

Biost 517
Applied Biostatistics I
.....
Scott S. Emerson, M.D., Ph.D.
Professor of Biostatistics
University of Washington

Lecture 9:
Probability

October 28, 2011

1

© 2002, 2003, 2005 Scott S. Emerson, M.D., Ph.D.

Lecture Outline
.....

- Overview
- Properties of Probabilities
- Distributions of Random Variables
 - Binomial, Poisson, Exponential, Normal
- Inference Based on Summary Measures
- Sampling Distributions

2

Use of Probability
.....

- Recall that in statistical analysis we use probability in two ways
 - Probability models describe nondeterministic nature of measurements
 - Summary measures define “tend to”
 - Probability is used to quantify the uncertainty in the results (conclusions) of our statistical analysis

3

Distributions of Statistics
.....

- Probability theory relevant to quantifying uncertainty in our conclusions
 - Distributional theory for estimators and statistics across replicated experiments
 - Properties of expectation, variance, covariance
 - Asymptotic (large sample) results: Normal distribution
 - Criteria for “best” estimators and statistics
 - Bias (expectation), variance

4

Special Distributions

.....

- Common probability distributions used in probability models
 - Binomial, Poisson, Exponential, (Normal), etc.

- Probability distributions that are important for the sampling distribution of statistics
 - Normal
 - (Other sampling distns: t, F, chi squared)

5

Properties of Probabilities

.....

6

Axioms of Probability

.....

- Probabilities are between 0 and 1
 - $0 \leq Pr(A) \leq 1$

- Probability of the sample space is 1
 - $Pr(\Omega) = 1$ (something has to happen)

- Probabilities of disjoint events add
 - If events A and B cannot happen simultaneously, then

$$Pr(A \text{ or } B) = Pr(A) + Pr(B)$$

7

Properties of Probability

.....

- Probability of the complement
 - $Pr(A^c) = 1 - Pr(A)$

- Probability of two simultaneous events
 - $Pr(A \text{ and } B) \leq Pr(A)$

- Probability of at least one event
 - $Pr(A) \leq Pr(A \text{ or } B) \leq Pr(A) + Pr(B)$
 - $Pr(A \text{ or } B) = Pr(A) + Pr(B) - Pr(A \text{ and } B)$

- Probability of a partition
 - $Pr(A) = Pr(A \text{ and } B) + Pr(A \text{ and } B^c)$

8

Conditional Probability

.....

- Probability of event A given B occurred
 - $Pr(A | B) = Pr(A \text{ and } B) / Pr(B)$
- Conditional probabilities are probabilities
 - $0 \leq Pr(A | B) \leq 1$
 - $Pr(B | B) = 1$
 - If $Pr(A_1 \text{ and } A_2 \text{ and } B) = 0$ (mut excl given B), then
 $Pr(A_1 \text{ or } A_2 | B) = Pr(A_1 | B) + Pr(A_2 | B)$
 - $Pr(A^c | B) = 1 - Pr(A | B)$

9

Bayes' Rule

.....

- Derived properties
 - $Pr(A \text{ and } B) = Pr(A | B) \times Pr(B)$
 - $Pr(A \text{ and } B) = Pr(B | A) \times Pr(A)$
 - $Pr(B) = Pr(B | A) \times Pr(A) + Pr(B | A^c) \times Pr(A^c)$
- Bayes' Rule

$$Pr(A | B) = \frac{Pr(A \text{ and } B)}{Pr(B)} = \frac{Pr(B | A)Pr(A)}{Pr(B | A)Pr(A) + Pr(B | A^c)Pr(A^c)}$$

10

Independent Events

.....

- Definition: Events A, B are independent if
 - $Pr(A \text{ and } B) = Pr(A) \times Pr(B)$
- Properties
 - $Pr(A | B) = Pr(A)$
 - $Pr(B | A) = Pr(B)$
- NOTE: "Independence" and "mutually exclusive" are in some sense opposites

11

Distributions of Random Variables

.....

12

Random Variables

.....

- The idea of making some measurement
- Measurement value can vary
- Probability distribution describes the relative likelihood of observing particular values
- Once we know the probability distribution, we know everything

13

Distribution Functions

.....

- Specification depends on type of variable
- Categorical variables
 - List the probability for each possible value
 - A formula available for binomial, poisson
 - If ordered, can give cumulative probabilities (cdf)
- Continuous variables
 - Density or cumulative distribution function
 - Formulas available for density and/of cdf
 - Sometimes one is easier than the other

14

Bernoulli Distribution

.....

- A binary (0-1) random variable
 - E.g., sex, vital status
 - Probability distribution involves only the parameter p :
 $0 \leq p \leq 1$
 - $Pr(X = 1) = p$
 - $Pr(X = 0) = 1-p$
 - Mean $E(X) = p$
 - Variance $Var(X) = p(1-p)$
- A binary random variable must have a Bernoulli distribution

15

Binomial Distribution

.....

- Counts the number of events in n independent trials of the same experiment
 - The sum of n independent Bernoulli variables
 - Probability distribution has parameters n, p
 - For $k = 0, 1, 2, \dots, n$

$$Pr(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

- Mean $E(X) = np$
- Variance $Var(X) = np(1-p)$

16

Poisson Distribution

.....

- Counts the events occurring at a constant rate λ in a specified time (and space) t
 - Independent intervals of time and space
 - E.g., Heart attacks in Seattle in 2005
 - Probability distribution has parameter $\lambda > 0$
 - For $k = 0, 1, 2, 3, 4, \dots$

$$\Pr(X = k) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}$$

- Mean $E(X) = \lambda t$
- Variance $Var(X) = \lambda t$

- Poisson approx to Binomial for low p

17

Exponential Distribution

.....

- Continuous positive random variable having constant hazard
 - Sometimes used in time-to-event
 - “Memorylessness” precludes wide application
 - Probability distribution has hazard $\lambda > 0$

$$\Pr(X \leq t) = 1 - e^{-\lambda t}$$

- Mean $E(X) = 1/\lambda$
- Variance $Var(X) = 1/\lambda^2$

- Poisson interarrival times are Exponential

18

Normal (Gaussian) Distribution

.....

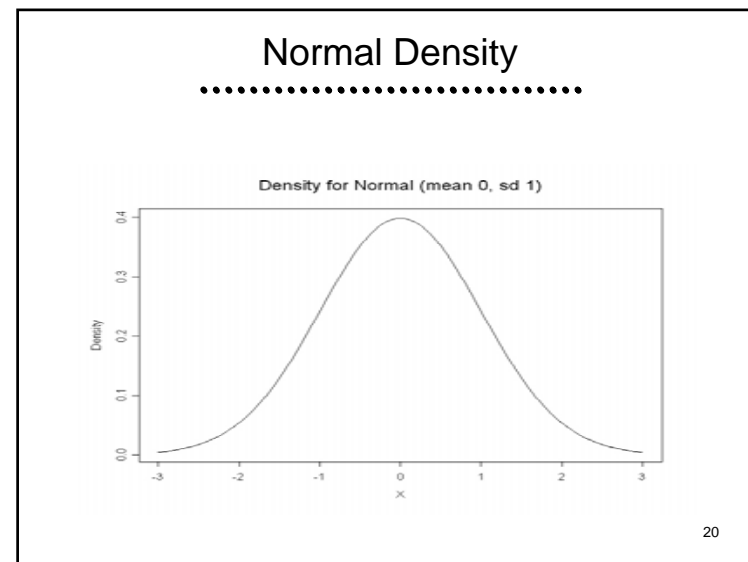
- The “bell-shaped curve”
- Arises from sums of multiple effects
 - (due to the central limit theorem (CLT))
- Density and cdf from

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

$$F(x) = \int_{-\infty}^x f(u)du$$

- Mean $E(X) = \mu$
- Variance $Var(X) = \sigma^2$

19

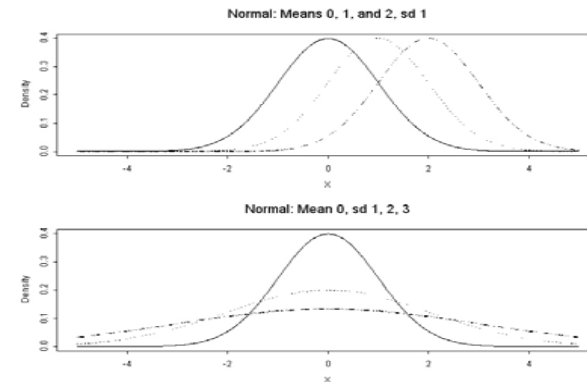


Linearity of Normal Distribution

- If $X \sim N(\text{mean } \mu, \text{var } \sigma^2)$, then for any constants a, b
 - Transformation $Y = aX + b$ is also normally distributed with
 - Mean $E(Y) = a\mu + b$
 - Variance $\text{Var}(Y) = a^2\sigma^2$

21

Varying Mean, Std Deviations



22

Standardizing Transformation

- Important special linear transformation
 - $Z = (X - \mu) / \sigma \sim N(0,1)$
 - (and Z is very common notation here)
 - Allows computation of probabilities
- The density for the normal distribution cannot be integrated in closed form, instead use
 - numerical integration,
 - tables in the back of any statistics text tabulate probabilities for $Pr(Z \leq c)$ (or equivalent)
 - statistical software

23

Stata Commands

- "normal"
 - "norm (c)" gives $Pr(Z \leq c)$ for $Z \sim N(0, 1)$
- Examples: Z normal with mean 0, std dev 1
 - $Pr(Z \leq 1.07)$: "display norm (1.07)"
 - $Pr(Z \geq 0.37)$: "display 1 - norm (0.37)"
 - $Pr(0.37 \leq Z \leq 1.07)$:
 - "disp norm (1.07) - norm (0.37)"

24

Stata Commands

.....

- For other normal distributions: Standardize
- Examples: X normal with mean 3, std dev 2
 - $Z = (X - 3) / 2$ is standard normal
 - $\Pr(X \leq 1.07)$:
 - `"disp norm ((1.07 - 3) / 2)"`
 - $\Pr(X \geq 0.37)$:
 - `"disp 1 - norm ((0.37 - 3) / 2)"`
- Examples: X normal with mean -4, std dev 5
 - $Z = (X - (-4)) / 5 = (X + 4) / 5$ is standard normal
 - $\Pr(X \geq 1.07)$:
 - `"disp 1 - norm ((1.07 + 4) / 5)"`

25

Sums of Independent Variables

.....

- Sums of independent variables (no matter the distribution)
 - For independent random variables
 - $X \sim (\text{mean } \mu, \text{var } \sigma^2)$, and
 - $Y \sim (\text{mean } \tau, \text{var } \theta^2)$
 - The sum $X + Y$ has moments (mean and variance)
 - mean = $\mu + \tau$, and
 - variance = $\sigma^2 + \theta^2$

26

Sums of Normal Variables

.....

- Sums of independent normally distributed variables are also normally distributed
 - For independent random variables
 - $X \sim N(\text{mean } \mu, \text{var } \sigma^2)$, and
 - $Y \sim N(\text{mean } \tau, \text{var } \theta^2)$
 - The sum $X + Y$ is *normally distributed* with
 - mean = $\mu + \tau$, and
 - variance = $\sigma^2 + \theta^2$

27

Inference Based on Summary Measures

.....

28

Real World

.....

- Most often, we do not know the probability distribution for a random variable

- Scientific questions correspond to answering questions about unknown distributions
 - Prediction of individual measurements
 - Typical measurements, spread, etc.
 - Comparison of probability distributions across groups
 - Tendencies to be larger or smaller
 - Tendencies toward greater spread
 - Etc.

29

Parameters of Distributions

.....

- Generally we must define what we mean by “typical”, “spread”, etc.

- Usually we use some summary measure of the population’s data in a manner analogous to using descriptive statistics for a sample
 - Mean, median, mode
 - Proportion above some threshold
 - Variance

- Summary measures of population distributions are often called “parameters”

30

Defining “tends to”

.....

- Tendencies toward “larger” and “smaller” measurements based on comparing parameters of the distribution

- E.g., Scientific question:
 - Does the population of smokers tend to have smaller life expectancies than nonsmokers?

- Corresponding statistical question:
 - Is the median (or mean, etc.) life expectancy smaller for smokers compared to nonsmokers?

31

Justification

.....

- Do statistical questions about parameters really answer the scientific question: Yes and No
 - Certainly if parameters differ across populations, then the distributions differ

 - But: Two populations may have different distributions, but the same mean (median, ...)
 - We need to be sure that a parameter used for inference captures what we care about

32

Expectation of Variables

- Expectation of a random variable is simply its average value over the population
 - “Average” = “Mean” = “Expected Value”
 - For random variable X , this is written $E(X)$
 - For the same reasons that we used sample means to describe a sample, the population mean is often used to
 - Describe “typical” values in the population
 - Compare distributions across populations

33

Other Population Parameters

- Median: $Mdn(X)$ is just the value such that half the population is above it and half below it
 - More generally we can describe any quantile (percentile) for the distribution of X in the population
- Proportion of the population exceeding some threshold
- Variance: $Var(X)$ is just the variance of the measurement across the entire population

34

Scientific Uses of the Mean

- Allows prediction of totals for population
 - E.g., average health care costs can be used to predict total costs for larger populations
- Often sensitive to a wide variety of differences in distributions
 - Sensitive to outliers as well as shifts of the entire distribution

35

Statistical Uses of the Mean

- Estimates of the mean are easy to use
- From mathematical theory we know
 - How to estimate the mean in an unbiased and/or consistent manner
 - How to quantify the precision of the estimates
- Easy to compute and combine across studies
 - Often the most efficient parameter estimated
 - Bernoulli, binomial, Poisson, exponential, normal

36

Scientific Uses of the Variance

.....

- Some scientific questions are related to the variability of the data
 - Quality control

37

Statistical Uses of the Variance

.....

- Because we often try to make inference about the population mean, we also tend to be interested in the population variance
 - The population variance can often be used to calculate the precision of our inference about the population mean

38

Sampling Distributions

.....

39

Sampling Distributions

.....

- We often quantify our confidence in study conclusions according to what would happen in replications of the experiment
- The distribution of a statistic across such replications is called the “sampling distribution” of the statistic
- E.g., Having observed 20% of smokers dying within a year, we might ask the probability of observing other results if we repeated the experiment many times

40

Statistical Importance

.....

- Both “frequentist” and “Bayesian” inference need sampling distributions
- Frequentist inference:
 - How often does data like this (or more extreme) happen under a particular hypothesis?
 - Purely a question about the sampling distribution
- Bayesian inference:
 - Given our data, what is the probability that a particular hypothesis is true?
 - Uses sampling distribution along with “prior distribution” (prevalence) in Bayes’ rule

41

Bias and Variability of Statistics

.....

- Often the exact sampling distribution for an estimator or test statistic is unknown or too difficult to compute
- Theory about expectations allows us to describe general tendencies of statistics
 - Unbiasedness: If we repeated the study many times, would the resulting statistics tend to be centered on the true parameter?
 - Variability: How variable would the estimates be across replicated experiments?

42

Using Properties of Expectation

.....

- We sometimes want to combine statistics reported in the literature
 - Perform *ad hoc* adjustment for covariates
 - Perform *ad hoc* meta-analyses across studies
- The basic properties of expectation often allow us to perform such informal analyses

43

Linearity of Expectation

.....

- For any
 - random variables X , Y , and
 - constants a , b , c
$$E(aX + bY + c) = aE(X) + bE(Y) + c$$
- (There is no requirement of independence or identical distribution)

44

Sample Mean is Unbiased

.....

- Across replicated studies, the average sample mean will tend to be the true mean
 - Population in which the average value is μ
 - Many studies each with n measurements
 - In each study calculate the sample mean
 - Average of those sample means tend to μ

45

Properties of Variances

.....

- The properties of expectation also help us describe the variability of statistics
- Recall that variance is the average squared distance from the mean
 - The fact that it is an average allows us to use the properties of expectation
 - The fact that it involves squaring the data means that we have to be careful as we use those properties

46

Variance: Linear Transformations

.....

- For any
 - random variable X , and
 - constants a , c

$$\text{Var} (aX + c) = a^2 \text{Var} (X)$$

- (Recall that the variance involves subtracting off the mean and then squaring)

47

Variance of Sums of Variables

.....

- Variance of sums (and differences) of random variables
 - For any random variables X , Y with
 - $\text{Corr} (X, Y) = \rho$ (the Greek letter rho)

$$\text{Var} (X + Y) = \text{Var} (X) + \text{Var} (Y) + 2\rho\sqrt{\text{Var} (X) \text{Var} (Y)}$$

$$\text{Var} (X - Y) = \text{Var} (X) + \text{Var} (Y) - 2\rho\sqrt{\text{Var} (X) \text{Var} (Y)}$$

48

Sums of Independent Variables

.....

- Variance of sums (and differences) of independent random variables
 - For any independent random variables X, Y
 - (so $Corr(X, Y) = 0$)

$$Var(X + Y) = Var(X) + Var(Y)$$

$$Var(X - Y) = Var(X) + Var(Y)$$

- (recall that multiplying Y by -1 would multiply the variance by the square of -1)

49

Variability of Sample Means

.....

- Computed across replicate experiments
 - Suppose we sample n independent measurements from a population in which
 - the average value in the population is μ , and
 - the variance in the population is σ^2
 - Across a large number of replicated studies
 - the average sample mean will be μ , and
 - the variance of the sample means will be σ^2 / n

50

Correlated Observations

.....

- If the measurements are not independent of one another, then the preceding formula does not hold
 - Positively correlated measurements lead to greater variability of a group's sample means
 - Repeated measurements made on the same individual, family, etc. tend to be positively correlated
 - Negatively correlated measurements lead to less variable sample means for a group

51

Standard Errors of Statistics

.....

- Recall that we usually find it easier to use the square root of variances
 - Variance is in squared units
- “Standard errors” are the square root of the variance of a statistic
 - The major motivation for the nomenclature is to distinguish it from the standard deviation in the population of measurements

52

Properties of Standard Errors

.....

- Must be derived from variances
- We compute standard errors by first finding the sampling variance of the statistic
- When transforming or combining statistics
 - Convert standard errors to variances by squaring
 - Use properties of variances
 - (These properties are derived from expectations)
 - Convert resulting variance back to a standard error

53

Using Standard Errors

.....

- Knowing the bias and standard error of statistics to estimate sampling distribution
 - From properties of standard deviations
 - At least 89% of a population must be within 3 standard deviations of its mean (Lec 4, slide 20)
 - Thus, for any given statistic, we would expect the statistic to be within 3 standard errors of its mean in at least 89% of all studies
- (We would, however, prefer more precise information about the sampling distribution)

54

Asymptotic Theory

.....

- When the exact sampling distribution is unknown or too difficult to compute
 - Theoretical results allow statistical inference in the setting of sufficiently large sample sizes
 - Most of this asymptotic theory is based on some form of a central limit theorem for the distribution of a sum or arithmetic mean
 - “Asymptotics” = as the sample size becomes infinite

55

Central Limit Theorem

.....

- Sample means are asymptotically normally distributed

Data (X_1, X_2, \dots, X_n) are
 independent,
 identically distributed,
 with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$

For large n : $\bar{X} \sim N\left(\text{mean } \mu, \text{var } \frac{\sigma^2}{n}\right)$

56

Ex: Binomial Distribution

.....

- Recall that Binomial random variables can be viewed as the sum of n independent, identically distributed Bernoulli variables

Independent Bernoulli random variables

$$Y_1, \dots, Y_n \text{ with } \begin{cases} \Pr(Y_i = 1) = p \\ \Pr(Y_i = 0) = 1 - p \end{cases}$$

$$X = (Y_1 + \dots + Y_n) \sim \text{Binomial}(n, p)$$

57

Ex: Binomial Distribution

.....

- Application of CLT: Normal approximation to binomial

Central limit theorem : For large n ,

$$\frac{X}{n} = \frac{1}{n}(Y_1 + \dots + Y_n) = \bar{Y} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Properties of normal :

$$X \sim N(np, np(1-p))$$

58

Ex: How Large is Large?

.....

- “Rules of Thumb” for the normal approximation to the binomial:
 - Based on ensuring not many “outliers”
 - $np > 5$ and $n(1-p) > 5$; OR
 - $np(1-p) > 5$

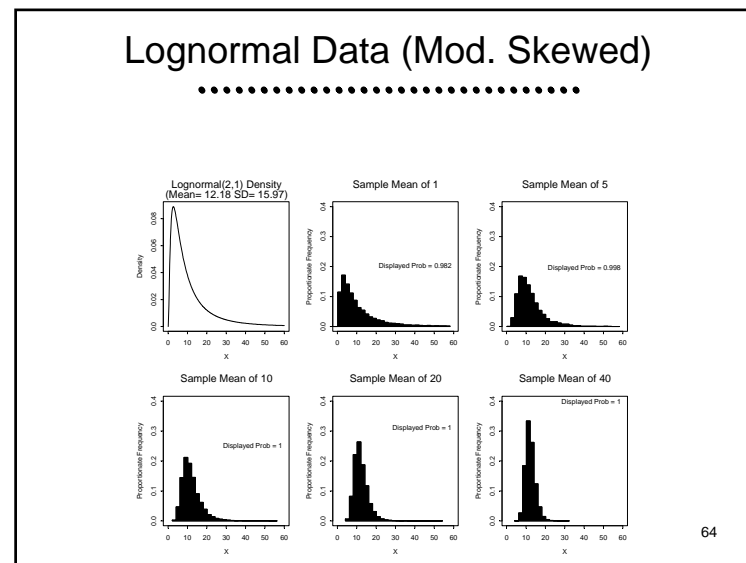
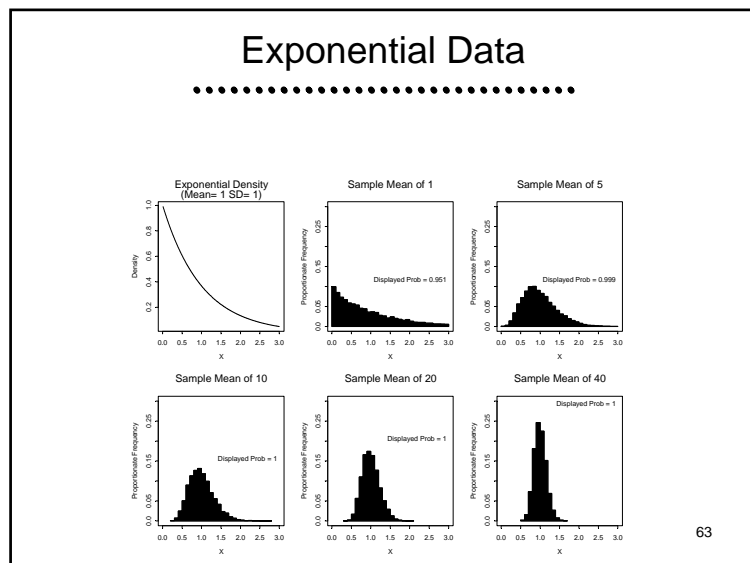
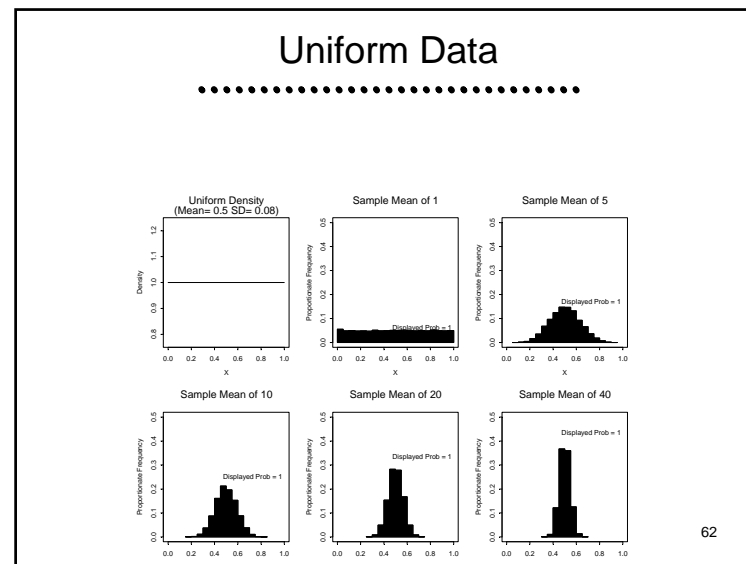
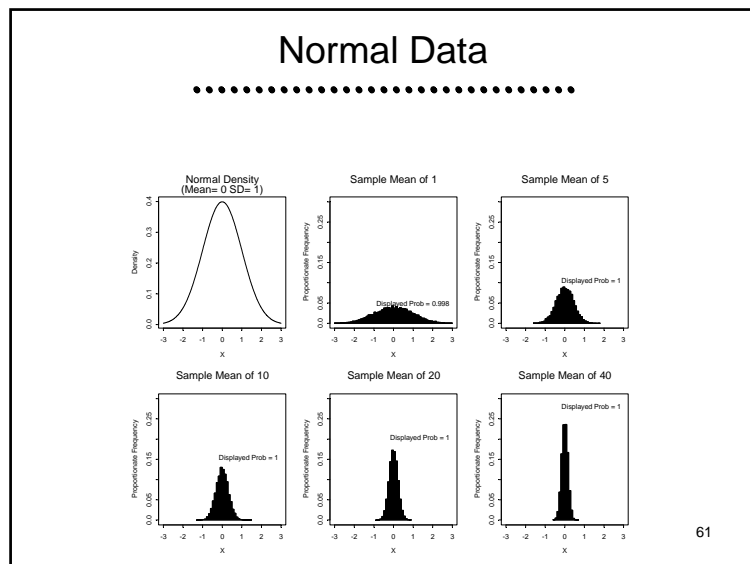
59

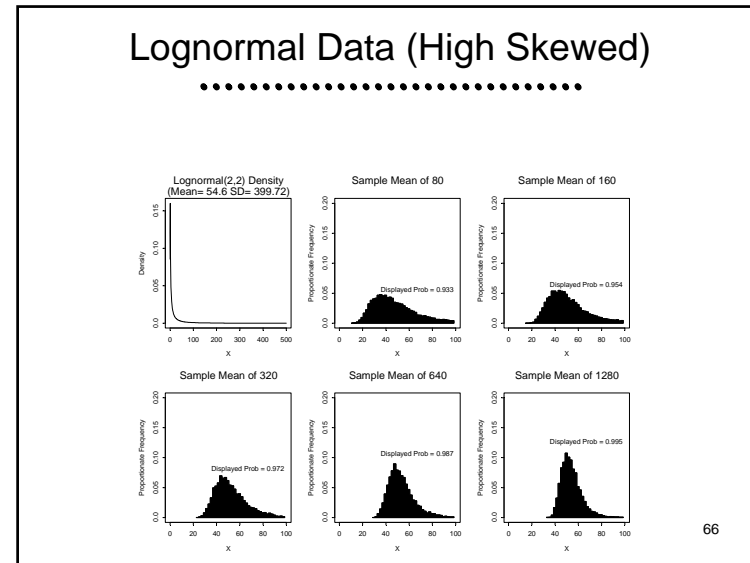
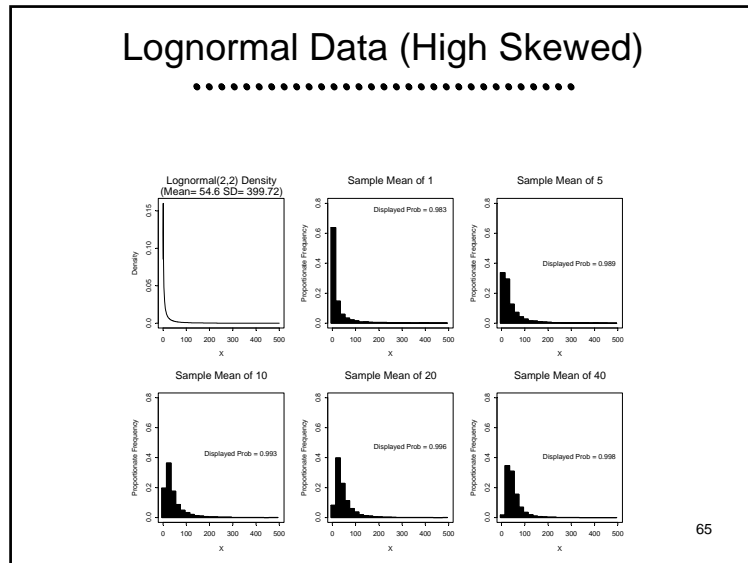
Limitations of CLT

.....

- The definition of “large” sample size will be dependent upon the original distribution
 - Distributions with heavy tails require larger sample sizes for a good approximation
- That having been said:
 - It is often surprising how small “large” is
 - A sample size of 30 – 40 will suffice for most data sets commonly encountered in practice
 - (see Lumley, et al., *Annual Rev. Pub Health*, 2002)

60





- ### Other Central Limit Theorems
-
- CLTs exist in some other settings
 - Independent, but not identically distributed
 - Correlated observations
 - Transformations of sample means

 - We will generally leave it to those who write computer programs to get the formulas right
 - But we must be aware of the additional conditions beyond finite variance
 - And we will have to tell the computer programs of the need to use the special formulas
- 67