Normal Distributions

1. Normal distributions are bell–shaped – unimodal and symmetric – and continuous. They're only mathematical approximations of Histograms. Since Normal distributions are smooth curves ~ continuous, P(X = *a*) = 0 [area under the *line* x = *a* is ZERO!] so that, theoretically, **P(X ≥ *a*) ~ P(X > *a*)**. That is why the calculator command doesn't make a distinction between P(X > 72in) and P(X ≥ 72in.)!

2. All Normal Distributions are uniquely determined by their Mean [~ Median, by the way] and s.d.

3. Forward problem: When X is known and the probability, P is asked.

Use **2nd + VARS [DISTR command] –> OPTION 2: Normalcdf(Left Limit, Right Limit, μ, σ)** if
– the X–value [variable] is given / known
– A Proportion / % / Probability / Area / Percentile is asked.
– Use proper Notation: **P(X ≥ a)** or **P(X ≤ b)** or **P(c ≤ X ≤ d)**

– Specific commands:
· For P(X > a), use **Normalcdf(*a*, 9999, μ, σ)**
· For P(X < *b*), use **Normalcdf(–9999, *b*, μ, σ)**
· For P(c < X < d), use **Normalcdf(*c, d,* μ, σ)**

How do we know if the Normalcdf command is to be used? When X is known and the probability, P is asked.

4. Backward problem: When the probability, P, is known, and the X–value is asked.

Use **2nd + VARS [DISTR command] –> OPTION 3: InvNorm (Left Area <u>in decimals</u>, μ, σ)** if
– A Proportion / % / Probability / Area / Percentile is given / known
– The X–value [variable] is unknown/ asked.
– Use proper notation: **P(X < *a*) = p%**

How do we know if the InvNorm command is to be used? When the probability, P, is known, and the X–value is asked.

5. The terms Probability, Proportion, Percentage, Relative Frequency are used *interchangeably* and refer to the same idea.

6. Understand what **Percentile** means: If X corresponds to the p–th percentile, then p% of ALL values [incomes, weights, heights, length of pregnancies, etc.] are ≤ X. P = P(X ≤ X–value)

Percentiles denote the **left–area** of the distribution.

7. A problem that *asks for* the Percentile is a forward problem [because a Proportion / Probability / Percentage / *left* Area is sought, for a given X–value].
A problem that *gives* a Percentile involves a backward problem [because an X–value is sought for a given Proportion / Probability / Percentage / *left* Area].

8. The **Expectations** while solving any problem are:

· Define a variable – say, X – and state its distribution [Centre / Shape / Spread: X ~ N(μ, σ)]. **This is NOT optional!**
· Use Probability Notation to describe the Q. **This is NOT optional!**
· Draw a graph, label and shade the appropriate region. **This is NOT optional!**
· Use a calculator command to solve the problem: it is NOT required to write down the command!

Note: At times, you may prefer to visualize / illustrate the problem *before* you interpret the Q in probability notation. Thats OK!

9. Z–scores give the number of s.d. an X–value is from the mean: Z = (X – μ) / σ. Clearly, Values above the Mean have positive Z–scores; Values below the Mean, Negative

10. **Power Tip!** Students, at times, wonder *when* to use Z–scores. Here's a good Rule of Thumb: when the Q uses the phrase: ***s.d. from the mean*.**

11. Z–scores and the **Empirical Rule**:
· ~68% of Z–scores lie between –1 and +1. [Why? Because ~68% of all X–values lie <u>within 1 s.d. of the mean</u>!].
· ~95% of Z–scores lie between –2 and +2. [Why? Because ~95% of all X–values lie <u>within 2 s.d. of the mean</u>!].
· ~99.7% of Z–scores lie between –3 and +3. [Why? Because ~99.7% of all X–values <u>lie within 3 s.d. of the mean</u>!].

E.g. For X ~ r.v. denoting heights of US males (inches) ~ N(72, 2.5) **<–––––– Study this example. You <u>need</u> to understand these translations for your HW!**

- a height that is Z = +1 s.d. *above* the mean is 72 **+ 1**·2.5 = 74.5
- a height Z = 1.5s.d *below* the mean is of 72 **– 1.5**·2.5 = 68.25
- a height of 66 is Z = (X – μ) / σ = (66 – 72)/2.5 = –2.4 ~2.4 s.d. *below* the mean. **AP 5 Note!** This height is quite rare: it lies *beyond* 2.d. from the Mean!
- a height of 78 is Z = (X – μ) / σ = (80 – 72)/2.5 = +3.2 ~3.2 s.d. *above* the mean **AP 5 Note!** This height is *very* rare: it lies *beyond* 3.d. from the Mean!

12. An outcome is said to **Rare** or **Unusual** or **Extreme** if the probability of its occurrence is <u><</u> 5%. Alternately, Rare Events occur at beyond the 5th or 95th percentiles. For any distribution, Z–scores beyond 2 are regarded as rare; for symmetric distributions, Z–scores beyond 1.645 are regarded as rare...since theyd occur < 5% of the time. On the other hand, In general, Z–scores closer to 0 (Zero) are less rare than otherwise since they relate to values or outcomes *close to the Mean* ~ representative value!

13. Z–scores *are* affected by extreme–value or outliers since it relies on the Mean and s.d., which, in turn, are influenced by extreme observations.

Comparing distributions in terms of Z–scores, and Percentiles

- Given 2 data–sets, we can determine the Percentiles corresponding to a given value, *a*, for both sets: P1 = P(X ≤ *a*) and P2 = P(Y ≤ *a*)

Recommended: for ANY data–set. Percentiles only depend on the *relative* position of the numbers, not on the values themselves...so they arent affected by extreme values!

- Given only the Mean and s.d. of 2 data–sets, we can determine the Z–scores corresponding to a given value, *a*, for both sets: Zx = (*a* – Mean1) / s.d.1 and Zy = (*a* – Mean 2) / s.d.2

  Recommended: for ANY symmetric data–set **or** when the entire data–set is unknown **or** when *only* the Mean and s.d. are provided.

Example 1

On a certain edition of the SAT, the Math scores were approximately normal with a mean of 500 and a s.d. Of 75. For *that* edition:
(i) Find the percentage of scores between 600 and 700.
(ii) Find the probability that an individual scores above 730.
(iii) Find the proportion of scores below 400.
(iv) A student gets a score of 710. Should one be impressed? Determine this using 4 methods: 2 by calculating probabilities, 2 using Z-scores!
(v) Find the percentile that a score of 780 corresponds to.
(vi) Find the score that corresponds to the 65th percentile.
(vii) Find the cut-off score that is the bottom 20% of SAT Math scores.
(viii) Find the score that separates the top 10% from the rest.
(ix) Find the 2 scores that constitute the middle 30% of SAT Math scores. What is the *spread* of the middle 30% of scores? **This is the more general version of the IQR problem, which is the spread of the middle 50% of scores!**
(x) Between what 2 scores did virtually all Math SAT scores fall between?
(xi) Which score lies 1.5s.d. *below* the mean?
(xii) What score***s*** lie 1.75 s.d. *from* the Mean?
(xiii) **Use the Empirical Rule.** 68% of SAT scores lie between what 2 scores? 95% of SAT scores lie between what 2 scores? 99.7% of SAT scores lie between what 2 scores?
(xiv) What scores might be unusually high? Unusually low?
(xv) Calculate the IQR of SAT scores.
(xvi) If 5895 students took the SAT, *how many* scores were below 600 or exceeded 700?

Before you attempt the Qs, ask yourself if the Q is
a forwards problem [X known => Normalcdf] **or**
a backwards problem [X unknown => InvNorm]**!**

Solution.

Note: I am <u>not</u> too particular about getting the inequality < or ≤ correct…since it isn't a big deal since P(X ≤ *a*) = P(X < *a*)

Let X be a r.v. denoting SAT scores. X ~ N(500, 75).
(i) P(600 < X < 700)
Sketch a figure to label the Mean, 500, and s.d., 75, with a region suitably shaded between 600 and 700.
P = Normalcdf(600, 700, 500, 75) = 8.73% [**Note:** it is NOT required to write down the command.]
8.73% of SAT Math scores lie between 600 and 700.

(ii) P(X > 730)
Sketch a figure to label the Mean, 500, and s.d., 75, with a right region suitably shaded above 730.
P = Normalcdf(730, 99999, 500, 75) = 0.1%. [**Note:** it is NOT required to write down the command.]
The probability that an individual scores above 730 on the Math SAT is 0.1%.
(iii) P(X < 400)
Sketch a figure to label the Mean, 500, and s.d., 75, with a left region suitably shaded to the left of 400.
P = Normalcdf(-99999, 400, 500, 75) = 9.12% [**Note:** it is NOT required to write down the command.]
9.12% of SAT Math scores lie below 400.
(iv) Method I P(X $\geq$ 710) = 0.255%
Sketch a figure to label the Mean, 500, and s.d., 75, with a region suitably shaded.
Since P = 0.255% << 5%, we find the score *very* impressive!
Method II P(X $\leq$ 710) = 99.75%
Sketch a figure to label the Mean, 500, and s.d., 75, with a region suitably shaded.
Since P = 99.75% >> 95%, we find the score *very* impressive!
Method III $Z = (X - \mu) / \sigma = (710 - 500)/75 = 2.8$
Since Z = 2.8 >> 1.645, we find the score *very* impressive!
Method IV A score that is 1.645s.d. above the mean would be impressive since that would happen < 5% of the time:
$\mu + 1.645\sigma = 500 + 1.645 \cdot 75 = 623.375$
Since 710 >> 623.375, it is *very* impressive.
(v) We need to find what % of scores are $\leq$ 780:
P(X $\leq$ 780)
Sketch a figure to label the Mean, 500, and s.d., 75, with a left region suitably shaded below 780.
P = Normalcdf(-9999, 780, 500, 75) = 99.99% [**Note:** it is NOT required to write down the command.]
A score of 780 corresponds to the 99.99th percentile: 99.99% of SAT Math scores lie at 780 or below.
(vi) We need to find a score, *a*, such that 65% of SAT scores are $\leq$ *a*: this is a *backwards* problem...so we use the **InvNorm** command!
Sketch a figure to label the Mean, 500, and s.d., 75, with a left region suitably shaded for 0.65 with *a* on the axis.
P(X < *a*) = 0.65 => *a* = InvNorm(0.65, 500, 75) = 528.89 [**Note:** it is NOT required to write down the command.]
(vii) We need a score *a* such that P(X < *a*) = 0.2: this is a *backwards* problem...so we use the **InvNorm** command!
Sketch a figure to label the Mean, 500, and s.d., 75, with a left region suitably shaded for 0.2 with *a* on the axis.
P(X < *a*) = 0.2 => *a* = InvNorm(0.20, 500, 75) = 436.87 [**Note:** it is NOT required to write down the command.]
(viii) We need a score *a* such that P(X < *a*) = 0.9 [since the calculator can *only* process left Area / %]: this is a *backwards* problem...so we use the **InvNorm** command!
Sketch a figure to label the Mean, 500, and s.d., 75, with a left region suitably shaded for 0.9 with *a* on the axis.
P(X < *a*) = 0.9 => *a* = InvNorm(0.90, 500, 75) = 596.11 [**Note:** it is NOT required to write down the command.]

(ix) If 30% of scores are in the middle, then 70% are left over and distributed *symmetrically* on the left and on the right => we need the 35th percentile and (30 + 35) = 65th percentile.
As in (v), We need to find 2 scores, *a* and *b*, such that 35% of SAT scores are $\leq a$ and 65% of scores are $\leq b$: this is a *backwards* problem...so we use the **InvNorm** command!
Sketch 2 figures to label the Mean, 500, and s.d., 75, with a left region suitably shaded for 0.35 with *a* on the axis, and with a left region suitably shaded for 0.65 with *b* on the axis.
P(X < *a*) = 0.35 => *a* = InvNorm(0.35, 500, 75) = 471.10 [**Note:** it is NOT required to write down the command.]
P(X < *b*) = 0.65 => *b* = InvNorm(0.65, 500, 75) = 528.89 [**Note:** it is NOT required to write down the command.]
The 2 scores that constitute the middle 30% of SAT Math scores are 480 and 530. The *spread* of the middle 30% of scores is 530 – 480 = 50.
(x) In a Normal distribution, since 99.7% of observations lie within 3 s.d. of the Mean [$\mu \pm 3\sigma$], in this case, virtually all Math SAT scores shall within 3.s.d of 500 i.e. between [500 – 3·75, 500 + 3·75] = [275, 725]
(xi) The SAT score that lies 1.5s.d. below the Mean is $\mu - 1.5\sigma$ = 500 – 1.5(75) = 387.5.
(xii) The 2 scores that lie 1.75s.d. from the Mean are: $\mu \pm 1.75\sigma$ i.e. [500 – 1.75·75, 500 + 1.75·75] ~ [370, 630]
(xiii) According to the Empirical Rule,

1. 68% of scores lie within 1s.d. of the mean $\mu \pm 1\sigma$ i.e. [500 – 1·75, 500 + 1·75] ~ [425, 575]
2. 95% of scores lie within 2s.d. of the mean $\mu \pm 2\sigma$ i.e. [500 – 2·75, 500 + 2·75] ~ [350, 650]
3. 99.7% of scores lie within 3s.d. of the mean $\mu \pm 3\sigma$ i.e. [500 – 3·75, 500 + 3·75] ~ [275, 725]

(xiv) Since most [~95%] of scores like between 350 and 650 [**see (xiii)**], scores *beyond* those limits might be regarded as unusually low / high, respectively!
Note! You might also use 99.7% limits of 275 and 725...or you may use the $\mu \pm 1.645\sigma$ limits pf [376.625, 623.375].
(xv) IQR = $75^{th} - 25^{th}$ percentiles. We need to find 2 scores, *a* and *b*, such that 25% of SAT scores are $\leq a$ and 75% of scores are $\leq b$: this is a *backwards* problem...so we use the **InvNorm** command!
Sketch 2 figures to label the Mean, 500, and s.d., 75, with a left region suitably shaded for 0.25 with *a* on the axis, and with a left region suitably shaded for 0.75 with *b* on the axis.
P(X < *a*) = 0.25 => *a* = InvNorm(0.25, 500, 75) = 449.4133 [**Note:** it is NOT required to write down the command.]
P(X < *b*) = 0.75 => *b* = InvNorm(0.75, 500, 75) = 550.5867 [**Note:** it is NOT required to write down the command.]
The IQR of 550.5867 – 449.4133 = 101.1734 is the *spread* of the middle 50% of scores
(xvi) First, we find: P(X < 600) = 90.87% [Sketch a figure to label the Mean, 500, and s.d., 75 and shade suitably...]
Next, P(X > 700) = 0.38%
Required P = 90.87% + 0.38% = 91.25%
If 5895 students took the SAT, then 91.25%·(5000) = 4562 scored below 600 or above 700.

Example 2

The blood glucose level (BGL) of 1200 patients being tested for diabetes (under the age of 50) was found (after a 12–hour fast), to be roughly normal with a mean of 85 mg of glucose per deciliter of blood & s.d 25 mg/dl.
1. Between what BGLs would <u>most</u> of the observations lie? Write a simple sentence to <u>clearly</u>

explain your choice.
2. *How many* patients have BGLs below 10mg/dl?
3. A BGL of 141.25 mg/dl is how many s.d. from the mean?
4. What BGL corresponds to the 16^{th} *percentile*? Interpret this.
5. Calculate and *interpret* the Z–score corresponding to a BGL of 41.25 mg/dl.
6. How many of the patients had BGLs *less than* 50 mg/dl or *more than* 105 mg/dl?
7. Calculate the IQR of the BGLs. Interpret this in simple English.
8a) What interval of BGLs captures all observations lying *within* 2.75 s.d. of the mean?

Solution.

Let X be a r.v. denoting BGLs (in mg/dl).
X ~ N(85, 25).
1. Since BGL are normally distributed, most [95%] of the BGLs shall lie within 2s.d. of the mean BGL of 85: i.e. between 85 – 2·25 = 35mg/dl and 85 + 2·25 = 135mg/dl.
2. Note: we used common–sense to resolve the Q: *if* we knew what % of individuals, in general, had a BGL of < 10, then observing that there are 1200 of them, we could determine the *Number* of individuals satisfying the condition!
P(X < 10) = 0.001349
– Sketch, *label* and shade a figure to illustrate the situation.
Therefore, 0.1349%·1200 = 1.6 ~ About 2 individuals would have a BGL below 10mg/dl.
3. The phrasing reveals it to be Z–score problem!
Z = (X – μ) / σ = (141.25 – 85)/25 = 2.25
A BGL of 141.25mg/dl is 2.25 s.d. above the mean BGL of 85.
4. P(X ≤ *a*) = 0.16
– Sketch, *label* and shade a figure to illustrate the situation.
a = 60.14
A BGL of 60.14mg/dl corresponds to the 16th percentile indicating that 16% of the patients had a BGL of [**or** 16% of the BGLs were] 60.14mg/dl or less.
5. Z = (X – μ) / σ = (41.25 – 85)/25 = –1.75
A BGL of 41.25mg/dl lies 1.75s.d. below the Mean BGL of 85mg/dl.
6. P(X < 50 or X > 105)
= P(X < 50) + P(X > 105)
= 8.07% + 21.18%
= 29.26%
– Sketch, *label* and shade a figure to illustrate the situation.
Therefore, 29.26%·1200 = 351.12 About 351 patients would have a BGL below 50mg/dl or more than 105.
7. P(X < Q1) = 0.25 and P(X < Q3) = 0.75
– Sketch and *label* a figure to illustrate the situation.
Q1 = 68.13
Q3 = 101.86
The IQR, the range of the middle 50% of BGLs, is Q3 – Q1 = 33.72 mg/dl.
8a) An interval of [85 – 2.75·25, 85 + 2.75·25] = [16.25, 153.75] would capture all BGLs within 2.75s.d. from the mean of 85mg/dl.
b) P(16.25 < X < 153.75) = 99.4%
– Sketch, *label* and shade a figure to illustrate the situation.
Therefore, 99.4%·1200 ~ 1193 individuals have a BGL between 16.25 and 153.74mg/dl.

c) An interval of [85 – 1.75·25, 85 + 1.75·25] = [41.25, 128.75] would capture all BGLs within 1.75s.d. from the mean of 85mg/dl.

Method I

P(X < 41.25) + P(X > 128.75) = 4% + 4% [can you see why?!] = 8%

– Sketch, *label* and shade a figure to illustrate the situation.

Method II

P(X < 41.25) + P(X > 128.75)

= 1 – P(41.25 < X < 128.75)

= 8%

– Sketch, *label* and shade a figure to illustrate the situation.

Therefore, 8% of individuals have a BGL within 1.75s.d. from the mean.

9. As in 4. Do it yourselves!

– Interpret the Q in Probability Notation.

– Sketch, *label* and shade a figure to illustrate the situation.

Example 3

In the 1910s, the batting averages of baseball players was approximately normal with a mean of 266 and a standard deviation of 37.1. In the 1940s, the batting averages were also approximately normal with a mean of 267 and a standard deviation of 32.6.

Define a suitable random variables and state their distribution using Notation.

(i) In 1910, what proportion of batting averages exceeded 300?

(ii) In 1940, batting average corresponded to the 80th percentile?

(iii) In 1910, what percentage of batting averages were below 200? Is this rare? Explain.

(iv) In 1940, find the percentile that a batting average of 320 corresponds to. Interpret it in context.

(v) In 1910, find the batting average that separated the bottom 10% from the rest.

(vi) Find the 2 batting averages – in 1940 – that constituted the middle 50%. Then, calculate and interpret the IQR. **Tip!** Which 2 percentiles do we need to find?

(vii) In 1910, what batting averages might be unusually high? Unusually low? **Tip!** Unusual ~ Rare!

(viii) In 1940, which batting average lie 1.645s.d. above the mean? **Tip!** Translate it from English to algebra. This is an English → Algebra translation problem and does not involve NormalCdf or InvNorm.

(ix) What 2 batting averages in 1910 lie 1.96 s.d. from the Mean?

(x) Ty Cobb's batting average in 1911 was 420 while Ted William's was 406 in 1941. Relative to their decades, who was the superior baseball player? Explain.

Solution. Let X and Y be a r.v. denoting the batting averages of baseball players in the 1910s and 1940s, respectively. X ~ N(266, 37.1) and Y ~ N(267, 32.6).

(i) P(X > 300) = 17.97% using NormalCdf. Sketch, label info and shade.

(ii) P(Y ≤ *a*) = 0.8 → Sketch, label info and shade: *a* = 294.43 using InvNorm.

(iii) $P(X \leq 200) = 3.76\%$ using NormalCdf. Sketch, label info and shade. Since $P = 3.76\% < 5\%$, this would be rare.

(iv)$P(Y \leq 320) = 94.8\%$ using NormalCdf. Sketch, label info and shade. A batting average of 320 corresponds to the 94.8th percentile indicating that 94.8% of batting averages in 1940 were 320 or less **OR** 94.8% of players in 1940 had batting averages of 320 or less.

(v) $P(X \leq a) = 0.1 \rightarrow$ Sketch, label info and shade: $a = 218.45$ using InvNorm.

(vi) $P(Y \leq a) = 0.25$ and $P(Y \leq b) = 0.75 \rightarrow$ Sketch, label info and shade: $a = 245.01$ and $b = 288.98$ using InvNorm. IQR = 43.97, which is the **spread** of the middle 50% of batting averages.

(vii) $P(X \leq a) = 0.05$ and $P(X \leq b) = 0.95 \rightarrow$ Sketch, label info and shade: $a = 204.97$ and $b = 327.02$ using InvNorm.

(viii) $Y = \mu + 1.645\sigma = 320.63$.

(ix) $Y = \mu \pm 1.96\sigma = [193.28, 338.72]$. Simply substitute $\mu = 266$ and $\sigma = 37.1$ into the expression!
(x) Ty Cobb's percentile: $P(X \leq 420) = 99.99983435\%$ **[haha!]** whereas Ted William's percentile: $P(Y \leq 406) = 99.99899443\%$ **[haha!]**. Relative to their decades, Ty Cobb's the MUCH [!] better player since he did better than a greater [!] proportion of his peers.

Example 4.

a) Given $X \sim N(\mu, \sigma)$ lies in the 10th percentile, how many s.d. from the mean of the distribution is X?
b) Given $X \sim N(\mu, \sigma)$ is 1.28s.d. below the mean of the distribution, calculate its percentile.
c) Given $X \sim N(\mu, \sigma)$ lies in the 80th percentile, how many s.d. from the mean of the distribution is X?
d) Given $X \sim N(\mu, \sigma)$ is 2.326s.d. above the mean of the distribution, what proportion of values exceed X?
Expectations. You **must** sketch a figure, label the values given + label μ, σ for a suitable distribution. Then write a probability statement...and answer the Q.

Solution.

Skills you needed to answer this Q: migrating between percentiles and X-values using the N(0, 1) distribution, distinguishing between forwards and backwards problems, recognizing that calculator operates on percentiles alone, that X-values below the mean -> negative Z-scores.

a) $P(X \leq a) = 0.10 \rightarrow a = -1.28$
b) $P(X \leq -1.28) = 0.10$
c) $P(X \leq a) = 0.80 \rightarrow a = 0.8416$
d) $P(X \geq 2.326) = 0.01$

Example 5.

a) Suppose crocodile egg-lengths were approximately Normal.
i) A wildlife zoologist measures the weight of a crocodile egg and finds it to be 0.2856s.d. below the Mean weight of that species. What % of eggs are **heavier** than this egg? **Tip!** Is *a* on the left-side or right? It should be obvious.
ii) A wildlife zoologist measures the weight of a crocodile egg and finds it be heavier than 21% of the eggs of that species. How many s.d. from the mean is the weight of this egg? **Tip!** Is *a* on the left-side or right? Read the English carefully → think!

b) A herpetologist finds that snake lengths of a certain species to be approximately Normal.
i) If a certain snake of that species has a length that is 1.06s.d. above the Mean length of that species, calculate the percentile length of the snake. **Tip!** Is *a* on the left-side or right? It should be obvious.
ii) If a certain snake of a certain species is shorter than only 3% of snakes of that species, how many s.d. from the mean is this snake's length? **Tip!** Is *a* on the left-side or right? Read the English carefully → think!
Expectations. You **must** sketch a figure when necessary, label the values given + label μ, σ for a suitable distribution. Then write a probability statement...and answer the Q.

Solution.

Skills you needed to answer this Q: migrating between percentiles and X-values using the N(0, 1) distribution, distinguishing between forwards and backwards problems, recognizing that calculator operates on percentiles alone, that X-values below the mean -> negative Z-scores.
a) (i) $P(Z > -0.2865) = 61.28\%$; **(ii)** $P(Z \leq a) = 0.21 \rightarrow a = -0.8064$
b) (i) $P(Z < 1.06) = 85.54\%$; **(ii)** $P(Z \leq a) = 0.97 \rightarrow a = 1.88$

Example 6.

1. Write an algebraic expression relating X, μ and σ to translate: A weight that is 1.25s.d.below the mean.
2. Write an algebraic expression relating X, μ and σ to translate: A height is 0.5912s.d. above the mean.
3. Write an algebraic expression relating X, μ and σ to translate: A cholesterol level is 0.3456 s.d. below the mean.
4. Write an algebraic expression relating X, μ and σ to translate: A batting average is 1.25s.d. below the mean.
5. Sketch a figure and label the information: Given X = 130, μ = 168, Z-score = -0.6025, σ = unknown. Write an algebraic expression relating X, μ and σ. Solve for the missing parameter.
6. A weight is 1.25s.d. below the mean. If the mean weight is 121lbs and the s.d. is 6.5lbs, find the weight.
7. A height of 68in is is 0.3456s.d. above the mean. If the s.d. of the population is 3.1in, find the mean.
8. A test-score of 130 is 0.6025s.d. below the mean score of 168. Find the s.d.
9. A test score of 792 is 1.2149s.d. above the mean score. If the s.d. is 162, find the Mean.
10. Sketch a figure and label the information: Given X = 792 which lies in the 37th percentile, μ = unknown, σ = 162. Solve for the missing parameter. **Tip!** What should you find 1st?
11. Sketch a figure and label the information: Given X = 1256 which separates the top 37% of values from the rest, μ = unknown, σ = 162. Write an algebraic expression relating X, μ and σ. Solve for the missing parameter. **Tip!** What should you find 1st?

12. Sketch a figure and label the information: Given X = 45 which lies in the top 17% of values. μ = 31, σ = unknown. Solve for the missing parameter. **Tip!** What should you find 1st?
13. Sketch **2 figures** and label the information: Given X = 81 which lies in the 26th

percentile; $X = 141$ which corresponds to the top 6% of values, μ = unknown, σ = unknown. Write 2 equations involving μ and σ…and stop! **Note:** Don't solve them.

14. Sketch **2 figures** and label the information: Given $X = 36$ which lies in the 56^{th} percentile; $X = 94$ which corresponds to the 81^{st} percentile, μ = unknown, σ = unknown. Write 2 equations involving μ and σ…and stop! **Note:** Don't solve them.

15. Sketch **2 figures** and label the information: Given $X = 350$ which corresponds to the top 81% of values; $X = 419$ which corresponds to the 38^{th} percentile, μ = unknown, σ = unknown. Write 2 equations involving μ and σ…and stop! **Note:** Don't solve them.

Solution.

Skills you needed to answer this Q: calculating Z-scores for percentiles, recognizing that calculator operates on percentiles alone, that X-values below the mean → negative Z-scores.

1. $X = \mu + Z\sigma = \mu - 1.25\sigma$

2. $X = \mu + Z\sigma = \mu + 0.5912\sigma$

3-4. Do these yourselves.

5. $X = \mu + Z\sigma \rightarrow 130 = 168 - 0.6025\sigma \rightarrow \sigma = 63.07$

6. $X = \mu + Z\sigma = 121 - 1.25 \cdot 6.5 = 112.875$

7. $X = \mu + Z\sigma \rightarrow 68 = \mu + 0.3456 \cdot 3.1 \rightarrow \mu = 66.9286$in

8. $X = \mu + Z\sigma \rightarrow 130 = 168 - 0.6025\sigma \rightarrow \sigma = 63.07$

9. $X = \mu + Z\sigma \rightarrow 792 = \mu + 1.2149 \cdot 162 \rightarrow 595.1862$

10. $X = \mu + Z\sigma \rightarrow 792 = \mu - 0.3318 \cdot 162 \rightarrow 845.7516$

11. $X = \mu + Z\sigma \rightarrow 1256 = \mu + 0.3318 \cdot 162 \rightarrow \sigma = 1202.2484$

12. $X = \mu + Z\sigma \rightarrow 45 = 31 + 0.9541\sigma \rightarrow \sigma = 14.6725$

13. Sketch 2 figures, $X \sim N(\mu, \sigma)$ labeling all the given information. Using $X = \mu + Z\sigma$:
$81 = \mu - 0.66433\sigma$; $141 = \mu + 1.5548\sigma$

14. Sketch 2 figures, $X \sim N(\mu, \sigma)$ labeling all the given information. Using $X = \mu + Z\sigma$:
$36 = \mu + 0.1510\sigma$; $94 = \mu + 0.8779\sigma$

15. Sketch 2 figures, $X \sim N(\mu, \sigma)$ labeling all the given information. Using $X = \mu + Z\sigma$:
$350 = \mu - 0.8779\sigma$; $419 = \mu - 0.3055\sigma$

Example 7.

1. A height is 1.6921s.d. above the mean height. If the mean height is is 67.5in and the s.d. is 2.5lbs, find the height.

2. A mileage of 26mpg is 0.4936s.d. below the mean mileage. If the s.d. of the mileage of vehicles in that category is 2.63mpg, find the mean.

3. HDL cholesterol levels of 60 corresponds to the top 29% of cholesterol levels for 20-29 year old females. Assuming normality, if the mean HDL cholesterol level is 53, find the s.d.

4. The lengths of human pregnancies is approximately normal with a s.d. of 16days. If 84.13% of pregnancies last longer than 250days, find the mean gestation period.

Use for 5-6. The mean serum high-density lipoprotein (HDL) cholesterol of females 20-29 years old is approximately normal.

5. An HDL in the 42^{nd} percentile lies how many s.d. from the mean?

6. An HDL cholesterol level is 1.28s.d. below the mean. Calculate its percentile.

Solution.

1. $X = \mu + Z\sigma = 67.5 + 1.6921 \cdot 2.5 = 71.73025$in

2. $X = \mu + Z\sigma \rightarrow 26 = \mu - 0.4936 \cdot 2.63 = 27.2982$mpg
3. $X = \mu + Z\sigma \rightarrow 60 = 53 + 0.5534 \cdot \sigma \rightarrow \sigma = 12.6491$ **Sketch, label figure.**
4. $X = \mu + Z\sigma \rightarrow 250 = \mu - 0.9998 \cdot 16 \rightarrow \mu = 266$days **Sketch, label figure.**
5. $P(Z \leq a) = 0.42 \rightarrow a = -0.2019$ **Sketch, label figure.**
6. $P(Z \leq -1.28) \approx 0.10$ **Sketch, label figure.**