## 8. Converting the percentiles under the normal curve back to original scores (Part 2)

I found some GMAT (graduate management admissions test) information on the internet:

–GMAT is scaled from 200 to 800 on a "bell shaped curve" (normal)
–Median GMAT is 500
–99th percentile is 750

Two questions:

(a) What's the standard deviation for the GMAT?
(b) You scored a 660, what's your percentile? (you need the answer for part a to complete part b)

(a) The 99th percentile corresponds to a Z=+2.33 or there are 2.33 standard (deviation) units between a Z=0 (score 500, the median or mean) and a Z=+2.33 (score 750)

Divide the 250 point difference by 2.33 and you get $\sigma$ = 107.296 points

Drawing a picture might help you to see how that works.

(b) To find your percentile, it's easy now

$$Z = \frac{660 - 500}{107.3} = +1.49$$

A Z=1.49 is around the 93rd percentile, congratulations, that score will get you consideration at the best MBA programs in the US.

## 9. When is a Z score Large?

In some situations a Z-score can be used to identify outliers (extreme observations). It depends.

greater than +1.65 or less than –1.65 is large to some people (identifies the top and bottom 5%)
greater than +1.96 or less than –1.96 (identifies the top and bottom 2.5% , commonly used in research)
greater than +2.33 or less than –2.33 (identifies the top and bottom 1%)
greater than +2.58 or less than –2.58 (national merit scholarship winners on the SAT)
greater than +3 or less than –3 (a 1600 on the SAT is a Z=+3.0 a 400 on the SAT is a Z=-3.0)
greater than +6 or less than –6 (basketball player Shaquille O'Neal has a Z=+6.0 or "six sigma")

The table in your textbook stops at +3.90 and –3.90, effectively saying once Z scores get past this size, you are looking at an extreme observation.

## 10. Assessing Normality

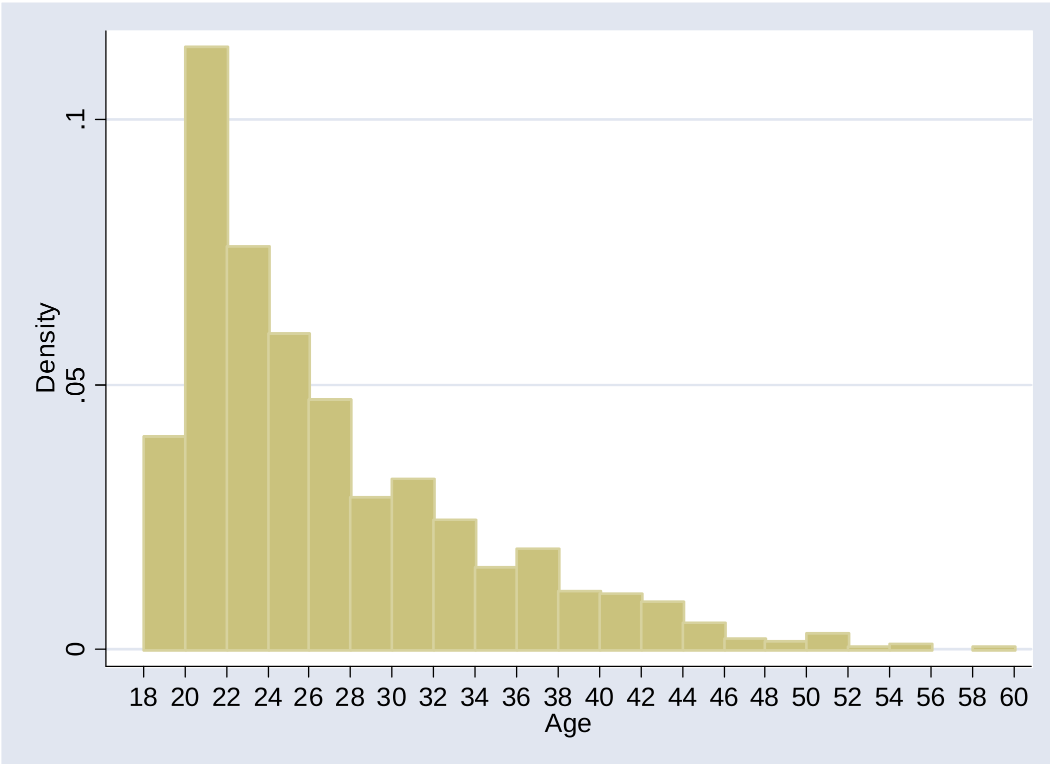
**A. Use some common sense:** if the normal curve implies nonsense results (for example, that people have negative incomes, or that some women have a negative number of children), the normal curve doesn't apply and using the normal curve will give the wrong answer.

Example: The average age at death for an American solider during Operation Iraqi Freedom (through September 7th, 2004) is approximately 26.4 years with a standard deviation of 7.1 years. Given this information, can you calculate the age of a soldier at the 5th percentile? How about the 90th percentile?

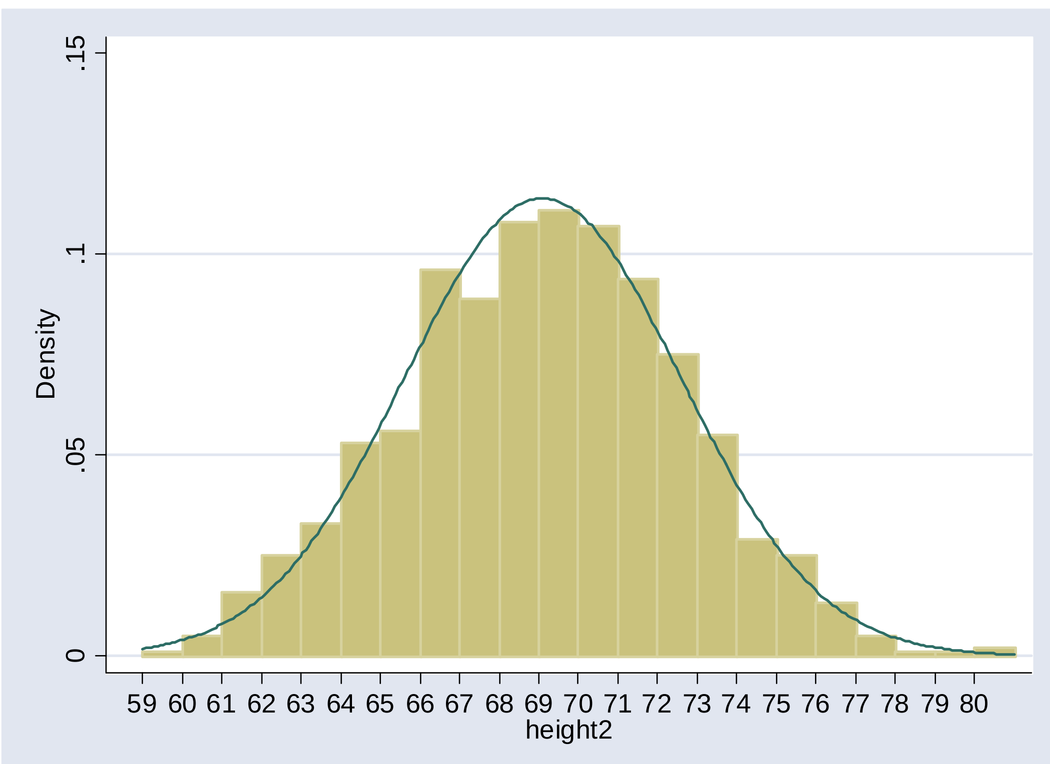
Can't do it, it's not normal. The 5th percentile has a Z= -1.65. That would translate into 11.7 years BELOW the average or 14.7 years.

**B. Construct a histogram: if the data look like a normal curve, the normal curve probably applies; otherwise, it does not.**

Here's the Operation Iraqi Freedom data, it's strongly (left/right? negatively/positively?) skewed:




Example: A sample of 1,000 American men. A histogram with a normal drawn on it:



It's not perfectly normal, but very close, certainly much closer than the Operation Iraqi Freedom ages. You can check some statistics if they are available:

height2

| | Percentiles | Smallest | | |
|---|---|---|---|---|
| 1% | 61.09424 | 59.50402 | | |
| 5% | 63.11546 | 60.32792 | | |
| 10% | 64.39524 | 60.73294 | Obs | 1000 |
| 25% | 66.69721 | 60.81042 | Sum of Wgt. | 1000 |
| 50% | 69.181 | | Mean | 69.09834 |
| | | Largest | Std. Dev. | 3.5 |
| 75% | 71.50625 | 78.18031 | | |
| 90% | 73.54923 | 79.31596 | Variance | 12.25 |
| 95% | 74.85286 | 80.08887 | Skewness | -.008467 |
| 99% | 76.89695 | 80.5591 | Kurtosis | 2.781324 |

The median and the mean are nearly equal. (Note: Skewness is nearly 0, with a Kurtosis (peakedness) of about 3, you don't need to know these, they are advanced topics, but they can help you determine normality)

Compare it with Iraqi Freedom

Age

| | Percentiles | Smallest | | |
|---|---|---|---|---|
| 1% | 18 | 18 | | |
| 5% | 19 | 18 | | |
| 10% | 20 | 18 | Obs | 999 |
| 25% | 21 | 18 | Sum of Wgt. | 999 |
| 50% | 24 | | Mean | 26.38238 |
| | | Largest | Std. Dev. | 7.066952 |
| 75% | 30 | 52 | | |
| 90% | 37 | 54 | Variance | 49.94181 |
| 95% | 41 | 55 | Skewness | 1.303998 |
| 99% | 50 | 59 | Kurtosis | 4.452734 |

**C. Do the data fall in a 68-95-99.7% pattern? If yes, normality is probably being met.**

We could a check on the men, I'll round their heights to half inches:

| heightround | Freq. | Percent | Cum. |
|---|---|---|---|
| 59.5 | 1 | 0.10 | 0.10 |
| 60.5 | 2 | 0.20 | 0.30 |
| 61 | 9 | 0.90 | 1.20 |
| 61.5 | 7 | 0.70 | 1.90 |
| 62 | 7 | 0.70 | 2.60 |
| 62.5 | 9 | 0.90 | 3.50 |
| 63 | 19 | 1.90 | 5.40 |
| 63.5 | 17 | 1.70 | 7.10 |
| 64 | 18 | 1.80 | 8.90 |
| 64.5 | 24 | 2.40 | 11.30 |
| 65 | 34 | 3.40 | 14.70 |
| 65.5 | 29 | 2.90 | 17.60 |
| 66 | 31 | 3.10 | 20.70 |
| 66.5 | 47 | 4.70 | 25.40 |
| 67 | 52 | 5.20 | 30.60 |
| 67.5 | 42 | 4.20 | 34.80 |
| 68 | 52 | 5.20 | 40.00 |
| 68.5 | 50 | 5.00 | 45.00 |
| 69 | 53 | 5.30 | 50.30 |
| 69.5 | 59 | 5.90 | 56.20 |
| 70 | 64 | 6.40 | 62.60 |
| 70.5 | 47 | 4.70 | 67.30 |

| | | | |
|---|---|---|---|
| 71 | 58 | 5.80 | 73.10 |
| 71.5 | 41 | 4.10 | 77.20 |
| 72 | 42 | 4.20 | 81.40 |
| 72.5 | 39 | 3.90 | 85.30 |
| 73 | 27 | 2.70 | 88.00 |
| 73.5 | 32 | 3.20 | 91.20 |
| 74 | 18 | 1.80 | 93.00 |
| 74.5 | 16 | 1.60 | 94.60 |
| 75 | 16 | 1.60 | 96.20 |
| 75.5 | 12 | 1.20 | 97.40 |
| 76 | 6 | 0.60 | 98.00 |
| 76.5 | 9 | 0.90 | 98.90 |
| 77 | 2 | 0.20 | 99.10 |
| 77.5 | 4 | 0.40 | 99.50 |
| 78 | 2 | 0.20 | 99.70 |
| 79.5 | 1 | 0.10 | 99.80 |
| 80 | 1 | 0.10 | 99.90 |
| 80.5 | 1 | 0.10 | 100.00 |
| Total | 1,000 | 100.00 | |

**D. NEVER ASSUME THAT A VARIABLE IS NORMAL.**

Recall this example

Example: The average age at death for an American solider during Operation Iraqi Freedom (through September 7th, 2004) is approximately 26.4 years with a standard deviation of 7.1 years. Given this information, can you calculate the age of a soldier at the 5th percentile? How about the 90th percentile?

Just because you have a mean and standard deviation does not allow you to automatically assume a variable is normal.

Either you will be told it’s normal or you will be given enough information (like a histogram) to safely assume it’s normal.