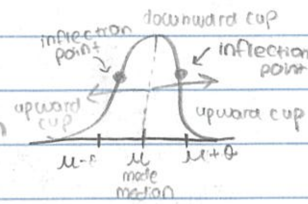
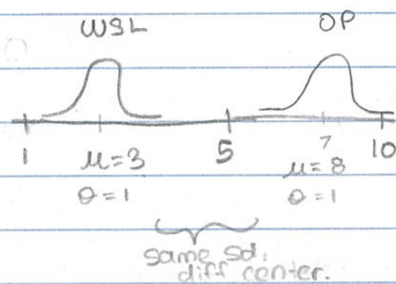
1/12/24

6.1 The Normal Curve

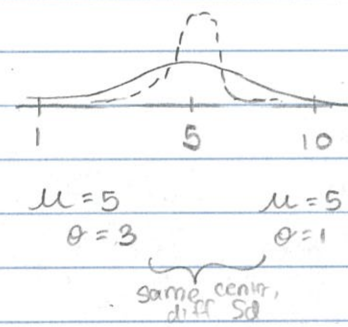
Features:

1) Bell-shaped, highest point is the mode (most common)
2) unimodal
3) x-axis is a number-line
4) Symmetric, meaning highest also is median
5) Mean/median/mode same spot!
6) Never touches/crosses x-axis: asymptote
7) Inflection points (changes from cupping down to up) is always one standard deviation from the mean ($\mu-\theta$ and $\mu+\theta$)



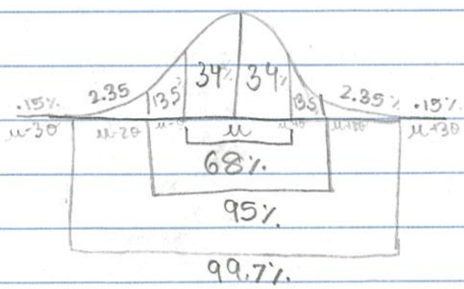


VS

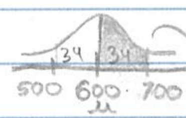


Empirical Rule ☆

- 68% one sd
- 95% two sd
- 99.7% three sd



Example: Sunshine radios are normally distributed w/ mean 600 hrs, SD 100 hrs



→ 34% would last 600-700 hrs

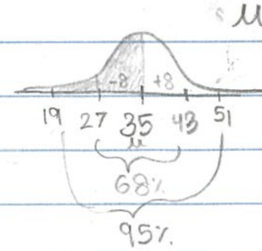
$P(600 < x < 700) = .34$

Ex a 34% chance that a SR would last 600-700 hours

Note: $(600 < x < 700)$ and $(600 \leq x \leq 700)$ $\leq$ doesn't matter in continuous at all same thing

Example 2: Wheat (p275)

$\mu = 35, \theta = 8$, $P(19-35)$?



x = bushel

$34 + 13.5 = \boxed{47.5\%}$

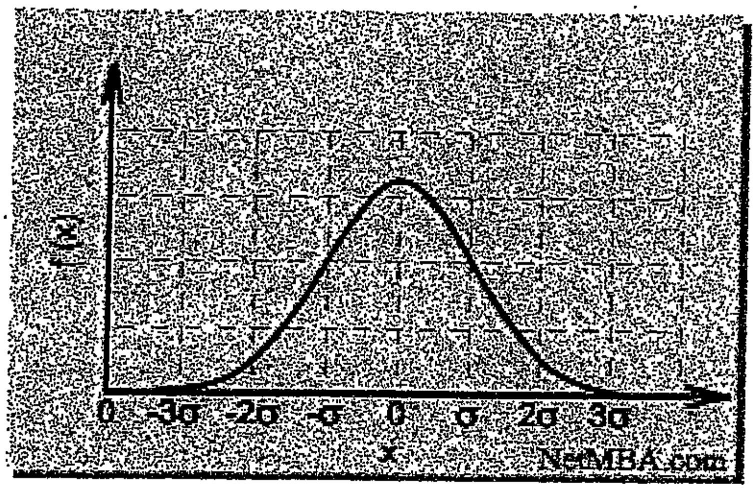
or $95/2 = 47.5\%$

Z-score

$$Z = \frac{x-\mu}{\theta}$$

7 The Normal Distribution (bell curve)

In many natural processes, random variation conforms to a particular probability distribution known as the normal distribution, which is the most commonly observed probability distributions. The normal curve was first used in the 1700's by French mathematicians and early 1800's by German mathematician and physicist Karl Gauss. The curve is known as the Gaussian distribution and is also sometimes called a bell curve.



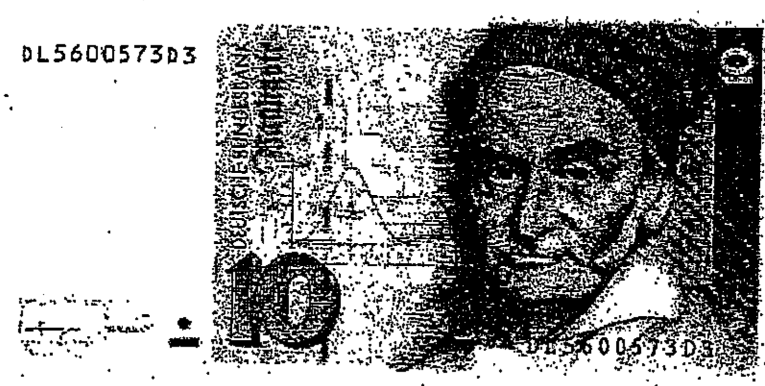
This curve is for a data set having a mean of zero and standard deviation of one. The normal distribution curve is described by this probability density function:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$ never need

<u>Normal (bell) curves have the following characteristics;</u>

- symmetric
- unimodal
- extend to +/– infinity
- area under the curve equals one (1)

A normal distribution can be completely described by two parameters, the mean and the standard deviation.



The **empirical rule** states that for a normal distribution:

- 68% of the data will fall within 1 standard deviation of the mean
- 95% of the data will fall within 2 standard deviations of the mean
- Almost all (99.7%) of the data will fall within 3 standard deviations of the mean

Earlier we used the language " -- percent of the data" falls within one or more standard deviations of the mean. Since probability has a range of 0% to 100%, we can also interpret the normal distribution as the probability of an event happening.

$x = z\theta + \mu$

Is it Normal?

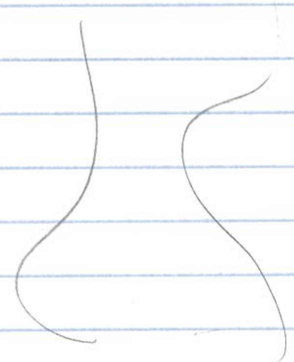
1) Histogram: if normal, should be roughly bell-shaped
2) Outliers: shouldn't be more than one
   - box + whisker plot, find IQR, and see outlier boundaries (Q3 + 1.5(IQR) and Q1 - 1.5(IQR))
3) Skewness w/ Pearsons index, shouldn't be greater than 1 or less than -1
   - $\frac{3(\bar{x} - \text{median})}{s}$
4) Normal quantile plot, if straight line its normal

   ↓

   concave down is skewed left graph (not normal)

   concave up is skewed right graph (not normal)

   T-curve is very straight <u>except</u> for ends (tails)



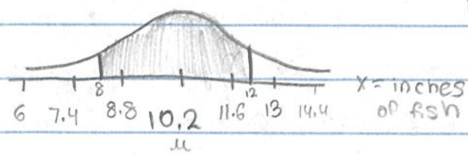
1/24/24

6.4

Warm-up

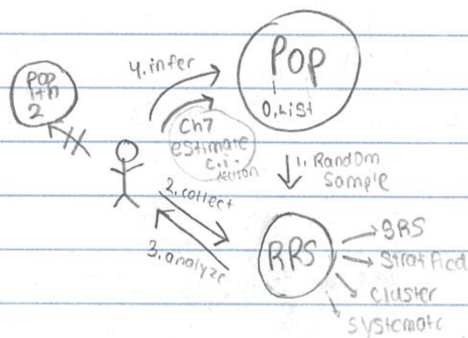
- Childrens fishing pond, fish sizes normally distributed, $\mu = 10.2$, $\sigma = 1.4$

$P(8 < x < 12) = 0.8427$



Ex a 84.27% chance that a random fish is between 8 and 12 inches long

---

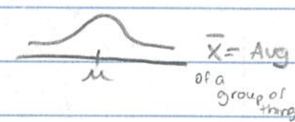


$\boxed{\text{Statistic:}}$ numerical descriptive measure of a sample ($\bar{x}$, s, s^2, r, $y = a + bx$, $\frac{r}{n} = \hat{p}$)

$\boxed{\text{Parameter:}}$ numerical descriptive measure of a popltn (μ, σ, σ, $p = r/n$)

- Statistics are used for inferences about population parameters
  - ↳ w/ estimation (confidence intervals) and decision (hypothesis testing)

$\boxed{\bar{X} \text{ graph:}}$ avg of a group of things (not a fish)



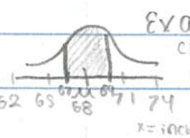
- $P(8 < \bar{x} < 12) = 0.9978$ (compare w/ top of page)

- Theorem 1: given x is random variable w/ normal distr
  - a) $\bar{x}$ distribution is a normal distribution
  - b) mean of the $\bar{x}$ distr is $\boxed{\mu}$ (means are the same)
  - c) Sd of $\bar{x}$ distr is $\boxed{\sigma/\sqrt{n}}$ (Standard error) (diff sd)
    - how big sample sizes are (eg 5 fishes)

- $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$
- sample size can be any size
- fine print: original distr has to be normal
  - the $\bar{x}$ graph gets narrower, (the sd decreases greatly (when a group of things))
  - (as sample size ↑, gets narrower) ☆

Example 1: $\mu = 68$ in, sd 3 in.

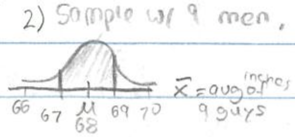
1) Prob a 18 yr-old man selected between 67 and 69 in tall? $P(67 < x < 69) = 0.2611$



Ex a 26.11% chance a guy is between 67 and 69 inches

sd changes, μ same! →

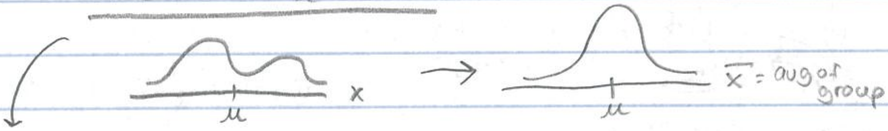
2) Sample w/ 9 men, $P(67 < \bar{x} < 69) = 0.6829$, Ex a 68.29% chance that the avg height of a sample of nine men is between 67 + 69 inches



Unknown Original Distribution

$\boxed{\text{Central Limit Theorem:}}$ $\bar{x}$ of any distr

- if x has <u>any distribution</u> w/ mean μ + sd σ, the sample mean $\bar{x}$ based on sample of size n will have a distr that approaches distr of a normal random variable w/ mean μ + sd $\sigma/\sqrt{n}$ as n increases w/o limit
- ★ x can have <u>any distribution</u>, but $\bar{x}$ will <u>still come out to be normal</u>



only works if sample size is large ($n \geq 30$)

↑ sample size = more normal + narrow

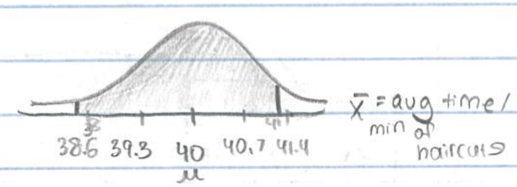
$\sigma_{\bar{x}} = \sigma/\sqrt{n}$, $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

same μ as og (no matter the og distr)

<u>Example:</u> μ is 40 min, sd is 4.2 min, 36 visits, ? prob that avg time it takes is between 38 and 41 minutes?

- $n = 36 > 30$, thus CLT invoked
- $\sigma_{\bar{x}} = 4.2/\sqrt{36} = 0.7$
- $\boxed{P(38 < \bar{x} < 41) = .9213}$

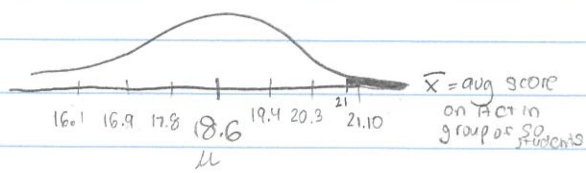
Ex a 92.13% chance the barber will have an average haircut time between 38 and 41 min.

→ 

1/31/24

$\mu = 18.6$, $sd = 5.9$, SRS of 50 students, what is probability that the mean score $\bar{x}$ of these students is 21 or higher?

$sd = \frac{5.9}{\sqrt{50}} = 0.8344$



$P(\bar{x} > 21) = 0.0020$

Ex a 0.2% chance the average ACT score in a group of 50 students is 21 or higher

$n = 50 > 30$, $\therefore$ CLT invoked

1/24/23

SPC (statistical process control)

AAO control charts + book notes (277-281)

- Process = chain of steps that turns inputs → outputs
  - control charts see if a process is in or out of control
  - common-cause variation is due to day to day factors affecting process = normal variation (printer runs out of paper)
  - special cause variation: sudden, unpredictable events (blackout, car crash)
- Control chart tracks variation
  - center line = target (finish time target)
  - control limits set 3 sds above + below center line
  - assume normal distribution
  - if falls out control limit, means it's out of control (should only happen .03 % of the time)
  - can reveal causes of variation
  - data over an equally spaced time intervals or in sequential order
  - in control = same probability distribution at successive points in time
  - out of control = doesn't follow target probability distribution

Steps:

1) Find μ and σ (by using past data or using specific "target" values for $x + \sigma$)
2) Create graph where vertical axis is x-value + horizontal axis is time
3) Draw horizontal line at height μ, and dashed control-limit lines at $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$
4) Plot x on graph in time sequence order + use line segments to connect the points

Out of control signals:	Prob of false alarm	"Think"
I 1) One point falls beyond $\pm 3\sigma$ →	$P = 0.003$	→ blow-up!
II 2) A run of 9 consecutive points on one side of the center line (0.004 chance)	$P = 0.004$ $(0.5)^9$ ↗ 50% chance above/below	slow-drift out
III 3) At least 2 out of 3 consecutive points lie beyond $\pm 2\sigma$ level on same side of line (0.004)	$P = 0.004$	blow-up/slow drift

↓ (also if it keeps getting out of control w/ time)

Rule 1) One point out of control
Rule 2) 2/3 between 2+3 sds on same side
Rule 3) 4/5 between 1/2 sd same side
Rule 4) 8+ points on one side

1/17/24

Z-score

of standard deviations from the mean

Bio
$\mu = 70$
$x = 74$
$\sigma = 1.5$

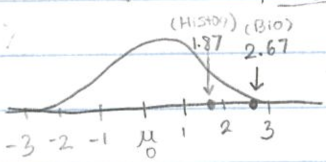
History
$\mu = 86$
$x = 91$
$\sigma = 2.67$

$$\boxed{Z = \frac{x-\mu}{\sigma}} \text{ or } \frac{x-\bar{x}}{s}$$

can use to compare among diff popltns

$Z_{Bio} = \frac{74-70}{1.5} = \frac{4}{1.5} = \boxed{2.67}$

$Z_{History} = \frac{91-86}{2.67} = \frac{5}{2.67} = \boxed{1.87}$

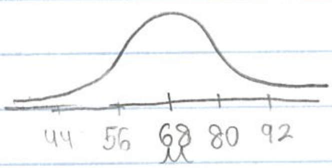


SNC (Standard normal curve)
Z = stdevs

Percentile: if 83%, then 83% have a score same or less than you

Z-table show what percent is left

Tuna: $\mu = 68$ lbs, $\sigma = 12$ lbs

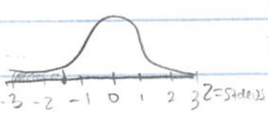


★ Calc:
normalcdf (not pdf)
- if upper/lower limit not given go up/down 4 sds

A) $P(X<50) =$ $Z_{50} = \frac{50-68}{12}$

$Z_{50} = \frac{-18}{12}$

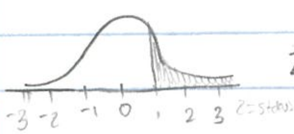
$Z_{50} = -1.5$



$\boxed{P(X<50) = 6.68\%}$

There is a 6.68% chance a fish weighs less than 50 lbs

B) $P(X>80)$



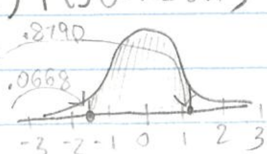
$Z_{80} = \frac{80-68}{12}$

$Z_{80} = \frac{12}{12}$

$Z_{80} = 1$ $\boxed{P(X>80) = 16\%}$

There's a 16% chance a fish will weigh more than 80 lbs

C) $P(50<X<82)$



$Z_{50} = -1.5$

$Z_{82} = \frac{82-68}{12} = 1.167$

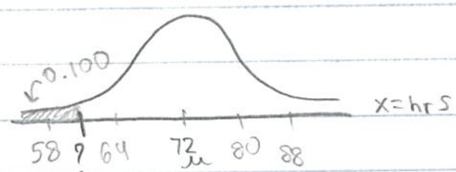
$0.8790 - 0.0668 = 0.8122$

$\boxed{P(50<X<82) = 81.22\%}$ + sentence

Inverse Normal (given %)

Malaria

$\mu = 72$ hrs, $\theta = 8$ hrs



$\boxed{z = -1.28}$

$-1.28 = \frac{x - 72}{8}$

$-10.24 = x - 72$

$\boxed{61.76 = x}$ round down

61 hrs b/c if go ↑ expose more

After how many hours should another pill be given so that: less than 10% unprotected?

Steps:

1) Dive into z-table + search for % and its z score!

or calc
↓
invnorm