2.5 Some Discrete Distributions

In this section, we will look at the distributions and related expected values and variances of two special discrete random variables called the binomial and Poisson.

Exercise 2.7 (The Binomial Function.) The binomial distribution function, often described by $B(n, \pi)$), is given by,

$$f(y) = \frac{n!}{y!(n-y)!} \times \pi^y \times (1-\pi)^{n-y}, \quad y = 0, 1, \ldots, n, 0 \leq \pi \leq 1.$$

where $\mu = n\pi$ and $\sigma^2 = n\pi(1-\pi)$.

See TI–83 Lab 2: binomial probability distribution, binomial cumulative distribution

1. *Factorial Notation.* For example, $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$.

   (a) $4! = 4 \times 3 \times 2 \times 1 =$ (circle one) **12** / **24** / **36**. In general, for any positive integer n, $n!$ is read "n factorial" and is defined by

   $$n! = n(n-1)\cdots 3(2)(1).$$

   Also, $0! = 1$.

   (b) $\frac{4!}{2!} =$ (circle none, one or more)

   i. $\frac{4\times3\times2\times1}{2\times1}$
   ii. $\frac{24}{2}$
   iii. 4×3
   iv. 12

   (c) $\frac{4!}{2!3!} =$ (circle none, one or more)

   i. $\frac{4\times3\times2\times1}{(2\times1)(3\times2\times1)}$
   ii. $\frac{4}{2\times1}$
   iii. $\frac{4\times3}{3\times2\times1}$
   iv. 2

2. *Binomial formula: Seed Germination.* A seed germinates 40% ($\pi = 0.4$) of the time. Assume each seed germination is independent of one another and, in general, this problem obeys the conditions of a binomial experiment. Ten ($n = 10$) are planted.

   (a) All ten seeds may germinate ($y = 10$). However, only nine seeds might germinate ($y = 9$). In fact, all of the different number of seeds, out of ten, that could germinate is (circle one)

i. $y = 1, 2, \ldots 10$
ii. $y = 0, 1, 2, \ldots 10$

(b) Since the chance a seed germinates is 40% and each seed germination is independent of one another, the chance all ten seeds germinate is

$\overbrace{0.40 \times 0.40 \times \cdots \times 0.40}^{10} =$ (circle one) $\mathbf{10^{-4}}$ / $\mathbf{10^{-5}}$ / $\mathbf{10^{-6}}$.

(c) Since the chance a seed germinates is 40%, (and so the chance it does not germinate must be 60%) and each seed germination is independent of one another, the chance all ten seeds do *not* germinate is

$\overbrace{0.60 \times 0.60 \times \cdots \times 0.60}^{10} =$ (circle one) **0.004** / **0.005** / **0.006**.

(d) The chance *one* of the ten seeds germinates, assuming the first seed germinates (and the next nine do not germinate), is

$0.40 \times \overbrace{0.60 \times \cdots \times 0.60}^{9} =$ (circle one) **0.004** / **0.005** / **0.006**.

(e) The chance *one* of the ten seeds germinates, assuming the *second* seed germinates (and the first, as well as the third to tenth seeds do not germinate) is

$0.60 \times 0.40 \times \overbrace{0.60 \times \cdots \times 0.60}^{8} =$ (circle one) **0.004** / **0.005** / **0.006**.

(f) The chance *one* of the ten seeds germinates, adding all the possible ways that this can be done, is

$10 \times 0.60 \times 0.40 \times \overbrace{0.60 \times \cdots \times 0.60}^{8} =$ (circle one) **0.040** / **0.050** / **0.060**.

(g) The chance *one* of the ten seeds germinates, 0.040, can be calculated using the formula (circle none, one or more)

i. $f(1) = 10 \times 0.4 \times 0.6^9$.
ii. $f(1) = \frac{10!}{1!9!} \times 0.4^1 \times 0.6^9$.
iii. $f(1) = \frac{10!}{1!(10-1)!} \times 0.4^1 \times 0.6^9$.
iv. $f(1) = \frac{n!}{y!(n-y)!} \times \pi^y \times (1-\pi)^{n-y}$ where $n = 10$, $y = 1$ and $\pi = 0.4$ (the binomial formula).

(h) The chance *two* of the ten seeds germinates, (circle none, one or more)

i. $f(2) = \frac{n!}{y!(n-y)!} \times \pi^y \times (1-\pi)^{n-y}$ where $n = 10$, $y = 2$ and $\pi = 0.4$.
ii. $f(2) = \frac{10!}{2!(10-2)!} \times 0.4^2 \times 0.6^8$.
iii. $f(2) = \frac{10!}{2!8!} \times 0.4^2 \times 0.6^8$.
iv. $f(2) = 45 \times 0.16 \times 0.017 \approx 0.121$.

(Use your calculator: 2nd DISTR 0:binompdf(10, 0.4, 2) ENTER.)

(i) **True** / **False** The chance of *seven* of the ten seeds germinating is $\frac{n!}{y!(n-y)!} \times \pi^y \times (1-\pi)^{n-y}$ where $n = 10$, $y = 7$ and $\pi = 0.4$.

(j) The tabular form of this distribution is given by,

y	0	1	2	3	4	5	6	7	8	9	10
chance	0.006	0.040	0.121	0.215	0.251	0.201	0.111	0.043	0.011	0.002	0.000

where, for example, the chance of 6 seeds germinating out of 10 is 0.111. In a similar way, the chance of 4 of 10 germinating is
(circle one) **0.121** / **0.215** / **0.251**.
(Use your calculator to construct the entire distribution table: STAT EDIT ENTER; type 0,1, ...,10 into L_1; define L_2 = 2nd DISTR 0:binompdf(10, 0.4, L_1) ENTER.)

(k) The graphical form of the probability distribution is given by,

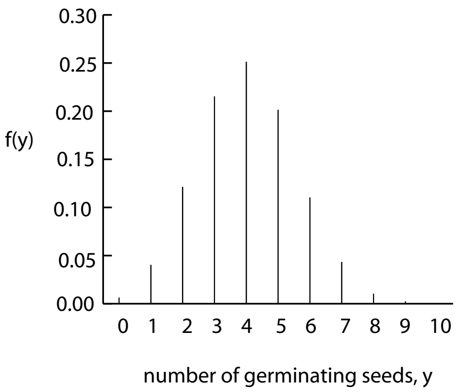

Figure 2.10 (Example Binomial Distribution)

The most likely number of seeds that will germinate out of ten is
(circle one) **one** / **two** / **three** / **four**.
(Use your calculator to draw the distribution: STAT EDIT ENTER; type 0,1, ...,10 into L_1; define L_2 = 2nd DISTR 0:binompdf(10, 0.4, L_1) ENTER; set up window: WINDOW 0 10 1 -0.1 0.3 0.1 1; pick histogram: 2nd STAT PLOT ENTER pick histogram, make Xlist L_1 and Freq L_2; then GRAPH TRACE.)

(l) The binomial distribution, in this case, is
(circle one) **skewed left** / **skewed right** / **more or less symmetric**.

(m) In general, the binomial distribution
(circle one) **will always be** / **sometimes, but not always, be** symmetric.

(n) The chance of getting 5 or more seed germinations is
(circle one) **0.201** / **0.367** / **0.633**

(o) *Mean (Expected Value).* Multiplying the sample size, $n = 10$, by the parameter, $\pi = 0.4$, $n\pi = 10(0.4) = 4$ (circle one) **gives / does not give** the same value as $E(Y) = \mu$.

(p) *Standard Deviation (Variance).* $n\pi(1-\pi) =$ (circle one) **2.4 / 0.251 / 4** The formula $n\pi(1-\pi)$ (circle one) **gives / does not give** the same value as $V(Y) = \sigma^2$.

3. *Another Example: Cancer Tumors.* A tumor is cancerous 11% ($\pi = 0.11$) of the time; four ($n = 4$) are inspected. Assume this problem obeys the conditions of a binomial experiment.

   (a) The chance three tumors are cancerous is $\frac{n!}{y!(n-y)!} \times \pi^y \times (1-\pi)^{n-y}$ where $n = 4$, $y = 3$ and $\pi = 0.11$, in other words $f(3) =$ (circle one) **0.005 / 0.011 / 0.040**.

   (b) The chance *at most* three tumors are cancerous is $\Pr\{Y \leq 3\} =$ (circle one) **0.995 / 0.997 / 0.999**.

   (c) The *expected* number of cancer tumors is $\mu = n\pi = 4(0.11) =$ (circle one) **0.44 / 0.51 / 0.62**.

   (d) The *variance* in the number of cancer tumors is $\sigma^2 = n\pi(1-\pi) =$ (circle one) **0.15 / 0.40 / 0.51**.

4. *Identifying when to use the Binomial formula: Seed Germination.* Match the five general conditions of a binomial experiment on the left with how these conditions appear in this question.

general conditions	seed example
(i) There are n trials, where n is fixed in advance of the experiment.	(i) There is a 40% chance any seed germinates.
(ii) The trials are identical and two possible outcomes: success (S) or failure (F).	(ii) Each germination is independent of one another.
(iii) The trials are independent of one another.	(iii) There are 10 seeds.
(iv) The probability of success is π and remains constant from one trial to the next.	(iv) The seeds are identical and can either germinate or not.
(v) The number of successes, y, is the measurement of interest.	(v) The number of seeds, y which germinate is measurement of interest.

Match the five items that describe the seed example with the five conditions of a binomial experiment. For example, the binomial experiment condition, "There are n trials, where n is fixed in advance of the experiment." matches, in particular, the seed example "There are 10 seeds."

Seed Example	(i)	(ii)	(iii)	(iv)	(v)
Binomial Experiment Conditions					

Exercise 2.8 (Poisson Distributions.) The Poisson distribution function is given by,

$$f(y) = \frac{e^{-\lambda}\lambda^y}{y!}, \quad y = 0, 1, 2, \ldots$$

where $e = 2.71828\ldots$ and where $\mu = \lambda$ and $\sigma^2 = \lambda$.

See TI–83 Lab 2: Poisson probability distribution

1. *Photons.* A piece of iron is bombarded with electrons and, as a consequence, releases a number of photons. A number, y, of the photon particles released hit a magnetic detection field that surround the piece of iron being tested. It is found that an *average* (or expectation) of $\lambda = 5$ particles hit the magnetic detection field *per microsecond*. Assume the chance that y particles hit the field *per microsecond* is given by the Poisson distribution function.

   (a) The chance that 2 particles hit the field per microsecond is $f(2) = \frac{e^{-\lambda}\lambda^y}{y!} = \frac{e^{-5}5^2}{2!} \approx$ (circle one) **0.06** / **0.07** / **0.08**.
   (Use your calculator: 2nd DISTR B:poissonpdf(5,2) ENTER.)

   (b) The chance that 0 particles hit the field per microsecond is
   $f(0) = \frac{e^{-\lambda}\lambda^y}{y!} = \frac{e^{-5}5^0}{0!} \approx$ (circle one) **0.007** / **0.008** / **0.009**.

   (c) If an average of 5 particles hit the field every one microsecond time interval, then, in a *two* microsecond time interval, an average of $2 \times 5 = 10$ particles will hit the field. In a similar way, in a *six* microsecond time interval, an average of
   (circle one) **25** / **30** / **35** particles will hit the field.

   (d) Since an average of $\lambda = 2 \times 5 = 10$ particles hit the field in a two microsecond time interval, the chance that 3 particles hit the field in a two microsecond time interval, is
   $f(3) = \frac{e^{-\lambda}\lambda^y}{y!} = \frac{e^{-10}10^3}{3!} \approx$ (circle one) **0.007** / **0.008** / **0.009**.

   (e) The chance that $y = 21$ particles hit the field in a four microsecond time interval, is
   $f(21) = \frac{e^{-\lambda}\lambda^y}{y!} = \frac{e^{-20}20^{21}}{21!} \approx$ (circle one) **0.073** / **0.085** / **0.091**.

   (f) *Expected Value and Standard Deviation.* The variance in the number of particles hitting the field is *equal* to the average number (expected number) of particles hitting the field, λ. Consequently, since the expected number of particles hitting the field is
   (circle one) **3** / **4** / **5**,
   the standard deviation must be
   (circle one) $\boldsymbol{\sqrt{\lambda}}$ / $\boldsymbol{\lambda^2}$ / $\boldsymbol{\frac{1}{2}\lambda^2}$.

2. *Bacterial Colonies.* It is found that, on *average*, $\lambda = 8$ bacterial colonies *per cm*^{2} are found on an agar plate. Assume the chance that there are y bacterial colonies *per cm*^{2} is given by the Poisson distribution function.

   (a) The chance that 2 bacterial colonies are found per cm^2 is
   $f(2) = \frac{e^{-\lambda}\lambda^y}{y!} = \frac{e^{-8}8^2}{2!} \approx$ (circle one) **0.011** / **0.073** / **0.085**.
   (Use your calculator: 2nd DISTR B:poissonpdf(8,2) ENTER.)

(b) The chance that no (0) bacterial colonies are found per cm^2 is $f(0) = \frac{e^{-\lambda}\lambda^y}{y!} = \frac{e^{-8}8^0}{0!} \approx$ (circle one) **0.0003 / 0.0008 / 0.0009**.

(c) In a *six* cm^2 area, an average of (circle one) **25 / 30 / 48** bacterial colonies will be found.

3. *Poisson Approximation Of The Binomial: Photons Again.* The Poisson distribution can be used to approximate the binomial distribution by letting $\lambda = n\pi$. This is fairly good approximation if $n \geq 100$ and $n\pi \leq 10$.

   (a) If $n = 2000$ particles are released by the iron per microsecond, and there is a chance $\pi = 0.005$ that a particle hits the surrounding field per microsecond, then $\lambda = 2000(0.005) = 10$. In a similar way, if $n = 1000$ particles are released by the iron, and there is a chance $\pi = 0.01$ that a particle hits the surrounding field, then $\lambda = 1000(0.01) =$ (circle one) **5 / 10 / 15**.

   (b) If $n = 2000$ particles are released by the iron per microsecond, and there is a chance $\pi = 0.005$ that a particle hits the surrounding field per microsecond, then, using the Poisson distribution, the chance that 15 particles hit the field in a one microsecond period, is $\frac{e^{-\lambda}\lambda^y}{y!} = \frac{e^{-10}10^{15}}{15!} \approx 0.0347$. In a similar way, if $n = 1000$ particles are released by the iron, and there is a chance $\pi = 0.01$ that a particle hits the surrounding field, using the Poisson distribution, the chance that 12 particles hit the field in a one microsecond period, is (circle one) **0.07 / 0.09 / 0.11**.

   (c) As a review, if $n = 2000$ particles are released by the iron per microsecond, and there is a chance $\pi = 0.005$ that a particle hits the surrounding field per microsecond, then, using the *Binomial* distribution, the chance that 15 particles hit the field in a one microsecond period, is $\frac{n!}{y!(n-y)!} \times \pi^y \times (1-\pi)^{n-y} = \frac{2000!}{15!(2000-15)!} \times (0.005)^15 \times (0.995)^{1985}$. In a similar way, if $n = 1000$ particles are released by the iron, and there is a chance $\pi = 0.01$ that a particle hits the surrounding field, using the Binomial distribution, the chance that 12 particles hit the field in a one microsecond period, is $\frac{n!}{y!(n-y)!} \times \pi^y \times (1-\pi)^{n-y}$ where (circle one)

      i. $\boldsymbol{n = 12, y = 1000, \pi = 0.005}$
      ii. $\boldsymbol{n = 1000, y = 12, \pi = 0.01}$
      iii. $\boldsymbol{n = 0.005, y = 12, \pi = 1000}$

4. *Identifying when to use the Poisson Distribution: The Poisson Process.*

   (a) The bacterial colony example could be described in the following way(s). (circle none, one or more)

      i. The probability of finding one bacterial colony in an infinitesimally small area on an agar plate is proportional to the size of this area.

ii. The probability of finding *more than* one bacterial colony in an infinitesimally small area on an agar plate is negligible. (This condition, combined with the first, essentially implies two bacterial colonies can*not* be found one on top of the other.)

iii. Finding a bacterial colony in each area is an independent occurrence.

(b) In fact, the bacterial colony example obeys the three conditions of a *Poisson process*, which is defined in general below. Do *not* circle anything yet; the question is a bit further down.

i. The probability that an event occurs in an infinitesimally small region is proportional to the size of this area.

ii. The probability that more than one event occurs in an infinitesimally small region is negligible.

iii. Events over disjoint regions are independent.

Match the three items that describe the bacterial colony example with the three conditions of a Poisson process. For example, the Poisson process condition, "Events over disjoint regions are independent." matches, in particular, the bacterial colony example "Finding bacterial colony in each area is an independent occurrence."

Bacterial Colony Example	(i)	(ii)	(iii)
Poisson Process Conditions			

2.6 Some Continuous Distributions

Of all the possible continuous distributions, we look at one special one, the *normal distribution*, denoted $N(\mu, \sigma^2)$, which is described by the following function,

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

with mean μ and standard deviation σ.

We will first look in detail at percentages for both a special case of the normal distribution, called the *standard* normal distribution and, then, we will then look at percentages for the general (nonstandard) normal distribution. We then look at percentiles for both these distributions. Finally, we look at z–scores and the log–normal distribution and at the (nonstandard) normal approximation of the binomial distribution.

Exercise 2.9 (Probabilities For Standard and Nonstandard Normal)

1. *Percentage, Standard Normal.* In Westville, in February, the temperature, Z, is assumed to be *standard* normally distributed with mean $\mu = 0^o$ and variance $\sigma^2 = 1^o$.

$\boxed{\text{See TI–83 Lab 2: normal distribution}}$

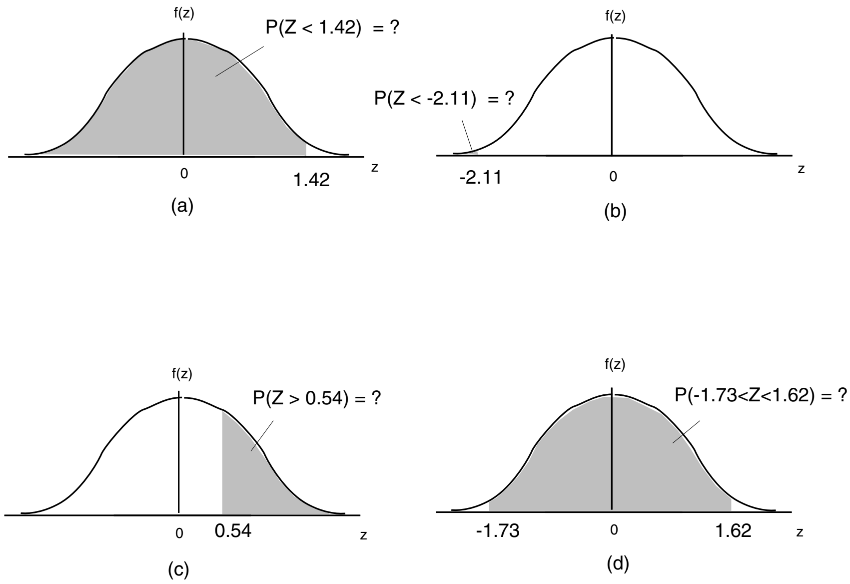

Figure 2.11 (Calculating Probabilities For The Standard Normal Distribution)

(a) The standard normal distribution, in (a) of the figure above, say, is (circle one) **skewed right** / **symmetric** / **skewed left**.

(b) Since the standard normal is much like a histogram, the total area under this curve is (circle one) **50%** / **75%** / **100%** / **150%**.

(c) The shape of this distribution is (circle one) **triangular** / **bell–shaped** / **rectangular**.

(d) This distribution is centered at (circle one) $\boldsymbol{\mu = 0^o}$ / $\boldsymbol{\mu = 1^o}$.

(e) According to the Empirical Rule, 68% of the distribution is within one standard deviation of the average. In this case, about (circle one) **99.7%** / **95%** / **68%** of the temperatures are within $\sigma = 1^o$ of the average temperature, 0^o.

(f) Since this distribution is symmetric, (circle one) **25%** / **50%** / **75%** of the temperatures are above (to the right) of 0^o.

(g) The probability of the temperature being less than 1.42^o is (circle one) **greater than** / **about the same as** / **smaller than** 0.50. Use (a) in the figure above.

(h) The probability the temperature is less than 1.42^o, $P(Z < 1.42) =$ (circle one) **0.9222** / **0.0174** / **0.2946** / **0.9056** See (a) above.
(Use your calculator: 2nd DISTR 2:normalcdf(− 2nd EE 99, 1.42,0,1).)

(i) $P(Z < -2.11) =$ (circle one) **0.9222** / **0.0174** / **0.2946** / **0.9056**.
(Use 2nd DISTR 2:normalcdf($-$ 2nd EE 99, -2.11,0,1).)

(j) $P(Z > 0.54) =$ (circle one) **0.9222** / **0.0174** / **0.2946** / **0.9056**.
(Use 2nd DISTR 2:normalcdf(0.54, 2nd EE 99).)

(k) $P(-1.73 < Z < 1.62) =$ (circle one) **0.9222** / **0.0174** / **0.2946** / **0.9056**.
(Use 2nd DISTR 2:normalcdf(-1.73, 1.62,0,1).)

(l) **True** / **False** The probability the temperature is *exactly* 1.42^o, say, is *zero*. This is because the probability is equal to the *area* under the bell–shaped curve and there is *no* area "under" the "line" at 1.42^o.

(m) **True** / **False** $P(Z < 1.42^o) = P(Z \leq 1.42^o)$.

2. *Nonstandard Normal, A First Look.* It has been found that IQ scores can be distributed by a nonstandard normal distribution. The following figure compares the two normal distributions for the 16 year olds and 20 year olds.

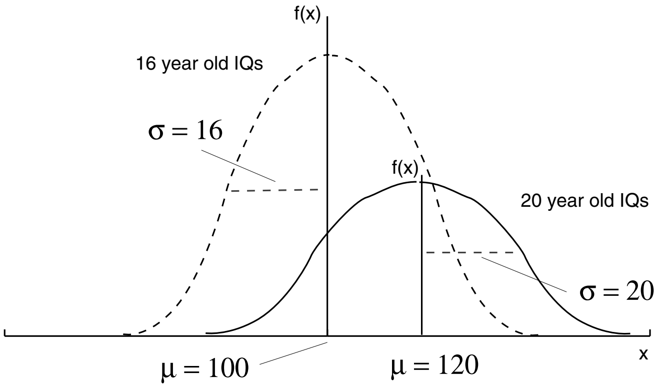

Figure 2.12 (Nonstandard Normal Distributions of IQ Scores)

$\boxed{\text{See TI–83 Lab 2: normal distribution}}$

(a) The mean IQ score for the 20 year olds is
(circle one) **100** / **120** / **124** / **136**.

(b) The average (or mean) IQ score for the 16 year olds is
(circle one) **100** / **120** / **124** / **136**.

(c) The standard deviation in the IQ score for the 20 year olds
(circle one) **16** / **20** / **24** / **36**.

(d) The standard deviation in the IQ score for the 16 year olds is
(circle one) **16** / **20** / **24** / **36**.

(e) The normal distribution for the 20 year old IQ scores is (circle one) **broader than** / **as wide as** / **narrower than** the normal distribution for the 16 year old IQ scores.

(f) The normal distribution for the 20 year old IQ scores is (circle one) **shorter than** / **as tall as** / **taller than** than the normal distribution for the 16 year old IQ scores.

(g) The total area (probability) under the normal distribution for the 20 year old IQ scores is (circle one) **smaller than** / **the same as** / **larger than** the area under the normal distribution for the 16 year old IQ scores.

(h) **True** / **False** Neither the normal distribution for the IQ scores for the 20 year old IQ scores nor the 16 year old IQ scores is a *standard* normal because neither have mean zero, $\mu = 0$, and standard deviation 1, $\sigma = 1$. Both, however, have the same general "bell–shaped" distribution.

(i) There is (circle one) **one** / **two** / **many** / **an infinity** of *non*standard normal distributions. The standard normal is one special case of the family of (nonstandard) normal distributions where $\mu = 0$ and $\sigma = 1$.

(j) (Review: Statistic, Parameter, Sample and Population.) The two nonstandard normal curves given here are examples of chance distributions for the (circle one) **sample** / **population**. Consequently, the average IQ score of 100 for the 16 year olds, μ, is an example of the value of a (circle one) **statistic** / **parameter**. Also, the SD in the IQ score of 20 for the 20 year olds, σ, is an example of the value of a (circle one) **statistic** / **parameter**.

(k) A random variable which has a normal distribution is a (circle one) **continuous** / **discrete** random variable.

(l) (Review) In a "real" situation, we probably (circle one) **do** / **do not** know the everything about the IQ scores for all 16 and 20 year olds, as would be required in constructing these two nonstandard normal curves. We probably (circle one) **do** / **do not** know, for sure, the population (true, actual) average or SD in the IQ scores. We could guess the parameter values of the average and SD and then check our guesses by comparing these guesses with the values of the statistical values of the average and SD–if parameter and statistic values are close, then our guesses of the parameter values are (circle one) **good** / **bad** ones.

3. *Percentages, Nonstandard Normal.*

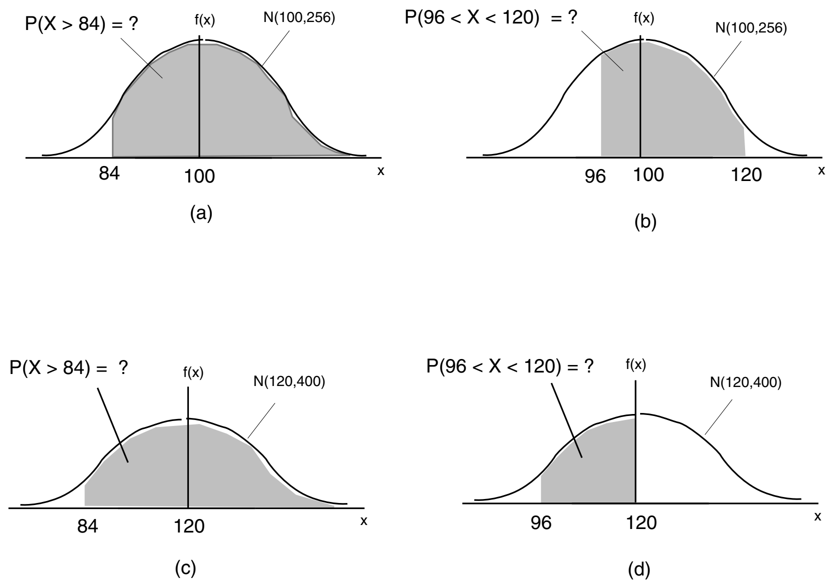

Figure 2.13 (Nonstandard Normal Distributions)

(a) The upper two (of the four) normal curves above represent the IQ scores for *sixteen* year olds. Both are nonstandard normal curves because the (circle none, one or more)

i. the average is 100 and the SD is 16.
ii. neither the average is 0, nor is the SD equal to 1.
iii. the average is 16 and the SD is 100.
iv. the average is 0 and the SD is 1.

The *lower* two normal curves above represent the IQ scores for *twenty* year olds ($\mu = 120, \sigma = 20$).

(b) Since the sixteen year old distribution is symmetric, (circle one) **25%** / **50%** / **75%** of the IQ scores are above (to the right) of 100.

(c) The probability of the IQ scores being less than 84, $P(Y < 84)$, for the sixteen year old distribution is (circle one) **greater than** / **about the same as** / **smaller than** 0.50.

(d) $P(Y < 84) =$ (circle one) **0.8413** / **0.1587** / **−0.1587**
(Use 2nd DISTR 2:normalcdf(− 2nd EE 99, 84, 100, 16).)

(e) Consider the following table of probabilities and possible values of probabilities.

Column I	Column II
(a) $P(Y > 84)$, "sixteen year old" normal	**(a)** 0.4931
(b) $P(96 < Y < 120)$, "sixteen year old" normal	**(b)** 0.9641
(c) $P(Y > 84)$, "twenty year old" normal	**(c)** 0.8413
(d) $P(96 < Y < 120)$, "twenty year old" normal	**(d)** 0.3849

Using your calculator and the figure above, match the four items in column I with the items in column II.

Column I	(a)	(b)	(c)	(d)
Column II				

(f) **True** / **False** $P(Y < 84) = P(Z < 84)$, where Y is a *nonstandard* normal random variable and Z is a *standard* random variable.

(g) According to the Empirical Rule, (circle one) **99.7%** / **95%** / **68%** of the temperatures are within $\sigma = 16$ of the average IQ score for the 16 year olds, 100.

(h) According to the Empirical Rule, 95% of the distribution falls within two standard deviations of the mean. So, (circle one) **99.7%** / **95%** / **68%** of the temperatures are within $2\sigma = 40$ of the mean IQ score for the 20 year olds, 120.

Exercise 2.10 (Percentiles, Standard and Nonstandard Normal.)

1. *Percentiles, Standard Normal.*

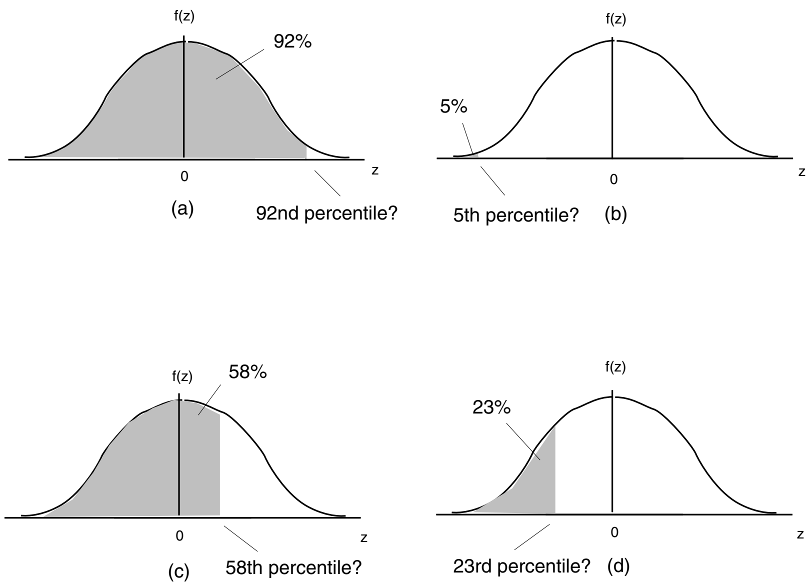

Figure 2.14 (Percentiles of a Standard Normal Distribution)

(a) **True** / **False** The 50th percentile is that temperature such that there is a 50% chance of temperatures being below (and also above) this temperature.

(b) Since the standard normal distribution is symmetric, centered at 0^o, and contains "100%" of the probability, the 50th percentile must be (circle one) **below** 0^o / **equal to** 0^o / **above** 0^o.

(c) The median must be (circle one) **below** 0^o / **equal to** 0^o / **above** 0^o.

(d) **True** / **False** The 75th percentile is that temperature such that there is a 75% chance of the temperatures being below this temperature and so a 25% chance of the temperatures being above this temperature.

(e) The 75th percentile must be
(circle one) **below** 0^o / **equal to** 0^o / **above** 0^o.

(f) The third quartile must be
(circle one) **below** 0^o / **equal to** 0^o / **above** 0^o.

(g) The 92nd percentile must be
(circle one) **below** 0^o / **equal to** 0^o / **above** 0^o. Use (a) in the figure above.

(h) The 92nd percentile is (circle one) $\mathbf{1.41^o}$ / $\mathbf{-1.65^o}$ / $\mathbf{0.20^o}$ / $\mathbf{-0.74^o}$.
(Use 2nd DISTR 3:invNorm(0.92).)

(i) The 5th percentile is (circle one) $\mathbf{1.42^o}$ / $\mathbf{-1.65^o}$ / $\mathbf{0.20^o}$ / $\mathbf{-0.74^o}$. Use your calculator and (b) in the figure above.

(j) The 58th percentile is (circle one) $\mathbf{1.42^o}$ / $\mathbf{-1.65^o}$ / $\mathbf{0.20^o}$ / $\mathbf{-0.74^o}$. Use your calculator and (c) in the figure above.

(k) The 23rd percentile is (circle one) $\mathbf{1.42^o}$ / $\mathbf{-1.65^o}$ / $\mathbf{0.20^o}$ / $\mathbf{-0.74^o}$. Use your calculator and (d) in the figure above.

(l) That temperature such that 77% of the temperatures are *above* this temperature is (circle one) $\mathbf{1.42^o}$ / $\mathbf{-1.65^o}$ / $\mathbf{0.20^o}$ / $\mathbf{-0.74^o}$.

2. *Percentiles, Nonstandard Normal.* It has been found that IQ scores can be distributed by a nonstandard normal distribution. The following figure compares the two normal distributions for the 16 year olds and 20 year olds.

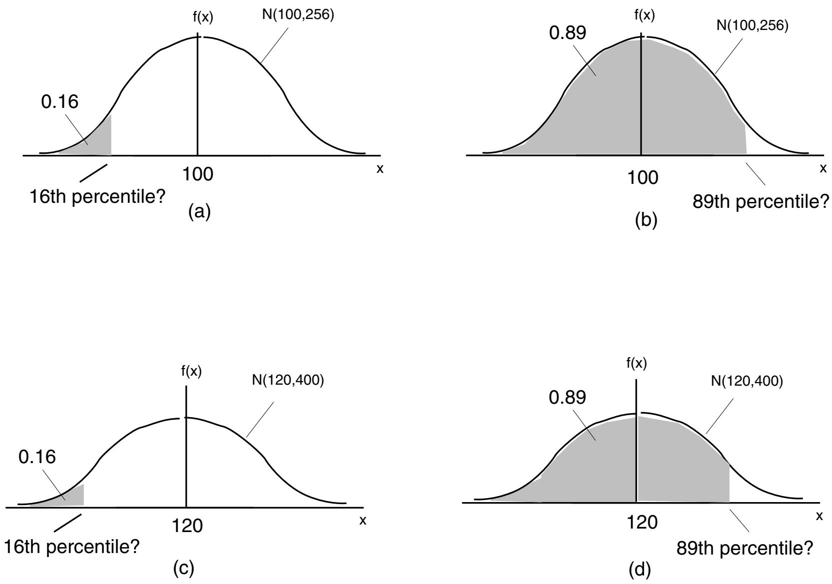

Figure 2.15 (Percentiles of Normal Distributions)

(a) Since the normal distribution for the sixteen year olds is symmetric, centered at 100, the 50th percentile must be (circle one) **below** 100 / **equal to** 100 / **above** 100.

(b) **True** / **False** The 75th percentile is that IQ score such that there is a 75% chance of the IQ scores being below this IQ score and so a 25% chance of the IQ scores being above this IQ score.

(c) The 75th percentile for the sixteen year olds must be (circle one) **below** 100 / **equal to** 100 / **above** 100.

(d) The 75th percentile for the sixteen year olds is (circle one) **103.5** / **106.5** / **110.8** / **122.3**. (Use 2nd DISTR 3:invNorm(0.75, 100, 16).)

(e) The 50th percentile for the normal distribution for the twenty year olds is (circle one) **below** 120 / **equal to** 120 / **above** 120.

(f) The 75th percentile for the twenty year olds must be (circle one) **below** 120 / **equal to** 120 / **above** 120.

(g) The 75th percentile for the twenty year olds is (circle one) **133.5** / **106.5** / **125.4** / **142.3**. Use your calculator.

(h) Consider the following table of percentiles and possible values of percentiles.

Column I	Column II
(a) 16th percentile, sixteen year olds	**(a)** 119.6
(b) 89th percentile, sixteen year olds	**(b)** 84.1
(c) 16th percentile, twenty year olds	**(c)** 144.5
(d) 89th percentile, twenty year olds	**(d)** 100.1

Using your calculator and the figure above, match the four items in column I with the items in column II.

Column I	(a)	(b)	(c)	(d)
Column II				

Exercise 2.11 (z–score and log–normal.)

1. *Z–score.* $Z = \frac{Y-\mu}{\sigma}$.

   (a) The *z–score* for an IQ of 110 for the 20 year olds is a measure of how far below 110 is from the average 120 in SD units,

   $$z = \frac{110 - 120}{20} = -0.5.$$

   Similarly, the z–score for an IQ of 110 for the 16 year olds is given by $z = \frac{110-100}{16}$ = (circle one) **−0.625 / 0.625 / 1.25**.

   (b) Although both the 20 year old and 16 year old scored the same, 110, on an IQ test, the 16 year old is clearly brighter relative to his/her age group than is the 20 year old relative his/her age group because (circle none, one or more)

   i. the 16 year old SD, 16, is smaller than the 20 year old SD, 20.
   ii. the 16 year old z–score, 0.625, is bigger than 20 year old z–score, -0.5.
   iii. both got the same z–score.
   iv. 110 is greater than the average 16 year old score, 100, but less than the 20 year old average score, 120.

2. *Log–Normal Distribution: Soil–Water Fluxes.* Y is log–normal if $\ln Y$ is normal. Consider the following sample of ten soil–water fluxes.

   $$0.306, 0.363, 0.437, 0.787, 0.899, 1.272, 1.424, 1.634, 1.682, 5.128$$

   We will take the natural log of these numbers and see if the resulting set of numbers obeys the Empirical Rule (and, hence, follow a normal distribution) and thus show that these numbers follow the log–normal distribution.

   (a) Use your calculator and take the ln of these ten measurements. This gives (circle one)

i. 0.306, 0.363, 0.437, 0.787, 0.899, 1.272, 1.424, 1.634, 1.682, 5.128

ii. -1.185, -1.103, -0.829, -0.239, -0.106, 0.241, 0.354, 0.491, 0.520, 1.635

(Type STAT EDIT, then type ten fluxes into L_1 and define L_2 as $\ln(L_1)$.)

(b) The average (mean) of the ten fluxes is -0.013 and the standard deviation is (circle one) **0.853** / **0.925** / **1.253**.

(c) According to the Empirical rule, about 68% or $0.68 \times 10 \approx 7$ values *should be* within *one* standard deviation of the average, or, in other words, in the interval $(-0.013 - 0.853, -0.013 + 0.853) = (-0.866, 0.840)$. In fact, there are (circle one) **6** / **7** / **8** fluxes in this interval, or a percentage of $\frac{7}{10} \times 100 =$ (circle one) **60%** / **70%** / **80%**.

(d) According to the Empirical rule, about 95% or 9 or 10 values *should be* within *two* SDs of the average, or, in other words, in the interval $(-0.013 - 2(0.853), -0.013 + 2(0.853)) = (-1.719, 1.693)$. In fact, there are (circle one) **8** / **9** / **10** fluxes in this interval, or a percentage of $\frac{10}{10} \times 100 =$ (circle one) **80%** / **90%** / **100%**.

(e) Since the Empirical Rule is based on the assumption the population has a normal distribution and since it appears that the ln of the soil–water flux data, Y, obeys the Empirical Rule, this indicates $\ln Y$ (circle one) **does** / **does not** follow a normal distribution.

Exercise 2.12 (Normal Approximation To Binomial Distribution: Caterpillar Survival) Forty percent ($\pi = 0.4$) of caterpillars die at birth. If there are $n = 10$ caterpillars and Y represents the number of deaths at birth (of the 10 caterpillars), the functional form of the probability is given by (assuming this problem obeys the conditions of a binomial experiment),

$$C_{10,y}(0.4)^y(0.6)^{10-y}, \quad y = 0, 1, 2, \ldots 10$$

Calculate $\Pr(Y \geq 5)$ both exactly and using a normal approximation.

1. *Review: Binomial, Exactly.*

(a) The mean number of deaths at birth is given by $\mu = n\pi$ (circle one) **4** / **2.4** / **3.0**.

(b) The standard deviation is given by $\sigma = \sqrt{n\pi(1-\pi)}$ (circle one) **4** / **2.4** / **3.0** / **1.55**.

(c) $\Pr(Y \geq 5) =$ (circle one) **0.367** / **0.289** / **0.577**.
(Use your calculator; subtract 2nd DISTR A:binomcdf(10, 0.4, 4) from one (1).)

2. *Normal Approximation.* Consider a graph of the binomial and a normal approximation to this distribution below.

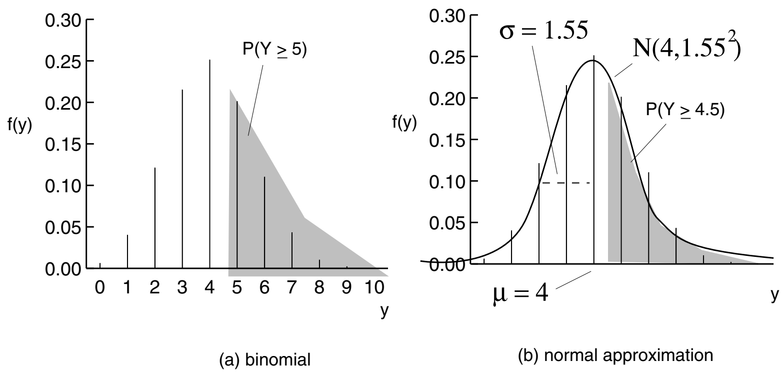

Figure 2.16 (Binomial and Normal Approximation)

(a) **True** / **False** The binomial is a discrete distribution, whereas the normal approximation is a continuous distribution. We will be approximating a discrete distribution by a continuous one. In particular, we plan to approximate the shaded $P(Y \geq 5)$ from the binomial with the shaded $P(Y \geq 4.5)$ from the normal.

(b) It would *appear*, simply by looking at the two graphs above, as though the binomial is (circle one) **skewed** / **symmetric**. This is "good" because we plan to approximate a (not necessarily, but, in this case, apparently) symmetric binomial with an *always* symmetric normal distribution.

(c) To check to see if the binomial is symmetric "enough", we must show that both

$$np \geq 5 \quad \text{and} \quad n(1-\pi) \geq 5.$$

In fact, these conditions are violated in the following way (circle one)

i. $n\pi \geq 5$ and $n(1-\pi) \geq 5$
ii. $n\pi < 5$ and $n(1-\pi) \geq 5$
iii. $n\pi \geq 5$ and $n(1-\pi) < 5$
iv. $n\pi < 5$ and $n(1-\pi) < 5$

and so the binomial, in this case, is actually *not* symmetric enough to be approximated by the normal.

(d) However, in spite of violating the conditions required for symmetry, we will proceed to approximate the binomial with a normal. The normal we will use to approximate the binomial with will be a nonstandard normal with mean equal to the mean of the binomial, $\mu = n\pi = 10(4)$, and standard deviation equal to the standard deviation of the binomial, $\sigma = \sqrt{n\pi(1-\pi)}$ = (circle one) **2.4** / **1.55**.

(e) **True** / **False** The nonstandard normal distribution we will use to approximate the binomial distribution has a mean of 4 and a standard deviation of 1.55.

(f) If Y in normal where $\mu = 4$ and $\sigma = 1.55$, then
$\Pr(Y \geq 5) =$ (circle one) **0.374** / **0.259**.
(Use 2nd DISTR 2:normalcdf(5, 2nd EE 99, 4, 1.55).)

(g) The normal approximation, $P(Y \geq 5) = 0.259$, is (circle one) **smaller than** / **about the same as** / **larger than** the exact binomial value, $\Pr(Y \geq 5) = 0.367$ and so this is a bad normal approximation to the binomial.

(h) To improve the *continuous* normal approximation to the *discrete* binomial, a *continuity correction* factor is introduced. In this case, 0.5 is subtracted from 5 and the revised normal approximation becomes
$\Pr(Y \geq 4.5) =$ (circle one) **0.374** / **0.259**.
(Use 2nd DISTR 2:normalcdf(4.5, 2nd EE 99, 4, 1.55).)

3. *Another Example.* Fifty–six percent of Arctic–line trees survive for 50 years. Compute the chance that at most 14 of 35 trees will survive for 50 years (assuming this problem obeys the binomial experiment conditions).

(a) *Binomial Exactly.* $\Pr(Y \leq 14) \approx$ (circle one) **0.042** / **0.056** / **0.077**.
(Use your calculator: 2nd DISTR A:binomcdf(35, 0.56, 14).)

(b) *Normal Approximation.* $\Pr(Y \leq 14) \approx \Pr(Y \leq 14.5) \approx$ (circle one) **0.041** / **0.056** / **0.077**.
(Use 2nd DISTR 2:normalcdf(−2nd EE 99, 14.5, 35(0.56), $\sqrt{35(0.56)(0.44)}$).)

2.7 The Critical Values Of A Probability Distribution

Exercise 2.13 (Critical Values.) Roughly, *critical values* are the "reverse" of *percentiles*. The α–level ($0 \leq \alpha \leq 1$) critical value of a distribution, $f(y)$, is that value of $y(\alpha)$ such that $100\alpha\%$ of the distribution is at least as large as $y(\alpha)$.

1. *Normal.* The 95th percentile, for example, is often called the critical value at the 0.05 level and denoted $z(0.05)$. The 5th percentile, on the other hand, is often called the critical value at the 0.95 level and is also denoted $z(0.95)$.

(a) The 99th percentile is (check none, one or more)

i. critical value at the 0.01 level

ii. $z(0.01)$

iii. critical value at the 0.99 level

iv. $z(0.99)$

(b) The $z(0.02) =$ (circle one) **−2.05** / **0.35** / **2.05**.
(Type 2nd DISTR 3:invNorm(0.98) ENTER.)

(c) The $z(0.98) =$ (circle one) **−2.05** / **0.35** / **2.05**.
(Type 2nd DISTR 3:invNorm(0.02) ENTER.)

(d) The amount of area (probability) between the critical value at the 0.90 level and the critical value at the 0.02 level is (circle one) **0.80** / **0.09** / **0.88**.

(e) The amount of area (probability) between the critical value at the 0.90 level and the critical value at the 0.01 level is (circle one) **0.80** / **0.09** / **0.89**.

(f) The amount of area (probability) between the $z(0.90)$ and $z(0.10)$ is (circle one) **0.80** / **0.09** / **0.88**.

2. *Discrete: Number of Ears of Corn.* The number of ears of corn, Y, on a typical corn plant has the following probability distribution.

y	0	2	4	6	8	10
$f(y)$	0.17	0.21	0.18	0.11	0.16	0.17

(a) The critical value at the 0.17 level is
(circle one) **0** / **2** / **8** / **10**.

(b) The critical value at the 0.33 level is
(circle one) **0** / **2** / **8** / **10**.

(c) The critical value at 0.05 (circle one) **does** / **does not** exist.

(d) The critical value at the 0.20 level is
(circle one) **0** / **6** / **8** / **not in the data set**.

(e) The critical value at the 0.43 level is
(circle one) **0** / **6** / **8** / **10**.

2.8 Using Computers To Describe Populations

We are going to mostly use the TI–83 calculator throughout this course. We will also use, to a lesser extent, the statistical software package, SAS.