Homework 7, Due March 18 2010

Chapter 6, questions 6.6, 6.10, 6.26, 6.28

Remark: 6.28 is a follow-on to 6.27. You are not requested to hand in 6.27 as well, but probably you should work through 6.27 for yourself, before tackling 6.28.
Here is another example:

For mean, the mean height is 70 inches and the standard deviation is 4 inches. *What proportion of men are between 64 and 72 inches?*
Solution:

For $x = 72$, we calculate $z = \frac{72-70}{4} = 0.5$. But corresponding to 0.5, the probability is .6915, i.e. 69.15% of men are of height 72 inches or less.

For $x = 64$, we calculate $z = \frac{64-70}{4} = -1.5$. This corresponds to a probability of .0668, i.e. 6.68% of men are of height 64 inches or less.

The difference, .6915 – .0668 = .6247 or about 62.5%, represents the proportion of men whose height lie between 64 and 72 inches.
Another type of question is *given the probability, find the value that this corresponds to*.

For example: In the SAT the mean is 500 and the standard deviation is 100. *What score do you have to get to be in the top 15% of all students?*
In this case, we need to know the $z$-score corresponding to an “area under the curve” of 0.85 (0.85 being the left-tail probability, i.e. 1–0.15 where 0.15 is the right-tail probability).

By looking up the table, the answer is 1.04 (to two decimal places, e.g. for $z = 1.03, 1.04, 1.05$ we get probabilities 0.8485, 0.8508, 0.8531 — .8508 is the closest to exactly 0.85 and this corresponds to a $z = 1.04$).

Thus

$$z = \frac{x - \mu}{\sigma} = \frac{x - 500}{100} = 1.04.$$ 

Translates to an actual score of

$$x = 500 + 100 \times 1.04 = 604.$$
Using $z$-scores to compare distributions on different scales

An example: Comparing performances of athletes in different events

In the 2005 World Track and Field Championships in Helsinki, the women’s javelin event was won in 71.70 m. (world record) while the women’s high jump event was won in 2.02 m. Which was the better performance?
One way to judge this is to convert both performances to a *standardized score* by the formula

\[ z = \frac{x - \bar{x}}{s} \]

where

- \( x \) is the actual performance of a given athlete in an event
- \( \bar{x} \) is the mean over all athletes in that event
- \( s \) is the SD over all athletes in that event

For multi-event competitions such as the men’s decathlon or the women’s heptathlon, we might also consider adding the \( z \)-scores over different events, to create a combined score for each athlete.
A simplified but realistic example:

Three athlete compete in the 100 m dash, the shot put and the long jump. Shown are their results, and also the means and standard deviations over many competitions. Who should get the gold medal, and which one performance stands out as the best in the competition?

<table>
<thead>
<tr>
<th>Competitor</th>
<th>100 m</th>
<th>Shot Put</th>
<th>Long Jump</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10.1 sec.</td>
<td>66’</td>
<td>26’</td>
</tr>
<tr>
<td>B</td>
<td>9.9 sec.</td>
<td>60’</td>
<td>27’</td>
</tr>
<tr>
<td>C</td>
<td>10.3 sec.</td>
<td>63’</td>
<td>27’ 3”</td>
</tr>
<tr>
<td>Mean</td>
<td>10.0 sec.</td>
<td>60’</td>
<td>26’</td>
</tr>
<tr>
<td>SD</td>
<td>0.2 sec.</td>
<td>3’</td>
<td>6”</td>
</tr>
</tbody>
</table>
Compute the $z$ scores —

<table>
<thead>
<tr>
<th>Competitor</th>
<th>100 m</th>
<th>Shot Put</th>
<th>Long Jump</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-0.5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0.5</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>-1.5</td>
<td>1</td>
<td>2.5</td>
</tr>
</tbody>
</table>

So C’s performance in the long jump is the best of the whole competition.

But B’s total score (2.5) deserves the gold medal ahead of A (1.5) and C (2.0).
A more complicated example of normal probability calculations

Part 1. A manufacturer of washing powder sells its powder in 32 ounce cartons. However the machine that fills the cartons is not precisely accurate — assume that the amount actually delivered to the carton has a mean that can be set by the manufacturer, but the standard deviation is 0.3 ounces. Assume the curve is normal.

If the machine was set to deliver (a mean of) exactly 32 ounces, half the cartons would be underweight. To avoid this, the company sets the mean to be 32.5 ounces. In that case, what proportion of cartons sold are underweight?
Solution. The $z$-score is
\[ z = \frac{32 - 32.5}{0.3} = -1.6667 \]
From the normal tables, the area under the curve, to the left of $-1.6667$, is about 0.048.

Therefore, under this regime, about 4.8% of all cartons sold are underweight.
Part 2. The company lawyers advise that 4.8% underweight cartons is too many, and advise it should be no more than 2%. To achieve this, what should the setting of the machine be?

Solution. Using the table, the $z$ value associated with a 2% area under the curve is $z = -2.05$.

Therefore, we have to solve for $\mu$ in the equation

$$z = \frac{32 - \mu}{\sigma} = \frac{32 - \mu}{0.3} = -2.05.$$  

Hence

$$\mu = 32 + 0.3 \times 2.05 = 32.615.$$  

The machine should be set to 32.62 ounces.
*Part 3.* The company president says this is giving away too much powder and instructs that the machine setting be no higher than 32.3 ounces. To achieve the 2% rate of underweight boxes, the standard deviation $\sigma$ will have to be reduced. What value of $\sigma$ is needed to comply with both the company president’s and the lawyers’ requirement?

*Answer.* We still have $z = -2.05$, but now we fix $\mu = 32.3$. Solve

$$z = \frac{32 - 32.3}{\sigma} = -2.05.$$  

We find $\sigma = \frac{0.3}{2.05} = 0.1463$. 