Stat 324 –
Introduction to Statistics for Engineers

LECTURE 4: DEFINING RANDOM VARIABLES AND 2 COMMON DISTRIBUTIONS

Random Variable Big Ideas

A **Random Variable** (RV) associates a numerical value with each outcome of an expirament.
It is customary to denote random variables with uppercase letters. The reason it is called "random" is because we don't know the value observed until the experiment is completed.

e.g X=weight of an ant chosen at random from Claire's Ant Farm

Y= number of heads in 3 tosses of a coin

Once the random process that defines an RV is performed, we call the result a realization of the RV. Realizations of RVs are usually denoted by lower-case letters, like x,y, etc

Think of an RV as representing a population and a sample as a collection of realizations of that RV

The **Probability Distribution** of a random variable consists of the RV's possible values along with the probabilities that each realization will occur. Depending on the type of RV (discrete vs continuous), the descriptions of the possible values and probabilities can take different forms.

Discrete RV Example

Discrete RVs only take a countable number of values. If the values are arranged in order, there is a gap between each value and the next. (The set of possible values may be infinite)

e.g. Number of Items (X) purchased by a customer, Number of pages (Y) in a book

probability distributions for discrete RVs are called **probability mass functions (pmfs:p(x))**, and consist of lists of the values that can be taken by the RV, together with the probabilities of each value.

Ex1: Define Y to be the number of heads obtained in three tosses of a coin. Create a pmf for X.

Outcome	P(Outcome)	Value of X
HHH	.5^3=.125	3
HHT	.5^3=.125	2
HTH	.5^3=.125	2
THH	.5^3=.125	2
TTH	.5^3=.125	1
THT	.5^3=.125	1
HTT	.5^3=.125	1
TTT	.5^3=.125	0

pmf

Value of X	Probability of X P(X=x)
0	P(X=0) = .125
1	P(X=1)=3/8=0.375
2	P(X=2) = 3/8 = .375
3	P(X=3)=1/8=0.125
Total	1

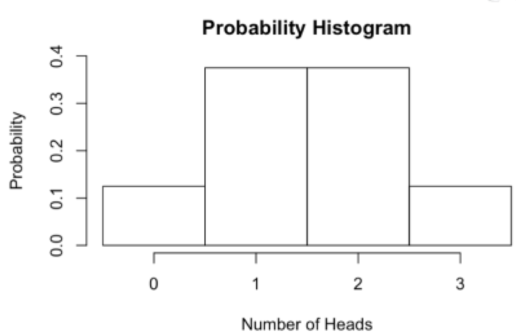
probability that X realizes to 3

Probability Histograms of Discrete RVs

A **probability histogram** can be used to display a pmf when the possible values of a discrete random variable are evenly spaced.

The area [and height in this case] of the rectangle centered at a value x is equal to P(X=x). We sum up the area of the rectangles to get the probability for a range of values.

Ex 1: Y=Number of Heads



b. use it to find P(0<X<3). What does this value mean?

$P(0 < X < 3)$

$= P(X=1) + P(X=2)$

$= .375 + .375$

$= .75$

Numerical Properties of Discrete RVs

The **expectation or expected value** of a RV X, denoted E(X) or μ_X is like the _____mean_____ of the population. It represents the mean of an infinite number of _realizations_ of X (infinite number of replications of experiment).

Consider the example of tossing 3 coins and and counting the number of heads. X=number of heads with a simulation

Here I tossed 3 coins and counted the number of heads 100 times.

```
(samp1<-rbinom(100, 3, .5))
 [1] 2 1 0 1 1 2 3 1 1 3 1 3 1 3 2 1 2 2 2 0 1 1 0 2 2 1 1 1 1 2 2 2 0 1 1 1 1 1 1 2 1 2 1 1 1 2 2 1 2 1 2
[52] 0 2 3 3 0 2 2 1 1 2 3 3 1 0 2 1 0 0 2 0 3 0 1 2 0 1 1 2 2 1 1 1 3 1 2 2 2 1 1 2 1 1 0 1 3 1 1 1 2
mean(samp1)
1] 1.4
```

Here I tossed 3 coins And counted the number Of heads 1000 times

```
samp2<-rbinom(1000, 3, .5)
samp2[1:100]
 [1] 1 2 2 2 0 1 2 1 1 1 2 2 2 2 2 2 0 1 2 1 2 1 3 1 1 1 2 1 1 1 1 2 1 2 2 2 2 2 1 3 1 2 3 2 2 2 1 2 1 3 2
[52] 1 1 2 1 2 2 2 3 1 2 2 2 1 2 2 1 1 1 1 1 2 2 1 3 2 1 0 2 1 0 3 1 1 2 2 2 0 1 2 2 3 2 0 2 0 1 1 1 1
mean(samp2)
1] 1.449
```

Here I tossed 3 coins And counted the number Of heads 1000000 times

```
> samp4<-rbinom(1000000, 3, .5)
> samp4[1:100]
  [1] 2 1 1 0 1 3 1 0 2 0 2 1 3 2 2 0 1 3 3 2 0 3 1 1 3 2 2 1 1 2 1 2 1 1 0 2 2 1 2 2 1 2 1 3 1 0 2 3 2 3 1
 [52] 1 0 1 2 2 0 1 2 3 3 2 2 1 2 1 3 1 1 2 3 2 2 1 2 0 2 1 3 1 2 3 0 2 1 3 0 1 1 2 3 1 2 1 0 2 1 1 2 1
> mean(samp4)
[1] 1.498812
```

Numerical Properties of Discrete RVs

The **expectation or expected value** of a RV X, denoted E(X) or μ_X is like the mean of the population. It represents the mean of an infinite number of realizations of X (infinite number of replications of experiment).

$$\mu_X = E(X) = \sum_x \overset{\text{each possible outcome}}{x} * \overset{\text{\# of outcomes}}{P(X = x)}$$

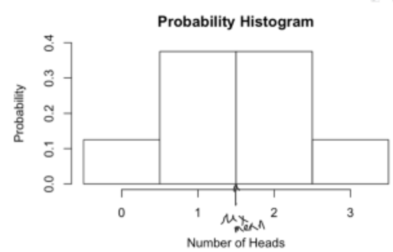
Ex 1: Consider the example of tossing 3 coins and and counting the number of heads. Y=number of heads

$\mu_Y = E(Y) = 0(.125) + 1(.375) + 2(.375) + 3(.125) = 1.5$

wont get 1.5 heads in expriment
only long term

Notice, the mean is the point on the horizontal axis at which the graph of the pmf _balances_____.

Value of Y	Probability of Y P(Y=y)
0	P(Y=0)=1/8=.5^3=0.125
1	P(Y=1)=3/8=0.375
2	P(Y=2)=3/8=0.375
3	P(Y=3)=1/8=0.125
Total	1



Numerical Properties of Discrete RVs

The **variance** of a RV X, denoted VAR(X) or σ_X^2 is like the variance of a _population_. It represents the variance of an infinite number of realizations of X (infinite number of replications of experiment).

The **standard deviation** of a RV X, denoted SD(X) or σ_X is the _square root_ of the population variance.

Again, consider the example of tossing 3 coins and and counting the number of heads. X=number of heads with a simulation

```
> #remember, R calculates the sample variance
> var(samp1); sd(samp1)
[1] 0.7272727
[1] 0.8528029
> var(samp2); sd(samp2)
[1] 0.7341331
[1] 0.8568157
> var(samp3); sd(samp3)
[1] 0.7363815
[1] 0.8581267
> var(samp4); sd(samp4)
[1] 0.7507133
[1] 0.8664372
```

```
> length(samp1);length(samp2);length(samp3);length(samp4)
[1] 100
[1] 1000
[1] 10000
[1] 1000000
```

Our sample estimates for the population sd and variance from a sample of 1000000 realizations of X are: estamite

$$var_X = \sigma_X^2 = 0.7507$$ And

$$sd_X = \widehat{\sigma_X} = 0.866$$

But, lets use our pmf to calculate them _exactly_!

Numerical Properties of Discrete RVs

$$\frac{\sum (x-\bar{x})^2}{n}$$

The **variance** of a RV X, denoted VAR(X) or σ_X^2 is like the variance of a population.

$$\sigma_X^2 = VAR(X) = \sum_x (x - E(X))^2 * P(X = x) = \sum_x x^2 * P(X = x) - E(X)^2$$

subtract each value we expect from mean * probablty of it happeng

$(\mu_X = E(X))$ population mean

Ex 1: Consider the example of tossing 3 coins and and counting the number of heads. X=number of heads

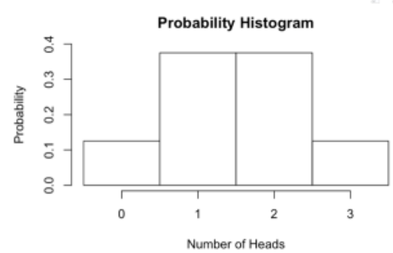
number, we expect

$\sigma_Y^2 = VAR(Y) =$ $(0-1.5)^2 * .125 + (1-1.5)^2 * .375 + (2-1.5)^2 * .375 + (3-1.5)^2 * .125 = .75$

$= \sqrt{.75} = .866$

And $\sigma_X = \sqrt{.75} = .866$

Value of X	Probability of X P(X=X)
0	P(Y=0)=1/8=.5^3=0.125
1	P(Y=1)=3/8=0.375
2	P(Y=2)=3/8=0.375
3	P(Y=3)=1/8=0.125
Total	1



A Famous Discrete RV : Binomial

Our Tossing Coin Example where we counted the number of Heads observed in the three flips is a special type of RV called a Binomial

individual coin flip called bernoulli trial multiple make a Bernoulli

A **Binomial Random Process** has the following properties:

counts number

1. The random process consists of n identical sub-processes (Bernoulli trials)
   - Each experiment had 3 identical flips

   a. Each Bernoulli trial is a RV and results in one of two possible outcomes (1 "success" or 0:"failure")

   b. The probability of a success on any single Bernoulli trial is the same for every trial, and is denoted π.
      - $Y_i \sim Bern(\pi = 0.5)$ with pmf: $P(Y = 1) = 0.5, P(Y = 0) = 0.5$
2. The trials are independent. The outcome of any trial doesn't affect the outcome of any other

The **Binomial** RV **B,** is the total number of sucesses achieved in n trials of a binomial random process with probability π of success on any given trial. We denote such a RV as $B \sim Bin(n, \pi)$.
The individual trial RVs:Y are Bernoulli.
In our coin tossing experiment $Y \sim Bern(\pi = 0.5)$ $and\ X \sim Bin(n = 3, \pi = 0.5)$.

prob of head

trials prob of sucess

A Famous Discrete RV : Binomial

Ex2: Based on several years of testing, it is determined that 96% of circuit boards are fully operational. A warehouse contains a very large population of boards. If 4 are selected at random, the distribution of X=the number of operational boards in that sample of 4 would be described by a binomial RV.

Why? Check the 4 [5] assumptions of a binomial process and identify the Bernoulli RVs (Y).

n = 4 fixed number of trials (checking a circuit board is trial)

bernoulli trials that are identical?

case 0 : not operational

case 1 : operational

is probability the same for each trial?

$\pi = 0.96$

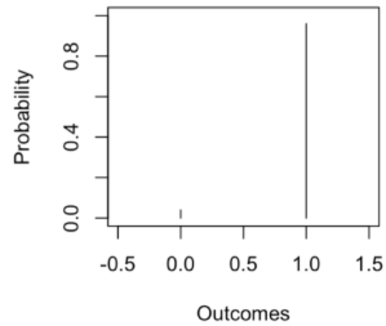
circuit board = $Y_i \sim Bern(0.96)$

are trials independent? OK

binomial = total # of sucesses

A Famous Discrete RV : Binomial

Ex 2: Y=the number of operational boards in a sample of 1 would be described by a Bernoulli. Y ~ Bern(0.96)



Pmf:

Y	P(Y=y)
0	0.04
1	0.96

Ex: What is the mean and variance for Y ~ Bern(0.96) ?

$E(Y) = (0 * .04) + (1 * .96) = .96$

$VAR(Y) = \pi$ and $Var(Y) = \pi(1-\pi) = .96(1-.96)$

In general, when Y is Bern(π): $E(Y) = \pi$ and $\mathrm{Var}(Y) =$ $\pi(1-\pi)$

A Famous Discrete RV : Binomial

Ex2: Based on several years of testing, it is determined that 96% of circuit boards are fully operational. X=the number of operational boards in that sample of 4 would be described by a binomial.

X = # of operational boards

Create a pmf for X

Values of X P(X=x)

x	P(X=x)
0	$(0.04)^4$
1	$4(0.96 * (0.04)^3)$
2	
3	
4	

$P(X=0) = NNNN = 0.04 * 0.04 * 0.04 * 0.04$

$P(X=1) = \left.\begin{matrix} TNNN \\ NTNN \\ NNTN \\ NNNT \end{matrix}\right\} = (0.96 * (0.04)^3) * 4$

$\binom{n}{b} \pi^b (1-\pi)^{n-b} = P(X=b)$

with n trials
and b successes

A Famous Discrete RV : Binomial

$\binom{4}{1} = \frac{4!}{1!(4-1)!} = \frac{4\cdot 3\cdot \cancel{2\cdot 1}}{1\cdot \cancel{3\cdot 2\cdot 1}} = 4$

The **Binomial** RV **B,** is the total number of successes achieved in n trials of a binomial random process with probability π of success on any given trial. We denote such a RV as $B \sim Bin(n, \pi)$.

The probability of observing b successes is:

$$p(b) = \binom{n}{b}\pi^b(1-\pi)^{n-b} = \frac{n!}{b!(n-b)!}\pi^b(1-\pi)^{n-b}$$

Where

$\binom{n}{b} = \frac{n!}{b!(n-b)!}$ calculates the number of outcomes with b S's and n! is the product of all numbers from n to 1. (By definition 0!=1).

A Famous Discrete RV : Binomial

Ex2 cont. X=the number of operational boards in that sample of 4 would be described by a binomial

$$P(X=2) = \binom{4}{2}(.96)^2 * (0.04)^2 =$$

$$\frac{4!}{2!\,2!}(.96)^2 * (0.04)^2 =$$

$$\frac{4*3*2*1}{2*1*2*1}(.96)^2 * (.04)^2 =$$

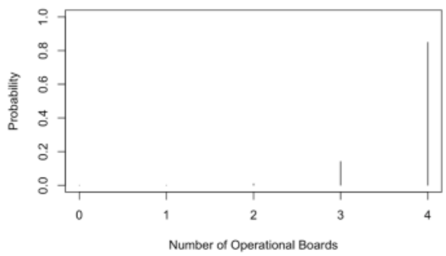
```
> dbinom(2,4, .96 )    ##This calculates P(X=2)
[1] 0.00884736
```

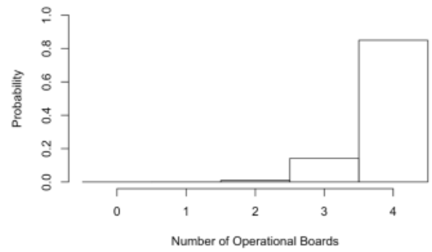
A Famous Discrete RV : Binomial

Ex 2: X=the number of operational boards in that sample of 4 would be described by a binomial. X ~ $Bin(4, 0.96)$

Pmf:

X	P(X=x)
0	0.00000256
1	0.00024576
2	0.00884736
3	0.1415578
4	0.8493466





Ex: What is the mean and variance for X ~ B(4, 0.96)?

$\mu_X = E(X) = n\pi = 4 * .96 = 3.84 \quad \sigma^2_X = n\pi(1-\pi) = 0.1536$

$E(X) = n\pi$ and $Var(X) = n\pi(1-\pi)$

Binomial Ex 3

X = # invoices recieve discount

A large industrial firm allows a discount on any invoice that is paid within 30 days. Of all invoices, 10% receive the discount. In a company audit, 12 invoices were sampled at random.

P(success) = recieve discount

$P(s) = .10$

$n > 12$

independant 1 invoice doesent effect others

total sum

a. What is the probability that exactly 4 of the invoices receive the discount?

Fixed trials

$\frac{12!}{4!\,8!} = \frac{12\cdot 11\cdot 10\cdot 9\cdot \cancel{8!}}{4\cdot 3\cdot 2\cdot 1\cdot \cancel{8!}}$

?

$499 - 5 = 495$

$P(X=4)$

$\binom{12}{4}(.10)^4 * (.90)^8 = 495(.10)^4 * (.90)^8 = .0213$

b. What is the probability that fewer than three of them receive the discount?

$P(X=2) + P(X=1) + P(X=0)$

$\binom{12}{2}(.10)^2(.90)^{10} + \binom{12}{1}(.10)^1(.90)^{11} + \binom{12}{0}(.10)^0(.90)^{12} = .88913$

//choose() sum dbinom()

c. What is the probability that at least one of them receive the discount?

opposite of one

1 − getting none

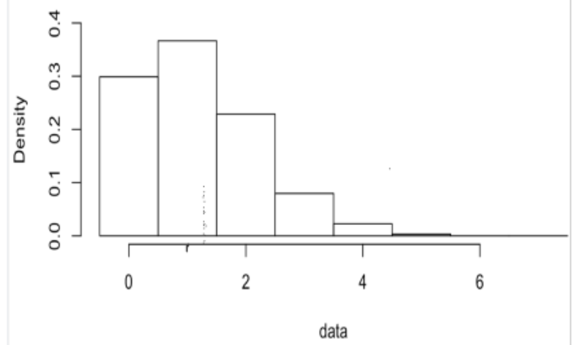
$P(X \geq 1) = 1 - P(X=0) \qquad 1 - \binom{12}{0}(.10)^0(.90)^{12} = 1 - 1 * 1 * (.90)^{12} = .7176$

Binomial Ex 3 Continued:

a. Find the expected number that receive the discount in 12 invoices.

$E(X) = 12(.10) = 1.2$

b. Find the standard deviation of the number that receive the discount.

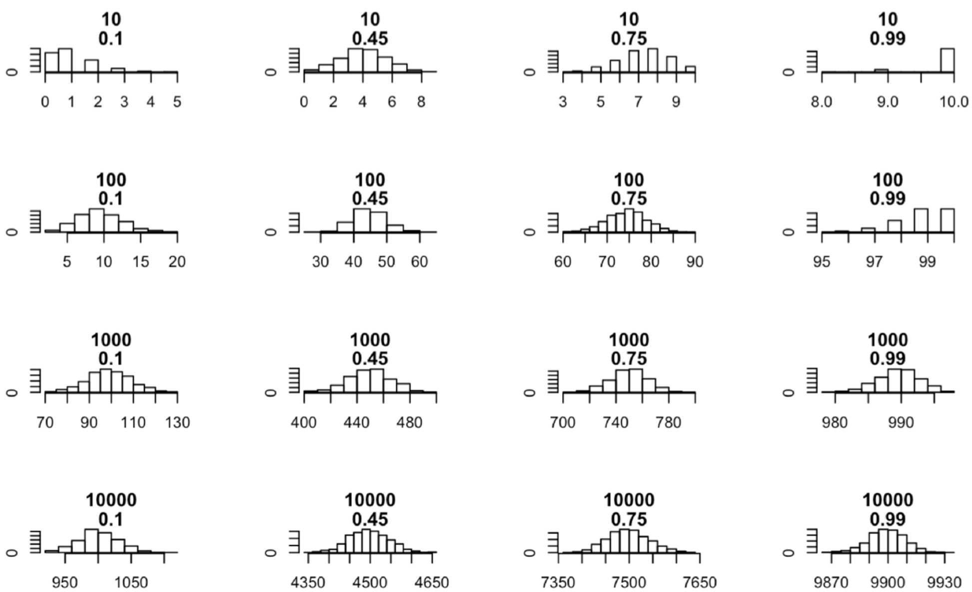


SD

$\sigma^2 = n\pi(1-\pi) = 12(.10)(.90)$

$\sigma_x = \sqrt{12(.10)(.90)} = 1.04$

Binomial Probability Histograms for different n and p



Take Aways:

*for small number of trials (n):

Distribution is more obviously ___________ and

____

*for larger n:

For a wide range of p, we have a more

____

curve and prob of any one specific outcome is

____

Continuous RVs Example

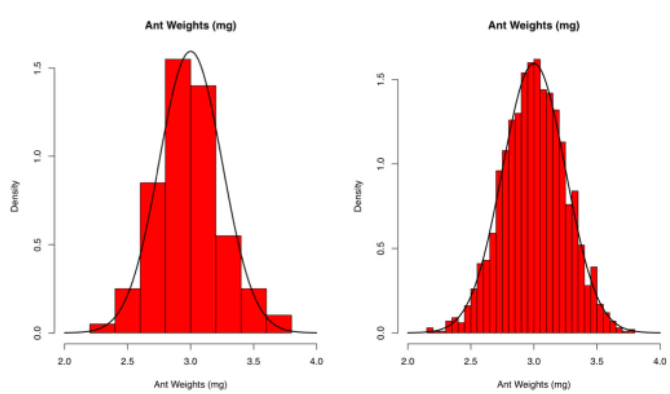
Continous RVs take values in specified range.

A continuous RV can be though of as the limit of a discrete RV (as possible values get infinitely close to each other)

e.g. Z: emissions in grams of particulates per gallon of fuel consumed

Y: resistance of a resistor labeled 100 Ω

Density histogram shown (y axis Scaled so total area =1)



Continuous RVs Example

Probability distributions for continuous RVs are called **probability density functions (pdfs:f(x))**, and consist of ranges of values the RV can take, together with a function that lives on those ranges.

The ___area___ under the function between any two possible realizations of the RV determines the probability that the RV will realize to a value in that range.

The probability of a truly continuous RV being a distinct value is zero. P(X=x)=0 , however X ___can be that value___

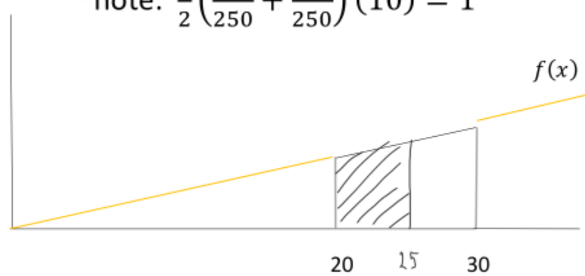
ex: P(X=27.3 g of emissions)=0, but 27.3 ___could___ a measured emission value.

Continuous RVs Example

Ex: Elongation (in %) of steel plates treated with aluminum are random with probability density function:

$$f(x) = \begin{cases} \frac{x}{250}, & 20 < x < 30 \\ 0, & otherwise \end{cases}$$

*note: $\frac{1}{2}\left(\frac{30}{250} + \frac{20}{250}\right)(10) = 1$



What proportion of steel plates have elongations less than 25%?

Ie, what is the probability of a randomly chosen steel plate treated with aluminum having an elongation less than 25%?

$P(20 \leq X \leq 25)$

$P(20 < X < 25)$ dosent matter because continuous

$\frac{1}{2}\left(\frac{20}{250} + \frac{25}{250}\right) 5$

$\frac{1}{2}\left(\frac{45}{250}\right) 5 = 0.45$

$\frac{1}{500}(225) = 0.45$

Most commonly used continuous RVs: Gaussian (Normal)

If X is Normal or "Gaussian", it has a pdf of: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

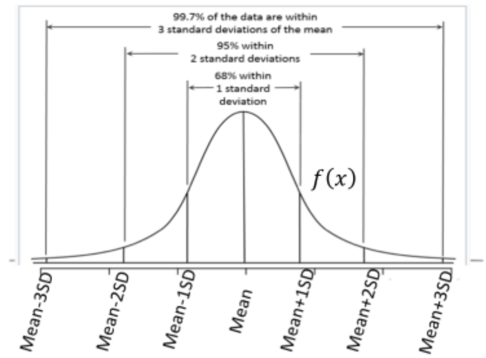
where $\mathrm{E}(X) = \mu$ and $VAR(X) = \sigma^2$

$X \sim N(\mu, \sigma^2)$

The Normal distribution:

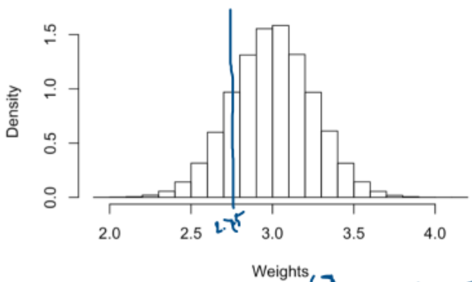
- Is a good approximation for many real-life random processes
- Can realize any value between $-\infty$, ∞.
- Is symmetric around μ
- The inflection points are at $\mu \pm 1\sigma$
- The total area under the curve is 1
- The area under the curve between
  - $(\mu - \sigma, \mu + \sigma)$ is about .68
  - $(\mu - 2\sigma, \mu + 2\sigma)$ is about .95
  - $(\mu - 3\sigma, \mu + 3\sigma)$ is about .997

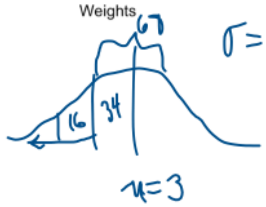
  (For all normal populations)



Most commonly used continuous RVs: Gaussian (Normal) cont.

Suppose the weights of ants in a large population Y is well approximated by a Normal distribution with mean 3 and standard deviation 0.25. $Y \sim N(3, 0.25^2)$.





$\sigma = .25$

a. What is the probability that a randomly selected ant has a weight less than 2.75 mg?

$P(Y < 2.75) = 16\%$ $\qquad z = \frac{2.75 - 3}{.25}$

b. Outside what values are the most extreme 5% of weights?

$Y < 2.5$ and $Y > 3.5$ is most extreme 5%

$3 - 2(.25)$

$3 + 2(.25$

Most commonly used continuous RVs: Gaussian (Normal) cont.

Standardized Units (z-scores): $\quad z = \frac{x - \mu}{\sigma}$. $\quad$ (*Z score calculation can be applied to non-normal X)

value of interest

Translate back to population units: $\quad \mathbf{x} =$ $z\sigma + \mu$

- Z scores tell use how many standar deviations an observation is above/below the population mean
- Can compare standardized scores across distributions to compare scores within their relative distributions

e.g. Scores on the SAT are well approximated by N(1500, 300^2) and scores on the ACT are well approximated by N(21, 5^2). Ann scored an 1800 and Tom scored 24 on the ACT. Who performed better relative to their fellow test takers?

$z_A = \frac{1800 - 1500}{300} = 1 \qquad z_T = \frac{24 - 21}{5} = 0.6$

Most commonly used continuous RVs: Gaussian (Normal) cont.

The **Standard Normal** distribution :

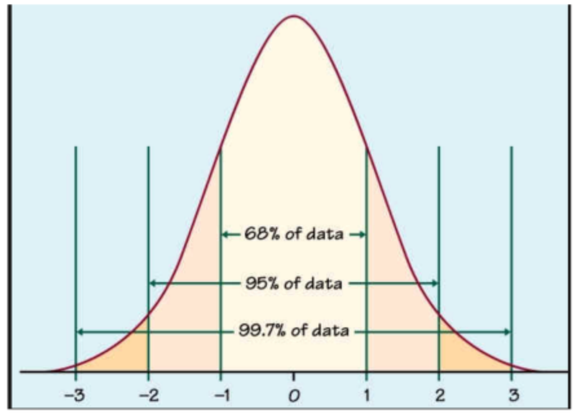
$$Z = \frac{X-\mu}{\sigma} \sim N(\underline{\ 0\ }, \underline{\ 1\ })$$

- Can realize any value between $\underline{-\infty}$, $\underline{\infty}$.
- Is symmetric around $\underline{E(Z)=0}$
- Has standard deviation of 1
- The total area under the curve is 1
- The area under the curve between
  - $(-1,1)$ is about $\underline{.68}$
  - $(-2,2)$ is about $\underline{.95}$
  - $(-3,3)$ is about $\underline{.997}$

Can be used with tables to compute probability
For ranges where endpoints are not multiples of
1,2,3 sd when data is normally distributed.

Standardized Units (z-scores)

$$Z = \frac{X-\mu}{\sigma}$$



Someone did a bunch of integration for us and calculated the probability below many standardized values. (Full table on Canvas)

How to read table:

- Values on Top Row and left column combine to be a zscore.
- Value in table where column and row intersect is the percent or values below that zscore.

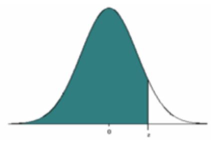
Ex1: Percent of scores below a zscore of 1 is

$P(z \le 1) = .8413$

Ex2: Percent of scores below a zscore of 1.24 is

$P(z \le 1.24) = .8925$

Table of Standard Normal Probabilities for Positive Z-scores



CDF for Z scores:
$F(z) = P(Z \le z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857

Suppose the weights of ants in a large population Y is well approximated by a Normal distribution with mean 3 and standard deviation 0.25. $Y \sim N(3, 0.25^2)$. DRAW THE PICTURE!

a. What is the probability that a randomly selected ant has a weight less than 2.8 mg?

$P(X < 2.8) = P\left(Z < \frac{2.8-3}{.25}\right) = P(Z < -.8) = \boxed{.2119}$

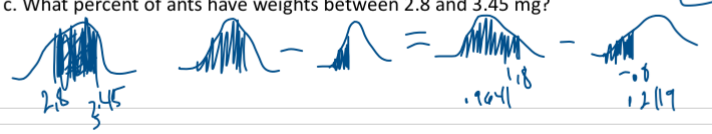
b. What percent of ant weights are above 3.45 mg?

$P(X > 3.45) = P\left(Z > \frac{3.45-3}{.25}\right) = P(Z > 1.8) =$

$= 1 - P(Z < 1.8)$

$= 1 - .9641 = \boxed{0.359}$

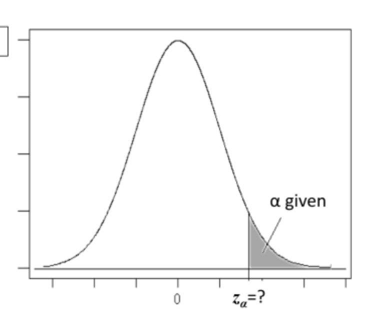
c. What percent of ants have weights between 2.8 and 3.45 mg?



$= .9641 - .2119 = \boxed{0.7522}$

Percentile and and z Critical Value

Data Values ←→ zscore ←→ Percent Below ←→ Percent Above



k^{th} percentile/quantile (e.g. 80[th]), K% of data at or below, (100-k)% above

Z-critical value z_a: for given a, find z_a such that the **right** tail area $P(Z \ge z_a) = a$

Suppose the weights of ants in a large population Y is well approximated by a Normal distribution with mean 3 and standard deviation 0.25. $Y \sim N(3, 0.25^2)$.

a. What is the critical value z value above which 20% of Z scores fall?

$Z_{.2} = 0.84$

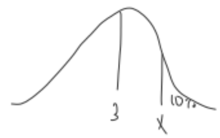
b. Above what ant weight is the most extreme 10% of weights?

.90 below

$Z_{0.1} = 1.28 = \frac{x-3}{0.2}$

$x = 1.28(.25) + 3$

$\boxed{x = 3.32}$



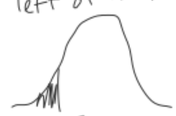
c. What weight is at the first quartile of ant weights?

(−) value because left of mean

0.25 inside table

↑

$Q_1 = 25\%$ below $= Z_{.75}$



$Z_{.75} = -.675$

$-.675 = \frac{x - 3}{.25}$

$x = -.675(.25) + 3$

$= 2.83$

More Continuous Distributions to Come:

- Student's t-distribution
- F-distribution
- χ^2-distribution

For Tomorrow

- Homework 2 will be posted today (Thursday). Take a look and get started and post any questions you have on Pizza.

- Continue working on Quiz 1 due Friday 11:59pm.