Chapter 3 Probability Distributions and Statistics

Section 3.1 Random Variables and Histograms

The value of the result of the probability experiment is called a **random variable**.

If you roll a die we can let X be the number of dots showing,

If we have a hand of three cards, X could be the number of clubs in the hand,

These examples are all finite discrete random variables.

Example
Roll a single die and count the number of rolls until a 6 comes up.

outcome	Y

Finally, the random variable can be continuous:

Example

You are dealt a hand of 3 cards. Find the probability distribution table for the number of hearts. Graph this in a histogram

3.2 Expected Value

The expected value for the variable X in a probability distribution is

$$E(X)=X_1 \cdot P(X_1)+X_2 \cdot P(X_2)+...+X_n \cdot P(X_n)$$

Example
You are dealt a hand of 3 cards. What is the expected number of hearts?

Example
A sample of mini boxes of raisin bran cereal was selected and the number of raisins in each box was counted. The results are shown in the table below.

# of boxes	13	14	15	16	17
# of raisins	10	14	19	16	12

Determine the appropriate random variable X and display the data in a probability histogram. What is the expected value of X?

What is the MEAN of X?

What X value occurred the most often?

What X value is in the middle?

Histograms and Averages:

The MEAN (expected value) is where the histogram "balances"

The MODE is the tallest rectangle.

The MEDIAN is where the area is cut in half.

Example

Find the mean, median and mode of the following test scores:

77, 46, 98, 87, 84, 62, 71, 80, 66, 59,
79, 89, 52, 94, 77, 72, 85, 90, 64, 70

```
L2        L3       L4      3  EDIT CALC TESTS    1-Var Stats L3
.0625     77       ------     1:1-Var Stats
.25       46                  2:2-Var Stats
.375      98                  3:Med-Med
.25       87                  4:LinReg(ax+b)
.0625     84                  5:QuadReg
------    62                  6:CubicReg
          71                  7↓QuartReg
L3 = {77,46,98,87...

1-Var Stats            1-Var Stats
x̄=75.1                 ↑n=20
Σx=1502                 minX=46
Σx²=116532             Q₁=65
Sx=14.01465398         Med=77
σx=13.65979502         Q₃=86
↓n=20                  maxX=98
■
```

Mean =   Median =

Mode =   Range =

Another way to measure spread?   QUARTILES

46 52 59 62 64 66 70 71 72 77 77 79 80 84 85 87 89 90 94 98

Q1 =   Q3 =   IQR =

Box and whisker plot

Example

A game consists of choosing two bills at random from a bag containing 7 one dollar bills and 3 ten dollar bills. The player gets to keep the money picked. How much should be charged to play this game to keep it "fair" (expected value of zero)?

Example

From a group of 2 women and 5 men a delegation of 2 is chosen. Find the expected number of women in the delegation.

3.3 Measures of Spread

3 3 3 3 3   2 2 3 4 4   0 0 5 5 5

POPULATION VARIANCE,

POPULATION STANDARD DEVIATION,

What do we mean by population?

This means everyone, so if ALL the members of the population are used to find the mean, we use the symbols μ and σ.

The mean from SAMPLE uses the symbol $\bar{x}$.

SAMPLE STANDARD DEVIATION,

When we have the probability, it is assumed we had the entire population to base it on, so it is appropriate to use μ and σ.

Example
Find the mean and standard deviation for the following distribution:

If X is a binomial random variable associated with a binomial experiment consisting of n trials with probability of success p and probability of failure q=1-p, then the mean (expected value) and standard deviation associated with the experiment are:

$$\mu=np \quad \sigma=\sqrt{npq}$$

Example

Let the random variable X be the number of girls in a 6 child family. Find the probability distribution table, the mean and standard deviation for the number of girls in the family.

X	P(X)

Example

A flash drive has a 0.8% probability of being defective. In a shipment of 1000 flash drives, what is the mean and standard deviation in the number of defective flash drives?

3.4 The Normal Distribution

Continuous random variables can take on any value.

Let t = time in seconds to run a race

Let w = weight of kitten in kg

Let L = length of a week-old bean plant

Let X = value where pointer lands.

$0 \leq X < 1$ and $P(0 \leq X < 1) = 1$

What is $P(X= \frac{1}{2})$?

What is $P(0 \leq X < \frac{1}{4})$?

What is $P(0.75 \leq X \leq 0.80)$?

Define a PROBABILITY DENSITY FUNCTION

Discrete finite variables - graph the probability as a histogram.

Each rectangle has a base of width 1 (centered on X) and the height was P(X).

So the area, length x height was the probability that X occurred.

AREA above our X value will be the probability that get that X value.

If we want to find the probability of a range of X values, we would add up the areas over the range of X values.

When we graph the probability distribution for a continuous variable we find a smooth curve.

Many natural and social phenomena produce a continuous distribution with a bell-shaped curve.
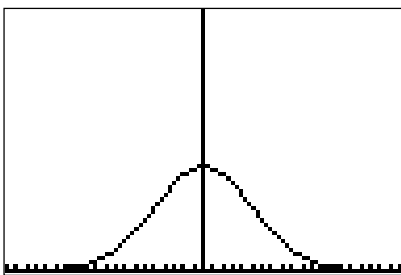
Every bell-shaped (NORMAL) curve has the following properties:
- Its peak occurs directly above the mean, μ
- The curve is symmetric about a vertical line through μThe curve never touches the x-axis. It extends indefinitely in both directions.
- The area between the curve and the x-axis is always 1 (total probability is 1).
- The curve switches from concave down to concave up one standard deviation away from the mean.

The shape of the curve is completely determined by μ and σ,

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
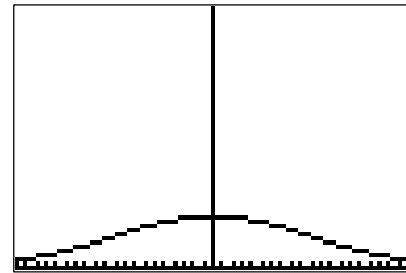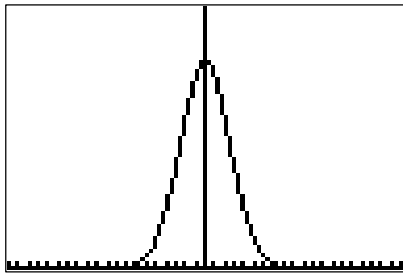
The probability that a data value will fall between x=a and x=b is given by the area under the curve between x=a and x=b.

The standard normal curve has $\mu = 0$ and $\sigma = 1$. Use Z, NOT X.

Example
On a standard normal curve, what is the probability that a data value is between -1 and 1, P(-1<z<1)?

Example

Suppose that X is a normal random variable with $\mu=50$ and $\sigma=10$.

a) What is the probability that X<30?

b) What is the probability that 35<X<65?

c) What is the probability that X > 55?

Example

An instructor wants to "curve" the grades in his class. The class mean at the end of the semester is 73 with a standard deviation of 12. He decides that the top 12% of the class should get an A, the next 24% should get a B, the next 36% a C, the next 18% a D and the last 10% of the class will get an F. What are the cutoffs for the grades?

A cutoff:

B cutoff:

C cutoff:

D cutoff:

Example
A machine that fills quart milk cartons is set to average 32.2 oz with a standard deviation of 1.2 oz.

a) What is the probability that a filled carton will have more than 32 oz?

b) If the store receives 500 quart milk cartons, how many will have more than 32 ounces?

c) What is the volume of milk that corresponds to the 25th percentile?

The mean height for 18 year old girls is 64.5 inches (50th percentile), with a standard deviation of 1.875 inches. These heights closely approximate the normal distribution. (this data is old)

a) What is the probability that a woman is shorter than 5' 3"?

b) In a group of 200 women, how many would you expect to be between 65" and 68"?

c) What is the probability that a woman is taller than 6'?

d) What height corresponds to the 90th percentile?