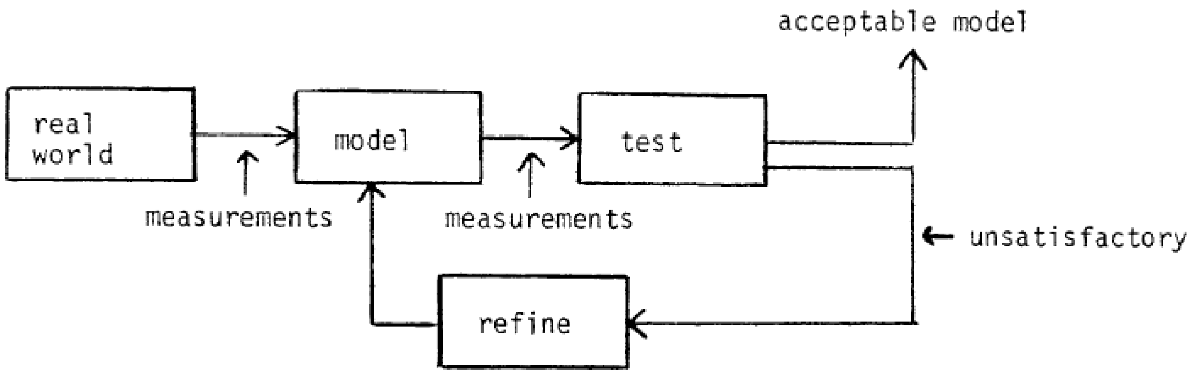
Name: ________________________

EMPIRICAL MODELING MODULE

INTRODUCTION

The essential parts of the model making process are introduced here in a diagrammatic fashion:



The block diagrams reveal some important characteristics of models:

a) They are built from "real world" situations; a model is a replica of the "real world."
b) The creation of a mathematical model involves at the initial stage the collection of appropriate data.
c) The acceptability of a model depends on "tests" : How do the model reproduce measurements? How well has the model predicted future events? In the last analysis, is the model useful?

Models range from the simple (one equation) to the complex, requiring computer processing. (A model of the national economy of the U.S., for example, can easily consists of hundreds if not thousands of equations.) This module works with models of one equation, serving as an introduction to the whole idea of modeling, such as the econometric models covered in Chapters 2 and 3 of the textbook. After completing this module you should:

a) Be exposed to model construction based on empirical data.
b) Become familiar the use of power/exponential functions in econometric modeling—a useful background for spatial modeling .
c) Understand the properties of power or exponential functions through log-linear transformation using semi-log graphs.

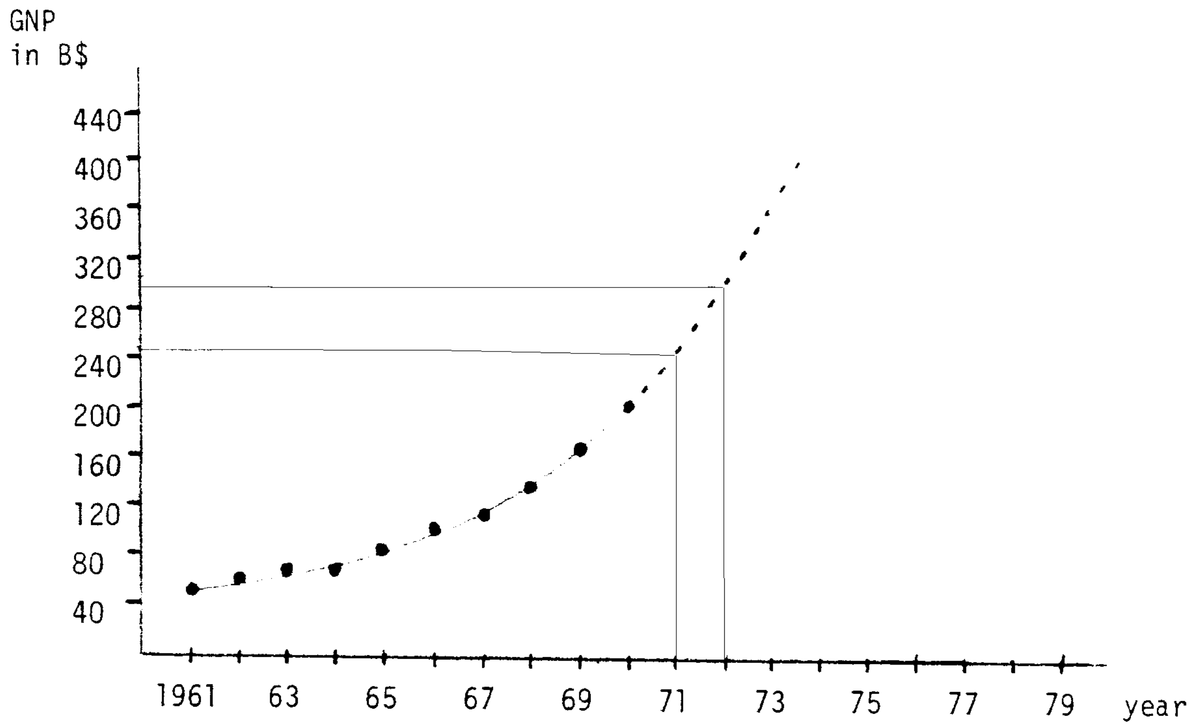
MODELING EMPIRICAL DATA

Illustration 1)

The Gross National Product (GNP) of a country is given below for ten years.

year	GNP (billions of dollars)
1	53
2	59
3	68
4	68
5	85
6	97
7	116
8	142
9	166
10	197

These measurements from the "real world" may be used to produce a model by



graphical means:

The dotted line of the graphical model fulfills one of the purposes of building the model: *prediction of the future*. For example, according to the graphical model, we may predict the GNP to be approximately \$240B for the 11th year, and the GNP to be approximately \$300B for the 12th year. Whether this graphical is acceptable or not can be gleaned by comparing the predicted GNPs with the measured GNPs. If the predicted compare "reasonably" well with the actual, we may consider the model to be acceptable; otherwise, we may reject the model.

The GNP-versus-year graph is in the familiar shape of an exponential curve, usually represented mathematically as: $y = a\, b^x$. The y and x, as usual, are the vertical and horizontal axes respectively. In our model, the GNP is measured by the vertical axis and the year (call it "t") is measured by the horizontal axis. We substitute GNP and "t" into the standard exponential curve formula: GNP = $a\ b^t$ We take the logarithm of both sides of the equation: *

$$\log GNP = \log a + t \log b$$

or equivalently,

$$\log GNP = (\log b)\ t + \log a$$

For simplicity, we specify t = 0 for the base year, t = 1 for the first year, t = 2 for the second year, ... , t = 10 for the 10th year, ... , t = 20 for the 20th year, ... , etc. The new equation, after taking logarithms, is in the standard form of a linear equation with slope m and intercept c:

$$y = m\ x + c$$

where log GNP is equivalent to y of the standard form, log b is equivalent to the slope, t is equivalent to x, and log a is equivalent to the y-axis intercept. If we plot on a graph paper where the vertical axis is in logarithmic scale and the horizontal axis is in the regular numeric scale, the exponential equation of the form $y = a\ b^x$, or its equivalent form, $\log y = \log a + (\log b)\ x$, should be a straight line of the form $y = m\ x + c$. Such a graph paper is called *semi-loq* because one of its axises is in logarithmic scale. The above points are reflected in part by the population projection discussion in Section II.A (Chapter 2) of the textbook.

---

* Remember for logarithms:

$\log 10 = 1$
$\log 100 = 2$
$\log 1000 = 3$ etc.
$\log (a)(b) = \log a + \log b$
$\log a^x = x (\log a)$

For further discussion on logarithmic scales, please see the section beginning on page 5.

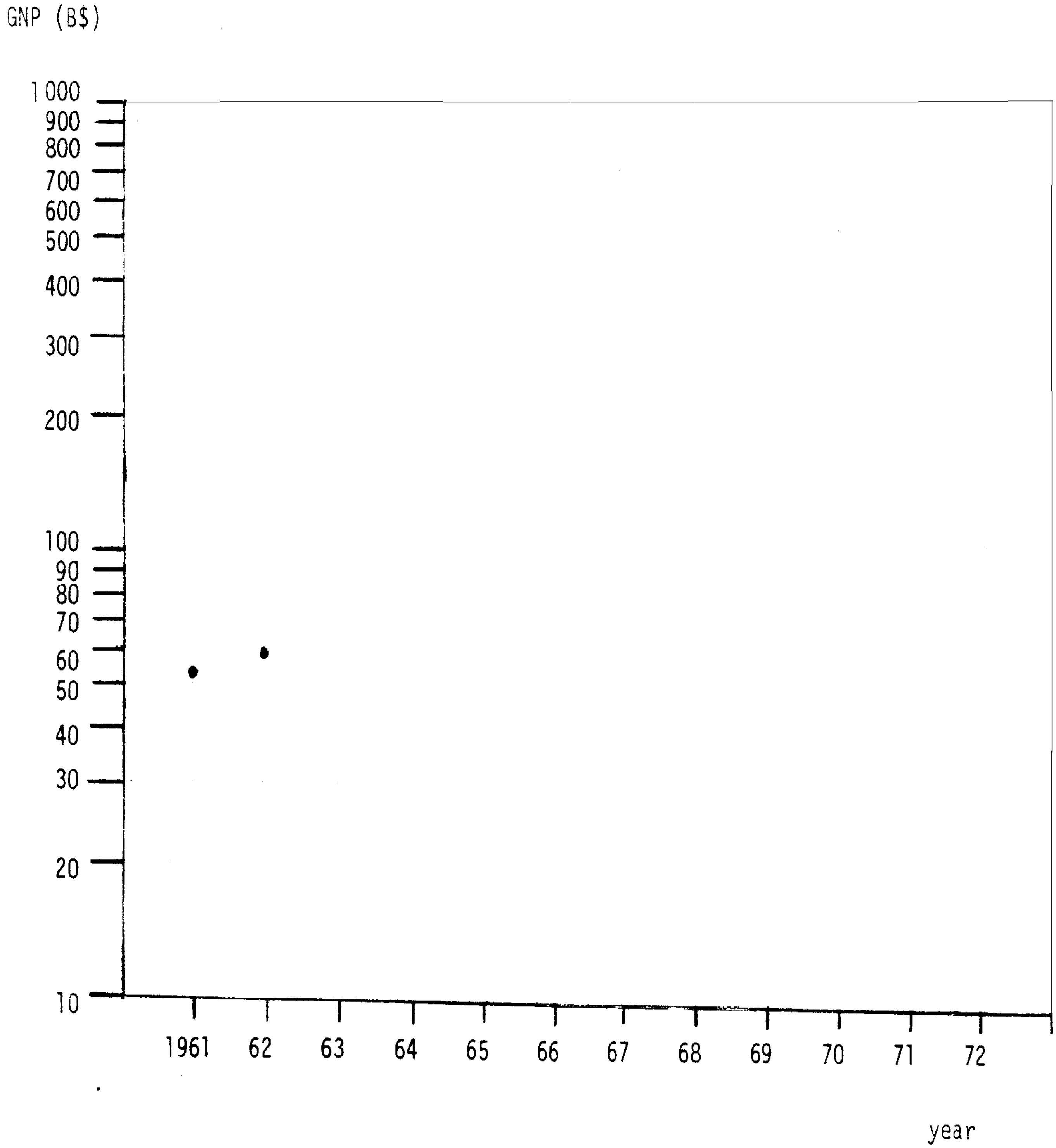
Illustration 2)

Plot the values of GNP and year on the semi-log graph:

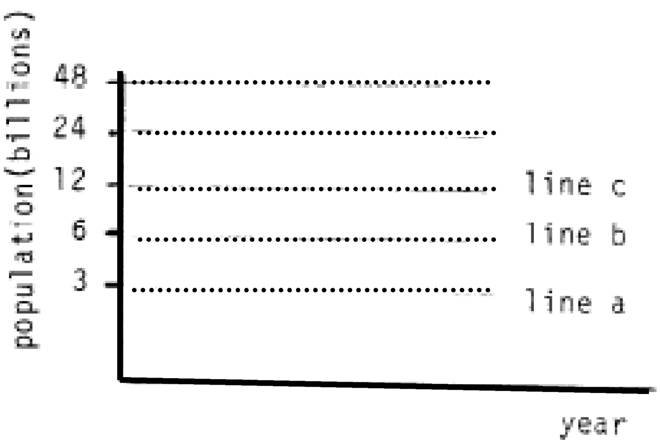
year	1	2	3	4	5	6	7	8	9	10
GNP	53	59	68	68	85	97	116	142	166	197

Note that vertical scale is spaced in logarithmic scale. For example, if we want to graph the value of 100 on the vertical scale, we don't have to look up the value of the log of 100 on a log table, because the mark at 100 on the vertical scale represents the log of

100. For further discussion on logarithm see the section beginning on page 5.

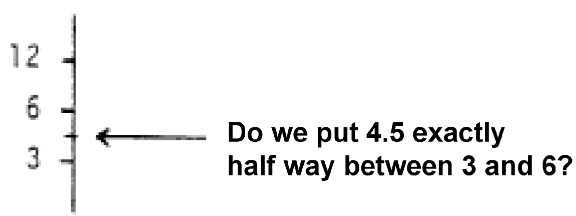


Compare this graph with the graph in Illustration I.



LOGARITHM1C SCALE

A graphic way to appreciate the power function is through its log-linear transformation. Consider the "unusual vertical scale"



If we want to locate or graph a population of 3 billions, we pick line a or if we want to locate or graph a population of 6 billions, we pick line b and so on. However, consider this question: If we want to graph a population of 4.5 billions, how do we place it on the vertical scale?
We know that:

$6 = 2^1(3)$
$12 = 2^1(6) = 2^2(3)$
$24 = 2^1(12) = 2^2(6) = 2^3(3)$

These relations show how the marks are placed on the vertical scale. For example, if we let the distance between "6" and "3" to be one inch, then we mark "12" one inch above "6" (the exponent 1 in 2^1 from $12 = 2^1(6)$); we mark "12" *two inches* above "3" (the exponent 2 in 2^2 from $12 = 2^2(3)$). In other words, it is the exponent in the relations that tells us how to mark the numbers.

-- (to be continued) --

-- (continued from last page) --

Back to the original question, where do we place 4.5? We know the $6 = 2^1(3)$ and $12 = 2^2\,(3)$ etc. What value of x such that $4.5 = 2^x\,(3)$ is true?

$4.5 / 3 = 2^x$ or $1.5 = 2^x$.

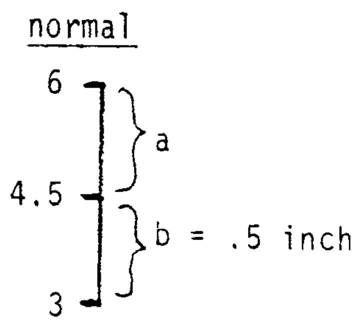
Take logarithm:

$\log 1.5 = x \log 2$

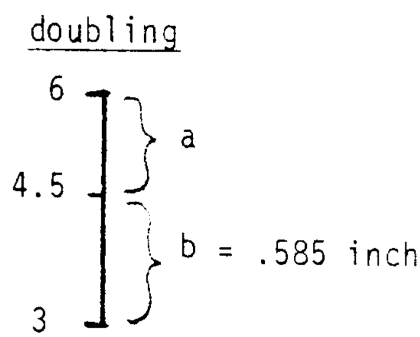
$0.17609 = x\,(0.30103)$

$0.585 = x$

If we place "6" *one* inch from "3" (the exponent in $6 = 2^1(3)$), and we place "4.5" *0.585* inch from "3" (the exponent in $4.5 = 2^{0.585}\,(3)$). Thus, in a *normal* scale, we place "4.5" midway between "6" and "3;" but on a *doubling scale*, we place "4.5" closer to "6" than "3."



Suppose a and b are equal distances. If a + b equals 1 inch, then b = 0.500 inch.

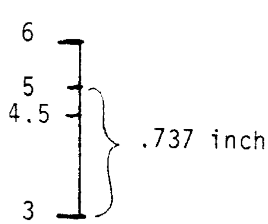


Here, b = 0.585 (a+b). If a + b equals 1 inch, then b = 0.585 inch.

Where do we place "5" on the doubling scale if the distance from "3" to "6" is one inch?

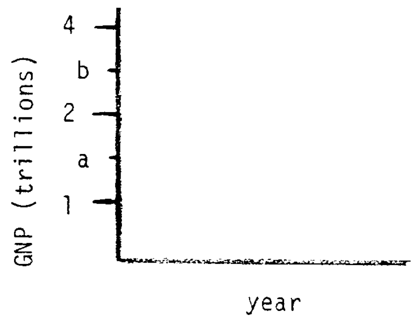
$5 = 2^x\,(3)$, or $0.737 = x$

We place "5" 0.737 inch above "3."



-- (continued from last page) --

Consider the vertical scale of the economic model.



What is the value (GNP) indicated by the marks a and b, where a is half between "1" and "2" and b is also half way between "2" and "4?" Note that ' this is a doubling scale.

We know that:

$2 = 2^1(1)$

If we assume the distance between "2" and "1" is one inch, then a $= 2^{½}$ (1). The ½ is in the exponent because "a" is midway between "2" and "1." "a" is ½ inch from "1."

$2^{½} = \sqrt{2}$
$a = \sqrt{2}(1)$
$a = 1.414.$

Therefore "a" represents 1,414 trillions.

For the mark "b" between "4" and "2":

$4 = 2^1 (2)$
$b = 2^{½} (2)$

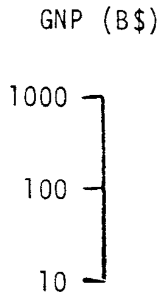
The "½" is in the exponent be because b is ½ inch from "2."

$b = \sqrt{2}(2)$, or $b = 2.828$

Therefore, b represents 2.828 trillions.

The vertical scale on page 4 of the module, unlike the doubling scale of the textbook, is a *multiple factor of 10* scale.

-- (continued from last page) --



$100 = 10^1\ (10)$

$1000 = 10^1\ (100) = 10^2\ (10)$

In the doubling scale, 2 is raised to an exponent; whereas in the multiple factor of 10 scale, 10 is raised to an exponent.

Where do we mark the value of 90 on the multiple of 10 scale if the distance from "10" to "100" is one inch?

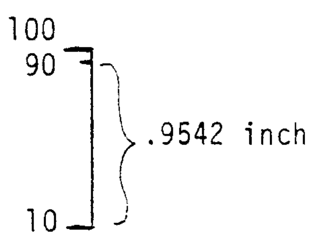
$90 = 10^x\ (10)$

$9 = 10^x$

$\log 9 = x \log 10$

$0.9542 = x$

The value for "90" should be marked 0.9542 inch above "10."



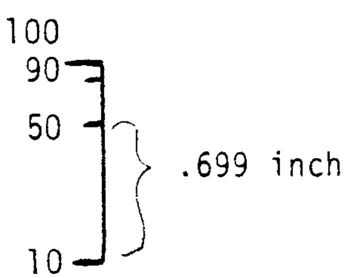
Where should 50 be marked?

$50 = 10^x\ (10)$

$5 = 10^x$

$\log 5 = x \log 10$

$0.699 = x$



USE OF POWER/EXPONENTIAL FUNCTIONS

Illustration 3)

Using the new notations, we calculate new values from the old values in illustration I.

Year	GNP (billions of dollars)	New: log GNP
1	53	1.72
2	59	1.77
3	68	1.83
4	68	1.83
5	85	1.93
6	97	1.99
7	116	2.06
8	142	2.15
9	166	2.22
10	197	2.29

The values of log GNP and year (t) are measurements from the "real world" and we want to build a “regression” model from these data. From the discussions and exercises of illustrations 1 and 2, we know that log GNP and year (t) have a linear relationship of the form:

$$\log GNP = a_1\, t + a_0$$

To "fit" a linear equation to these data points, we use the least squares method:
Given a set of points, $(X_1, Y_1), (X_2, Y_2), \ldots , (X_n, Y_n)$, we could derive an equation to "fit" these points of the form:

$$Y = a_1 X + a_0$$

where

$$a_0 = [(\Sigma_i Y_i)\, (\Sigma_i X_i^2) - (\Sigma_i X_i)\, (\Sigma_i X_i Y_i)] \,/\, [n\, \Sigma_i X_i^2 - (\Sigma_i X_i)^2]$$
$$a_1 = [n\, \Sigma_i X_i Y_i - (\Sigma_i X_i)(\Sigma_i Y_i)] \,/\, [n\, \Sigma_i X_i^2 - (\Sigma_i X_i)^2]$$

For our example:
$Y = \log GNP$
$X =$ year (t)
$n = 10$ (Because we have a set of 10 years, each with a log GNP observation)

The least-squares regression was done for the data points and we arrived at the equation:

logGNP = 0.064 t + 1.62

Please note that the formulas of the previous page were used, giving:
$a_0 = 1.62$
$a_1 = 0.064$
This is our least-squares regression model of the data. The asterisks are to emphasize that *logGNP* is the fitted or predicted value.

a) Using the given equation, what is the fitted value of log GNP for the year t = 1?

logGNP = 0.064 (1) + 1.62 = 1.684

The fitted value, *logGNP*, for t = 1 is 1.684. Compare this fitted value with the actual value of 1.72 for t = 1.

Complete the following Table:

year (t)	fitted (predicted)	actual
1	1.684	1.72
2	1.748	1.77
3		1.83
4		1.83
5		1.93
6		1.99
7		2.06
8		2.15
9		2.22
10	2.26	2.29
15	2.58	no data
20		“
25		“
30		“

The predicted *logGNP* for the country in year 15 is 2.58:

logGNP = 2.58

Taking anti-log we get *GNP* = 380. The least-squares equation model predicts a GNP of 380 billion dollars for year 15. Comparing this prediction with the actual measurements is one of the tests of the appropriateness of this model.

Please convert the fitted *logGNP* to *GNP* and compare with the actual GNP. It is best to show the information as two columns appended to the above Table, labeled “actual GNP (B$)” and “fitted GNP (B$)” respectively.

Illustration 4)

Illustration 1 presented the exponential growth model: $y = a\, b^{x}$. A related model called exponential decay has the form: $y = a\, b^{-x}$. This exponential decay model is very useful for modeling radioactive decay. The appropriate model for radioactive decay is:

$$N = N_0\, e^{-\lambda t}$$

where N has replaced Y of the standard form; N_0 has replaced a; e has replaced b; and λt has replaced x. Notice e is the natural base of logarithm, or e = 2.718.

N = number of radioactive nuclei at time t.
N_0 = The number of radioactive nuclei at time 0, i.e., it is the initial amount.
λ = The decay rate per nucleus; the value of λ varies for different radioactive substances.

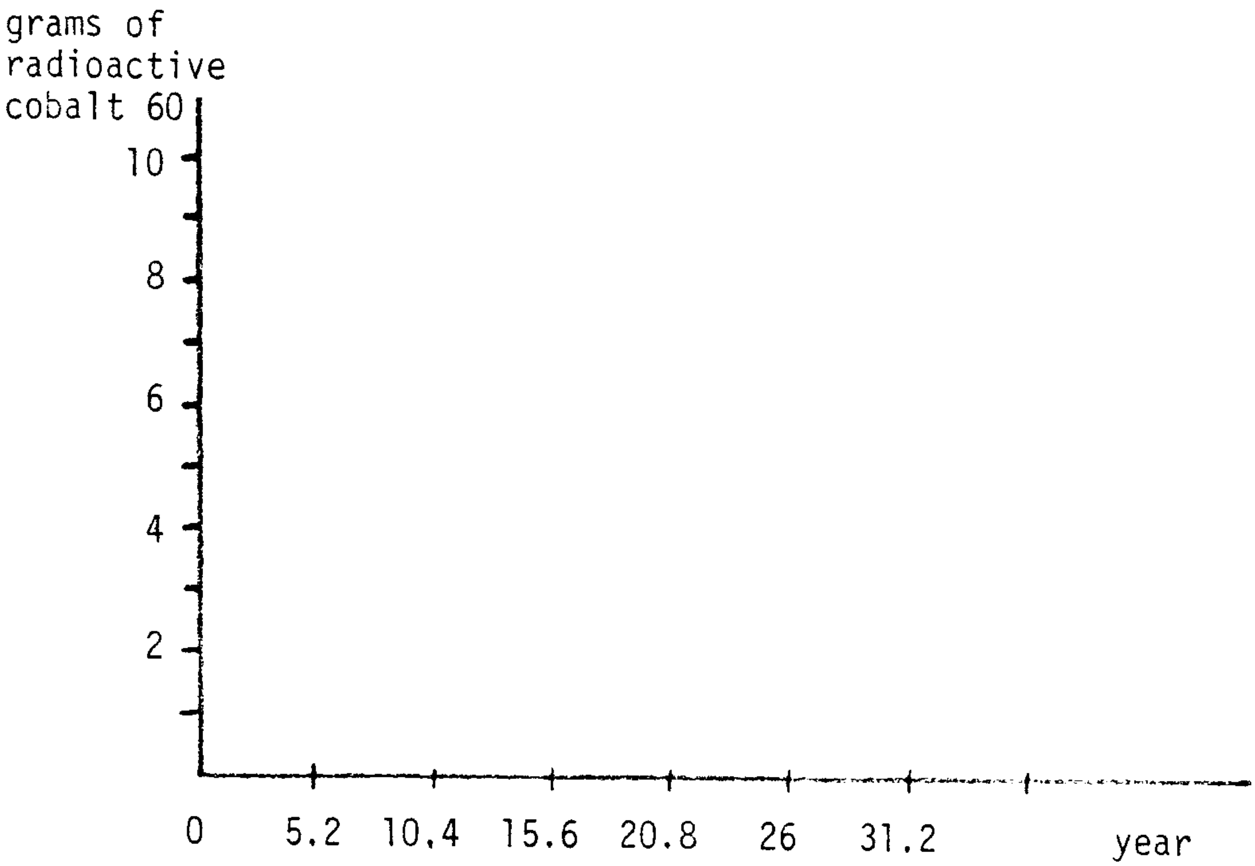
A useful concept is the half-life of a radioactive material. The half-life, $t_{1/2}$, is defined as the time after which the number of radioactive nuclei has decreased to half of its original value. For example, strontium 90 is a radioactive element with a half-life of 28.8 years. If we initially have 2 grams of strontium 90, then after 28.8 years we would have 1 gram of radioactive strontium.

year:	0 (initial year)	28.8	57.6	86.4	115.2	144
amount radioactive:	2 grams	1 g	½ g	1/4 g	1/8 g	1/16 g

Cobalt 60 is radioactive and has a half-life of 5.2 years. If we have initially 10 grams of cobalt, calculate the amount of remaining radioactive cobalt 60 after each time period:

year	0	5.2	10.4	15.6	20.8	26.0	31.2
amount radioactive	10 g	5	( )	( )	( )	( )	( )

Plot your results for cobalt 60 on the following graph. The shape of the resulting graph is typical of the general exponential decay model.



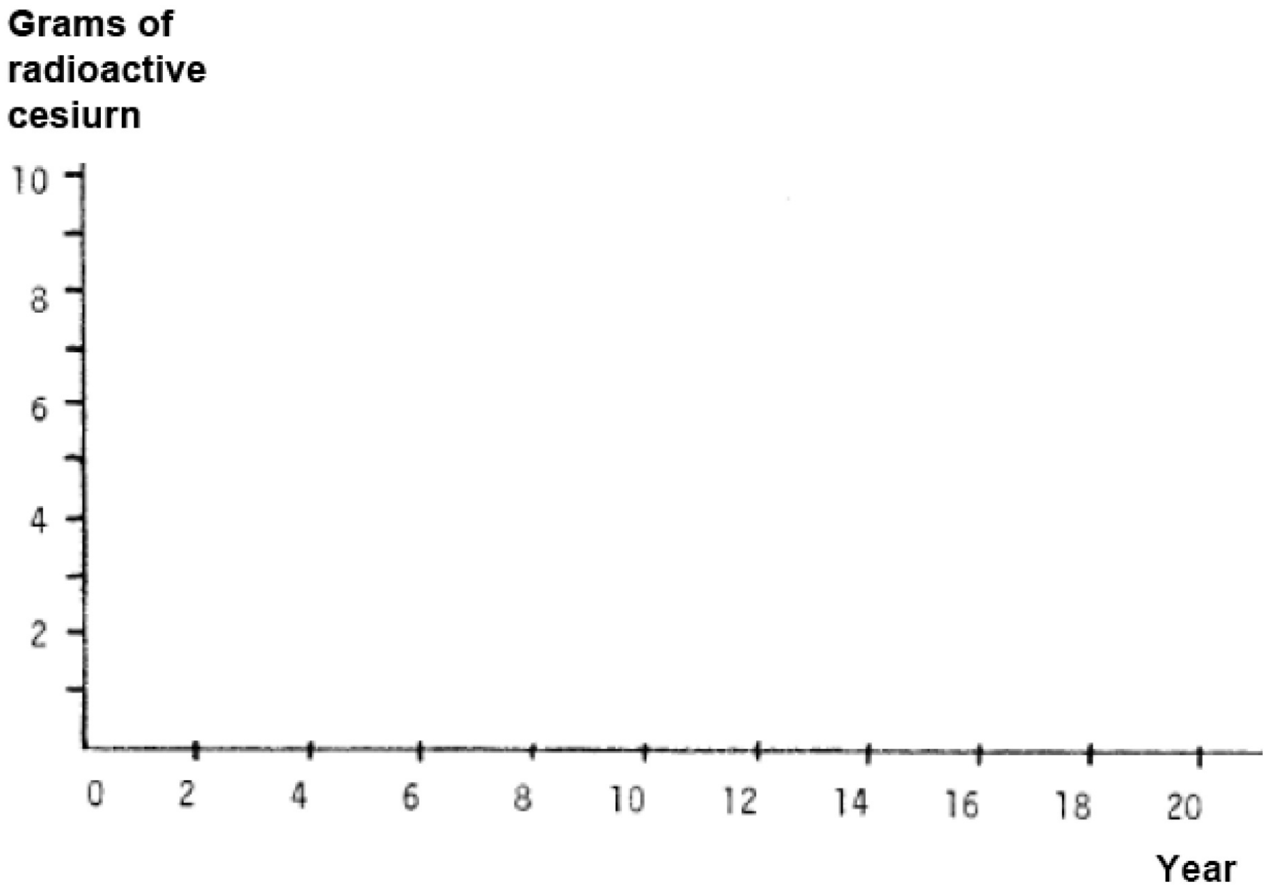
The half-life of cesium 134 is 2 years. If the initial amount of cesium 134 is 10 grams, plot its exponential decay on the blank Figure on page13.

For the same set of numbers (grams and year) for cesium, plot them on the graph on page 14 with grams of cesium 134 on the vertical axis. Compare this graph with the graph on page 13 for cesium and compare this graph with the graph on page 4.

Illustration 5)

Throughout history, humankind has always been concerned with the environment and energy. With the increased energy consumption in industrialized countries, these concerns started to be heightened in the late sixties and seventies. Here is an illustration that is built upon this theme, complementing Section III in Chapter 3 of the text.

The total world oil reserves estimated in 1977 is 600×10^9 barrels. Also, in 1977, it was estimated that the growth rate in demand is 5% per year. For 1977, the annual demand is 15×10^9 barrels. We proceed to build a model of world oil reserves.



Let

D_n = Demand in the year n.

$C_n = D_0 + D_1 + D_2 + \ldots + D_n$ = The cumulative demand through the year n.

For example, for C_4 we have the cumulative demand equal to $D_0 + D_1 + D_2 + D_3 + D_4$. Note that D_0 is the demand at the initial year 1977; D_1 is the demand for 1978; ... , and D_4 is the demand for 1981. Let r = The growth rate in demand per year. If D_0 is the demand in 1977 then:

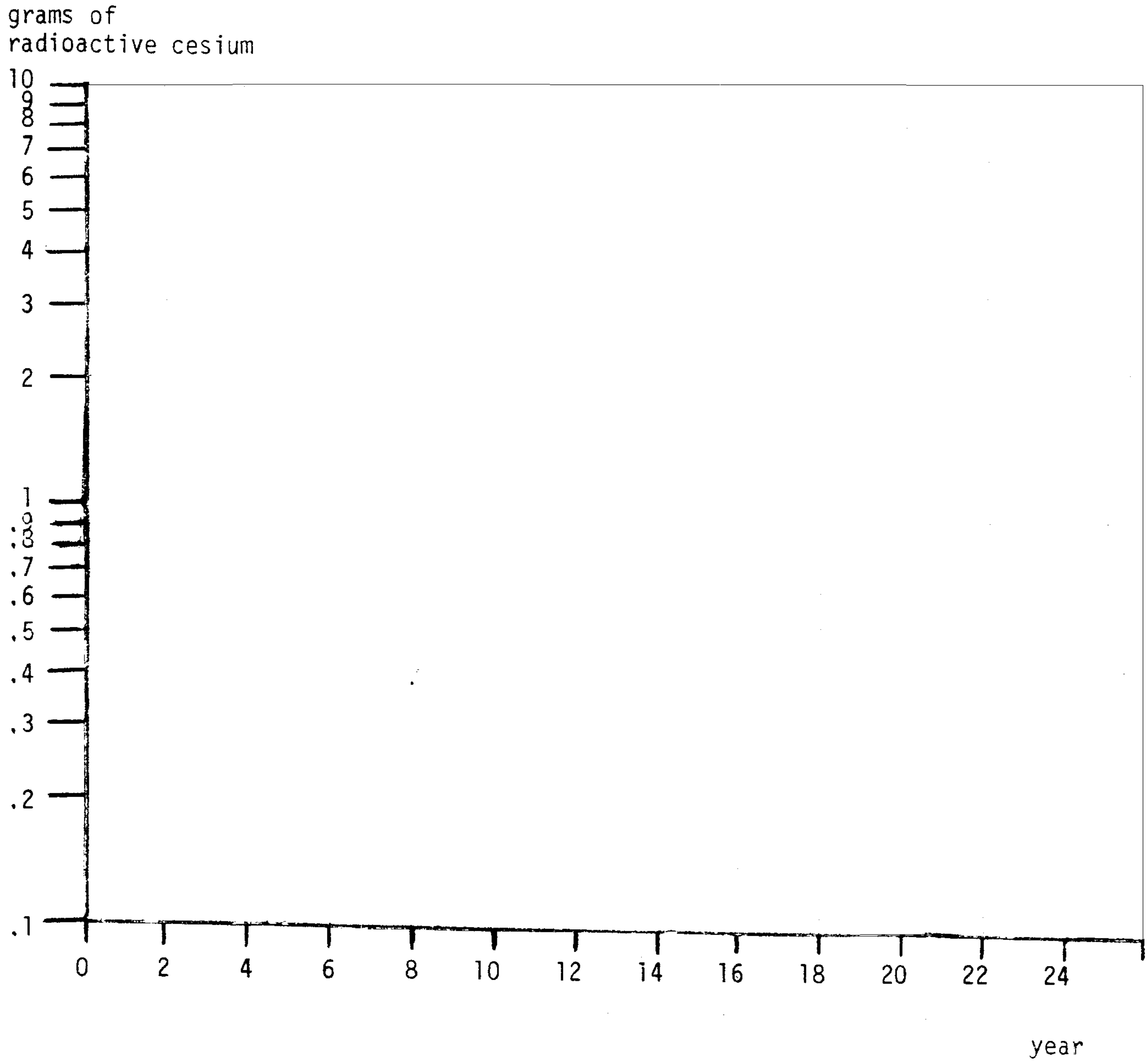
$D_1 = (1+ r)\ D_0$ is the demand in 1978.

$D_2 = (1+ r)\ D_1 = (1+ r)(1+ r)D_0 = (1+ r)^2 D_0$ is the demand in 1979.

. . .

$D_n = (1+ r)^n D_0$ is the demand in the nth year after 1977.

For people familiar with banking interest, this is the annual interest rate model. D_n is the "future" value of the initial deposit of D_0 after compounding for n years at an interest rate of r.



As mentioned, the cumulative demand for oil after n years is $C_n = D_0 + D_1 + D_2 + ... + D_n$, or

$C_n = D_0 + (1+ r) D_0 + (1+ r)^2 D_0 + (1+ r)^3 D_0 + ... + (1+ r)^n D_0$

$C_n = D_0 [1 + (1+ r) + (1+ r)^2 + (1+ r)^3 + ... + (1+ r)^n]$

Notice $1 + (1+ r) + (1+ r)^2 + (1+ r)^3 + ... + (1+ r)^n$ is in the form of $1 + x + x^2 + x^3 + ... + x^n$, which is called a geometric series.

The sum of $1 + x + x^2 + x^3 + ... + x^n$ is $(1 - x^{n+1}) / (1 - x)$. We could therefore write $1 + (1+ r) + (1+ r)^2 + (1+ r)^3 + ... + (1+ r)^n$ as $[1 - (1 + r)^{n+1}] / [1 - (1 + r)]$. Collecting terms to simplify, the final equation looks like $[(1 + r)^{n+1} - 1] / r$. Therefore $C_n = D_0 [1 + (1+ r) + (1+ r)^2 + (1+ r)^3 + ... + (1+ r)^n]$ becomes

The world's oil reserves run out when the cumulative demand, C_n, equals $600x10^9$. We know at 1977, the annual demand, D_0, is equal to $15x10^9$ barrels.

$$600x10^9 = 15x10^9 \, [(1 + r)^{n+1} - 1] / r$$

If we divide both sides of the equation by $15x10^9$, we get:

$$40 = [(1 + r)^{n+1} - 1] / r$$

This model tells us the year, from 1977, at which the world oil reserves are completely depleted if the growth rate of demand for oil remains at r, *provided* no new reserves are added. The qualifier, provided, is very important in this model because of the very "tricky" nature of determining oil reserves. The generally accepted definition for "proven oil reserve" is the "recoverable amount given the present market price of oil." If the price of crude oil goes up, then the "proven oil reserves" may increase because new source or recovery of existing oil becomes economically feasible. There are other modifiers to the model and we don't need to deal with them except to realize that models depend on numerous "*ifs*" and *assumptions*.

The value of $[(1 + r)^{n+1} - 1] / r$ at 40 is tabulated for various n's and r's. In other words, the table provides for a given depletion rate r the number of years, n, from 1977, when the total world oil reserves of $600x10^9$ runs out:

n (years from 1977)	r (in %)
33	1
32	1.25
31	1.5
29	2.0
27	2.5
26	3.0
23.5	4.0
21.5	5.0
20	6.0

The Table shows, for example, that if the world's annual growth rate in demand for oil is 1% then by the year 2010 (1977 + 33), the total proven reserves as determined in 1977 would run out; or, if annual growth rate is 6%, the proven reserves would run out by 1997 (1977 + 20).

Suppose the world's proven reserves is increased from 600×10^9 barrels to 1500×10^9 barrels. Tabulate a chart such as the above for various n's and r's to show the time for the depletion of the reserves.

Remember, the supply is depleted when the cumulative demand, C_n, equals 1500×10^9 barrels, and we assume again that the initial consumption, D_0, is $15 \, x10^9$ barrels:

$$1500 \times 10^9 = 15x10^9 \, [(1 + r)^{n+1} - 1] / r$$

Again, dividing both sides by $15x10^9$ yields

$$100 = [(1 + r)^{n+1} - 1] / r$$

Suppose r = 1%, find n such that the above equation is true.

$$100 = [(1 + 0.01)^{n+1} - 1] / 0.01$$

We multiply both sides of the equation by 0.01:

$$1 = (1 + 0.01)^{n+1} - 1$$

Add 1 to both sides:

$$2 = (1.01)^{n+1}$$

Take logarithm

$$\log 2 = (n+1) \log 1.01$$
$$0.30103 = (n+1)(0.00432)$$
$$69.68 = (n+1)$$
$$68.68 = n$$

In approximately 69 years from 1977, or year 2046, the world's oil reserves of the new 1500×10^9 barrels will run out, if the growth rate of demand is 1% annually.

Calculate n's for 2%, 3%, 4%, 5%, 6% annual growth rates. Starting with 1500×10^9, the number of years from 1977, for given r and for the initial depletion rate of $15x10^9$, we construct the following Table:

n (in years from 1977)	r (in %)
68.68	1.0
( )	2.0
( )	3.0
( )	4.0
( )	5.0
( )	6.0

Sample calculation for r = 2%:

$$1500 \times 10^9 = 15 \times 10^9 [(1 + r)^{n+1} - 1] / r$$
$$100 = [(1 + 0.02)^{n+1} - 1] / 0.02$$

solve for n. (Please show one of these calculations in detail, just as an example.)