Chapter 23

Statistical distributions of discrete random variables

A DISCRETE RANDOM VARIABLES

A ***random variable*** is a letter (variable) that represents all the possible outcomes of an event in number form.

Example:

X represents the result of the roll of a die.
G represents the position of a pig after rolling it *(no, not a real pig...)*
M represents the number of machine malfunctions over a 24 hour period
T represents the times at which runners finish a race.
E represents the distance by which a dart player misses the bulls eye.

A ***discrete random variable*** is a random variable that can take on a finite set of distinct values $x_1, x_2, x_2, \ldots$
Generally speaking, discrete random variables represent things that are ***counted***.

Examples *X* and *M* above. Notice that while *M* can take on very many different values, there is a finite set of possibilities. Therefore *M* is discrete.

By contrast a ***continuous random variable*** can take on any value in an interval.
Generally speaking, continuous random variables represent things that are ***measured***.
Examples *T* and *E* above.

(Notice that the "continuity" of something depends on the degree of precision of the measuring device. But even though we may not be able to measure a distinction, we assume that the underlying phenomenon is continuous and thus so is the random variable assigned to it.)

HW: 23A #1-4

B DISCRETE PROBABILITY DISTRIBUTIONS

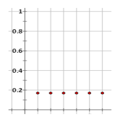
A random variable has an associated ***probability distribution*** that is a description of the probability that the variable will take on any particular value.

For example: X is the result of the roll of one die.
The variable X can take on one of six values. The probability of each is the same so the probability distribution function would look like:

x	$P(X=x)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

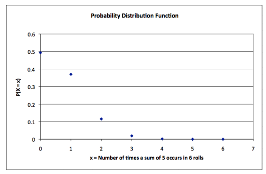
Notice that the capital X stands for the variable. Lower case x or x_i stands for the ***values that X can take on***.

We can also graph probability distribution functions:



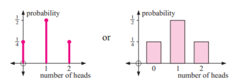
A more interesting example: X is the number of times you get a sum of 5 if you roll a pair of dice 6 times.
The variable X can take on any value from 0 to 6 with differing probabilities. The probability distribution function in table form would be:

x	$P(X=x)$
0	$\binom{6}{0}\left(\frac{1}{9}\right)^0\left(\frac{8}{9}\right)^6=\left(\frac{8}{9}\right)^6$
1	$\binom{6}{1}\left(\frac{1}{9}\right)^1\left(\frac{8}{9}\right)^5$
2	$\binom{6}{2}\left(\frac{1}{9}\right)^2\left(\frac{8}{9}\right)^4$
3	$\binom{6}{3}\left(\frac{1}{9}\right)^3\left(\frac{8}{9}\right)^3$
4	$\binom{6}{4}\left(\frac{1}{9}\right)^4\left(\frac{8}{9}\right)^2$
5	$\binom{6}{5}\left(\frac{1}{9}\right)^5\left(\frac{8}{9}\right)^1$
6	$\binom{6}{6}\left(\frac{1}{9}\right)^6\left(\frac{8}{9}\right)^0=\left(\frac{1}{9}\right)^6$



When the probability distribution is given as a function, $P(x)$, the function P is called the ***probability mass function*** of the random variable.

Can you graph the probability distribution function for the random variable that represents the number of head in two flips of a fair coin?



Properties of Discrete Probability Distributions

The probability of any individual outcome must obey:

$$0 \le p_i \le 1$$

The total probability must sum to one: $\sum_{i=1}^{n} p_i = 1$

Some practice:

A magazine store recorded the number of magazines purchased by its customers in one week. 23% purchased one magazine, 38% purchased two, 21% purchased three, 13% purchased four, and 5% purchased five.

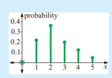
a What is the random variable?
b Make a probability table for the random variable.
c Graph the probability distribution using a spike graph.

a The random variable X is the number of magazines sold.
So, $x = 0, 1, 2, 3, 4$ or 5.

b

x_i	0	1	2	3	4	5
p_i	0.00	0.23	0.38	0.21	0.13	0.05

c



Show that the following are probability distribution functions:

a $P(x) = \dfrac{x^2+1}{34}$, $x = 1, 2, 3, 4$

b $P(x) = C^3_x\,(0.6)^x(0.4)^{3-x}$, $x = 0, 1, 2, 3$

a $P(1) = \frac{2}{34}$ $P(2) = \frac{5}{34}$ $P(3) = \frac{10}{34}$ $P(4) = \frac{17}{34}$

all of which obey $0 \leqslant P(x_i) \leqslant 1$, and $\sum P(x_i) = \frac{2}{34} + \frac{5}{34} + \frac{10}{34} + \frac{17}{34} = 1$

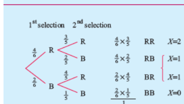
$\therefore$ $P(x)$ is a probability distribution function.

b For $P(x) = C^3_x\,(0.6)^x(0.4)^{3-x}$

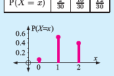
$$\begin{aligned}
P(0) &= C^3_0\,(0.6)^0(0.4)^3 &&= 1 \times 1 \times (0.4)^3 &&= 0.064\\
P(1) &= C^3_1\,(0.6)^1(0.4)^2 &&= 3 \times (0.6) \times (0.4)^2 &&= 0.288\\
P(2) &= C^3_2\,(0.6)^2(0.4)^1 &&= 3 \times (0.6)^2 \times (0.4) &&= 0.432\\
P(3) &= C^3_3\,(0.6)^3(0.4)^0 &&= 1 \times (0.6)^3 \times 1 &&= \underline{0.216}\\
& && &&\text{Total} \quad 1.000
\end{aligned}$$

All probabilities lie between 0 and 1 and $\sum P(x_i) = 1$.
So, $P(x)$ is a probability distribution function.

A bag contains 4 red and 2 blue marbles. Two marbles are randomly selected without replacement. If X denotes the number of reds selected, find the probability distribution of X.



x	0	1	2
$P(X=x)$	$\frac{2}{30}$	$\frac{16}{30}$	$\frac{12}{30}$



HW: 23B #1-11

C EXPECTATION

Expectation: That you will turn your IA in on Mon (Jan 30)
Exceptions: None

You roll a dice 120 times. How many times do you *expect* to roll a six?

Since the probability of rolling a 6 on any individual roll is $\frac{1}{6}$, and you are rolling 120 times, you would expect to get $\frac{1}{6}$ of 120 or 20 sixes.

Expectation

In n trials of an experiment with an outcome that has a probability of p of occurring, the expected number of those outcomes is np.

Note that this is not the same as the question:
What is the ***probability*** that you will roll 20 sixes?

$$\binom{120}{20}\left(\frac{1}{6}\right)^{20}\left(\frac{5}{6}\right)^{100} \approx 0.0973$$

Another example:
There are 20 bills in a box and you pay \$5 to pull one out (they are all crisp and clean so there is no way to distinguish one from another). There are 10 one dollar bills, 5 five dollar bills, and 3 ten dollar bills, and 2 twenty dollar bills.

Would you play this game?

What would you expect the outcome of the game to be?

i	Outcome x_i	*Value* of x_i	Probability of Outcome x_i	Likely payout of x_i
1	\$1	\$1	$P(X=1)=\frac{10}{20}=\frac{1}{2}$	$\frac{1}{2}\cdot$ \$1 = \$0.50
2	\$5	\$5	$P(X=5)=\frac{5}{20}=\frac{1}{4}$	$\frac{1}{4}\cdot$ \$5 = \$1.25
3	\$10	\$10	$P(X=10)=\frac{3}{20}$	$\frac{3}{20}\cdot$ \$10 = \$1.50
4	\$20	\$20	$P(X=20)=\frac{2}{20}=\frac{1}{10}$	$\frac{1}{10}\cdot$ \$20 = \$2.00
			Expected Value	\$5.25

Since the player pays \$5 per game, the expected gain for the player is \$0.25 ***over the long run***. A ***fair*** game is one where the expected value of the player's gain is 0.

Expected Value

A random variable X that takes on n values x_i each of which has probability p_i has an ***expected value*** of:

$$E(X)=\sum_{i=1}^{n} p_i x_i$$

Super important: Notice that $E(X)^2 \neq E(X^2)$ Why not?

Try one:

The table gives the probabilities that a customer buys x magazines from a shop.

x_i	0	1	2	3	4	5
p_i	0.00	0.23	0.38	0.21	0.13	0.05

What is the expected number of magazines that a customer would buy?

How do we interpret the fact that the result is not an integer?

<u>The Mean of a Discrete Random Variable</u>:
A discrete random variable can be thought of a population of values that occur with different relative frequencies. Recall the definition of the population mean in general for values x_i, each of which occurs with frequency f_i

$$\mu = \frac{\sum f_i x_i}{\sum f_i}$$

This can be written as: $\frac{f_1x_1 + f_2x_2 + f_3x_3 + ... + f_nx_n}{N}$

where N is the total number of samples. We can break apart the sum, and rewrite the population mean as

$$= x_1\left(\frac{f_1}{N}\right) + x_2\left(\frac{f_2}{N}\right) + x_3\left(\frac{f_3}{N}\right) + ... + x_n\left(\frac{f_n}{N}\right)$$
$$= x_1p_1 + x_2p_2 + x_3p_3 + ... + x_np_n$$
$$= \sum x_ip_i = E(X)$$

Mean of a Random Variable

The ***population mean*** of a random variable is its *expected value*.

HW: 23C #2-14 even

D THE BINOMIAL DISTRIBUTION

More on our favorite type of experiment.

A ***binomial random variable*** obeys the conditions of a binomial experiment:

Binomial Experiments

There are only two possible outcomes, ***success*** and ***failure.***

Thus, the probability of success is 1 minus the probability of failure.

Multiple trials of the experiment must be ***independent***.

Each trial must be one ***with replacement***.

Let X be the number of successful outcomes in n trials.

We have discussed the probability of getting r successes in n trials. The binomial distribution function is a description of the probabilities of all the possible number of successes, r, that can occur in n trials.

Recall that for a binomial experiment with a probability of success given by p, the probability of getting r successes in n trials is given by:

P(n successes in n trials) = $\binom{n}{r} p^r (1-p)^{n-r}$

Using the notation of random variables, we write:

$$P(X=r)=\binom{n}{r} p^r (1-p)^{n-r}$$

$P(X=r)$ is called the ***binomial probability distribution function***.

The ***expected*** or ***mean*** outcome of the experiment is $\mu = E(X) = np$

If X is a random variable of a binomial experiment with parameters n and p we say that X is ***distributed as*** B(n, p) and write $X \sim B(n, p)$

For Standard Level, you will be using a calculator in these types of problems.

You should know how to use n C r, binom**p**df(n, p, r) and binom**c**df(n, p, r)

While IB does not require it, you should also be able to calculate n C r values using the following technique which is very fast without a calculator:

$$\binom{n}{r}=\left(\frac{n}{1}\right)\cdot\left(\frac{n-1}{2}\right)\cdot\left(\frac{n-2}{3}\right)\cdot\cdots\cdot\left(\frac{n-r+1}{r-1}\right)\cdot\left(\frac{n-r}{r}\right)$$

Chapter 24

Statistical distributions of continuous random variables

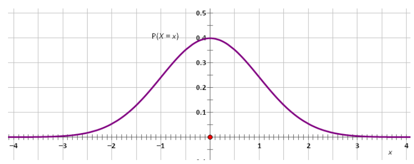
A CONTINUOUS PROBABILITY DENSITY FUNCTIONS

Recall that a ***continuous random variable*** can take on any value in an interval. Generally speaking, continuous random variables represent things that are ***measured***.

Consider a random variable *X* that represents the length of an adult arm span (as opposed to, say, the number of M&M's in a bag which would be discrete)

Like a discrete random variable, each possible arm span has a certain probability of occurring that can be represented as a ***probability density function (PDF)*** of the random variable *X*.

In most cases that we look at we can describe the PDF with a mathematical function and look at its associated graph. For example,



Recall the properties of discrete PDFs. What are the analogous conditions for continuous PDFs?

Properties of a Continuous Probability Density Function

Given a probability density function *f*(*x*) that describes P(*X* = *x*)

1) Any individual value of *x* must obey $0 \le f(x) \le 1$

2) The total probability on some interval, *a* to *b*, must equal one:

$$\int_a^b f(x)dx = 1$$

Technically, the probability of *X* taking on an exact value is zero! Why?

Suppose the probability of *X* being exactly *m* were > 0. Since there are an infinite number of possible values *m* in the continuous interval from *a* to *b*, there is no way the total probability between *a* and *b* could add to one. Thus, the probability of *X* being exactly *m* must be zero.

The ***value (height) of the PDF*** at some value *x* is ***not the same as*** the ***probability that the exact value of*** **x** ***will occur***.

In practice, we are always interested in the probability that a variable takes on a value within some range. Thus:

Properties of a Continuous Probability Density Function

The probability that the variable lies between the values *c* and *d* is given by the area under the PDF curve between *c* and *d*.

$$P(c \le X \le d) = \int_c^d f(x)dx$$

Recall that the mean of a ***discrete*** random variable is given by its expected value. What is the expected value of a continuous random variable?

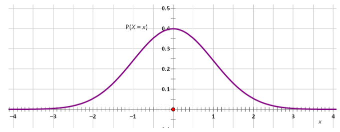
Expected Value and Mean of a Continuous Random Variable

$$E(X) = \int_a^b xf(x)dx = \mu$$

HW: 23D #1-6 all
24A #1-5 all

B NORMAL DISTRIBUTIONS

The most common continuous PDF is the bell-curve or ***normal distribution***. It describes phenomena that are symmetrically centered around a mean, with a bell shaped spread.



Mathematically, the PDF for a normal distribution is given by:

The Normal Probability Density Function

A ***normally distributed*** random variable *X* has a PDF given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad for -\infty \le x \le \infty$$

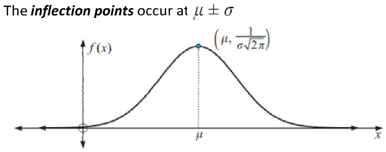
Where μ is the mean of the random variable and
σ is the standard deviation of the random variable

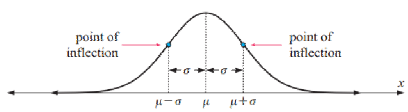
This looks complex, but it's really just a transformation of $g(x) = e^{-x^2}$

Properties of Normal PDFs

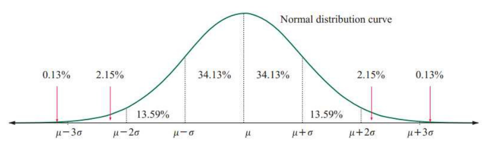
μ and σ are called the ***parameters*** of the curve.
They shift and stretch the curve respectively
For a given pair of parameters we say that $X \sim N(\mu,$
The total area underneath the curve is always one
The curve is centered at μ
The value of *f*(*x*) at μ is $\frac{1}{\sigma\sqrt{2\pi}}$

The ***inflection points*** occur at $\mu \pm \sigma$





Because it is so common, we need to ***memorize*** some facts about its proportions:



Notice that:
- $\approx 68.26\%$ of values lie between $\mu - \sigma$ and $\mu + \sigma$
- $\approx 95.44\%$ of values lie between $\mu - 2\sigma$ and $\mu + 2\sigma$
- $\approx 99.74\%$ of values lie between $\mu - 3\sigma$ and $\mu + 3\sigma$.

Knowing these allows us to easily answer such questions as:
The arm spans of 1000 individuals are normally distributed with a mean of 180 cm and standard deviation of 20 cm.
> How many people do you expect to have arm spans above 200 cm?
> How many people do you expect to have arm spans between 140 and 160 cm?
> ...and a million variations on these types of questions....

One can also find the mean and/or standard deviation given other information.

Fortunately, one can use a calculator to find probabilities for normal distributions:

TI-84 Help with Normal Distributions

For a normally distributed random variable with mean μ and standard deviation σ the probability that the variable lies between *c* and *d* is given by

normalcdf(*c*, *d*, μ , σ)

One can also find the value, *a*, such that P(*X* < *a*) = *p* by using the inverse normal PDF function:

invNorm(*p* μ , σ)

For both functions, if μ and σ are omitted, they are assumed to be 1 and 0 respectively (the ***Standard Normal Distribution***)

HW: 24B.1 #1-9 all
24B.2 #1-3 all

C THE STANDARD NORMAL DISTRIBUTION (Z-DISTRIBUTION)

The ***normal distribution*** describes the variability of many natural phenomena.

Suppose you go shopping and pick a random watermelon and a random apricot from their respective bins and ask the question which is the "better value". It makes no sense to compare the weights of the watermelon and the apricot. But we still might want to know which is the more "unique".

When two variables are normally distributed we can compare the relative "uniqueness" of two samples by finding the number of standard deviations that each is from the mean. Let's take an example:

For the watermelons, the mean weight is 6 kg and the standard deviation is 500 grams. The mean weight of the apricots is 80 grams and the standard deviation is 10 grams.
Suppose you choose a 6.2 kg watermelon and a 85 gram apricot.

How many standard deviations is your watermelon from the mean? $\frac{6.2-6.0}{0.5}=0.4$

How many standard deviations is your apricot from the mean? $\frac{85-80}{10}=0.5$

Z -Score

The number of standard deviations that a sample is from the mean is called the sample's ***z-score***.

It is calculated as $z=\frac{x-\mu}{\sigma}$

Recall the formula for the normal distribution?

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\quad for -\infty\le x\le\infty$$

If we make the substitution $z=\frac{x-\mu}{\sigma}$ we get a new distribution whose mean

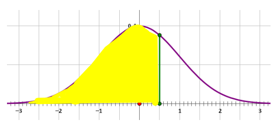
will be zero and whose standard deviation will be one. (Proof is an exercise for the reader)
Making this substitution yields the ***Standard Normal Distribution*** whose formula is:

Standard Normal Distribution

The probability density function for the *Z*-distribution is given by:

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\quad for -\infty\le x\le\infty$$

We can also ask the question, what fraction of the apricots would be less than or equal to the one I chose? This would be calculated as the fraction of the area under the Z-distribution at or below a *z*-score of 0.5

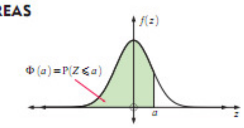


This area is the ***cumulative standard normal distribution function*** and is given by

$$\Phi(a)=P(Z\le a)=\int_{-\infty}^{a}f(z)dz=\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz$$

Notice that these areas depend only on the value of *a*. Prior to calculators, people used tables of values to look up cumulative *z*-values. Even now, you are expected to know how to use these tables.

STANDARD NORMAL CURVE AREAS



	the second decimal digit of a									
a	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

Many problems involving normal distributions can be "standardized" by making the substitution $z = \frac{x-\mu}{\sigma}$

In this way we can find probabilities using only the table of areas. For example:

Given that X is a normal variable with mean 62 and standard deviation 7, find without using technology:

a $P(X \leqslant 69)$ **b** $P(58.5 \leqslant X \leqslant 71.8)$

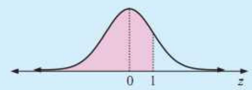
Interpret each result.

a $P(X \leqslant 69)$

$= P\left(\frac{X-62}{7} \leqslant \frac{69-62}{7}\right)$

$= P(Z \leqslant 1)$

≈ 0.841



There is an 84.1% chance that a randomly selected X-value is 69 or less.

b $P(58.5 \leqslant X \leqslant 71.8)$

$= P\left(\frac{58.5-62}{7} \leqslant \frac{X-62}{7} \leqslant \frac{71.8-62}{7}\right)$

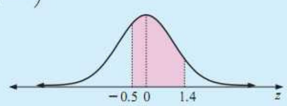
$= P(-0.5 \leqslant Z \leqslant 1.4)$

$= P(Z \leqslant 1.4) - P(Z \leqslant -0.5)$

$= P(Z \leqslant 1.4) - (1 - P(Z \leqslant 0.5))$

$= 0.9192 - (1 - 0.6915)$

≈ 0.611



There is a 61.1% chance that a randomly selected X-value is between 58.5 and 71.8 inclusive.

When a problem says "without technology" they mean transform the problem into the standard normal distribution, then look up the answer in a table of values.

HW: 24C.1	#1-6 all
24C.2	#1-3 all

D QUANTILES OR k-VALUES

Problems that involve ranking of things often use the idea of a percentile or quantile.

You have seen these in test scores. The 95th percentile represents the test score below which 95% of the sample fell. In other words, there is some value k such that $P(X \le k)$ is 0.95. The value k is the 95th ***quantile***. The quantile is also known as the percentile or the k-value.

Suppose that a set of test scores are normally distributed with a mean score of 48 and standard deviation of 10. What would be the score that corresponds to the 75th percentile?

$P(X \le k) = 0.75$? What score lies at the 75th percentile of these data? To answer this question, we can use the ***inverse normal distribution*** function on our calculators or use the ***inverse standard normal distribution*** given in the table below. Let's try the calculator first:

In general, the syntax is **invNorm**(k, μ, σ) where μ and σ are optional

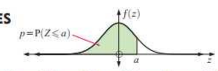
For our problem we use **invNorm**(.75, 48, 10) = 54.7. So a score of 54.7 represents the 75th percentile.

To use the tables, first find the value for 0.75 which is 0.6745. But remember that this is a z value which must be converted back to the initial distribution. Recall that

$$z = \frac{x - \mu}{\sigma} \quad \text{and that therefore } x = \mu + z \cdot \sigma$$

Thus our value is 48 + (0.6745)(10) = 54.7

INVERSE NORMAL PROBABILITIES



p	0	.001	.002	.003	.004	.005	.006	.007	.008	.009
0.50	0.0000	0.0025	0.0050	0.0075	0.0100	0.0125	0.0150	0.0175	0.0201	0.0226
0.51	0.0251	0.0276	0.0301	0.0326	0.0351	0.0376	0.0401	0.0426	0.0451	0.0476
0.52	0.0502	0.0527	0.0552	0.0577	0.0602	0.0627	0.0652	0.0677	0.0702	0.0728
0.53	0.0753	0.0778	0.0803	0.0828	0.0853	0.0878	0.0904	0.0929	0.0954	0.0979
0.54	0.1004	0.1030	0.1055	0.1080	0.1105	0.1130	0.1156	0.1181	0.1206	0.1231
0.55	0.1257	0.1282	0.1307	0.1332	0.1358	0.1383	0.1408	0.1434	0.1459	0.1484
0.56	0.1510	0.1535	0.1560	0.1586	0.1611	0.1637	0.1662	0.1687	0.1713	0.1738
0.57	0.1764	0.1789	0.1815	0.1840	0.1866	0.1891	0.1917	0.1942	[illegible]	0.1993
0.58	0.2019	0.2045	0.2070	0.2096	0.2121	0.2147	0.2173	0.2198	0.2224	0.2250
0.59	0.2275	0.2301	0.2327	0.2353	0.2378	0.2404	0.2430	0.2456	0.2482	0.2508
0.60	0.2533	0.2559	0.2585	0.2611	0.2637	0.2663	0.2689	0.2715	0.2741	0.2767
0.61	0.2793	0.2819	0.2845	0.2871	0.2898	0.2924	0.2950	0.2976	0.3002	0.3029
0.62	0.3055	0.3081	0.3107	0.3134	0.3160	0.3186	0.3213	0.3239	0.3266	0.3292
0.63	0.3319	0.3345	0.3372	0.3398	0.3425	0.3451	0.3478	0.3505	0.3531	0.3558
0.64	0.3585	0.3611	0.3638	0.3665	0.3692	0.3719	0.3745	0.3772	0.3799	0.3826
0.65	0.3853	0.3880	0.3907	0.3934	0.3961	0.3989	0.4016	0.4043	0.4070	0.4097
0.66	0.4125	0.4152	0.4179	0.4207	0.4234	0.4261	0.4289	0.4316	0.4344	0.4372
0.67	0.4399	0.4427	0.4454	0.4482	0.4510	0.4538	0.4565	0.4593	0.4621	0.4649
0.68	0.4677	0.4705	0.4733	0.4761	0.4789	0.4817	0.4845	0.4874	0.4902	0.4930
0.69	0.4959	0.4987	0.5015	0.5044	0.5072	0.5101	0.5129	0.5158	0.5187	0.5215
0.70	0.5244	0.5273	0.5302	0.5330	0.5359	0.5388	0.5417	0.5446	0.5476	0.5505
0.71	0.5534	0.5563	0.5592	0.5622	0.5651	0.5681	0.5710	0.5740	0.5769	0.5799
0.72	0.5828	0.5858	0.5888	0.5918	0.5948	0.5978	0.6008	0.6038	0.6068	0.6098
0.73	0.6128	0.6158	0.6189	0.6219	0.6250	0.6280	0.6311	0.6341	0.6372	0.6403
0.74	0.6433	0.6464	0.6495	0.6526	0.6557	0.6588	0.6620	0.6651	0.6682	0.6713
0.75	[illegible]	0.6776	0.6808	0.6840	0.6871	0.6903	0.6935	0.6967	0.6999	0.7031
0.76	0.7063	0.7095	0.7128	0.7160	0.7192	0.7225	0.7257	0.7290	0.7323	0.7356
0.77	0.7388	0.7421	0.7454	0.7488	0.7521	0.7554	0.7588	0.7621	0.7655	0.7688
0.78	0.7722	0.7756	0.7790	0.7824	0.7858	0.7892	0.7926	0.7961	0.7995	0.8030
0.79	0.8064	0.8099	0.8134	0.8169	0.8204	0.8239	0.8274	0.8310	0.8345	0.8381
0.80	0.8416	0.8452	0.8488	0.8524	0.8560	0.8596	0.8633	0.8669	0.8706	0.8742
0.81	0.8779	0.8816	0.8853	0.8890	0.8927	0.8965	0.9002	0.9040	0.9078	0.9116
0.82	0.9154	0.9192	0.9230	0.9269	0.9307	0.9346	0.9385	0.9424	0.9463	0.9502
0.83	0.9542	0.9581	0.9621	0.9661	0.9701	0.9741	0.9782	0.9822	0.9863	0.9904
0.84	0.9945	0.9986	1.0027	1.0069	1.0110	1.0152	1.0194	1.0237	1.0279	1.0322
0.85	1.0364	1.0407	1.0451	1.0494	1.0537	1.0581	1.0625	1.0669	1.0714	1.0758
0.86	1.0803	1.0848	1.0893	1.0939	1.0985	1.1031	1.1077	1.1123	1.1170	1.1217
0.87	1.1264	1.1311	1.1359	1.1407	1.1455	1.1503	1.1552	1.1601	1.1650	1.1700
0.88	1.1750	1.1800	1.1850	1.1901	1.1952	1.2004	1.2055	1.2107	1.2160	1.2212
0.89	1.2265	1.2319	1.2372	1.2426	1.2481	1.2536	1.2591	1.2646	1.2702	1.2759
0.90	1.2816	1.2873	1.2930	1.2988	1.3047	1.3106	1.3165	1.3225	1.3285	1.3346
0.91	1.3408	1.3469	1.3532	1.3595	1.3658	1.3722	1.3787	1.3852	1.3917	1.3984
0.92	1.4051	1.4118	1.4187	1.4255	1.4325	1.4395	1.4466	1.4538	1.4611	1.4684
0.93	1.4758	1.4833	1.4909	1.4985	1.5063	1.5141	1.5220	1.5301	1.5382	1.5464
0.94	1.5548	1.5632	1.5718	1.5805	1.5893	1.5982	1.6072	1.6164	1.6258	1.6352
0.95	1.6449	1.6546	1.6646	1.6747	1.6849	1.6954	1.7060	1.7169	1.7279	1.7392
0.96	1.7507	1.7624	1.7744	1.7866	1.7991	1.8119	1.8250	1.8384	1.8522	1.8663
0.97	1.8808	1.8957	1.9110	1.9268	1.9431	1.9600	1.9774	1.9954	2.0141	2.0335
0.98	2.0537	2.0749	2.0969	2.1201	2.1444	2.1701	2.1973	2.2262	2.2571	2.2904
0.99	2.3263	2.3656	2.4089	2.4573	2.5121	2.5758	2.6521	2.7478	2.8782	3.0902

Try one, using the table and your calculator

Find k for which $P(X \leqslant k) = 0.95$ given that $X \sim N(70, 10^2)$ and X is measured in mm.

$$P(X \leqslant k) = 0.95$$

$$\therefore \quad P\left(\frac{X-70}{10} \leqslant \frac{k-70}{10}\right) = 0.95$$

$$\therefore \quad P\left(Z \leqslant \frac{k-70}{10}\right) = 0.95$$

Using the inverse normal table we find:

$$\frac{k-70}{10} \approx 1.6449$$

$$\therefore \quad k \approx 86.4$$

So, approximately 95% of the values are expected to be 86.4 mm or less.

or

Using technology:

If $P(X \leqslant k) = 0.95$

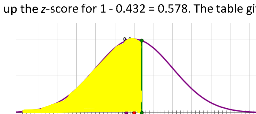
then $k \approx 86.5$.

```
invNorm(0.95,70,
10)
          86.44853626
```

There's a little twist when you have to use the table for percentiles less than 50%. Notice that it only shows percentiles ***above*** 50%. So how would we find the value of k to make $P(X \le k) = 0.432$ for $X \sim N(65, 25)$?

Short answer: Use symmetry!

a) Look up the z-score for 1 - 0.432 = 0.578. The table gives 0.1968



b) But this is the number of standard deviations to the ***right*** of the mean and we want the number to the ***left*** of the mean.

So we calculate 65 - (0.1968)(5) = 64.016

E APPLICATIONS OF THE NORMAL DISTRIBUTION

With an understanding of normal distributions, z-scores, and quantiles you answer ***many*** questions. Applications of the normal distribution are perhaps the most common use of mathematics that you will encounter.

So let's try a few:

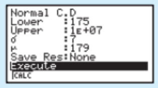
In 1972 the heights of rugby players were found to be normally distributed with mean 179 cm and standard deviation 7 cm. Find the probability that a randomly selected player in 1972 was:

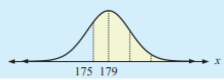
a at least 175 cm tall **b** between 170 cm and 190 cm.

If X is the height of a player then X is normally distributed with $\mu = 179$, $\sigma = 7$.

a $P(X \geqslant 175)$

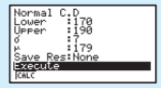
≈ 0.716 {using technology}

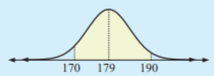




b $P(170 < X < 190)$

≈ 0.843 {using technology}





A university professor determines that 80% of this year's History candidates should pass the final examination. The examination results are expected to be normally distributed with mean 62 and standard deviation 13. Find the expected lowest score necessary to pass the examination.

Let the random variable X denote the final examination result, so $X \sim N(62, 13^2)$.

We need to find k such that $P(X \geqslant k) = 0.8$

$\therefore P(X \leqslant k) = 0.2$

$\therefore k = \text{invNorm}(0.2, 62, 13)$

$\therefore k \approx 51.1$

So, the minimum pass mark is more than 51%. If the final marks are given as integer percentages then the pass mark will be 52%.

A more indirect example (of which there are many variants)

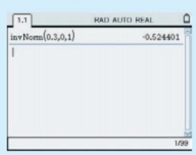
Find the mean and standard deviation of a normally distributed random variable X if $P(X \leqslant 20) = 0.3$ and $P(X \geqslant 50) = 0.2$.

Let the unknown mean and standard deviation be μ and σ respectively.

Now $P(X \leqslant 20) = 0.3$

and so $P\left(\frac{X-\mu}{\sigma} \leqslant \frac{20-\mu}{\sigma}\right) = 0.3$

$\therefore P\left(Z \leqslant \frac{20-\mu}{\sigma}\right) = 0.3$



$\therefore \frac{20-\mu}{\sigma} \approx -0.5244$

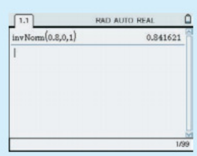
$\therefore 20 - \mu \approx -0.5244\sigma$ ..... (1)

and $P(X \geqslant 50) = 0.2$

$\therefore P(X \leqslant 50) = 0.8$

and so $P\left(\frac{X-\mu}{\sigma} \leqslant \frac{50-\mu}{\sigma}\right) = 0.8$

$\therefore P\left(Z \leqslant \frac{50-\mu}{\sigma}\right) = 0.8$



$\therefore \frac{50-\mu}{\sigma} \approx 0.8416$

$\therefore 50 - \mu \approx 0.8416\sigma$ ..... (2)

Solving (1) and (2) simultaneously we get $\mu \approx 31.5$, $\sigma \approx 22.0$.

Since we don't know the mean and standard deviation, we must convert the random variable to *z*-scores!

If you knew either the mean ***or*** the standard deviation, you would not need to solve simultaneous equations to find the other.

HW: 24D #1-4 all
24E #1-9 all

Unit Test coming Thursday, 2/16/12

Review problems for study

IB Questionbank Problems:	#1-16 all
H&H Review Sets 23A, 23B, 23C	p. 643
H&H Review Sets 24A, 24B, 24C	p. 664