There are two kinds of tools which range technicians can carry with them when they do research—mechanical and statistical. Some mechanical tools are simple. The simple ones may be crude, such as an ax handle for measuring stubble heights, or more refined, such as a meter stick. The complicated tools, like the capacitance meter, are designed for the task of taking measurements accurately and objectively, yet cheaply and easily. Statistical tools have a parallel classification. Some of the simple ones are crude, such as the use of the average to describe a population, and some are more refined—affixing the fiducial limits to the average. The complicated statistical methods are powerful tools for the analysis of data after the measurements have been taken. They increase the precision of research and facilitate the interpretation of experimental results and survey data.

Regression analyses are statistical methods which fit into the category just described. They are useful in their simplest form—linear regression—but even more useful with the complexity of multiple regression. Certain uses of regression are made-to-order for many problems in the field of range management. Studies of forage utilization, grazing management, reseeding, fertilization, brush control, chemical analyses, digestion trials, range economics—each of these brings to mind a large number of perfect set-ups for regression analyses.

Regression is perhaps better understood when illustrated than when defined. Most people are acquainted with the concept of regression even though they do not realize it is that. On the penny scale outside of the drug store is a table reading something like this:

<table>
<thead>
<tr>
<th>HEIGHT</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'2&quot;</td>
<td>130 lbs.</td>
</tr>
<tr>
<td>5'3&quot;</td>
<td>134</td>
</tr>
<tr>
<td>5'4&quot;</td>
<td>139</td>
</tr>
<tr>
<td>5'5&quot;</td>
<td>144</td>
</tr>
<tr>
<td>5'6&quot;</td>
<td>148</td>
</tr>
<tr>
<td>5'7&quot;</td>
<td>153</td>
</tr>
<tr>
<td>5'8&quot;</td>
<td>158</td>
</tr>
<tr>
<td>5'9&quot;</td>
<td>164</td>
</tr>
<tr>
<td>5'10&quot;</td>
<td>168</td>
</tr>
<tr>
<td>5'11&quot;</td>
<td>172</td>
</tr>
<tr>
<td>6'0&quot;</td>
<td>177</td>
</tr>
<tr>
<td>6'1&quot;</td>
<td>183</td>
</tr>
<tr>
<td>6'2&quot;</td>
<td>187</td>
</tr>
</tbody>
</table>

This table shows the regression of weight on height for the human adult male population. It is useful to express this graphically in the form of a regression line, which is the heavy diagonal line in Fig. 1. There is an independent variable: height. This is arrayed along the horizontal axis. The dependent variable, weight, is perpendicular to this. The idea of independence and dependence usually, but not necessarily, denotes cause and effect. The height which a person has attained, which in turn is dependent on other factors such as age, heredity, or childhood diet, determines to a large extent what he weighs.

The independent variables are fixed. They can be selected, that is, they do not have to be taken at random. Now, in our example, there is a distribution of weights for any selected height. The distribution has a mean that is actually the average weight for all men of this height.

The data in Fig. 1 make up a scatter diagram. Plotted are the heights and weights of 17 range technicians residing in the Berkeley, California area. The finer diagonal line is the regression line computed from the 17 positions on the graph. This line represents the average weight-per-height of these individuals better than any other straight line we could draw. The line was determined by the method of least squares, which simply means that if we sum the squares of each of the 17 distances from the points up to or down to the line, that sum is smaller than the sum for any other line.

With this knowledge of the characteristics of independent and dependent variables and the meaning of regression, little imagination is needed to visualize relationships between rate of seeding and density of grass, rate of nitrogen application and yield of forage, age of ruminant and digestibility, and so on for any biological attribute that is affected by measurable environmental phenomena.

As used in the height-weight example, regression is simple, but it can be extended to six or more uses. These are: (1) predicting an unknown value of a dependent variable by interpolation or extrapolation of data; (2) estimating a difficultly obtained value of a char-
acteristic more precisely by measuring an auxiliary variate which can be found quickly and cheaply; (3) reducing the error variance of designed experiments and increasing the relative information through statistical control; (4) determining the degree of independence of a factor; (5) testing the form of a relationship, whether linear or exponential (curvilinear); and (6) determining the proportion of the increase or decrease of a dependent variable which is contributed by each of several measurable independent factors.

The object is not to go into the statistical aspects of regression, but merely to point out, with appropriate examples, the ways in which regression can be applied to range research.

Predicting Unknown Values

The man at the carnival who can guess your weight within two pounds has had much experience at that game. He sizes up his sucker, height being an important factor in formulating his prediction, and allows a four pound leeway (confidence limit). The greater his length of experience, the narrower he can make his confidence limits and still operate his concession at a good profit.

A regression equation is a mass of past experience expressed in algebraic form. It can be used to predict what will happen under certain specified conditions. The accuracy of the prediction is entirely dependent upon the breadth of the past experience. Thus, the larger the sample used in the compilation of the equation, the narrower the confidence limits may be. Confidence limits are an expression of the number of chances out of a hundred that your prediction will be right.

Using the regression lines in Fig. 1, if you know of a man who is 6 feet tall, you can guess his weight to be 178 pounds plus or minus a few pounds. The wider this leeway, the greater proportion of time you will be right. Now if you learn that he is a Berkeley range technician he is likely to be five pounds heavier, but you can’t say this with too much confidence since there were only 17 men in this sample compared to the thousands involved in the other regression line. Moreover, it is safer to predict weights on range men between the heights of LR and ALH (interpolation) than beyond those extremes (extrapolation), for reasons given later.

Another way to use the regression line is to determine what the normal or expected weight should be, then compare this with the actual. If the deviation is great, it may be well to look for reasons. Note that comparison with the sample mean weight does not yield such insight. For example, HFH is only four pounds over the group or population average weight of 171 pounds but for his height he is 28 pounds over while RAM is 21 pounds under the population average and 35 pounds under the average for his height. If the data represented 17 pigs, the farmer who owned them might use this information and select the HFH, JIM, and ALH types for breeding stock and deworm those way off the line in the lower right corner.

The divergence of two regression lines is sometimes of interest. The reasons why, among these research men, the ectomorphs are more ectomorphic and the “heavyweights” are heavier than generally found in the adult American male population may be supplied with a few more independent variables or they may have Freudian explanations. At least, in this case, height does not explain everything. Most likely, the two lines in Fig. 1 are not significantly different.

Bona fide range problems find the most common and practical use of regression to be in its predictive value. Ordinarily you predict the unknown dependent variable when the size of the independent variable is stipulated. A graph using some familiar data on the grass density-brush seedling relationship (Schultz, Launchbaugh, and Biswell 1955) will illustrate how this can be done in reverse (Fig. 2). Each increment of density of rye-grass is associated with an increase in mortality of the brush seedlings. On the average, at 38 percent density, mortality is 100 percent. Suppose we would like to maintain about one seedling per 10 square feet for browse, what should we strive for in grass density? The answer can be interpolated from the graph.

Regression Estimates

Sometimes we are interested in a measurement which is extremely
USE OF REGRESSION IN RANGE RESEARCH

43

Y = 799451 - 8705X

X = PERCENT UNGRAZED

FIGURE 3. Percent use as determined by the weight of herbage removed in relation to percent of ungrazed plants of all important grasses on the Santa Rita Experimental Range, Tucson, Arizona. The circles are observations of ungrazed plants grouped by 5 percent intervals (from Roach, 1950).

costly to obtain with any degree of accuracy. Protein analyses, for example, take a lot of time and the equipment necessary to run them is not always available. Measures of range utilization are difficult to get in terms in which a range technician or a rancher may have confidence. Use of the linear regression estimate increases the precision and decreases the cost. By measuring an auxiliary variate which can be found quickly and cheaply, we can estimate more precisely the value of a characteristic obtained with great difficulty. We do not have any interest in this auxiliary variate itself; we are interested only as it is related to the dependent variate in question.

Let us illustrate this principle with two examples. It would be well to have a quick method of estimating crude protein in range forage. Ranchers could use it to decide when supplemental feeding should be started and the amount of supplement needed. A close relationship has been found between the crude protein in forage on longleaf-pine bluestem ranges in Louisiana and the free moisture in that forage (Campbell and Cassidy 1954). Free moisture is relatively easy to determine. This relationship can be expressed in the form of a regression line. When used in conjunction with the known protein requirements of beef cattle, the moisture content of the forage will indicate the amount of protein that must be supplemented.

The other example is one of estimating perennial grass utilization on semidesert ranges by percentage of ungrazed plants (Roach 1950). After establishing the relationship between the percent of plants grazed and the percent utilization as a whole (Fig. 3), the latter value can be estimated merely by counting the number of grazed and ungrazed plants on representative samples. Roach claims that this simple method, when compared with the height-weight method, cuts field time by half and office computation time by at least three fourths; and it gives utilization estimates within 5 percent of those gotten the other way. The percentage of ungrazed plants itself is of no interest except that it helps to estimate utilization.

**Statistical Control**

Plant physiologists who are bequeathed with unlimited funds have elaborate laboratories and greenhouses where nearly every essential feature of the environment can be controlled. Thus, an experiment can be reduced to only one variable such as growth. With complete control over all factors, there should, theoretically, be no unexplained error encountered in the experimentation. Range ecologists have two strikes against them—they never are bequeathed with unlimited funds and if they were, they would fall short in controlling most factors of the outdoor environment, as the rainmakers can attest. So their research is redolent with what is called experimental error.

While he cannot control the factors, the range ecologist can often measure them and these measurements enable him to use regression for increasing the information available in his data. This is statistical control. In fact, the ecologist may be better off than the physiologist because in many cases statistical control is more desirable than experimental control. First, the actual situation is studied, not one produced artificially; secondly, a far greater range of observations can be made which broadens the foundation for inference; and finally, one learns how two quantities instead of one vary, singly and together—the factorial approach.

All this can be illustrated with some data on brush seedlings that are competing with certain species of annual grasses (Table 1). The table is excerpted from one containing many other species of grasses and legumes.
The number of brush seedlings per 100 square feet remaining alive after growing one season with these annual grasses varied with the grass species—significantly. But the number of brush seedlings that came up on these plots was not the same. Stated anthropomorphically, it is harder for cereal rye to reduce 275 seedlings to 0 than for ryegrass to reduce 125 since 125 is already much closer to 0. (Biologically, this premise may be incorrect because intraspecific competition may be greater among the 275.) Because of the uncontrollable discrepancy in numbers, the original seedlings were counted and this auxiliary variate was used to adjust the fall numbers. Column 3, after the adjustment, shows the number of seedlings which would occur in fall if the same number had started on all the plots in the spring.

Some grasses grow better than others on different sites so grass species occur at varying densities. Would these species be equally good competitors if they grew at equal densities? Again, column 3 was adjusted so that column 4 shows the numbers of seedlings that would occur if the original seedling numbers were the same and the grass densities the same. Now, two variables have been controlled. We have divested the species of extraneous factors and perhaps we can compare them strictly on the basis of some physiological character such as transpiration ratio.

The method whereby these adjustments are made is called analysis of covariance. Other examples and explanations of the method are in the range literature (Pechanc 1941; Blaisdell 1953).

Test for Independence

One cannot always be sure that one factor is dependent on another even though logic says they are, and conversely, that two factors are independent when no cause and effect are indicated. It is a matter of degree. There is a statistical test which gives you reason to say, at a given level of significance, that one variable is independent of another, or that one is dependent on the other. We can say that in the height and weight example, weight of range technicians is dependent on their height, and be right in saying that unless a 1 out of 25 chance has come about.

Table 1. Number of brush seedlings established in competition with annual grasses

<table>
<thead>
<tr>
<th>Competing Species</th>
<th>Number of Brush Seedlings per 100 Sq. Ft.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) Spring Count</td>
</tr>
<tr>
<td>Cereal Rye</td>
<td>275</td>
</tr>
<tr>
<td>Red Brome</td>
<td>233</td>
</tr>
<tr>
<td>Annual Ryegrass</td>
<td>125</td>
</tr>
<tr>
<td>Rose Clover</td>
<td>152</td>
</tr>
<tr>
<td>Soft Chess</td>
<td>163</td>
</tr>
<tr>
<td>Check</td>
<td>177</td>
</tr>
</tbody>
</table>

The statement was made earlier that a relationship of dependency may exist but it may not necessarily mean that a cause and effect relationship exists. Any relationship can be a tool for prediction or analysis purposes but the decision as to whether it is causal or not comes only after detailed re

![Figure 4. Comparison of costs of controlled burns and of wild fire suppression in northern California, 1947-1948 (from Sampson and Burcham, 1954).](image-url)
search of the biological or logical kind rather than from mathemati-
cal manipulation of data.

An example of the question of dependence can be illustrated with
data taken from an article on costs of controlled brush burning
(Sampson and Burcham 1954). Fig. 4 is reproduced from this
article. The second curve from the
top implies that if a certain size
of area is controlled burned, say,
for example, 280 acres, it will cost
about $1.20 per acre. In short, the
cost per acre is directly dependent
on size of burn. The top curve
implies that this is also true of
wildfires. However, it is much
more reasonable to assume that the
size of the wildfire is dependent on
the amount of money devoted to
its suppression. Thus, while the
curve drawn for "cost of wildfire
suppression" is not invalidated for
predictive purposes, there may be
some theoretical objections to com-
paring it with the curve for "total
cost of controlled burns."

Test for Linearity of Regression

Not always do two variables go
off in the same direction as do
height and weight. Especially in
biological phenomena, sigmoid,
bell-shaped, or cyclic relationships
are found. With growth phenom-
ena a factor may increase the
rate of growth until that factor has
reached an optimum magnitude,
and then growth decelerates.
This relationship is also described
by a regression line but a curved
instead of a straight one. A curvi-
linear regression equation is diffi-
cult to compute, but from the prac-
tical as well as the academic view-
point it is important to know
whether a relationship is curvi-
linear or actually linear.

This can be illustrated with a
hypothetical fertilizer experiment
(Fig. 5). Six rates of nitrogen
were applied to replicated plots,
0 to 500 pounds per acre in 100
pound increments. Forage produc-
tion for each plot has been plotted
on the graph. The broken line
connects the averages for the treat-
ments. Linear regression tells us
—if we allow ourselves to extrap-
olate—that if we were to apply 1000
pounds of nitrogen per acre we
could increase our yield over the
control (N) by 315 percent; curvi-
linear regression tells us that we
could expect no yield at all after
applying 700 pounds of nitrogen.
Which is right?

A statistical test indicates that
the regression in this case is quad-
ratic (one-humped). One can
readily understand that if more
nitrogen is added without adding
additional amounts of phosphorus, potassium, or water, not only will
the extra nitrogen do no good but
it will actually even have a de-
pressing effect on the yield.

This test for linearity, then, is a
useful tool for many aspects of
range research. Straight line re-
lationships are the exception rather
than the rule where growth phe-
nomena are involved.

Multiple Regression

Why are two plants not exactly
alike even when they are of the
same species and grow only 12
inches apart? A host of factors
contribute to their difference.

Among these are the number of
hours head start one had as a seed-
ling, the difference in food reserves
in the seed, amount of time one is
in the shade of the other, and so on.
Conceivably 480 or more measure-
ments of that nature would explain
quite fully the variation between
the two plants. Let us extend this
idea to a current problem in game
range management.

It would be easy to determine
the production of available deer
browse if we knew that all brush
plants produced the same amount
of forage: simply count the plants
and multiply by the weight pro-
duced per plant. Chamise plants
in the brushlands of northern Cali-
forinia grow on varied sites and
their size and vigor vary so much
that an extremely large and costly
sample must be taken to yield a
reliable answer. Nor is it a simple
matter to clip the available forage
on a single bush. It might be better
to choose a number of other char-
acteristics of the plant, preferably
some that can be measured easily
and with objectivity, find out how
much each contributes to the total
variation in browse production be-
tween plants. The few character-
istics which contribute the most
can then be used in range investi-
gations to determine browse pro-
duction.

To illustrate, the following
"independent" variables were
measured on each of 32 chamise
plants: average height (X1), num-
ber of live stems (X2), foliage
crown diameter (X3), root crown
diameter (X4), length of average
leader (X5), and length of longest
leader (X6). The dependent vari-
able was available annual growth
(Y). Many hours of calculations
resulted in this equation:

\[
Y_{\text{(predicted)}} = -366.60 - 23.35X_1 + 7.60X_2 + 1.86X_3 - 37.96X_4 + 5.15X_5 + 69.77X_6.
\]

This equation is so impressive
that it cannot be presented in
graphical form since our compre-
hension is limited to three dimen-
sions. The variables X2, X3, X5,
and X6 are associated with vigor;
any increase in either of these will increase Y. The variable \( X_1 \) is negative; any height over 4 feet is superfluous for deer, since they browse only to that height. Indeed, shorter plants provide a top surface in addition to the sides where browse is available. The large and negative \( X_4 \) is a function of age. When all other factors are held constant, the plant with the larger root crown is the older, more decadent, and less productive. This fact was not readily discernible from routine methods of analyzing the data, e.g., correlation.

The six measurements account for only 38 percent of the variability. Most of the other 62 percent lies in the remaining 473 characteristics which could be measured and analyzed simultaneously with standard IBM machines.

It is not difficult to see how both fundamental and practical range problems can be clarified by the approach of multiple regression. The multitude of factors contributing to the processes which we call competition, soil formation, or plant succession, and the many forces and conditions that influence animal production, plant vigor, or the economics of range management are still too much a matter of subjective debate. If these factors are real, they are measurable; and, if measurable, they are interpretable. Multiple regression affords a good way to interpret them.

To conclude with the same whimsical example with which we began, multiple regression would permit us to predict with great accuracy the weights of Berkeley range technicians. For, besides height, their weights are dependent on age, amount of field work or desk work they do, size of their expense accounts, number of milk shakes they imbibe, and so on, until all the variation in their weights is accounted for.

Six ways in which regression can be used in range research have been discussed. There are others that may well be applicable in this field although not widely used. Numerous textbooks explain how to do the calculations but that has not been in the scope of this paper.

**LITERATURE CITED**


