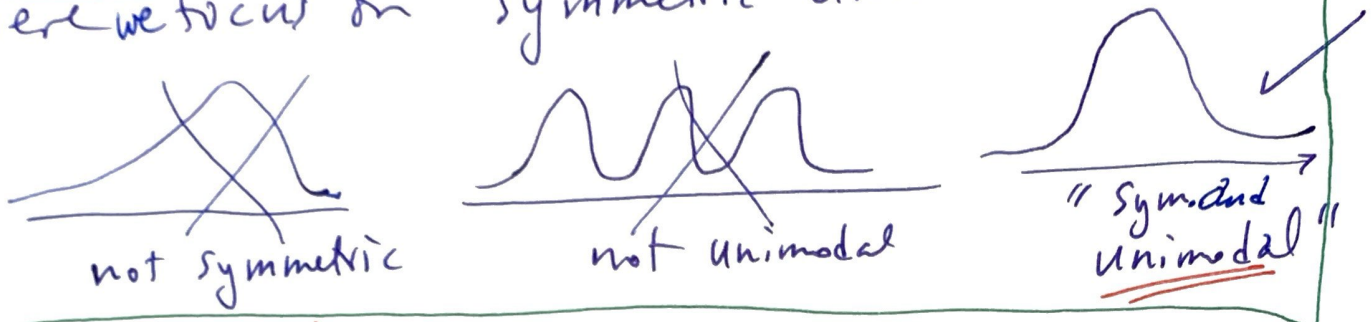
Chapter 7 The Normal Distribution (Bell curve)

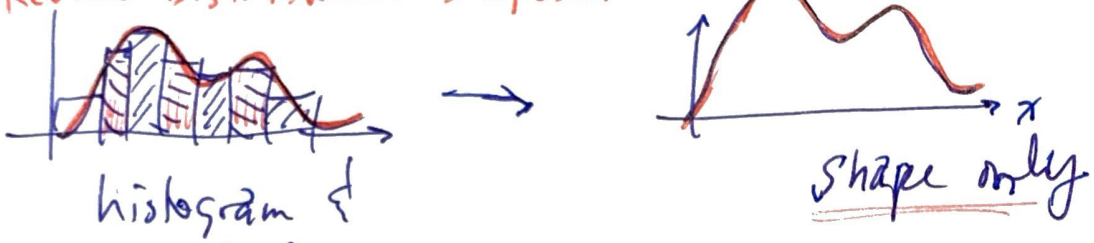
7.1 The "standard" Normal distribution

- Recall Data → Buckets → Histogram → Shape.

Here we focus on symmetric unimodal shapes

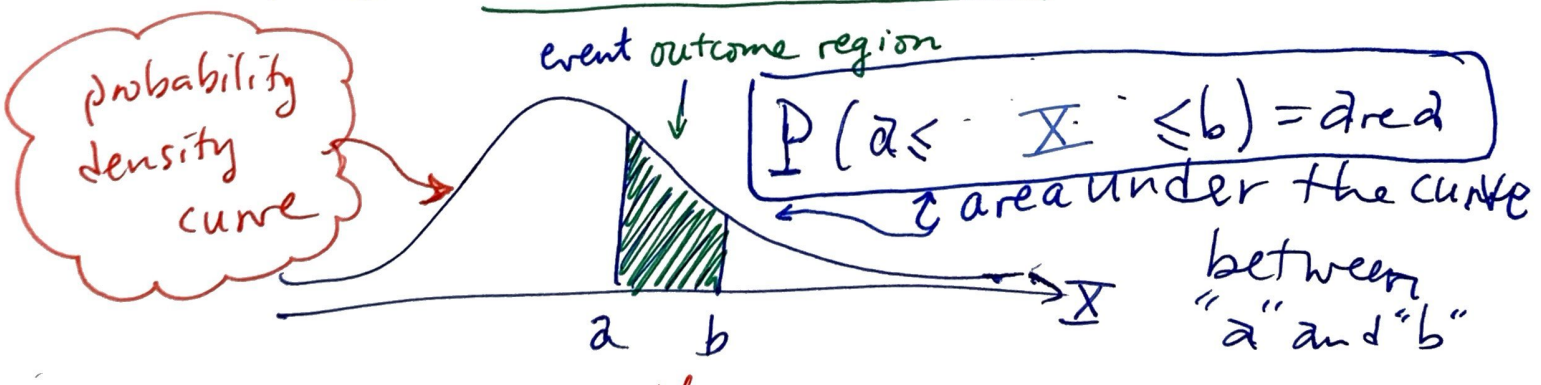


Review Distribution Shapes...



- Recall a probability "histogram" has a net area = 1

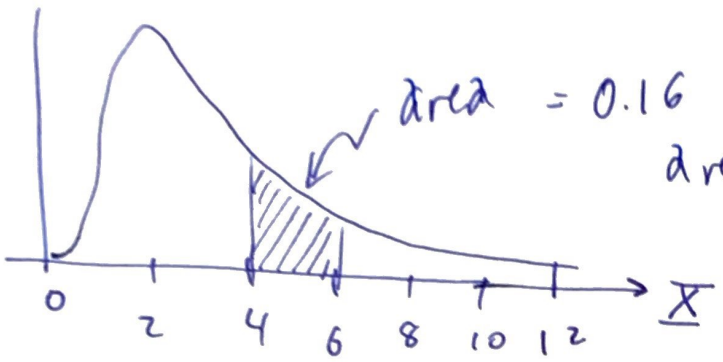
- This allows us to relate the probability of an outcome to the area of the region that the outcome defines -



a ✓ iable

Ex

Consider the prob. density curve for a population.



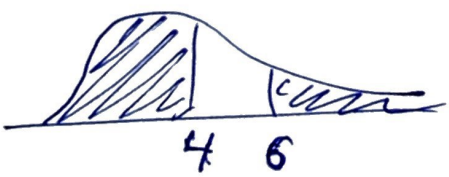
a) What proportion of the population's parameter lies between $\bar{X} = 4$ to 6 ?

$P(4 \le X \le 6) = \text{area} = \underline{0.16}$

b) If a subject from the population is chosen at random, what is the prob. that the value of the parameter being studied is between 4 & 6

ans: **0.16**

c) What proportion of the population is not between 4 & 6



$P(\text{not between } 4 \text{ \& } 6)$

$= 1 - P(4 \le x \le 6)$

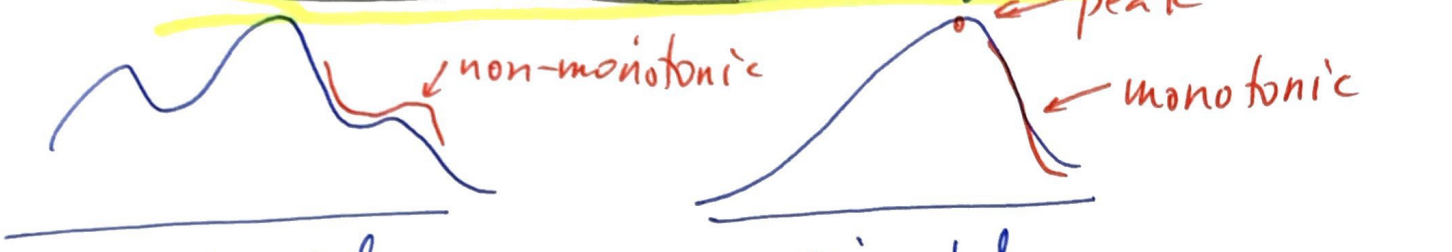
$= 1 - 0.16$

$= \boxed{0.84}$

⊛ Normal Distribution (Bell Curve)

Remarkably many important statistical surveys can be described using a normal distribution curve

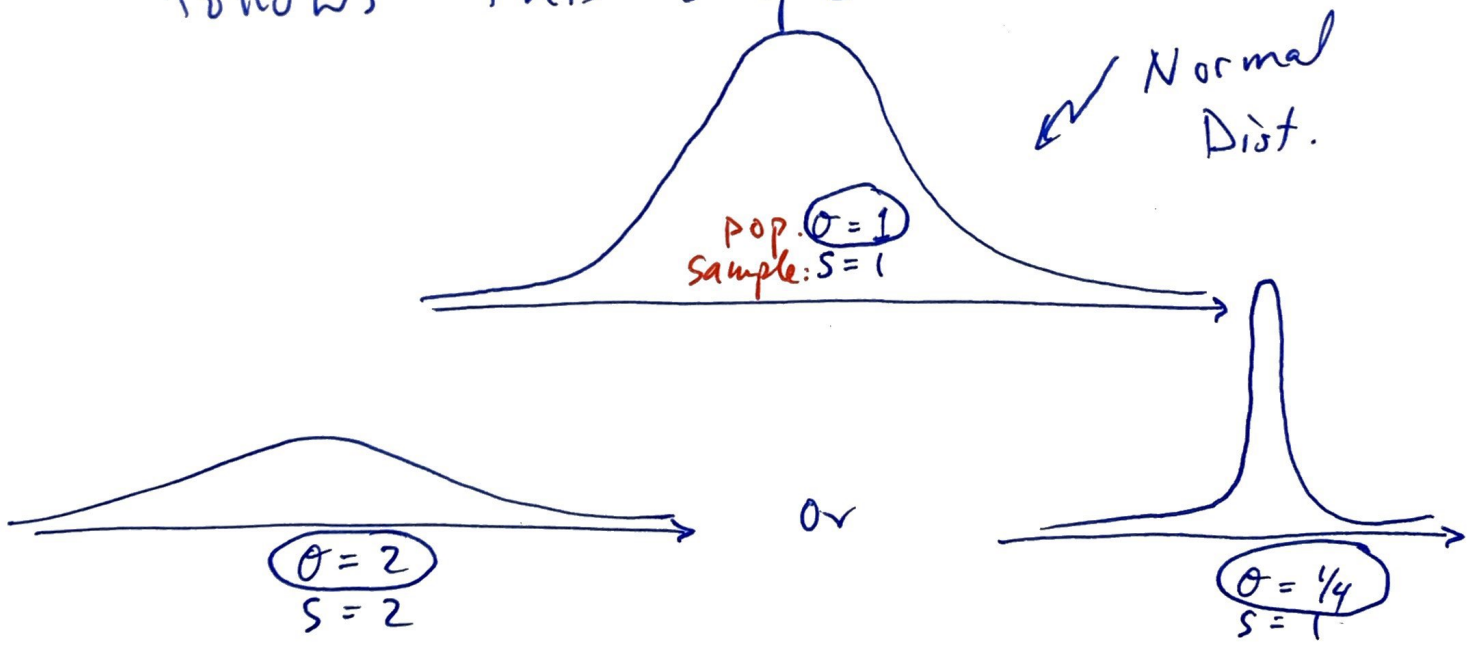
- these are unimodal, symmetric and monotonically decreasing from the peak.



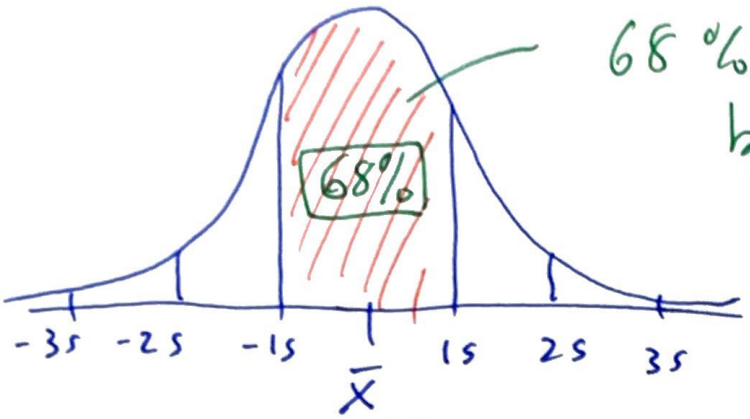
- unimodal
- non-symmetric
- non-monotonic

- unimodal
- non-sym.
- monotonic (uniform fall) no bumps before or after the peak

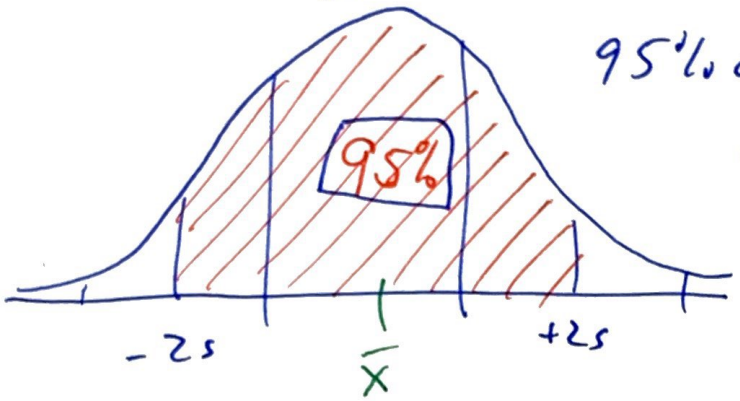
So our normal distribution follows this shape



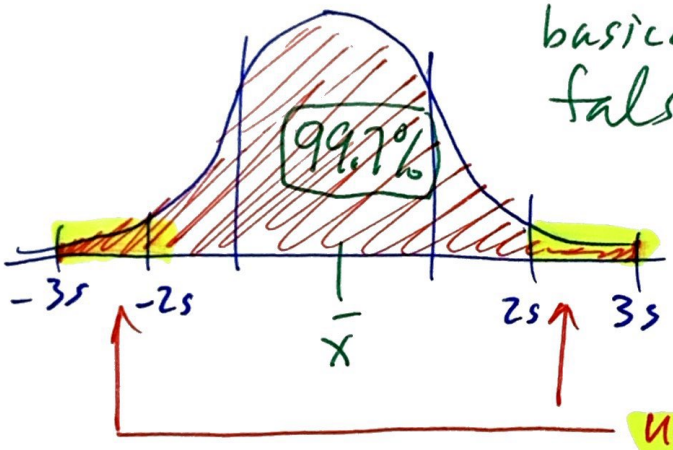
⊛ Rule-of-thumb applies (chebychevs)



68% of data falls between $(\bar{x}-s, \bar{x}+s)$



95% of the data falls between $(\bar{x}-2s, \bar{x}+2s)$ (usuall data)



basically all data fals between $(\bar{x}-3s, \bar{x}+3s)$

unusual to find data past $\pm 2s$

⊛ z-score

$$z = \frac{x - \bar{x}}{s} = \frac{x - \mu}{\sigma}$$

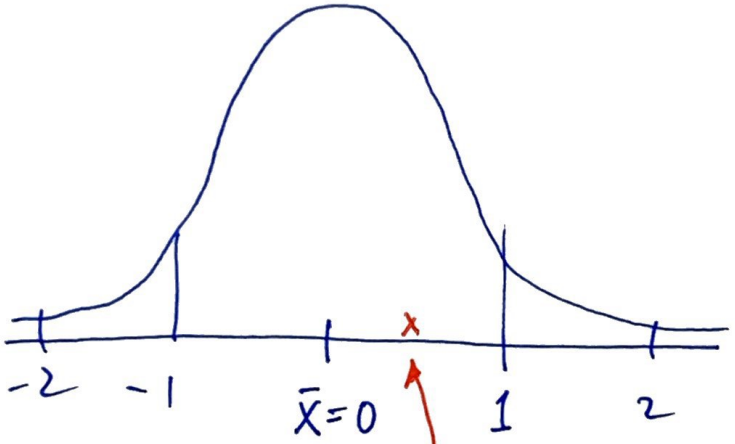
($\bar{x}$: sample data; μ: census results)

Each Data point in a set has a z-score

the z-score tells us how many std. dev's we are from the mean value.

⊛ "Standard" Normal Distribution

Here $\bar{x} = 0$ $(\mu = 0)$ and the std. dev $s = 1$ $(\sigma = 1)$.



ex! $x = 0.5$ in a std. norm dist.

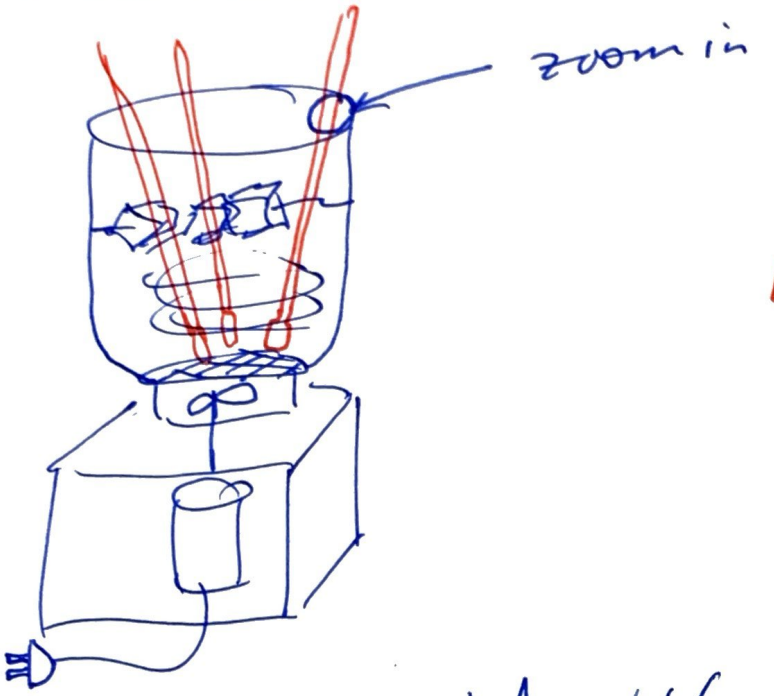
$$z = \frac{0.5 - 0}{1} = 0.5$$

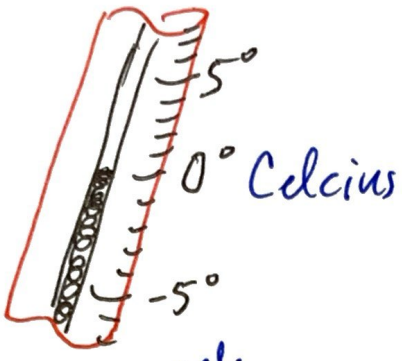
(x → 0.5, $\bar{x}$ → 0, s → 1)

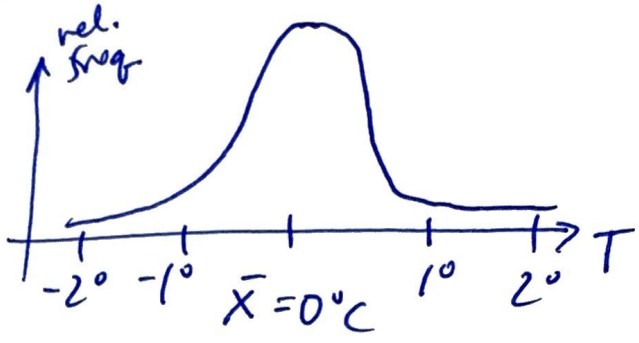
So the x in a std. normal dist is the z-score.

Ex Thermometer Factory

ACME thermometers produce products that have a mean of 0°C and a std. deviation of 1°C when the thermometer is placed in 0°C ice water





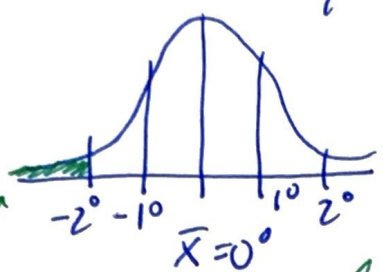


Model = $N(\bar{x}=0, s=1)$ = $N(0,1)$

Ex Using the Rule of thumb to approximate the areas under a normal distribution, find the probability that a randomly extracted bromine water thermometer measures below -2°C

Steps to ans prob. distribution questions

(i) Picture



(ii) shade in the area requested

(iii) calculate that area using the rule of thumb:

$0.0250 + 0.9500 + 0.0250 = 1.0000$

(iv) answer the question

$$P(\bar{X} < -2°) = 0.0250$$ ⊛

or 2.5% which is unusual.

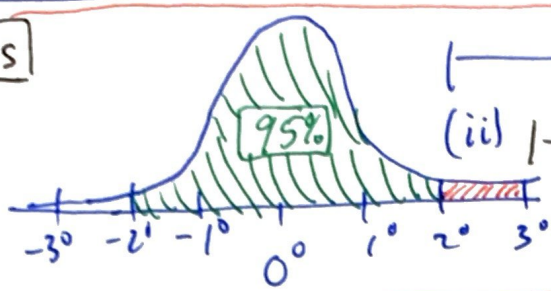
Comment: due to symmetry

$$P(\bar{X} > 2°) = 0.0250$$

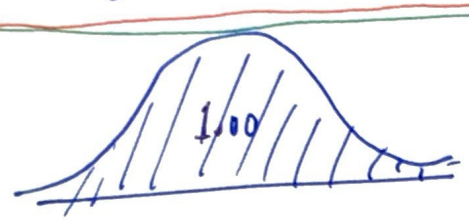
⊛ Rules-of-thumb are not exact answers - just approximations

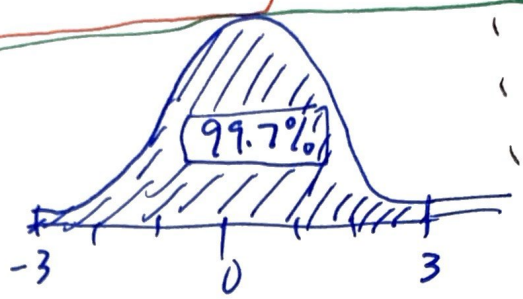
Ex What is the probability that the selected thermometer has a temp between 2°C & 3°C ?

(i) 2s

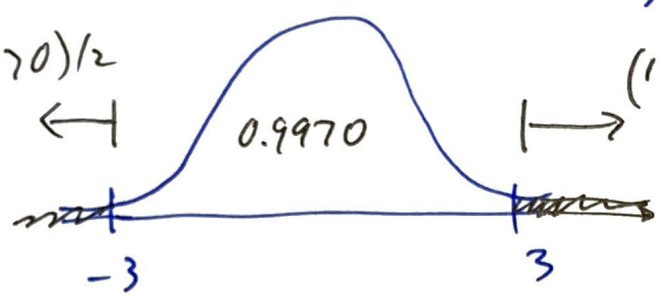


(iii) 3s





$(1.0000 - 0.9970)/2 = 0.0015$



$(1.0000 - 0.9970)/2 = 0.0030/2 = \boxed{0.0015}$ right wing > 3°C

So to answer the question

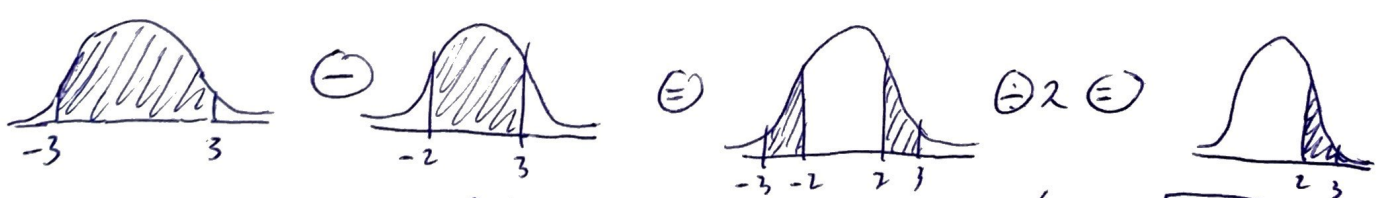
$$P(2°C < X < 3°C) = P(X > 2°) - P(X > 3°)$$

$$= 0.0250 - 0.0015$$

$$= 0.0235 \text{ or } \boxed{2.35\%}$$

chance the therm. is between 2° & 3°

Alternatively



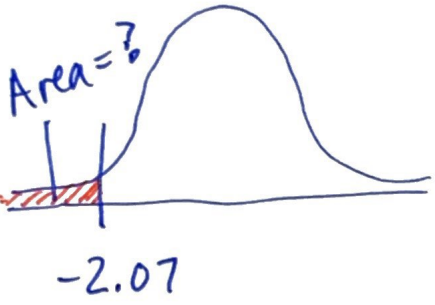
$0.9970 - 0.9500 = 0.047$ $/2 = \boxed{0.0235}$

(*) We now focus on more detailed questions

Assume you are a researcher in ice studies (pollutants effect on ice freezing). You need a much more accurate thermometer, and much more flexability in your probability calculations - more acc'y than Rule-of-thumb

Ex What is the probability that a polluted ice sample will freeze not at 0°C but rather at -2.07°C or less (Assume the ice follows $N(0,1)$)

(i) & (ii)



The rule of thumb will not be accurate enough.
We need the z-tables.

Pause the example and introduce the z-tables.

These tables describe in detail the area to the left of a given z-score.

{ For our $N(0,1)$ model z is the direct temperature readouts }

Resume values of negative z-score numbers

z	.00	.01	...	.06	.07	.08	.09
⋮							
-2.2							
-2.1							
-2.0	0.0228	0.0222	.....	0.0197	0.0192	0.0188	0.0183
-1.9							

answer: $P(Z \leq -2.07) = 0.0192$

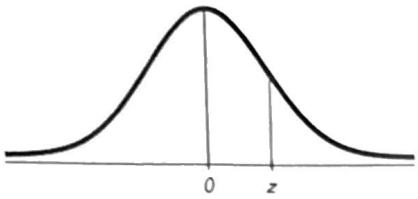
NEGATIVE z Scores

TABLE A-2 Standard Normal (z) Distribution: Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.50 and lower	.0001									
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051 *	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505 *	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

NOTE: For values of z below −3.49, use 0.0001 for the area.
*Use these common values that result from interpolation:

z score	Area
−1.645	0.0500
−2.575	0.0050



POSITIVE *z* Scores

TABLE A-2 (continued) Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495 *	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949 *	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.50 and up	.9999									

NOTE: For values of *z* above 3.49, use 0.9999 for the area.

*Use these common values that result from interpolation:

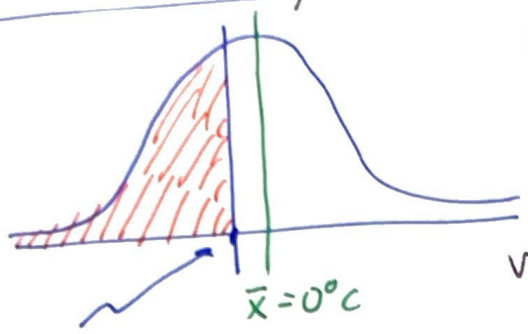
z score	Area
1.645	0.9500
2.575	0.9950

Common Critical Values

Confidence Level	Critical Value
0.90	1.645
0.95	1.96
0.99	2.575

EX (a) Find the probability that the thermometer reads $-0.28°C$ or lower, when in ice water.

(i) & (ii)

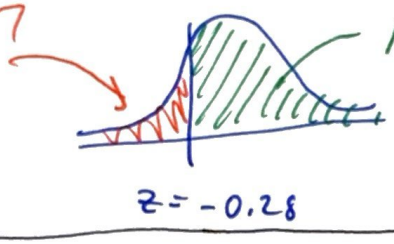


(iii) go to ~~z-table~~ and read row "0.2", col "0.08"

value ⇒ $\boxed{0.3897}$

answer $P(x < -0.28°C) = 0.3897$ or 38.97%

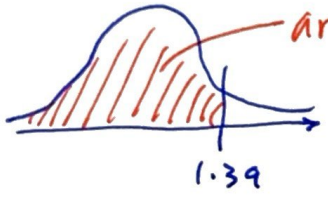
(b) Find the probability the readout reads greater than $-0.28°C$



$1.0000 - 0.3897$

$= \boxed{0.6103}$

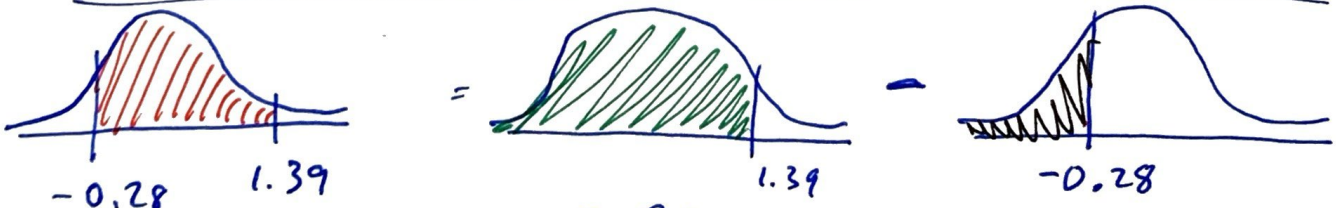
(c) Find the probability the read out say ≤ 1.39 or less



positive z-score table:

row 1.3, col 0.09 → value $\boxed{0.9177}$

(d) Find the probability that the readout is between -0.28 and 1.39

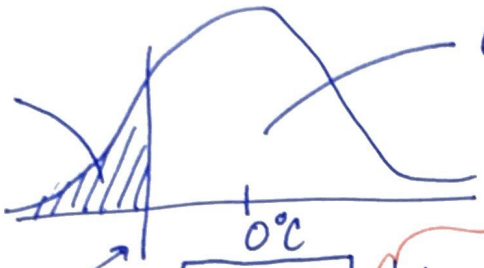


$= 0.9177 - 0.3897$

$= \boxed{0.5280}$ OR 52.8% chance readout is between $-0.28°C$ & $1.39°C$

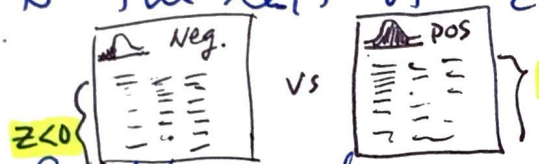
⊗ Reading the table in reverse.

Ex What z-score separates the lower 25% from the upper 75% of the thermometers



We go to the body of the z-table and look for the value 0.2500 and note row and column and then we form the z-score

- We know to look at the neg z-scores since area to the left of z=0.00 is less than 50%



So, z-scores for Neg. values page we see 0.2500 is in between two columns:

row	−0.6	col 0.07	and	col 0.08
		0.2514		0.2483

Ans: $z = -0.675$ Reverse Look Up

estimate between 0 and 9 depending on how close our desired number is to the given numb

25% of thermometers read −0.675°C or less.