WEEK 10 REVIEW – Statistics

1, 1.5, 2

Example

Determine the possible values for the given random variable and indicate if the random variable is finite discrete, infinite discrete or continuous.

(a) A hand of 5 cards is dealt from a standard deck of 52 cards. Let X be the number of clubs in the hand of cards.

$X = 0, 1, 2, 3, 4, 5$

finite discrete

(b) A kitten is weighed. Let X be the weight of the kitten in pounds.

$X > 0$

cont.

kms

weight/mass

length

time

Temp

2.674

1.29663

or. . . .

(c) A single card is drawn without replacement from a standard deck of 52 cards. Let X be the number of cards drawn until a red card is picked.

$X = \underbrace{1, 2, 3, \ldots 26}_{\text{Black cards}}, 27$

finite discrete

(d) A single card is drawn with replacement from a standard deck of 52 cards. Let X be the number of cards drawn until a red card is picked.

$X = 1, 2, 3, \ldots\ldots$ inf. discrete

Example the prob adds to 1

Use the probability distribution table below to answer the following questions.

X	-2	-1	0	1	2	3
P(X)	0.15	0.05	0.2	0	z	0.25

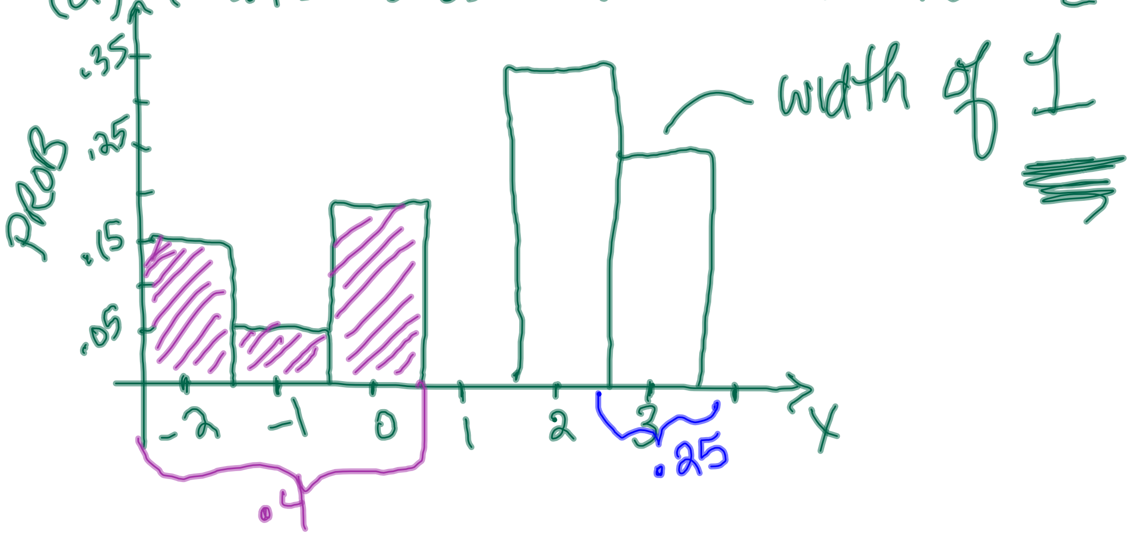
(a) Determine the value of z

(b) Represent this information in a histogram.

(c) $P(X \leq 0) = 0.4$

(d) $P(X > 2) = 0.25$

(a) $1 - 0.15 - 0.05 - 0.2 - 0 - .25 = .35 = z$



The expected value of a random variable X is given by

$$\boxed{E(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + \ldots x_n \cdot p_n}$$

where x_i represents the values that X can have and p_i is the probability that x_i occurs

Example

A raffle sells $\overbrace{\text{500 tickets at \$5}}^{2500}$ each. There is one first place prize awarded for \$1000, two second place prizes for \$100 each and 10 third place prizes at \$20 each. What is the expected value of a ticket in this raffle?

OUTCOME	X	P(X)
First	1000−5 =995	1/500
Second	100−5 =95	2/500
Third	20−5 =15	10/500
nothing	0−5=−5	487/500

$$E = (995)(1/500) + (95)(2/500) + (15)(10/500) + (-5)(487/500)$$
$$= -2.20$$
or −\$2.20

alt

$$E = (1000)(1/500) + (100)(2/500) + (20)(10/500) + (0)(487/500) - \underbrace{5}_{\text{cost of ticket}} = -2.20$$

Example

An Aggie ring is insured for \$1200. The annual premium on the ring insurance is \$15 and the probability that the ring will need to be replaced is 0.5%. What is the insurance company's expected gain on this policy?

outcome	X	P(X)
replace	−1200+15	0.005 ← $\frac{.5}{100}$
not replace	15	0.995

$$E = (-1185)(.005) + (15)(.995) = \$9$$

Example

One version of roulette has a wheel with 38 equally spaced numbers, 0 – 36 and 00. Half of the numbers from 1 – 36 are red and the other half are black (0 and 00 are green). Calculate the expected value of a $1 bet with the given payouts.

(a) Red. Pays $2.

(b) First dozen (1 – 12). Pays $3.

(c) 00. Pays $36.

a)

outcome	X	P(X)
red	2	18/38
not red	0	20/38

$$E = (2)(18/38) + (0)(20/38) - \underset{\text{to bet}}{\underline{\underline{1}}}$$

$$= -1/19$$

$$\Rightarrow -\$0.05$$

b)

outcome	X	P(X)
1-12	3	12/38
not 1-12	0	26/38

$$E = (3)(12/38) + (0)(26/38) - 1$$

$$= -1/19$$

$$= -\$0.05$$

c)

outcome	X	P(X)
00	36	1/38
not 00	0	37/38

$$E = (36)(1/38) + (0)(37/38) - 1$$

$$= -1/19$$

$$\approx -\$0.05$$

Example

A game consists of choosing two bills at random from a bag that contains two \$20 bills and ten \$1 bills. How much should be charged to play this game to keep it "fair" (expected value of zero)?

OUTCOME	X	P(X)
2 twenties	40	$\frac{C(2,2)C(10,0)}{C(12,2)} = \frac{1}{66}$
1 twenty and 1 one	21	$\frac{C(2,1)C(10,1)}{C(12,2)} = \frac{20}{66}$
2 ones	2	$\frac{C(2,0)C(10,2)}{C(12,2)} = \frac{45}{66}$

$$E = (40)(1/66) + (21)\left(\frac{20}{66}\right) + (2)\left(\frac{45}{66}\right) - \underset{\text{cost of ticket}}{\underline{t}} = 0$$

$$t = \$8.33$$

Example

A hand of 5 cards is dealt from a standard deck of 52 cards. Let X be the number of clubs in the hand of cards. Find the expected number of clubs.

OUTCOME	L1 X	L2 P(X)
0 clubs	0	C(13,0)C(39,5)/C(52,5) ≈ 0.2215
1 club	1	C(13,1)C(39,4)/C(52,5) ≈ 0.4114
2 clubs	2	C(13,2)C(39,3)/C(52,5) ≈ .2748
3 clubs	3	C(13,3)C(39,2)/C(52,5) ≈ 0.0815
4 clubs	4	C(13,4)C(39,1)/C(52,5) ≈ 0.0107
5 clubs	5	C(13,5)C(39,0)/C(52,5) ≈ 0.0005

Expected value ⟺ mean

1-var stats L1, L2

$$\bar{X} = 1.25 = E(X) = \mu$$

sample mean ↑ ↑ population mean

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

$$\text{“av”} = \frac{11}{10} = 1.1$$

Measures of Central Tendency

The **mean** of the n numbers $x_1, x_2, \ldots, x_n$ is $\mu = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$

The **median** of the n numbers $x_1, x_2, \ldots, x_n$ is the number in the middle when the n number are arranged in order of size and there are an odd number of values. When there are an even number of values, the median is the mean of the two middle numbers.

The **mode** of the n numbers $x_1, x_2, \ldots, x_n$ is the number that occurs the most often. If no number occurs more often than any other number, there is no mode. If two numbers both occur the most often, then there are two modes.

Example

Find the mean, median and mode for the given sets of numbers

(a) 6, 3, 0, 9, 1, 9, 6, 0, 6, 6, 1, 9, 0, 3, 6, 1, 1, 7, 8, 7, 7, 4, 2, 9, 9

L_3 (25 values)

1-var stats L_3

mean $= 120/25 = 4.8$

med $= 6$

mode is 6 and 9

(b) 6, 12, 3, 14, 9, 99

NO mode

L_4 (6 values)

$3 \quad 6 \quad \underbrace{9 \mid 12}_{} \quad 14 \quad 99$

$\dfrac{9+12}{2} = 10.5 =$ med

$$\text{mean} = \frac{\overset{\Sigma x}{143}}{6} \approx 23.8$$

The standard deviation is a measure of spread in a set of *n* numbers.

Example

Find the standard deviation and $\overset{\sigma_x^2}{\boxed{\text{variance}}}$ for the given sets of numbers

(a) 6, 3, 0, 9, 1, 9, 6, 0, 6, 6, 1, 9, 0, 3, 6, 1, 1, 7, 8, 7, 7, 4, 2, 9, 9

(b) 6, 12, 3, 14, 9, 99 L4

→ (a) in L3 . 1-Var stats L3

$\boxed{\sigma_x = \text{population std. dev} \approx 3.2373}$ Var $= \sigma_x^2 = 10.48$

S_x = sample std. dev ≈ 3.3040

(b) 1-Var stats L4

$\sigma_x \approx 33.8103$ Var $= \sigma_x^2 \approx 1143.14$

15, 20, 18, 18, 16, ...

Example

We are given the following data for the number of a certain magazine sold each week at a newsstand during the past year.

FREQ	# of weeks	5	4	8	11	9	15	L2
X	# of magazines	15	16	17	18	19	20	L1

(a) Represent this data in a frequency histogram.

(b) Find the mean, median, mode and standard deviation for this data.

1-var stats XLIST, FREQ LIST (PROB LIST)

mean = 18.1538

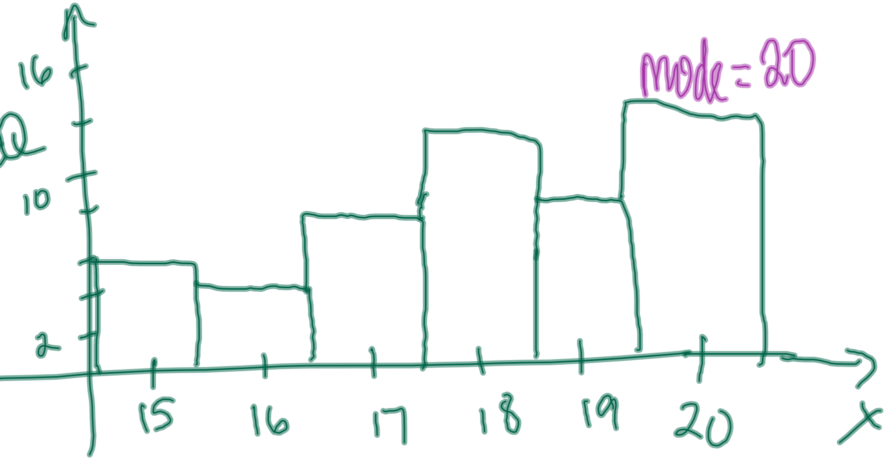
med = 18

n = 52

↑

a population

$\sigma = 1.6218$



Example

An group 200 apples were weighed and the following results were found. Note that x is the weight of an apple in grams.

L2 L1

Wt (gm)	# apples	midpt
$130 \le x < 140$	10	$(140+130)/2 = 135$
$140 \le x < 150$	95	145
$150 \le x < 160$	56	155
$160 \le x < 170$	33	165
$170 \le x < 180$	6	175

(n = 200)

Find the mean and standard deviation for this data.

1-Var Stats L1, L2 $\mu = 151.5, \sigma = 9.1515$

Example 7, 4, 3, ...

A sample of grapefruits is selected from a large shipment and the number of seeds in each grapefruit is counted. The following results were found

highest freq

FREQ	# grapefruits	6	7	8	9	10	L_2
X	# seeds	7	6	5	4	3	L_1

Find the mean, median, mode and standard deviation for this data.

What is X # grapefruits [✓ # seeds]
circle one

1-var stats L_1, L_2

$\bar{x} = 4.75$ $S = 1.4097$ med = 5

mode = 3

Example A PROB DIST ⇒ $\boxed{\sigma, \mu \text{ A POPULATION}}$

A hand of 5 cards is dealt from a standard deck of 52 cards. Find the standard deviation in the number of clubs in the hand of cards.

\# clubs in L_1, prob in L_2

1-var stats L_1, L_2

$\mu = 1.25$ $\sigma \approx 0.9295$ [note, no S displayed ☺]

The **odds in favor** of an event E occurring is the ratio of $P(E)$ to $P(E^c)$ or $\frac{P(E)}{P(E^c)} = \frac{P(E)}{1-P(E)}$

Given the odds in favor of an event E are a:b, the probability of E is given by $\frac{a}{a+b}$

Example

The odds in favor of selected horses in the 2009 Kentucky Derby were Backtalk 1:14 Homeboykris 1:60 Sidney's Candy 2:15

(a) What is the probability that Backtalk will win the race?

(b) What is the probability that Homeboykris will win the race?

(c) What is the probability that Sidney's Candy will win the race?

(a) $\frac{1}{1+14} = \frac{1}{15} \approx 6.7\%$

(c) $\frac{2}{2+15} = \frac{2}{17} \approx 13.3\%$

(b) $\frac{1}{1+60} = \frac{1}{61} \approx 1.6\%$

Example

Given that $P(A) = 0.4$, $P(B) = 0.6$, and $P(A \cap B) = 0.3$, find the odds in favor of

(a) A occurring

(b) only A occurring

2 to 3

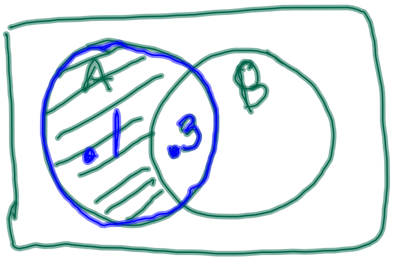
a) $\frac{P(A)}{1-P(A)} = \frac{(.4)}{1-.4} = \frac{.4}{.6} = \frac{2}{3} \Rightarrow 2:3$

$.4 : .6 \Rightarrow 2:3$

$P(A \cap B^c) = ? = .1$

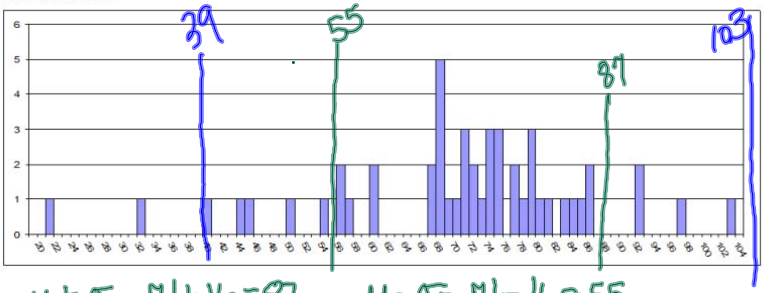
$\frac{.1}{1-.1} = \frac{.1}{.9} = \frac{1}{9} \Rightarrow 1:9$

1 to 9



Example

The following exam grades are from a class of 50 students. The mean is 71 and the standard deviation is 16. What is the probability that a randomly selected exam score is within 1 standard deviation of the mean? Within 2 standard deviations of the mean?



$\mu+\sigma = 71+16 = 87 \quad , \quad \mu-\sigma = 71-16 = 55$

$P(\mu-\sigma \leq X \leq \mu+\sigma) = P(55 \leq X \leq 87) = \dfrac{n(55 \leq X \leq 87)}{50 \text{ students}}$

$= 39/50 = 0.78$

$\mu+2\sigma = 71+2\cdot 16 = 103 \quad , \quad \mu-2\sigma = 71-2(16) = 39$

$P(\mu-2\sigma \leq X \leq \mu+2\sigma) = P(39 \leq X \leq 103) = 48/50 = .96$

To estimate the probability that a data value is within *k* standard deviations of the mean, use Chebychev's theorem,

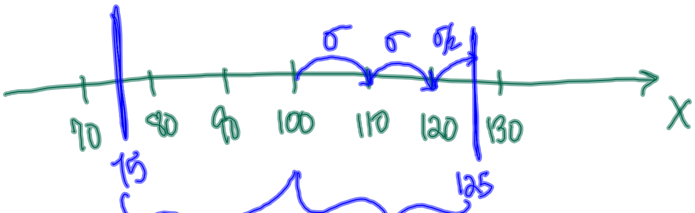
$$\underbrace{P(\mu - k\sigma \leq X \leq \mu + k\sigma)} \geq 1 - \boxed{\frac{1}{k^2}}$$

Example for $k>1$

A data distribution has a mean of 100 and a standard deviation of 10. Use Chebychev's theorem to find

(a) The probability that a data value is between 75 and 125.

(b) Find a value of c such that the probability that a data value is in the range 100-c to 100+c is 96%



$25 = 10\cdot k \qquad 25 = k\cdot 10$

$k = 2.5$

$P \geq 1 - 1/2.5^2 = 0.84$

(b) $.96 = 1 - 1/k^2 \Rightarrow \frac{1}{k^2} = 1 - .96 = 0.04$

$\sqrt{\frac{1}{k^2}} = \sqrt{.04} \Rightarrow \frac{1}{k} = .2 = \frac{1}{5} \Rightarrow k = 5$

go 5 std deviations or $5 \cdot 10 = 50$

$c = 50$ or

$P(100-50 \leq X \leq 100+50) \geq .96$