

## Probability - Chapter 5

---

given that among 12-14 year old boys, carbohydrate intake is normally distributed, with a mean of 124 and a standard deviation of 20 ...

5.6 What percentage of boys in this age range have carbohydrate intake above 140 g / 1000 cal ?

first standardize the value 140, calculate a z-score determining how many standard deviations 140 is above the mean of 124 ... in terms used in equation 5.4 ...

convert...  $X \sim N(\mu, \sigma^2)$  ...  $X \sim N(124, 400)$  to ...  $Z \sim N(0, 1)$

subtract the population mean from 140 and divide by the standard deviation ...

$$(140 - 124) / 20 = 16 / 20 = 0.8$$

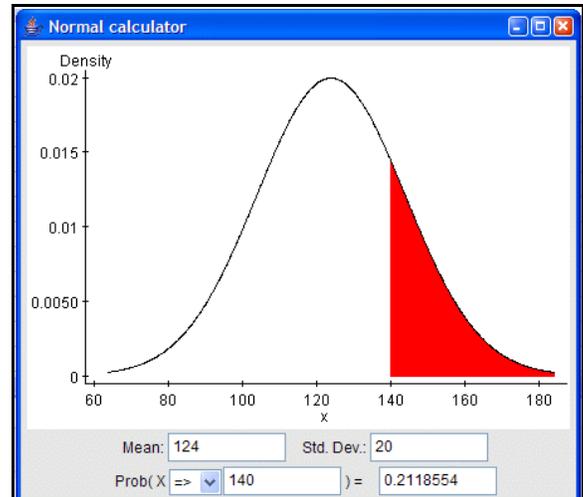
you can now find the value in column B in Table 3 of the appendix that corresponds to  $x = 0.8$  (in the leftmost column labeled  $x$  ... the values in column B correspond to the area under the normal distribution curve greater than 0.8 ...

$$1 - \Phi(0.8) = .2119.$$

21.2% of boys have carbohydrate intake above 140 g / 1000 cal

you can also use Statcrunch with the true mean and standard deviation

---



5.7 What percentage of boys in this age range have carbohydrate intake below 90g / 1000 cal ?

this problem is similar to 5.6 in that the first thing you must do is standardize the value 90. How many standard deviations below the mean of 124 is 97 ...

$$(90 - 124) / 20 = -34 / 20 = -1.7$$

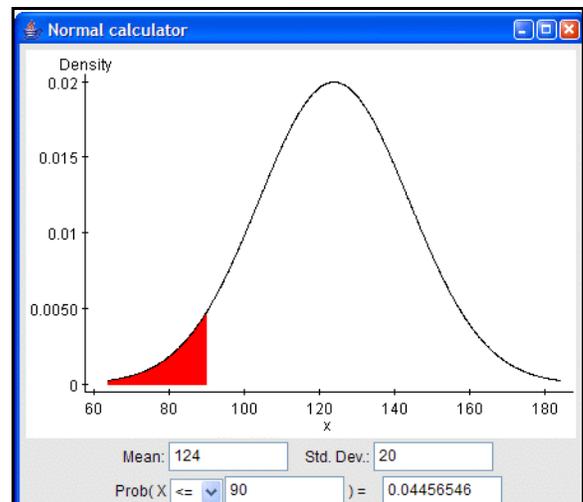
you can now find the value in column B in Table 3 of the appendix that corresponds to  $x = -1.7$  (in the leftmost column labeled  $x$ ) ... the values in column B correspond to the area under the normal distribution curve equal to or greater than -1.7 ... however, you will notice that the numbers in the column labeled  $x$  are all positive, but you also know that the normal curve is symmetric ... finding the area under the normal curve greater than 1.7 is the same as finding the same area less than -1.7 ...

$$1 - \Phi(1.7) = .0446$$

4.46% of boys will have carbohydrate intake below 90 g / 1000 cal

you can also use Statcrunch with the true mean and standard deviation

---



you are now told that you can assume that carbohydrate intake is normally distributed, with a mean of 121 and a standard deviation of 19 ...

5.8, 5.9

the solutions are the same as for problems 5.6 and 5.7 ... first standardize the given values, then use column B in Table 3 in the appendix ...

$$(140 - 121) / 19 = 19 / 19 = 1.0$$

$$1 - \Phi(1) = .1587$$

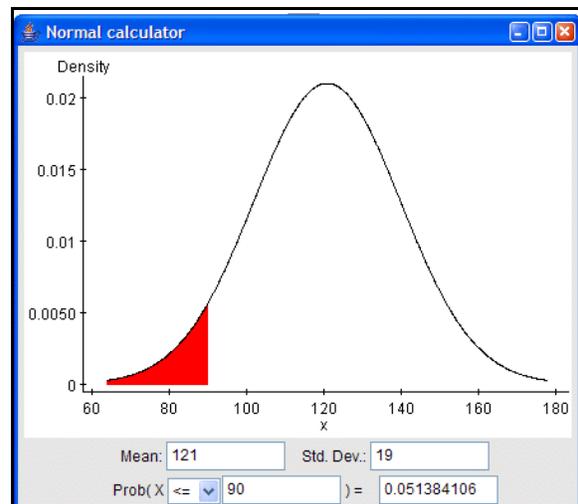
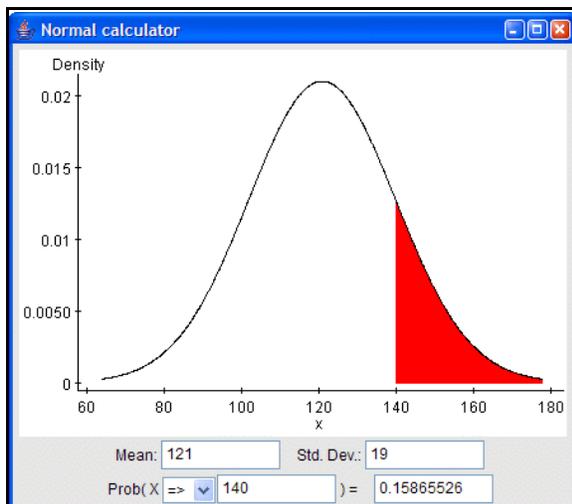
15.9% of these boys will have carbohydrate intake above 140 g / 1000 cal

$$(90 - 121) / 19 = -31 / 19 = -1.63$$

$$1 - \Phi(1.63) = .0516$$

5.2% of these boys will have carbohydrate intake below 90 g / 1000 cal

you can also use Statcrunch with the true mean and standard deviation for both problems



given that typical blood glucose levels are normally distributed with a mean of 90 mg/dl and a standard deviation of 38 mg/dl ...

5.41 if the normal range is 65-120 mg/dl, what percentage of values fall in the normal range?

... you can answer the question by standardizing the both values that define the normal range, computing z-scores and determining how many standard deviations both 65 and 120 deviate from the mean ... of 90 ... in terms used in equation 5.5 ...

$$(65 - 90) / 38 = -25 / 38 = -0.66$$

$$(120 - 90) / 38 = 30 / 38 = 0.79$$

since these values span the mean (with a z-score of 0), you can use column C in Table 3 of the appendix ...

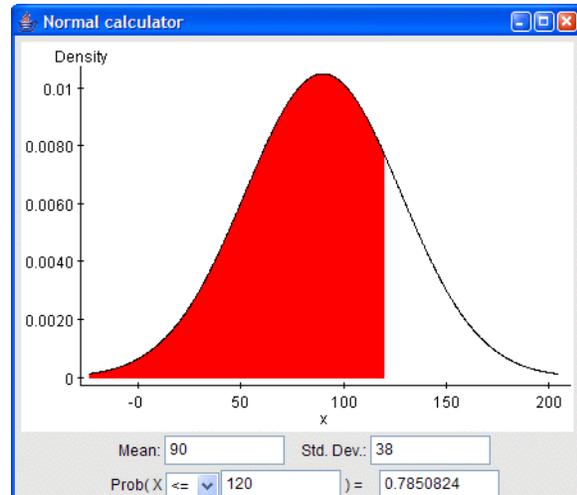
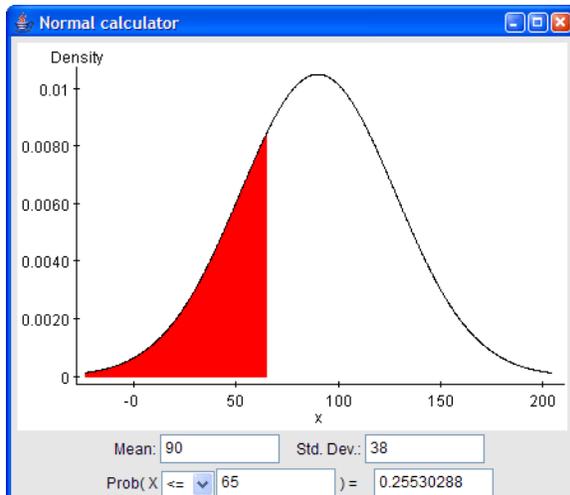
the area between z-scores of 0 and -0.66 is 0.2454

the area between z-scores of 0 and 0.79 is 0.2852

those two areas add up to 0.5306

53.1% of values fall within the normal range.

you could also use Statcrunch and the true mean and standard deviation, but you first compute  $P(X \leq 65)$  and subtract it from  $P(X \leq 120)$  ... or  $0.7851 - 0.2553 = 0.5298$  (differences due to rounding in the calculation of the z-scores in the "use a table in Rosner" method)



5.42 if abnormal values are those 1.5 times as high as the upper limit (of normal) ... what percentage of values fall within this range ...

abnormal is  $1.5 \times 120 = 180$  mg/dl ... you can answer the question by first standardizing the value 180, computing a z-score and determining how many standard deviations 180 is above the mean of 90 ...

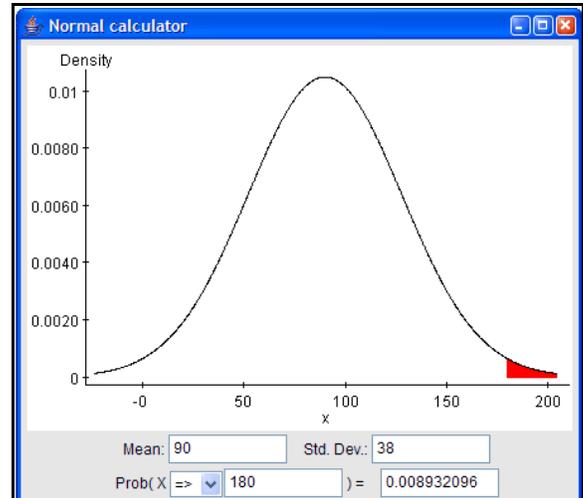
$$(180 - 90) / 38 = 90 / 38 = 2.37$$

... you can now find the value in column B in Table 3 of the appendix that corresponds to  $x = 2.37$  (in the leftmost column labeled  $x$ ) ... the values in column B correspond to the area under the normal distribution curve greater than 2.37 ...

$$1 - \Phi(2.37) = .0089$$

0.89% of values are abnormal

you can also use Statcrunch with the true mean and standard deviation



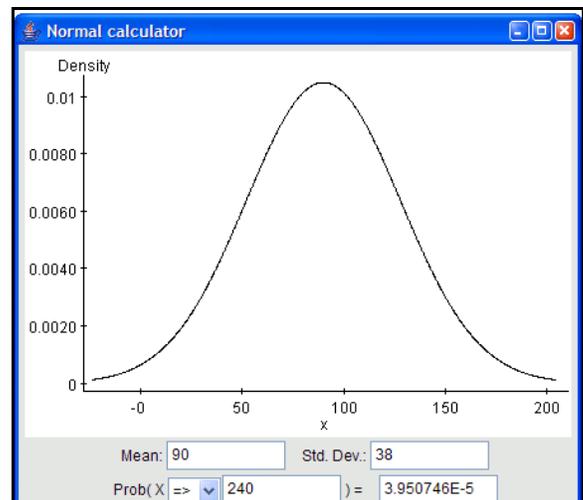
5.43 changing abnormal to  $2.0 \times 120 = 240$  mg/dl ... using the same methods as problem 5.43 ...

$$(240 - 90) / 38 = 150 / 38 = 3.95$$

$$1 - \Phi(3.95) = <.0001$$

less than 0.01% of values are abnormal

you can also use Statcrunch with the true mean and standard deviation



5.44 given that abnormal is defined as 1.5 times the upper limit of 120 (i.e. 180), what is the probability that two consecutive tests on an individual will be abnormal?

... in problem 5.42, the probability of a reading > 180 was found to be 0.0089 ... since the results of two consecutive tests are independent, you can use the multiplication rule ...

$$P(2 \text{ consecutive} > 180) = 0.0089 \times 0.0089 = 0.00008$$

5.45 given a study of 6,000 patients, what is the probability that 75 patients have a blood glucose level at least 1.5 times the upper limit of normal (180)?

... you can consider this as a binomial probability problem ... if  $N=6,000$  patients and  $P=0.0089$  (from problem 5.43), what is the probability of 75 patients being found as abnormal?

... looking at equation 5.15 ...

$$NPQ = 6000 \times .0089 \times (1-.0089) = 52.94$$

and the normal distribution can be used to approximate the binomial in this problem since  $NPQ > 5$ .

... calculate the mean and standard deviation...

$$\begin{aligned} \text{mean} &= NP = 6000 \times .0089 = 53.4 \\ \text{variance} &= NPQ = 6000 \times .0089 \times (1 - .0089) = 52.94 \\ \text{standard deviation} &= 7.28 \end{aligned}$$

you can answer the question by first 'standardizing' the value 75, i.e. determining how many standard deviations 75 is above the mean of 53.4 ... when using the normal approximation to the binomial, standardizing requires a continuity correction of either + or - 0.5 (equation 5.14, example 5.35) ... instead of 75, 74.5 is used ...

$$(74.5 - 53.4) / 7.28 = 21.1 / 7.28 = 2.90$$

find the value in column B in Table 3 of the appendix that corresponds to  $x = 2.90$  ... the values in column B correspond to the area under the normal distribution curve greater than 2.90 ...

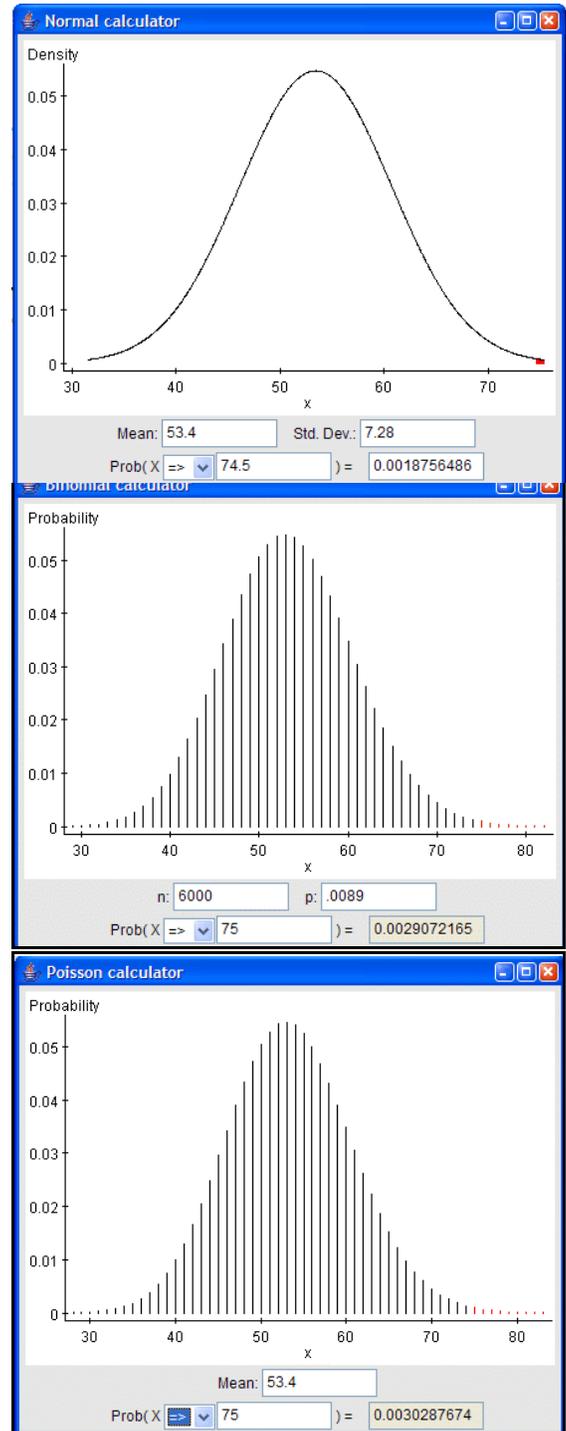
$$1 - \Phi(2.90) = 0.0019$$

75 abnormal patients is a rare event

three figures from Statcrunch are shown on the right ...

top uses the normal approximation to the binomial, but you need not compute a z-score, use the real data ... middle uses the binomial,  $N=6000$ ,  $P=0.0089$ , what is  $P(X \geq 75)$  ... bottom uses the Poisson since  $N$  is large and  $P$  is very small, the estimate of the mean is  $NP$ , what is  $P(X \geq 75)$

P using normal approximation: 0.0019  
 P using exact binomial: 0.0029  
 P using Poisson approximation the binomial: 0.0030



5.72 what distribution can be used to model cancers among cystic fibrosis patients?

this situation can be defined as a 'rare' event, so the Poisson distribution can be used

5.73 If 45.6 cancers are expected, is the occurrence of 37 cancers an unusually low number?

... use the normal approximation to the Poisson (as in example 5.36) .... the mean and variance are 45.6 (and standard deviation is 6.75) ...

... you can answer the question by first standardizing the value 37, computing a z-score and determining how many standard deviations 37 is above the mean of 45.6 ... however, using the normal approximation to the Poisson requires a continuity correction of either + or - 0.5 (equation 5.16, example 5.36) ... instead of 37, 37.5 is used ...

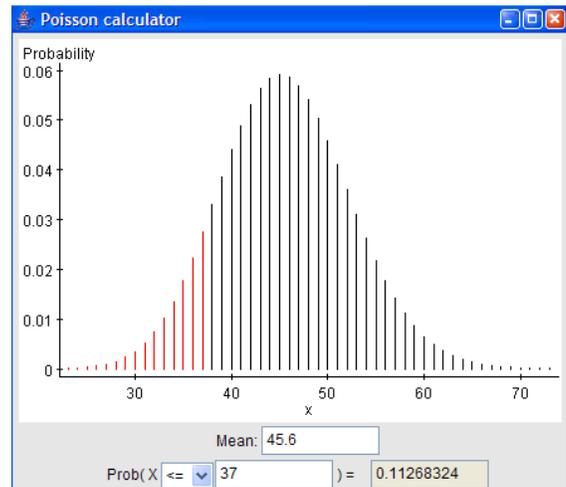
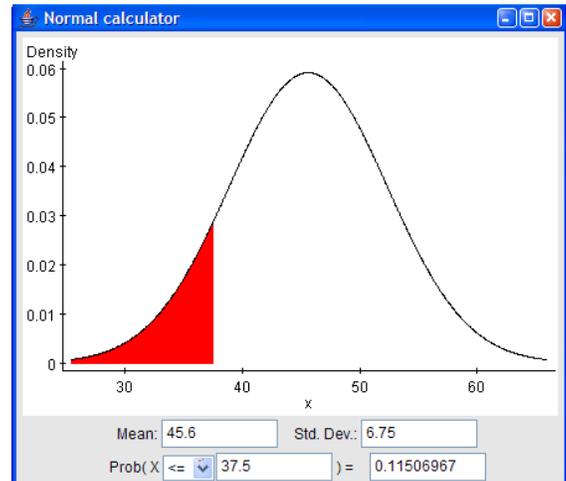
$$(37.5 - 45.6) / 6.75 = -8.1 / 6.75 = -1.2$$

... you can now find the value in column B in Table 3 of the appendix that corresponds to  $x = 2.90$  ... the value in column B correspond to the area under the normal distribution curve greater than 2.90 ...

$$1 - \Phi(1.2) = .1151$$

finding 37 cancers is not a 'rare' event

you can also use Statcrunch ... the figure on thetop shows the normal approximation method using the original data, not z-scores ... the figure on the bottom shows the result using the Poisson distribution, not the approximation



5.74 is the occurrence of 13 digestive tract cancers among cystic fibrosis patients when only 2 were expected a 'rare' event?

... you can use Table 2 in the appendix to answer this question ... with  $\mu = 2$ , what is  $P(X \geq 13)$  ... in Table 2, all probabilities with  $k \geq 13$  are 0 ... 13 or more cases a 'rare' event ...

you can also use Statcrunch

