Zhi
STAT C1000
Th 10/16/25

or use 1-Varstats.

mean: $\mu = \frac{1+3+5}{3} = \frac{9}{3} = \boxed{3}$

$\underbrace{N = 3}$

standard deviation, $\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$

$= \sqrt{\frac{(1-3)^2+(3-3)^2+(5-3)^2}{3}} = \sqrt{\frac{(-2)^2+2^2}{3}}$

$= \sqrt{\frac{4+4}{3}} = \sqrt{\frac{8}{3}}$

$= \sqrt{2.\bar{6}} \approx 1.6$

$\sigma = 1.632993162$

Example: From the population {1,3,5} list all possible samples of size $n = 3$, with replacement and calculate the mean of each sample. Find the mean, variance, and standard deviation of the original population. Find the mean, variance, and standard deviation of the sample means. Compare.

new dataset of 27 Sample means

$\boxed{L1}$

μ = mean of original population

σ = std. dev. of original population $\frac{1.63299}{\sqrt{3}} = .9428$

Sample size

27 possible Samples			mean of each sample, $\bar{x}$
1	1	1	1.000
1	1	3	1.667
1	1	5	2.333
1	3	1	1.667
1	3	3	2.333
1	3	5	3.000
1	5	1	2.333
1	5	3	3.000
1	5	5	3.667
3	1	1	1.667
3	1	3	2.333
3	1	5	3.000
3	3	1	2.333
3	3	3	3.000
3	3	5	3.667
3	5	1	3.000
3	5	3	3.667
3	5	5	4.333
5	1	1	2.333
5	1	3	3.000
5	1	5	3.667
5	3	1	3.000
5	3	3	3.667
5	3	5	4.333
5	5	1	3.667
5	5	3	4.333
5	5	5	5.000

μ 1-Varstats L1

mean of all the Sample means

$\boxed{\mu_{\bar{x}} = \boxed{3} = \mu}$

and σ standard deviation of all the Sample means

$\boxed{\sigma_{\bar{x}} = \boxed{.9428}} \boxed{= \frac{\sigma}{\sqrt{n}}} = \frac{1.6329}{\sqrt{3}}$

$\{1, 3, 5\}$ pick samples of size $n=3$, with replacement

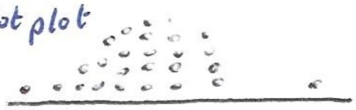
Samples	mean of sample, $\bar{x}$
1, 1, 1	$(1+1+1)/3 = 1$
1, 1, 3	$(1+1+3)/3 = 1.6$
1, 1, 5	$(1+1+5)/3 = 2.\bar{3}$
1, 3, 1	
1, 3, 3	
1, 3, 5	
1, 5, 1	
1, 5, 3	
1, 5, 5	
3, 1, 1	
3, 1, 3	
3, 1, 5	
3, 3, 1	
3, 3, 3	
3, 3, 5	
3, 5, 1	
3, 5, 3	
3, 5, 5	
5, 1, 1	
5, 1, 3	
5, 1, 5	
5, 3, 1	
5, 3, 3	
5, 3, 5	
5, 5, 1	
5, 5, 3	
5, 5, 5	

Zhi
STAT C1000
Th 10/16/25

Shows the spread of data

eg. dot plot

①



②

Sampling Distributions

A **Sampling Distribution** is the probability distribution of a sample statistic ($\bar{x}, s, n,$) that is formed when random samples of size n are repeatedly taken from a population. If the sample statistic is the mean, then the distribution is the "sampling distribution of the mean." ("shows where all the sample means lie")

Properties of the Sampling Distribution of Sample Means: (see separate page in Zhi's website)

The Mean of the sample means is $\boxed{\mu_{\bar{x}} = \mu}$ The standard error of the mean is $\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}} = \frac{\text{std. dev}}{\sqrt{\text{sample size}}}$

The Central Limit Theorem "30 or more" "at least 30"

memorise

1. If random samples of size n, where $n \geq 30$, are drawn from any population ("not normal") with mean μ and standard deviation σ, then the sampling distribution of sample means is approximately normal. The greater the sample size, the better the approximation.

2. If random samples of size n are drawn from a normally distributed population with mean μ and standard deviation σ, then the sampling distribution of sample means is normal from *any* sample size n.

z = the z-score

Previously we calculated the probability that a *z-score* lies in a certain interval by:

normalcdf(lower z , upper z , 0 , 1) gives the area between z_1 and z_2

x = the variable"

We also calculated the probability that an *x-value* lies in a certain interval by:

① normalcdf(lower x , upper x , μ , σ) gives the area between x_1 and x_2

or ② convert x values into z scores using $z = \frac{x-\mu}{\sigma}$, then use normalcdf ($z_1, z_2, 0, 1$)

* **Finding Probabilities using the Central Limit Theorem....** (for use with samples and sample means)

To find the probability that a sample mean, $\bar{x}$, lies in a certain interval:

We MUST use z scores, then normalcdf ($z_1, z_2, 0, 1$) because normalcdf does not allow us to adjust for sample size, n,

First convert the sample mean, $\bar{x}$, to a *z-score* using the new *z-score* formula:

So we have to adjust for sample size, n, ourselves, before using normalcdf command..

using a new z-score formula:

$$z = \frac{(\bar{x}-\mu)}{(\sigma/\sqrt{n})}$$

You MUST use parentheses!

tells us how many standard errors the sample mean is from true mean

vs. old: $z = \frac{x-\mu}{\sigma}$

then use $normalcdf(z_1, z_2)$ as before.

$$z = \frac{x - \mu}{\sigma}$$ tells us how many standard deviations x is from the mean

if x is less than the mean, z is −

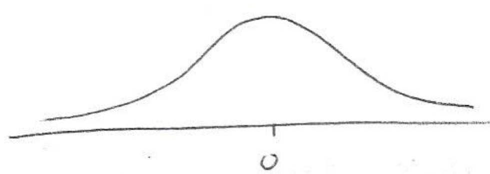
if x = the mean, z is 0

if x is more than the mean, z is +

* "z" is the "standard" normal distribution

has $\mu = 0$

$\sigma = 1$



normalcdf (lower value , upper value , μ, σ) gives the area between lower and upper val

* if using z scores:

normalcdf (low z, upper z, 0, 1)

Probability
= Proportion
= Area under normal curve

implied "all" i.e. population → "teen drivers" $\mu = 20.7$ variable, X — units: miles $\sigma = 6.5$

Example: Drivers age 16 to 19 drive a mean of 20.7 miles per day with standard deviation of 6.5 miles. ~~You randomly select 100 drivers age 16 to 19.~~

decimal or %

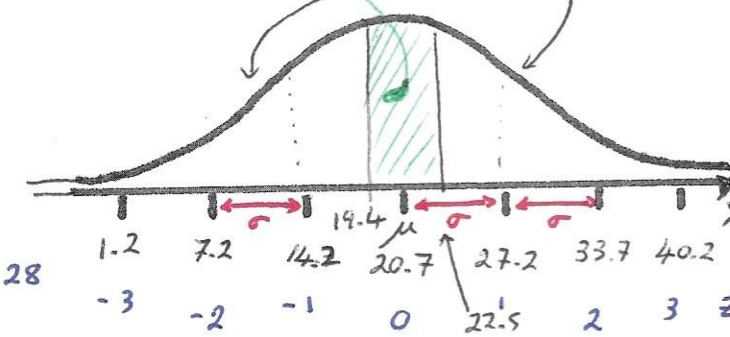
What is the probability that a randomly selected (one) 16 to 19 year old drives between 19.4 (x_1) and 22.5 (x_2) miles?

Shaded region = proportion = area

normalcdf (x_1, x_2, μ, σ)

= normalcdf (19.4, 22.5, 20.7, 6.5)

= $\boxed{.188}$ or $\boxed{18.8\%}$



or: convert $x_1 = 19.4$ and $x_2 = 22.5$ to z:

$$z_1 = \frac{(19.4 - 20.7)}{6.5} = -0.2 \qquad z_2 = \frac{(22.5 - 20.7)}{6.5} = +0.28$$

then normalcdf (-0.2, +0.28, 0, 1)

$n = 100$ If a random sample of 100 drivers age 16 to 19 is selected, what is the probability (= proportion = area) that the mean distance driven in the sample is between 19.4 and 22.5 miles?

$\underbrace{\text{mean distance driven in the sample}}_{\bar{x}}$

(same $\mu = 20.7$, $\sigma = 6.5$)

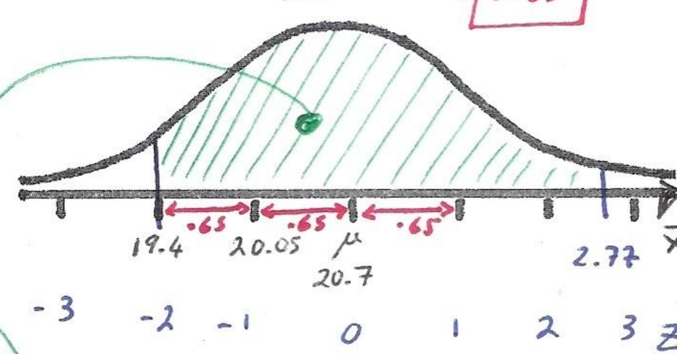
Samples.. Sample means.. use Central Limit Theorem..

"standard error" $= \frac{\sigma}{\sqrt{n}} = \frac{6.5}{\sqrt{100}} = \frac{6.5}{10} = \boxed{0.65}$

$$z_1 = \frac{(\bar{x}_1 - \mu)}{(\sigma/\sqrt{n})} \qquad z_2 = \frac{(\bar{x}_2 - \mu)}{(\sigma/\sqrt{n})}$$

$$z_1 = \frac{(19.4 - 20.7)}{(6.5/\sqrt{100})} \qquad z_2 = \frac{(22.5 - 20.7)}{(6.5/\sqrt{100})}$$

$$z_1 = -2.00 \qquad z_2 = 2.77$$



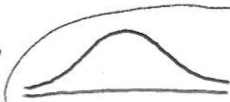
19.4 miles per day is 2 standard errors below true mean (20.7)

22.5 miles per day is 2.77 standard errors above true mean (20.7)

$P(19.4 < \bar{x} < 22.5)$

$= P(-2 < z < 2.77) =$ normalcdf (-2, 2.77, 0, 1) $= \boxed{.974}$ or $\boxed{97.4\%}$

Variable, X
↓ units: days



$Z =$

normalcdf, invnorm

* use ± 1000 for $\pm \infty$

$\mu = 268$ $\sigma = 15$

Example: Lengths of pregnancies are normally distributed with mean of 268 days and standard deviation of 15 days.

single (not Central Limit Theorem)

area $P(x < 260)$

What is the probability that a random woman's pregnancy will last less than 260 days?

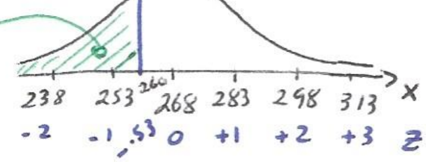
left

$$Z = \frac{x-\mu}{\sigma} = \frac{260-268}{15} = -0.5\bar{3}$$

$\boxed{29.7\%}$ or $\boxed{.297}$

29.7% of all pregnancies last less than 260 days

$P(X < 260)$
$= \text{normalcdf}(-\infty, 260, 268, 15)$
$= \text{normalcdf}(-\infty, -.53, 0, 1)$



Would it be unusual for a random pregnant woman to deliver in less than 260 days?

No, not unusual, because the probability is 29.7%, NOT < 5%

"unusual"

In a group of 25 pregnant women, how many would you expect to have a pregnancy lasting less than 260 days?

29.7% of them..

$= .297(25) = 7.4$ women ?

we expect $\boxed{7}$ (or 8)

$n = 25$

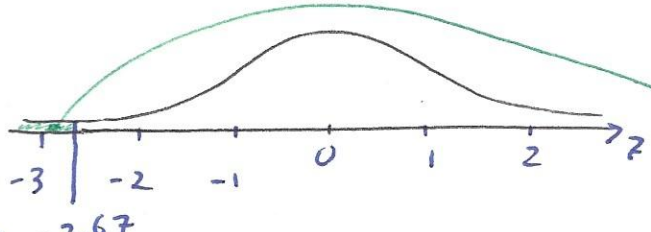
What is the probability that the mean length of pregnancies in a sample of 25 women will be less than 260 days?

left

$\bar{x}$

must use Central Limit Theorem

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{(260-268)}{(15/\sqrt{25})} = -2.67$$



$P(\bar{x} < 260)$
$= \text{normalcdf}(-\infty, -2.67, 0, 1)$
$= \boxed{.0038}$ or $\boxed{.38\%}$

A random sample of 25 pregnant women was put on a special diet during pregnancy. Suppose the mean length of their pregnancies is less than 260 days. Is this most likely due to chance or due to the special diet? Explain.

it is almost impossible (.38% almost 0) for a sample of 25 pregnant women to have a mean pregnancy length less than 260 days.

Diet.