

MA-250 Probability and Statistics

Nazar Khan

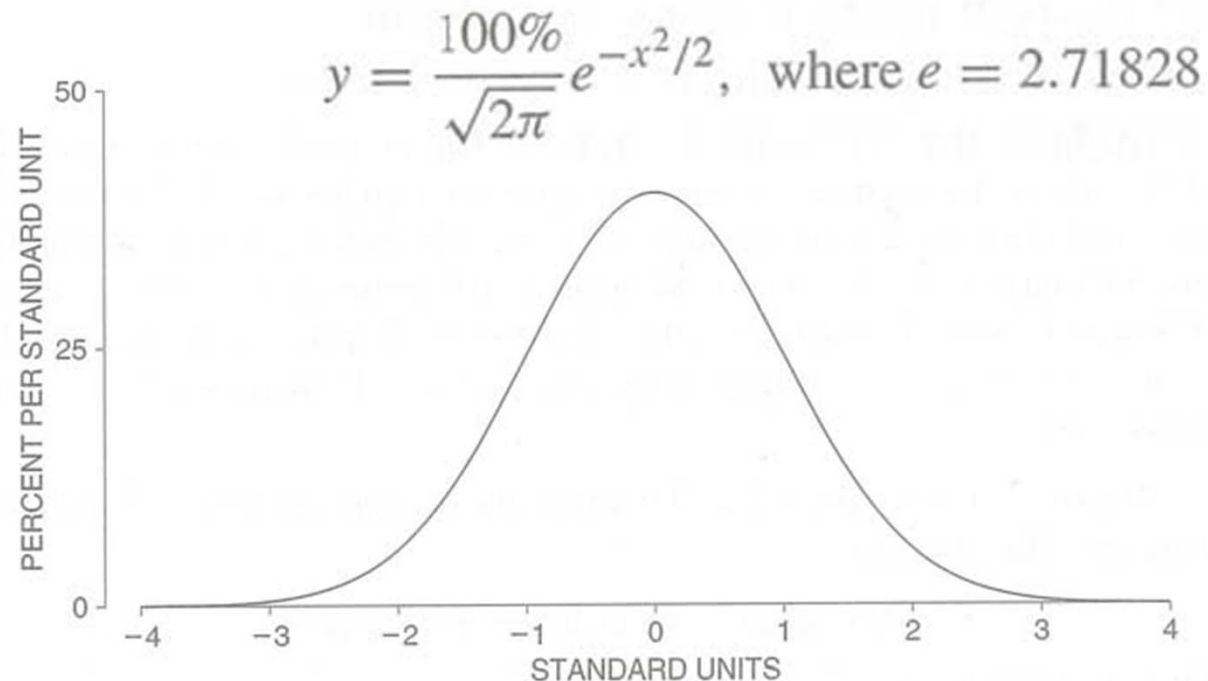
PUCIT

Lecture 4

The Normal Curve

- An approximation to data distribution that is **normally** quite accurate
 - Normally data follows such a distribution

The Normal Curve



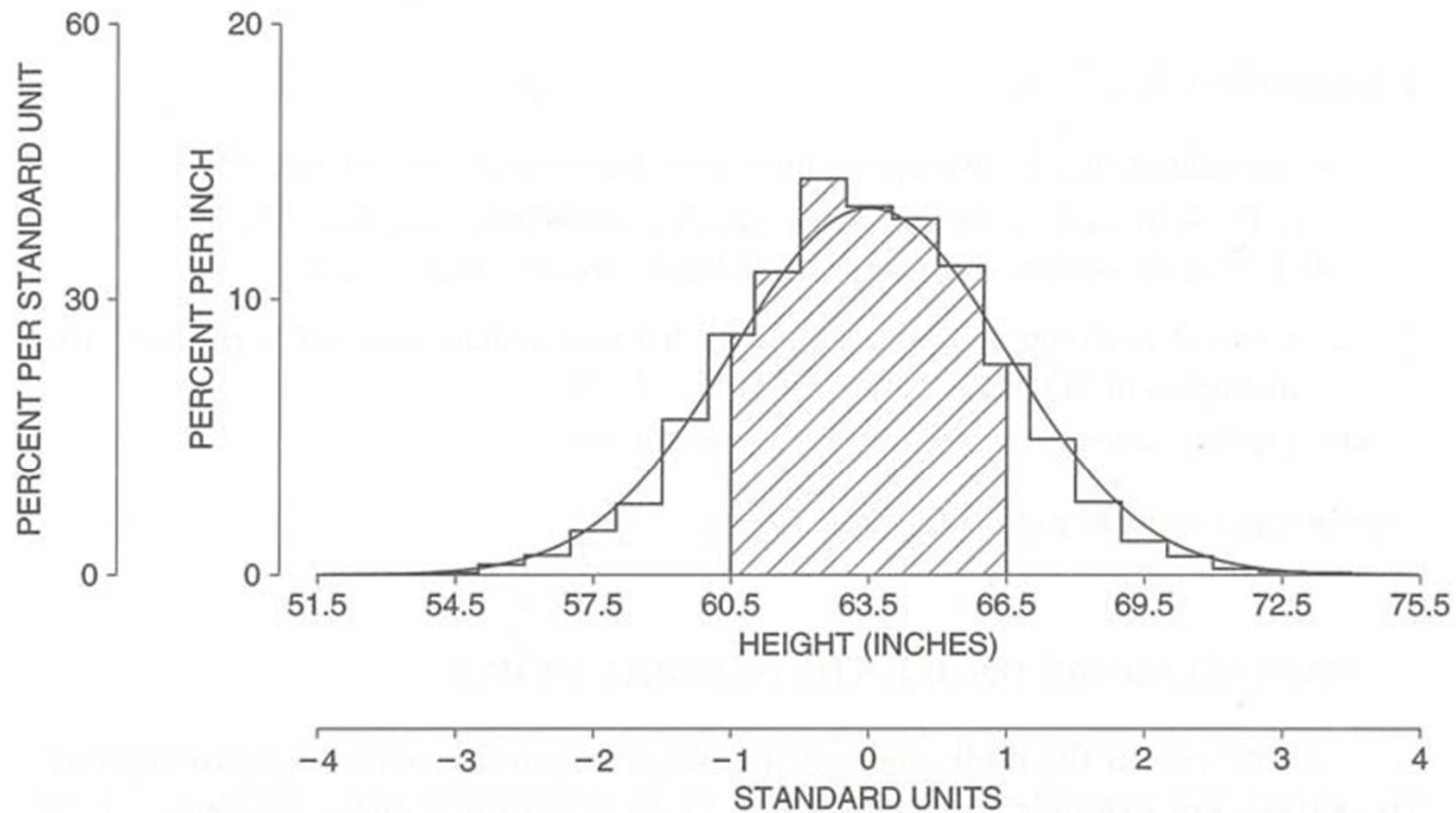
- the area under the normal curve between -1 and $+1$ is about 68%;
- the area under the normal curve between -2 and $+2$ is about 95%;
- the area under the normal curve between -3 and $+3$ is about 99.7%.

Standard Units

- Express the data in terms of standard deviation
- Converting a value X to standard units
 - $(X - \text{average}) / \text{SD}$

The Normal Approximation to Data

Figure 2. A histogram for heights of women compared to the normal curve. The area under the histogram between 60.5 inches and 66.5 inches (the percentage of women within one SD of average with respect to height) is about equal to the area between -1 and $+1$ under the curve—68%.

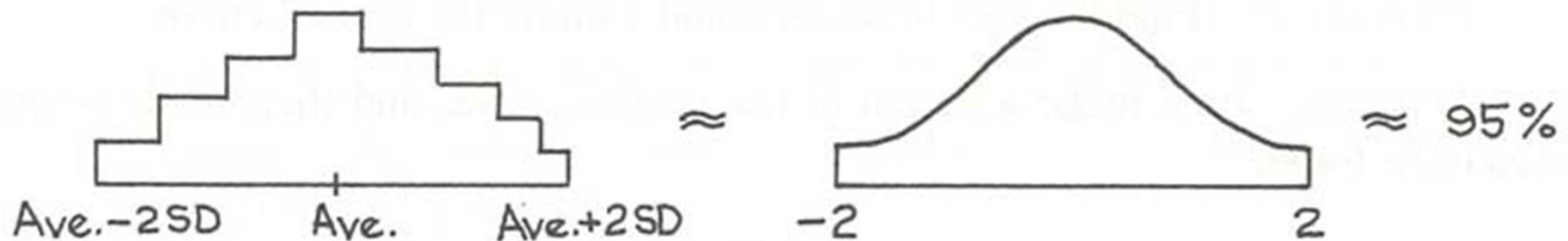


The Normal Approximation to Data

For many lists, roughly 95% of the entries are within 2 SDs of average. This is the range

average $-$ 2SDs to average $+$ 2SDs.

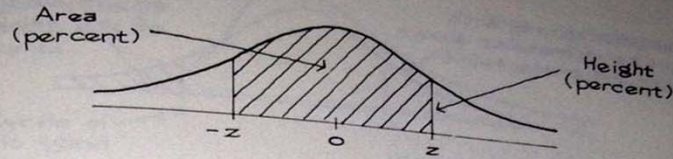
The reasoning is similar. If the histogram follows the normal curve, the area under the histogram will be about the same as the area under the curve. And the area under the curve between -2 and $+2$ is 95%:



On a certain exam, the average of the scores was 50 and the SD was 10.

- (a) Convert each of the following scores to standard units: 60, 45, 75.
- (b) Find the scores which in standard units are: 0, +1.5, -2.8.

Tables



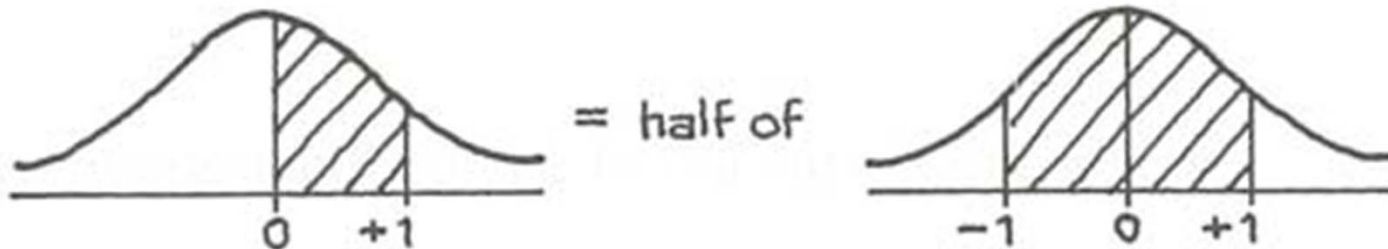
A NORMAL TABLE

z	Height	Area	z	Height	Area	z	Height	Area
0.00	39.89	0	1.50	12.95	86.64	3.00	0.443	99.730
0.05	39.84	3.99	1.55	12.00	87.89	3.05	0.381	99.771
0.10	39.69	7.97	1.60	11.09	89.04	3.10	0.327	99.806
0.15	39.45	11.92	1.65	10.23	90.11	3.15	0.279	99.837
0.20	39.10	15.85	1.70	9.40	91.09	3.20	0.238	99.863
0.25	38.67	19.74	1.75	8.63	91.99	3.25	0.203	99.885
0.30	38.14	23.58	1.80	7.90	92.81	3.30	0.172	99.903
0.35	37.52	27.37	1.85	7.21	93.57	3.35	0.146	99.919
0.40	36.83	31.08	1.90	6.56	94.26	3.40	0.123	99.933
0.45	36.05	34.73	1.95	5.96	94.88	3.45	0.104	99.944
0.50	35.21	38.29	2.00	5.40	95.45	3.50	0.087	99.953
0.55	34.29	41.77	2.05	4.88	95.96	3.55	0.073	99.961
0.60	33.32	45.15	2.10	4.40	96.43	3.60	0.061	99.968
0.65	32.30	48.43	2.15	3.96	96.84	3.65	0.051	99.974
0.70	31.23	51.61	2.20	3.55	97.22	3.70	0.042	99.978
0.75	30.11	54.67	2.25	3.17	97.56	3.75	0.035	99.982
0.80	28.97	57.63	2.30	2.83	97.86	3.80	0.029	99.986
0.85	27.80	60.47	2.35	2.52	98.12	3.85	0.024	99.988
0.90	26.61	63.19	2.40	2.24	98.36	3.90	0.020	99.990
0.95	25.41	65.79	2.45	1.98	98.57	3.95	0.016	99.992
1.00	24.20	68.27	2.50	1.75	98.76	4.00	0.013	99.9937
1.05	22.99	70.63	2.55	1.54	98.92	4.05	0.011	99.9949
1.10	21.79	72.87	2.60	1.36	99.07	4.10	0.009	99.9959
1.15	20.59	74.99	2.65	1.19	99.20	4.15	0.007	99.9967
1.20	19.42	76.99	2.70	1.04	99.31	4.20	0.006	99.9973
1.25	18.26	78.87	2.75	0.91	99.40	4.25	0.005	99.9979
1.30	17.14	80.64	2.80	0.79	99.49	4.30	0.004	99.9983
1.35	16.04	82.30	2.85	0.69	99.56	4.35	0.003	99.9986
1.40	14.97	83.85	2.90	0.60	99.63	4.40	0.002	99.9989
1.45	13.94	85.29	2.95	0.51	99.68	4.45	0.002	99.9991

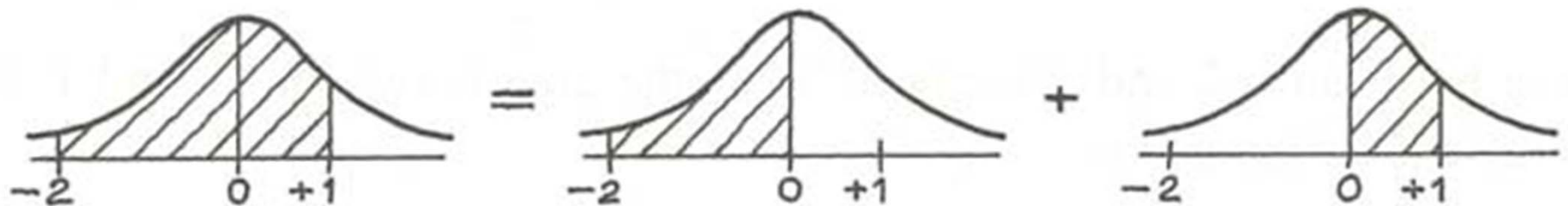
Finding Areas under Normal Curve

Find the area between 0 and 1 under the normal curve.

$$\frac{1}{2} \times 68\% = 34\%$$



- Find the area between 0 and 2 under the normal curve.
- Find the area between -2 and 1 under the normal curve.



- Find the area to the right of 1 under the normal curve.
- Find the area to the left of 2 under the normal curve.



- Find the area between 1 and 2 under the normal curve



Example 8. The heights of the men age 18 and over in HANES5 averaged 69 inches; the SD was 3 inches. Use the normal curve to estimate the percentage of these men with heights between 63 inches and 72 inches.

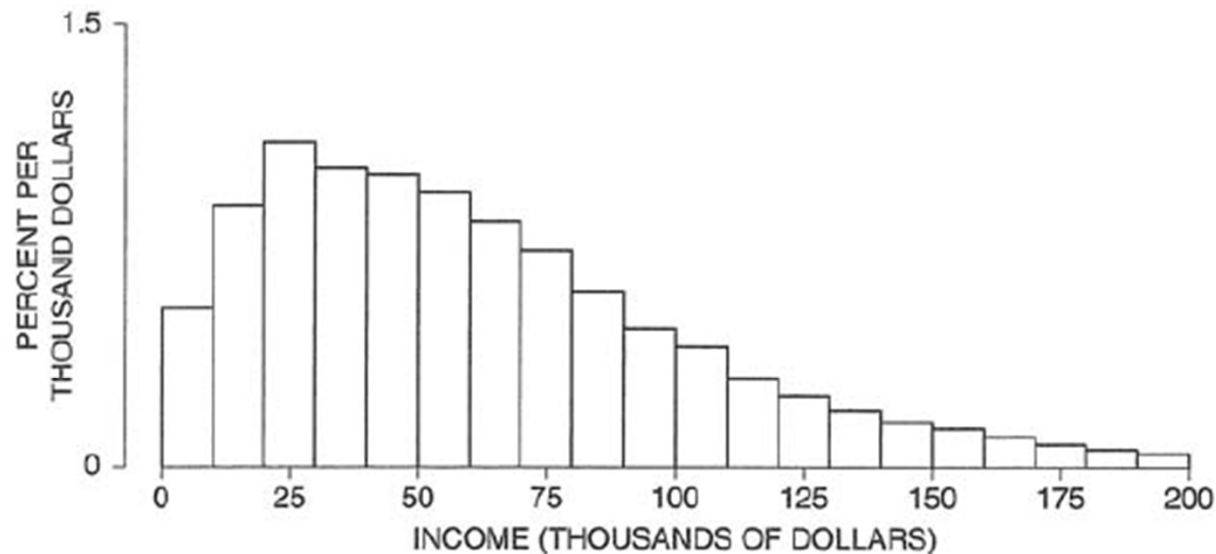
Key point:

Notice how the original data and its histogram can now be discarded.

This is the advantage of the normal approximation!

Normal approximation is not always accurate

Figure 5. Distribution of families by income: the U.S. in 2004.



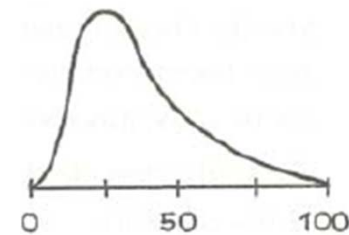
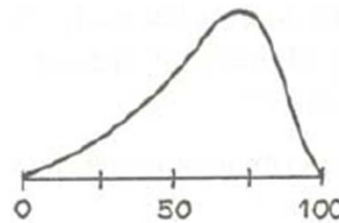
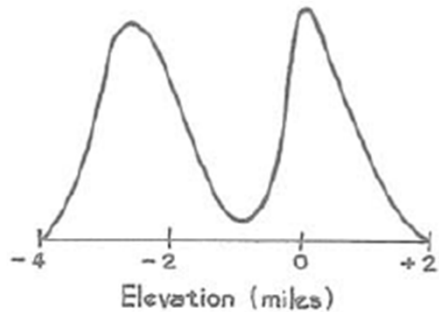
Source: March 2005 Current Population Survey; CD-ROM supplied by the Bureau of the Census.
Primary families.

The average income for the families in figure 5 was about \$60,000; the SD was about \$40,000.³ So the normal approximation suggests that about 7% of these families had negative incomes:

Verify this.

Percentiles

- Some histograms can not be accurately summarised by the normal curve.

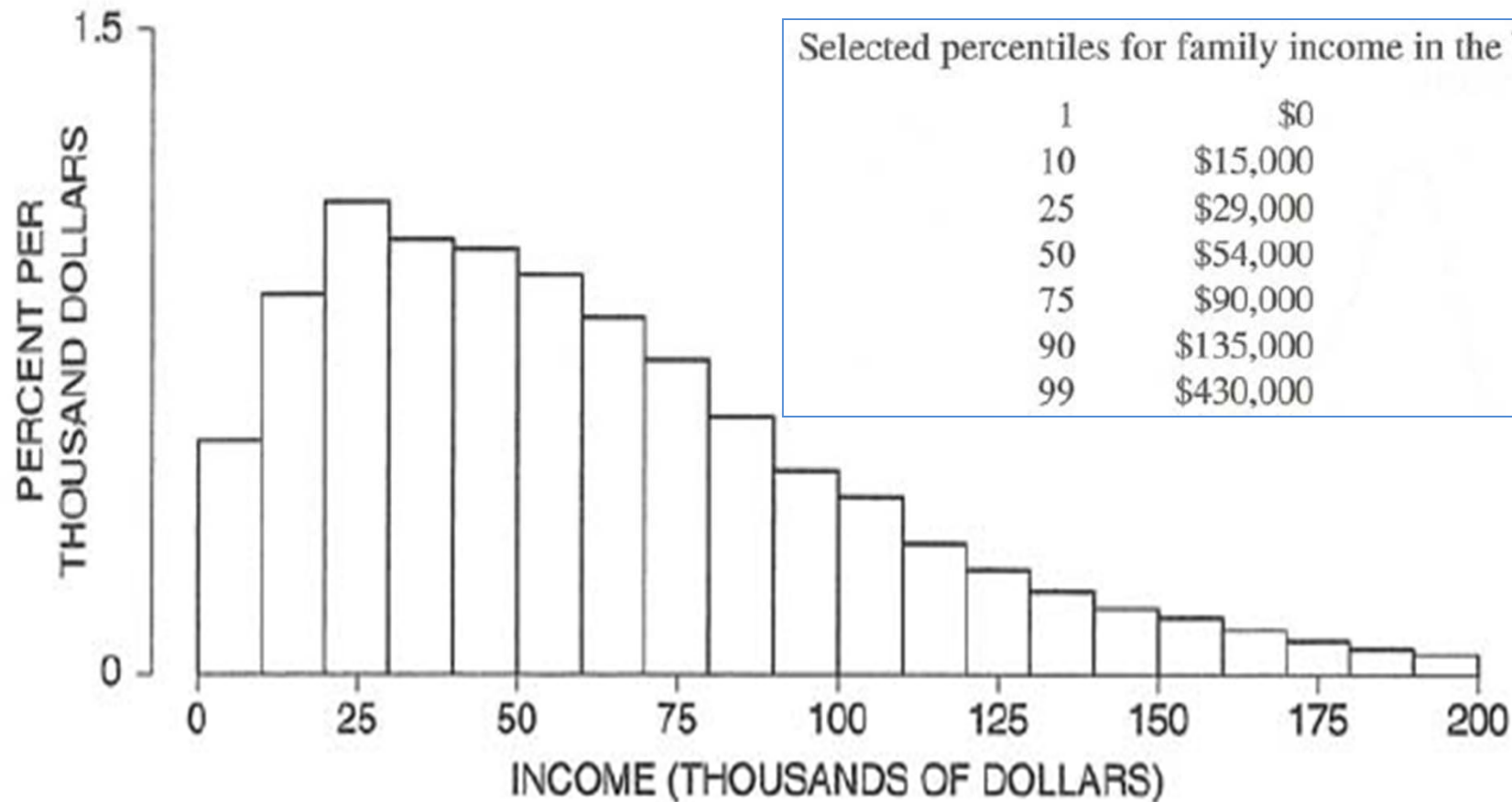


- Can they be summarised some other way?
 - Yes, using **percentiles**.
- If 10% of the data lies below x , then x is the 10th percentile.

Percentiles

- If 10% of the data lies below x , then x is the 10th percentile.
- If $k\%$ of the data lies below x , then x is the k^{th} percentile.
- A percentile is **not a percentage!**
- It is **a number that lies on the horizontal axis.**

Percentiles



- The 50th percentile is
- Is there a technical term for the 50th percentile?
 - Think about what the 50th percentile represents.

Percentiles

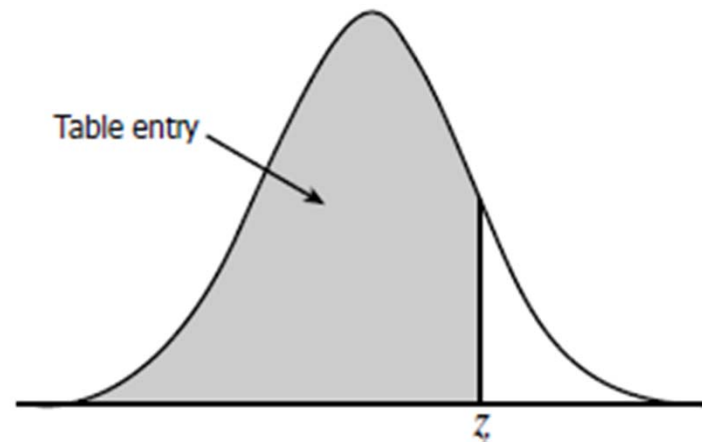
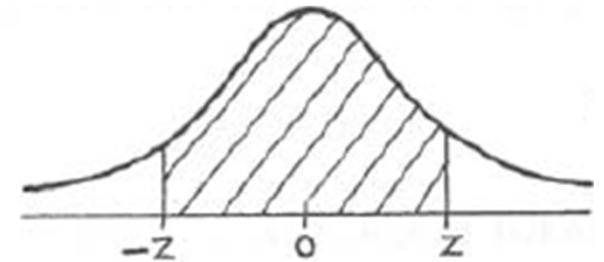
- How can we summarize spread of a histogram using percentiles?
 - Subtract the 25th percentile from the 75th percentile.
- This is called the **interquartile range**.
 - Is the interquartile range a percentage?

Percentiles

- Percentiles can summarize any histogram, whether it follows the normal curve or not.
- Question:
 - Assuming that a set of exam scores follows the normal curve with mean 63 and SD 10, what is the 95th percentile?

A note on reading Normal Tables

- The table in the book is not a standard table.
 - Gives area under the normal curve from $-z$ to z
- Standard tables give area from $-\infty$ to z



- So always note what a normal table represents before using area values from it.
- Is there a relationship between this z-score and percentiles?

Summary