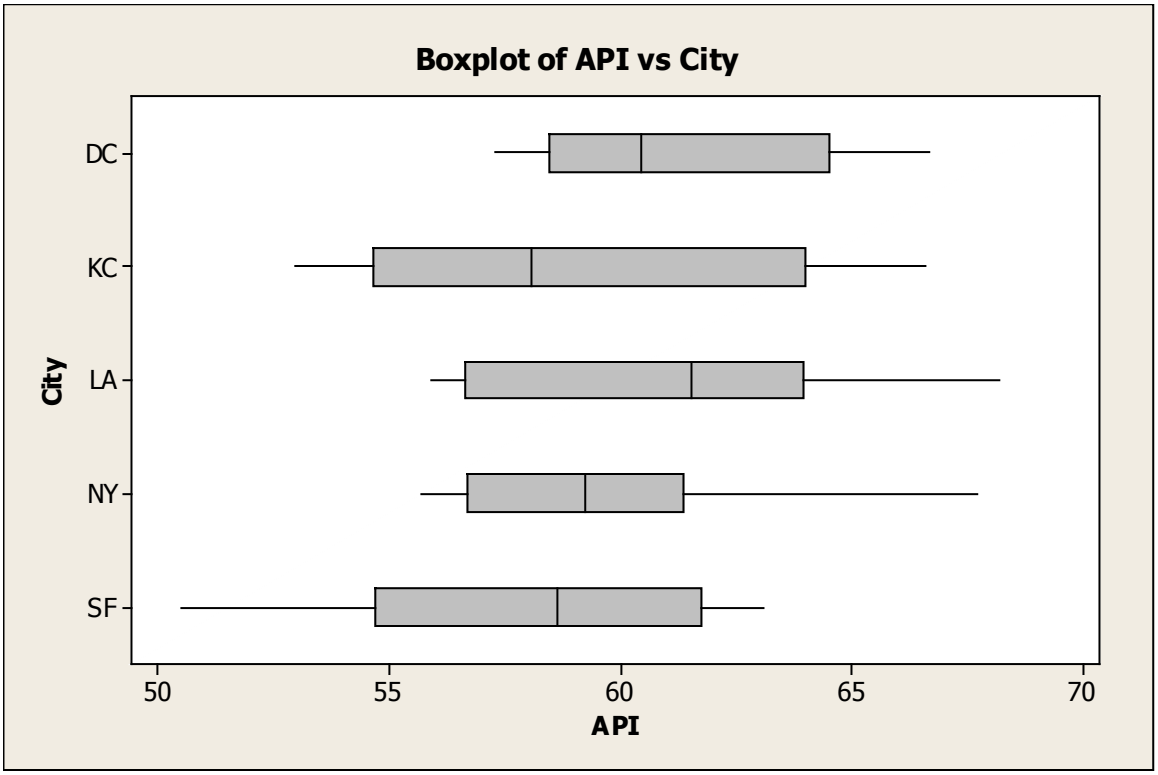
**Note the Stem and Leaf Plot on p.107 and given here:**

| | | |
|---|---|---|
| 4 | 47 | 0111 |
| 9 | 47 | 23333 |
| 17 | 47 | 45555555 |
| 23 | 47 | 666677 |
| 33 | 47 | 8889999999 |
| 36 | 48 | 011 |
| 47 | 48 | 22222222333 |
| (17) | 48 | 44444555555555555 |
| 58 | 48 | 666666666677777777 |
| 41 | 48 | 8888888999 |
| 31 | 49 | 000001111 |
| 22 | 49 | 22233 |
| 17 | 49 | 4455555 |
| 10 | 49 | 67777 |
| 5 | 49 | 89 |
| 3 | 50 | 111 |

## Example III. More on Solder Strengths – High and Low Temp's.




## Example IV. Air Pollution in 5 US Cities.



## Continuous Probability Distributions (Section 3.2, pp. 112-117)

**Again, as the sample size increases without bound ($n \to \infty$), the above density function h(x) approaches a continuous curve f(x), called a probability density function and denoted 'pdf'. In this case,**

**(a) for x in the sample space S, f(x) > 0, whereas f(x) = 0 for $x \notin S$;**

**(b) $\int_S f(x)\,dx = 1$;**

**(c) For $A \subset S$, $P(A) = \int_A f(x)\,dx$, so if the interval (a, b) is in S, then $P(a < X < b) = \int_a^b f(x)dx$**

**Ex. 3.2-1. X has a uniform distribution over the interval (0, 1), so f(x) = 1 for 0 < x < 1, and f(x) = 0 otherwise; we write X ~ U(0,1). Note the difference between the continuous uniform distribution in this chapter and the discrete uniform distribution in the last chapter.**

**Then for 0 < a < b < 1, we have**

$$P(a < X < b) = \int_a^b 1dx = x]_a^b = b - a.$$

**Examples are given on p.113 for choices of 'a' and 'b'. Notice that when a = b, we get P(a < X < a) = 0.**

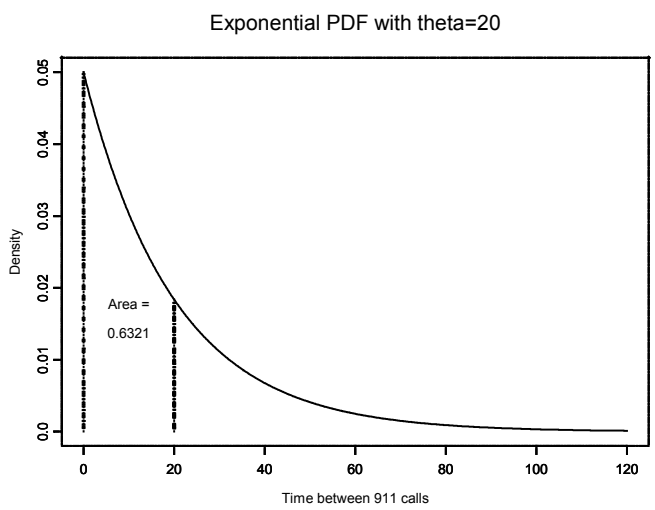
**Ex. 3.2-2. X has the pdf $f(x) = \frac{1}{20}e^{-x/20}$ for $x \geq 0$; in the next section, we'll say that X has an exponential distribution with parameter $\theta = 20$. Here, it represents the time between calls to the 911 operators. This pdf is a valid pdf since it is non-negative, and integrates to one:**

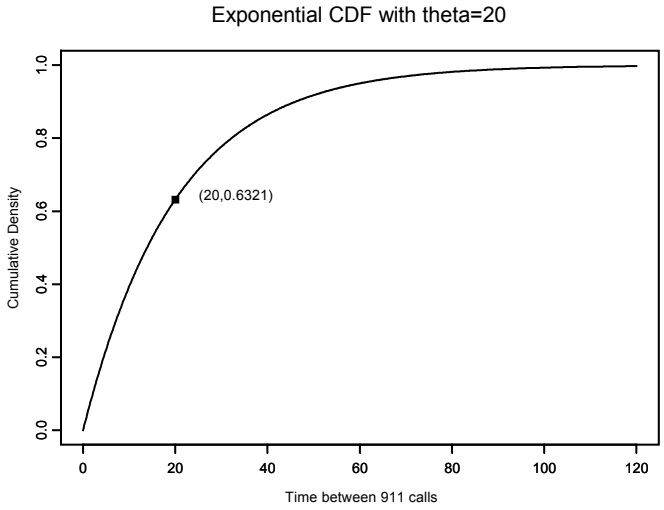
$$\int_0^\infty \frac{1}{20}e^{-x/20}dx = -e^{-x/20}]_0^\infty = 0 - (-1) = 1$$

**A graph of this pdf is below at left. Note that**

$$\mathbf{P(0 < X < 20)} = \int_0^{20} \frac{1}{20} e^{-x/20} dx = -e^{-x/20}\Big]_0^{20} = 1 - e^{-1} = 0.6321$$

**This probability is shown below at left (pdf plot) as the indicated *area* and at right (cdf plot) as the indicated *point*.**




**Another example is given at the top of p.114 and graph at the center of that page. Also, a more concise way to write the pdf is (p.114):**

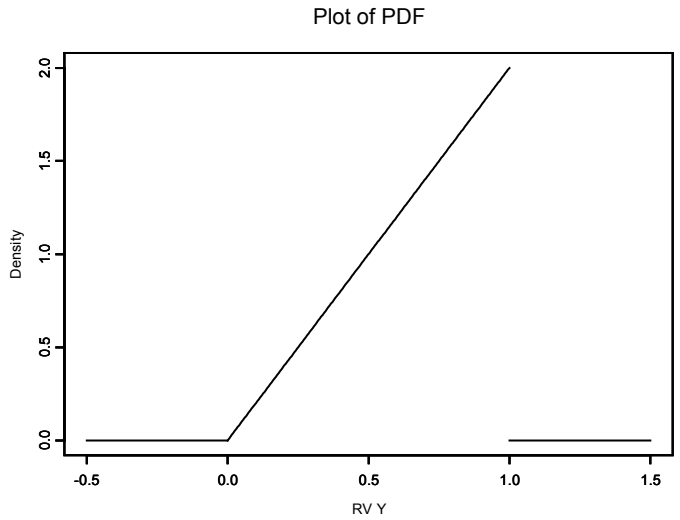
$$f(x) = \begin{cases} \frac{1}{20} e^{-x/20}, & 0 \le x < \infty \\ 0, & otherwise \end{cases}$$

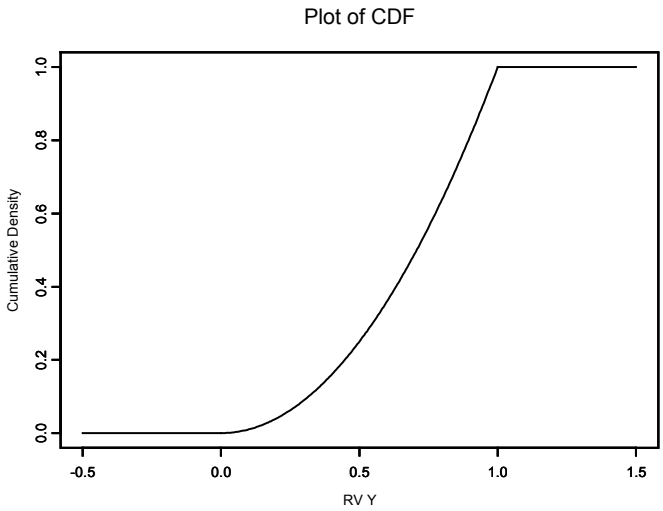
**The graph above at right is the cumulative distribution function (cdf):**

$$\mathbf{F(x) = P(X \le x)} = \int_{-\infty}^{x} f(t)dt$$

**In this example,** $F(x) = \int_0^x \frac{1}{20} e^{-t/20} dt = -e^{-t/20}\Big]_0^x = 1 - e^{-x/20}$ **; as mentioned, this cdf is plotted above at right. As indicated on p. 115, note that if you differentiate the cdf, you get the corresponding pdf.**

**Another example (see bottom of p.115): X has pdf $g(y) = 2y$ for $0 < y < 1$ (and $g(y) = 0$ otherwise), and cdf $G(y) = y^2$ for $0 < y < 1$. Also, $G(y) = 0$ for $x \le 0$ and $G(y) = 1$ for $y \ge 1$. See the graphs: the pdf is below at left and the cdf is below at right.**




**Next, we define the mean, variance, and SD of a continuous RV X:**

- **The mean is $\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$**
- **The variance is $\sigma^2 = Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$**
- **The variance short cut formula is $\sigma^2 = E(X^2) - \mu^2$**
- **The standard deviation (SD) is $\sigma = \sqrt{Var(X)}$**

**<u>Examples</u>: On pp.115-116, the text shows that for the pdf $g(y) = 2y$, the mean is $\mu = \frac{2}{3}$ and since $E(Y^2) = \frac{1}{2}$, the variance is $\sigma^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$.**

**Also, for $X \sim U(0,1)$ $f(x) = 1$ for $0 < x < 1$, $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$.**

For $p \in (0,1)$, the **$(100p)^{th}$ percentile** is the number $\pi_p$ such that

$$p = F(\pi_p) = \int_{-\infty}^{\pi_p} f(t)dt$$

The $50^{th}$ percentile is the median, the $25^{th}$ percentile is $Q_1$ (first quartile) and the $75^{th}$ percentile is $Q_3$ (third quartiles).

Some illustrations are on pp. 116-117. Also, for x in $(0,1)$, $g(y) = 2y$ so $G(y) = y^2$; by substituting ¼, ½, ¾ for p (in turn) in the above expression with this CDF, we obtain $Q_1 = \pi_{0.25} = ½ = 0.50$, median = $\pi_{0.50} = \frac{\sqrt{2}}{2} = 0.7071$, and $Q_3 = \pi_{0.75} = \frac{\sqrt{3}}{2} = 0.8660$.

## Special Continuous Distributions (Section 3.3, pp. 118-125)

In this section, we discuss four very important continuous distributions: the uniform, exponential, gamma, and chi-square.

**Uniform distribution.** Turning to p.125, ex.3.3-3, let $Y \sim U(0,1)$, and consider $W = a + (b-a)Y$. Since $\mu_Y = ½$ and $\sigma_Y^2 = \frac{1}{12}$, we expect $\mu_W = \frac{a+b}{2}$ and $\sigma_Y^2 = \frac{(b-a)^2}{12}$, but how is W distributed? Let's find the cdf of W: $G(w) = P(W \le w) = P(a + (b-a)Y \le w) = P(Y \le \frac{w-a}{b-a}) = \frac{w-a}{b-a}$. By differentiation, the pdf of W is $g(w) = \frac{1}{b-a}$ for $a < w < b$ (and $g(w) = 0$ otherwise). Thus, $W \sim U(a,b)$. The complete cdf is given on p.118; graphs are given in the center of p.119 for $a = 0.30$ and $b = 1.55$ – so that $f(w) = 0.80$ for $0.30 \le w \le 1.55$. An important application of the uniform distribution is random number generation (p.119).

**Exponential distribution.** For the parameter $\theta > 0$, X has the Exponential $EXP(\theta)$ distribution if its pdf is

$$f(x)=\begin{cases}0, & x<0\\ \frac{1}{\theta}e^{-x/\theta}, & 0\le x<\infty\end{cases}$$

**By integration, it follows that the cdf is**

$$F(x)=\begin{cases}0, & -\infty<x<0\\ 1-e^{-x/\theta}, & 0\le x<\infty\end{cases}$$

**Since F(x) = P(X ≤ x), the Survival function here for 0 ≤ x < ∞ is S(x) = P(X > x) = $e^{-x/\theta}$ ; S(x) = 1 for x < 0.**

**Using integration by parts, one obtains (p.120) that the mean and variance are μ = θ and $\sigma^2 = \theta^2$ (so the mean and SD equal θ). Also, the median is θ×ln(2) = 0.6931×θ. See the graph on p.120.**

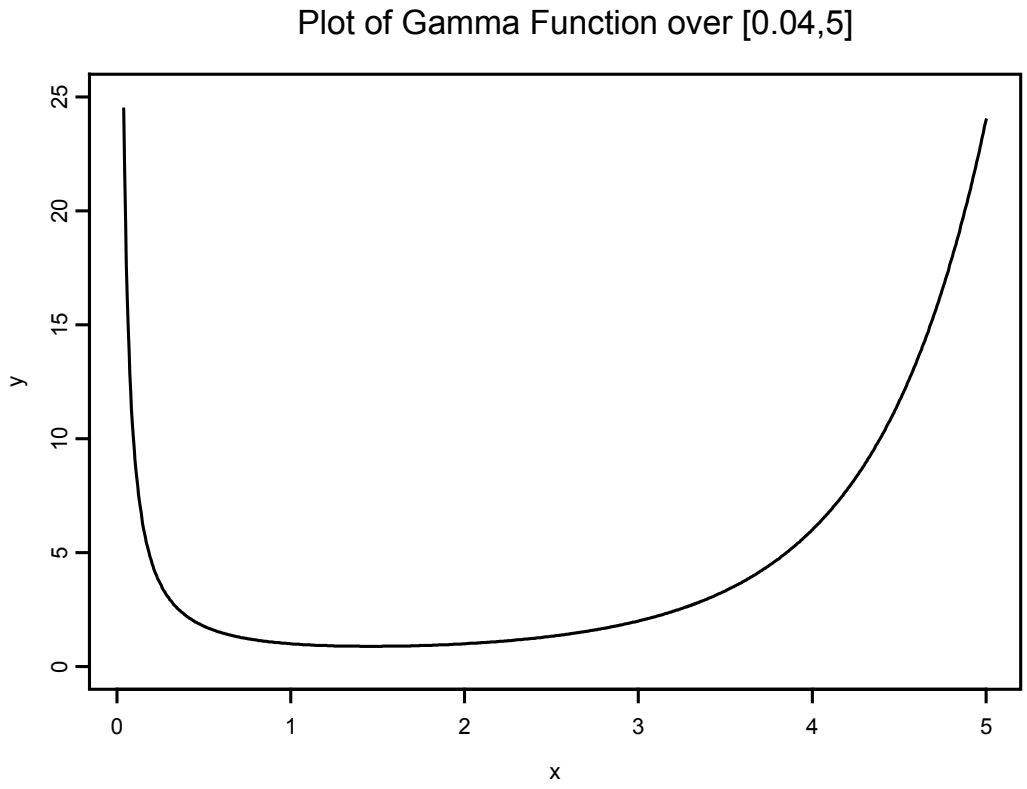
**Example 3.3-2 shows the connection between the Poisson and Exponential distributions: if the number of events per hour has a Poisson(λ = 20) distribution, then the waiting time until the first such event has the Exponential(θ = 3) distribution since 60/20 = 3 minutes.**

**Example 3.3-3 shows that for X an Exponential E(θ) RV,**

**P(X > 900 | X > 300) = P(X > 600)**

**This result is true in general: P(X > b | X > a) = P(X > b-a). This result follows easily and directly from the form of the Survival function for the Exponential distribution. Hopefully, this sounds familiar. Can you name this characteristic and draw a parallel between the Exponential distribution and a discrete distribution?**

**For t > 0, the *gamma function* is defined as Γ(t) = $\int_0^\infty y^{t-1}e^{-y}dy$ . It follows that Γ(t) = (t-1)Γ(t-1) and so for n an integer Γ(n) = (n-1)!**




**Gamma distribution. For $x < 0$, the Gamma$(\alpha, \theta)$ pdf is zero; for $0 \le x < \infty$ it is**

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta}$$

**Proving this is a valid pdf (i.e., *integrates to one*) follows by the change of variables $u = x/\theta$ and the definition of the gamma function. It's not hard to show that for $X \sim$ GAM$(\alpha, \theta)$, the mean is $\mu = \alpha\theta$, and since $E(X^2) = (\alpha+1)\alpha\theta^2$, the variance is $\sigma^2 = \alpha\theta^2$. To help us understand the roles of the two parameters, graphs are given on p.123.**

**Chi-Square distribution. The pdf at the bottom of p.123 (reproduced below) is for X a GAM$(\alpha = r/2, \theta = 2)$ RV, and it is the pdf of a chi-square RV with r degrees of freedom: for $0 \le x < \infty$,**

$$f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{(r/2)-1} e^{-x/2}$$

**We write $X \sim \chi^2(r)$, and "degrees of freedom" will be discussed later. The mean is $\mu = r$ and variance is $\sigma^2 = 2r$. Graphs are given in the center of p.124. The cdf has no closed-form solution, but Table IV on p.331 give the quantiles/percentiles for various values of r. To practice, work through Examples 3.3-4 and 3.3-5 on p.124.**

## The Normal Distribution (Section 3.4, pp. 126-133)

**One of the most popular distributions is the Normal distribution, and $X \sim N(\mu, \sigma^2)$ if its pdf is given by: for $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$,**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[ -\frac{(x-\mu)^2}{2\sigma^2} \right]$$

**The plot of this pdf gives the usual bell-shaped curve on p.127 centered at $\mu$ and with spread (SD) $\sigma$. If $X \sim N(\mu, \sigma^2)$, then the variable $Z = \frac{X-\mu}{\sigma}$ has the Standard Normal distribution, N(0,1). By exercise 2.2-8 on p.58, we know the mean and variance of Z are 0 and 1; Normality also carries over from X to Z. The CDF of Z cannot be given in closed form, but is tabled in Tables Va and Vb (pp.332-3).**

**Students should work through Examples 3.4-1 - 3.4-4 and understand how to work with Tables Va and Vb. To illustrate (see Example 3.4-3), if $X \sim N(3,16)$, then $P(2 < X < 5.2) = P(-0.25 < Z < 0.55) = 0.7088 - 0.4013 = 0.3075$. Also, the $95^{th}$ percentile is $3 + 1.645*4 = 9.58$.**

**Theorem 3.4-2 (p.129) states that the square of a Standard Normal RV has a $\chi_1^2$ distribution. We'll need this result later on. The proof is very instructive. Here, $X \sim N(\mu, \sigma^2)$, so $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ and $V = Z^2$. The trick is to (again) find the CDF and then PDF of V. The CDF is:**

$$G(v) = P(V \le v) = P(Z^2 \le v) = P(-\sqrt{v} \le Z \le \sqrt{v}) = \int_{-\sqrt{v}}^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

**Since the integrand above at right is an even function,**

$$G(v) = 2\int_{0}^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

**By differentiation, the PDF for $v \ge 0$ is**

$$g(v) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{v}} e^{-v/2} = \frac{1}{\Gamma(1/2)2^{1/2}} v^{1/2-1} e^{-v/2}$$

**Since this is the PDF of a $\chi_1^2$ RV, $V \sim \chi_1^2$.**

**On pp.130-3, the authors discuss q-q (quantile-quantile) plots – also called NPP's (Normal Probability Plots) by many software packages. A NPP is used to assess how "near" a given sample is to Normality; in these plots, quantiles of the normal distribution are plotted against the corresponding quantiles of the sample, and Normality is declared if the points are linear. See Example 3.4-6 on p.131 and the graph. In Example 3.4-7 on p.132, the authors simulate data from six distributions, and the NPP's are given on pp.132-3. Most notably, for (c), the pdf is skewed to the right and the NPP is concave down, and for (f), the t-distribution is 'heavier-tailed' than the Normal, and the NPP has obvious 'outliers' from the line in the tails.**

## Estimation in the Continuous Case (Section 3.5, pp. 134-141)

**We return to parameter estimation using the maximum likelihood approach. Whereas the method of moments approach equates the sample statistic to the population parameter, the likelihood approach finds the parameter estimator in the parameter space that maximizes the likelihood function. Six examples are given here.**

**Example 3.5-0.** For a random sample from the EXP(θ) distribution, the likelihood is $L(\theta) = (1/\theta)^n \exp\{-\Sigma x_k/\theta\}$, so the log-likelihood is

$$LL(\theta) = -n\ln(\theta) - \Sigma x_k/\theta$$

Then $\partial LL(\theta)/\partial\theta = -n/\theta + \Sigma x_k/\theta^2$. When we set this equal to zero and solve, we get the MLE $\hat{\vartheta} = \overline{X}$ (the sample mean). Note that this MLE is unbiased since (p.82) $E(\hat{\vartheta}) = E(\overline{X}) = \mu = \theta$.

**Example 3.5-1.** For a random sample from the $N(\theta, \sigma^2)$ distribution *with $\sigma^2$ known* and parameter space $\Omega = \Re$, the log-likelihood is

$$LL(\theta) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{\sum_k (x_k - \theta)^2}{2\sigma^2}$$

Differentiation with respect to θ, setting to zero, and solving again yields the MLE $\hat{\vartheta} = \overline{X}$ - substituting this value into the second derivative confirms the maximum (second derivative is negative).

**Example 3.5-2.** For a random sample from the GAM(α, θ) distribution *with α known* and parameter space $\Omega = [0, \infty)$ the log-likelihood is

$$LL(\theta) = -n\ln\Gamma(\alpha) - \alpha n\ln(\theta) + (\alpha - 1)\sum_k \log(x_k) - \frac{\sum_k x_k}{\theta}$$

Differentiation with respect to θ and setting to zero yields $\hat{\vartheta} = \frac{\overline{X}}{\alpha}$;

again the MLE is unbiased since $E(\hat{\vartheta}) = E(\frac{\overline{X}}{\alpha}) = \frac{\alpha\theta}{\alpha} = \theta$.

We can also maximize likelihood expressions involving *two parameters*, and obtain the MLE vector – sometimes this is easy and sometimes not. To illustrate when it's tough, return to the previous GAM(α, θ) example but with both parameters unknown (see **Example 3.5-4** on p.138). The likelihood is difficult to maximize because of the

**Gamma function. In this case, it is easy to use the method of moments estimators (see the bottom of p.138). Here's an easier ML example:**

**Example 3.5-3. For a random sample from the N($\theta_1, \theta_2$) distribution and parameter space $\Omega = (-\infty < \theta_1 < \infty) \times (0 < \theta_2 < \infty)$, the log-likelihood is**

$$LL(\theta_1, \theta_2) = -\frac{n}{2}\ln(2\pi\theta_2) - \frac{\sum_k (x_k - \theta_1)^2}{2\theta_2}$$

**Differentiating this expression with respect to $\theta_1$ and setting to 0 gives**

$$\frac{\sum_k (x_k - \theta_1)}{\theta_2} = 0$$

**Differentiating $LL(\theta_1, \theta_2)$ with respect to $\theta_2$ and setting to 0 gives**

$$-\frac{n}{2\theta_2} + \frac{\sum_k (x_k - \theta_1)^2}{2\theta_2^{\,2}} = 0$$

**Solving these two expressions simultaneously yields the MLE's** $\hat{\vartheta}_1 = \overline{X}$ **and** $\hat{\vartheta}_2 = \frac{\sum_k (X_k - \overline{X})^2}{n} = V$**. Notice that the MLE for the variance divides** $\sum_k (X_k - \overline{X})^2$ **by n and not (n-1) – as a result, it is a biased estimate of $\theta_2$ (the variance) as we now show:**

$$\Sigma(X_k - \theta_1)^2 = \Sigma(X_k - \overline{X})^2 + n(\overline{X} - \theta_1)^2$$

**so**

$$E[\Sigma(X_k - \theta_1)^2] = E[\Sigma(X_k - \overline{X})^2] + nE[(\overline{X} - \theta_1)^2]$$

**The left-hand term equals $n\theta_2$ and the right-most term equals $\theta_2$, so $E[\Sigma(X_k - \overline{X})^2] = (n-1)\theta_2$. Hence, $E(\hat{\vartheta}_2) = \theta_2 - (\theta_2/n)$, and the bias is equal to $-\theta_2/n$. On the other hand,** $S^2 = \frac{\sum_k (X_k - \overline{X})^2}{n-1}$ **is an unbiased estimator of $\theta_2$.**

**A big issue in Estimation is the performance of MLE's in general and comments are made to 'regular' situations. A non-regular case illustration is the U(0,θ) example on the bottom of p.138. Many estimators – including ML estimators – have approximately Normal distributions. On p.139, the authors discuss these results as related to Exponential, Normal, Gamma, and Binomial distributions. It's always the case for a random sample of size n where n is *large enough* that $\bar{X}$ has approximately a N(μ, σ²/n) distribution. A discussion of how big is *large enough* is given on pp.140-1, as are CI's for μ and p.**

## The Central Limit Theorem (Section 3.6, pp. 142-146)

**On p.142, we are reminded that E($\bar{X}$) = μ and Var($\bar{X}$) = $\frac{\sigma^2}{n}$ from p.82. Thus, we standardize $\bar{X}$ by subtracting μ and dividing by the standard error σ/√n to get W = $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{Y-n\mu}{\sigma\sqrt{n}}$ where Y = ΣX$_k$. Thus, E(W) = 0 and Var(W) = 1. The Central Limit Theorem states that when the X's come from a random sample of size n, that the distribution of W approaches a N(0,1) distribution as n →∞.**

**In Example 3.6-1, the random sample of size n = 25 is from a distribution with μ = 15 and σ = 2, so the probability the sample mean is between 14.4 and 15.6 is approximately P{$\frac{14.4-15}{2/\sqrt{25}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{15.6-15}{2/\sqrt{25}}$} ≈ Φ(1.5) - Φ(-1.5) = 0.9332 – 0.0668 = 0.8664.**

**The random sample in Example 3.6-2 is of size n = 20 and from the U(0,1) distribution so μ = ½, σ = 1/√12, nμ = 10, σ√n = $\sqrt{20/12} = \sqrt{5/3}$ It follows that P(Y ≤ 9.1) ≈ Φ(-0.70) = 0.2420, and P(8.5 ≤ Y ≤ 11.7) ≈ Φ(1.32) - Φ(-1.16) = 0.9066 – 0.1230 = 0.7836.**

**Example 3.6-3 is similar to the above except here we first must find the mean and variance: $\mu = \frac{1}{20}x^5\Big]_0^2 = 1.6$, $E(X^2) = \frac{1}{24}x^6\Big]_0^2 = 8/3$ so $\sigma^2 = 8/75$, and then proceed as on p.144 to find the requested probability.**

**On p.144, the authors claim that the Normal approximation is usually very good for n around 25 or 30, but sometimes this isn't right. If the parent population is Normal, then the sample average or sum is always Normal for all values of n. In Example 3.6-4, from the graphs on p.145, apparently n = 4 is large enough for this U(0,1) distribution. When the underlying distribution is very skewed – such as the $\chi_1^2$ distribution in Example 3.6-5 – we need a much larger sample size to get approximate Normality for the sum Y (see graphs on p.146).**

## Approximations for Discrete Distributions (Section 3.7, pp. 148-151)

**The Normal distribution also provides a good approximation for certain discrete distributions such as the Binomial(n,p). The reason the Normal distribution provides a good approximation is the CLT and since Y ~ BIN(n,p) and we think of Y as the sum of n iid Bernoulli RVs. Since $\mu_Y = np$ and $\sigma_Y^2 = np(1-p)$, the CLT states that**

$$W = \frac{Y - np}{\sqrt{np(1-p)}}$$

**is approximately a N(0,1) RV. The approximation is quite good provided both $np \geq 5$ and $n(1-p) \geq 5$, and provided we use the *continuity correction* illustrated below. As in Example 3.7-1, Y ~ BIN(n = 10, p = ½), since $np \geq 5$ and $n(1-p) \geq 5$, $\mu = 5$ and $\sigma^2 = 2.5$, then $P(Y = 7) = P(6.5 < Y < 7.5) \approx P\{\frac{6.5-5}{\sqrt{2.5}} < W < \frac{7.5-5}{\sqrt{2.5}}\} \approx \Phi(1.58) - \Phi(0.95) = 0.9429 - 0.8289 = 0.1140$. The actual value is P(Y = 7) =**

$\binom{10}{7}(½)^7(½)^3 = 0.1172$, so the Normal approximation is very close. We continue this setting in Example 3.7-3, and find

$$P(3 \le Y < 6) \approx P\{\frac{2.5-5}{\sqrt{2.5}} < W < \frac{5.5-5}{\sqrt{2.5}}\} \approx \Phi(0.32) - \Phi(-1.58)$$
$$= 0.6255 - 0.0571 = 0.5684$$

Just for practice, students are asked to work through Example 3.7-4 on pp.149-50 where n = 36 and p = ½ so $\mu = 18$ and $\sigma^2 = 9$. Explicit expression of the continuity correction is given in the center of p.150 where ½ is added or subtracted to k.

The Normal distribution also well approximates the Poisson($\lambda$) distribution so if $Y \sim POI(\lambda)$, then the RV $W = \frac{Y-\lambda}{\sqrt{\lambda}}$ approaches the standard Normal distribution. Thus, in Example 3.7-5 where $Y \sim POI(\lambda = 20)$, then

$$P(16 < Y \le 21) \approx P\{\frac{16.5-20}{\sqrt{20}} < W < \frac{21.5-20}{\sqrt{20}}\} \approx \Phi(0.34) - \Phi(-0.78)$$
$$= 0.6331 - 0.2177 = 0.4154$$

Again, very close to the correct value of 0.4226.