Statistics Notes 2

Before we go into normal distributions here are two important topics to be kept in mind. If A and B are mutually exclusive events then $P(A\ or\ B) = P(A) + P(B)$. (This rule can be extended to any number of mutually exclusive events.) If on the other hand A and B are not mutually exclusive events then the general rule is $P(A\ or\ B) = P(A) + P(B) - P(A\,\&\,B)$. (One way of telling that A and B are not mutually exclusive is that we show that $P(A\,\&\,B) \neq 0$ or equivalently P(A or B)<P(A)+P(B)). Now here is a question.

13. If P(A) = 0.9, P(B) =.8 and P(A *or* B) = 0.95 find P(A&B). Here P(A)+P(B) = 1.7 and because the probability P(E) or any event E must be such that $0 \le P(E) \le 1$. So $P(A\ or\ B) < P(A) + P(B)$. This means that the general rule applies. So $P(A\ or\ B) = P(A) + P(B) - P(A\,\&\,B)$. This gives $.95 = .9 + .8 - P(A\,\&\,B)$. Now you can solve this equation for P(A&B).

Suppose that X is a variable with mean m and standard deviation s and let x be an observed value of X. Then the z-score given by $z = \frac{x-m}{s}$ gives the number of standard deviations that *x* is away from m (this is because if $z = \frac{x-m}{s}$ then $x = m + zs$). The z-scores standardize the distribution of X in that the mean of the z-scores is zero and the the standard deviation of the z-scores is 1. (The z-scores of another variable y do have the same nice properties and this is often used for a kind of comparison.). Here is a useful example to indicate what the comparison really means:

14. Suppose that two graduating seniors, one a marketing major, and the other an accounting major are comparing job offers. The marketing major has a job offer for \$23,000 per year and the accounting major has a job offer for \$25,000. The means and standard deviations of the job offers in the two areas are as follows:

Marketing: mean = 22,500, standard deviation = 1000

Accounting: mean = 26000, standard deviation = 1500.

Looking at the z-scores of the offers.

For marketing: z = $\frac{23000-22500}{1000} = .5$.

For accounting: z = $\frac{25000-26000}{1500} = -.667$.

So the marketing major seems to be doing much better (with reference to his peers!) as compared to the accounting major who is getting below average pay in his field. So the comparison is not between the pays it is between their positions in their respective classes.

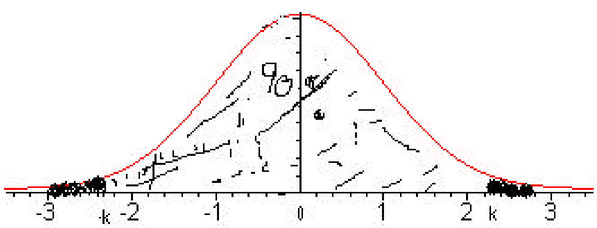
Now let X be a normally distributed random variable with mean m and standard deviation s. Then the z-scores of the values *x (*of X) given by $z = \frac{x-m}{s}$ would also be normally distributed i.e. their distribution curve would be a bell shaped curve that is centered at the mean 0, with standard deviation 1 and symmetric about the x-axis. Thus thanks to the z-scores every normal distribution can be represented by the standard

normal curve. The area under a normal curve (and hence under a standard normal curve) is 1. This indeed is the source of several interesting applications.
15. Use the standard normal table to find the area to the left of z = 1.32. (You already know how to find the area to the left, so I will let you do that. What you must remember is that this area is also the probability that a randomly picked observation has z- score less than (or equal to) 1.32.
16. Using the standard normal table find the following probabilities and sketch the associated areas:
(i). $P(0 \le z \le .79)$ (i.e. find the area under the normal curve between $z = 0$ and z=.79)
(ii). $P(-1.57 \le z \le 2.33)$
(iii). $P(z \ge 1.89)$ (Area to the right of z = 1.89, under the normal curve.)
17. The random variable X has a normal distribution with a mean of 200 and a standard deviation of 25.

(i) Find the probability that X is between 160 and 220. (In this case you would find the z-scores $z(150) = \frac{150-200}{25} = -2$ and $z(220) = \frac{220-200}{25} = .8$ So the problem of finding $P(160 \le X \le 220)$ has been transformed into $P(-2 \le z \le .8)$ which you can easily solve using the standard normal tables.

(ii) Find the probability that X greater than 240. (Here too, take the z- scores and make the problem into $P(z \ge \frac{240-200}{25})$

18. Find the z- score k so that $P(-k \le z \le k) = .9$. (A picture is in order here.



Looking at the picture we see that the area between –k and k is .9 so the area beyond k and –k (the one with paint smudges) is 0.1. But as the normal curve is symmetric about the y-axis the area to the right of k is equal to the area to the left of –k. This mean that the area to the right of k is = $\frac{.1}{2} = .05$. So $k = z_{.05}$ which you know how to calculate.)
So, if you are looking for a z- score *k* such that $1-a$ of the area is between –*k* and *k* then you are looking for $k = z_{\frac{a}{2}}$.
This leads to questions about the confidence intervals, which given time will come later.