STAT 155 Introductory Statistics

Lecture 5: Density Curves and Normal Distributions (I)
A variable has 5 values. The mean is 125 and the standard deviation is 25. If a sixth number of 125 is added in, what is the new standard deviation?
A Tool Kit for Describing Dist. of Single Quan. Var.

- Plot your data:
  - Stemplot, histogram, time plot, boxplot
- Look for overall pattern (shapes) + striking deviations (outliers)
- Calculate appropriate numerical summaries about center and spread
  - Mean, median, mode
  - Range, percentiles, quartiles, standard deviation
  - Five-number summary, boxplot
  - IQR and outliers
- Make things easier with a smooth curve
Density Curve

• To calculate probabilities, we define a probability density function \( f(x) \).
• The curve that plots \( f(x) \) is called the corresponding density curve.
• \( f(x) \) satisfies:
  – \( f(x) \geq 0 \);
  – The total area under the curve representing \( f(x) \) equals 1.
Density Curves

- Describe the overall shape of distributions
- Idealized mathematical models for distributions
- Show patterns that are accurate enough for practical purposes
- Always on or above the horizontal axis
- The total area under the curve is exactly 1
- Areas under the curve represent relative frequencies of observations
Median and mean
Histograms vs. Density Curves

- Histograms show either frequencies (counts) or relative frequencies (proportions) in each class interval.
- Density curves show the proportion of observations in any region by areas under the curve.
- You can think of density curves as refined histograms when there are huge amounts of data.
Histogram vs. Density Curve

![Histogram and Density Curve Graphs]

- Left: Histogram of Grade Equivalent Vocabulary Score with bars.
- Right: Density Curve with smooth line overlaying the histogram bars.

*Grade equivalent vocabulary score range from 2 to 12.*
Center of a Density Curve

- The **mode** of a distribution is the point where the curve is the highest. Highest Point.

- The **median** is the point where half of the area under the curve lies on the left and the other half on the right. Equal Areas Point.

- The **mean** $\mu$ is the point at which the curve would balance if made out of solid material. Balance Point.
Mean of a Density Curve
Spread of a Density Curve

- **Quartiles** can be found by dividing the area under the curve into four equal parts
  - $\frac{1}{4}$ of the area is to the left of the 1st quartile, $Q_1$
  - $\frac{3}{4}$ of the area is to the left of the 3rd quartile, $Q_3$

- The **standard deviation** of a density curve is denoted by $\sigma$.
  - Not easy to calculate
Normal Distribution

- Pictorially speaking, a Normal Distribution is a distribution that has a symmetric, unimodal and bell-shaped density curve.

- The mean and standard deviation completely specify the curve.

- The mean, median, and mode are the same.
• The height of a normal density curve at any point $x$ is given by

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

$\mu$ is the mean

$\sigma$ is the standard deviation
Change of Curvature
Example: The normal distribution is the most important distribution in Statistics. Typical normal curves with different sigma (standard deviation) values are shown below.
Examples with approximate Normal distributions

- Height
- Weight
- IQ scores
- Standardized test scores
- Body temperature
- Repeated measurement of same quantity
- ...
The 68-95-99.7 Rule
The 68-95-99.7 Rule
Example: Young Women’s Height

• The heights of young women are approximately normal with mean = 64.5 inches and std.dev. = 2.5 inches.
Example: Young Women’s Height

• % of young women between 62 and 67?
• % of young women lower than 62 or taller than 67?
• % between 59.5 and 62?
• % taller than 68.25?
Standardizing and z-Scores

• an observation \( x \) comes from a distribution with mean \( \mu \) and standard deviation \( \sigma \)

• The standardized value of \( x \) is defined as

\[
z = \frac{x - \mu}{\sigma},
\]

which is also called a **z-score**.

• A z-score indicates how many standard deviations the original observation is away from the mean, and in which direction.

• Mean and S.D. of the distribution of \( z \)?
Example: Young Women’s Height

- The heights of young women are approximately normal with mean = 64.5 inches and std.dev. = 2.5 inches.

- In our class, there is a female student who is 68.25 inches tall, what is her z-score?
Effects of Standardizing

• Standardizing is a linear transformation. What are $a$ and $b$?

• Effects on shape, center and spread.

• The standardized values for any distribution always have mean 0 and standard deviation 1.

• Linear transformation: normal into normal.
The standard normal distribution

- The **standard normal distribution** is the normal dist. with mean 0 and standard deviation 1, denoted as $N(0,1)$.
- $N(0,1)$ can be treated as a benchmark.
- Any normal distribution can be related to $N(0,1)$ by a linear transformation.
- $Z: N(0,1)$
- What is the distribution for $X=a+bZ$?
Take Home Message

• Density curve
  – Center, spread
• Normal distributions and normal curves
• The 68-95-99.7 rule for normal distributions
• Standardizing observations
• The standard normal distribution