**PROCESS OF OBTAINING STATA SAMPLE**

for obtaining a stratified sample.

The president of DePaul University

wants to conduct a survey to determine the community's

opinion regarding campus safety.

The president divides the DePaul community

In this example, we'll go over the procedure

into three groups-- resident students,

nonresident or commuting students, and staff-- including

faculty-- so that he can obtain a stratified sample.

Suppose there are 6,204 resident students, 13,304

nonresident students, and 2,401 staff

members for a total of 21,909 individuals in the population.

The resident students make up approximately 28%

of the population, the non-resident students make up

approximately 61% of the population,

and staff makes up approximately 11% of the population.

We want a stratified sample to represent this percentage

breakdown so that it looks just like the population, only

smaller.

The president wants to obtain a sample of size 100

with the number of individual selected

from each stratum weighted by the population size.

So we'll multiply the desired percentage

by the sample size of 100.

That would provide us with 28 resident students

in our sample, 61 non-resident students, and 11 staff members.

We simply multiply the percentage

by the sample size for each group.

To obtain the stratified sample, we'll

construct a simple random sample within each group.

We'll select 28 of the 6,204 resident students,

61 out of the 13,304 nonresident students,

and 11 out of the 2,401 staff members.

If you're going to do this with technology-- and we

will-- be careful not to use the same seed for all the groups

in the stratified sample because we

want the simple random samples within each stratum

to be independent of each other.

**So I'm going to use StatCrunch to select the random numbers.**

**I'm going to begin with the resident students.**

**We needed 28 students out of 6,204 students.**

**To select those numbers, press on Data, Simulate, Discrete**

**Uniform.**

**Again, I need 28 values, but I'm going to ask for 5 more**

**than that in case there are duplicates.**

**So 33 rows, 1 column, the minimum number**

**is 1, the maximum number 6,204, and I'm**

**going to use the dynamic seed.**

**Press compute.**

**And I'm going to get rid of this header**

**and replace it with Resident.**

**Now I will only use the first 28 of those values that**

**do not repeat.**

**I'll check them for repeats in a moment.**

Let's go ahead and generate the 61 non-resident students.

Data, Simulate, Discrete Uniform.

I needed 61.

I'm going to ask for 66, again, in case there are repeats.

The minimum value is 1.

The maximum value is 13,304.

I'll use another dynamic seed.

Press Compute.

And I'm going to change that header to Non-resident.

I'll check that for repeats in a bit as well.

We'll only use the first 61 that are non-repeating.

Now to generate the 11 staff.

Data, Simulate, Discrete Uniform.

I need 11, so I'll ask for 16 in one column.

The minimum value is 1.

The maximum value, 2,401.

I'm going to save myself some grief and where it says Prefix,

I'm going to put the column name.

This one's going to be called Staff,

and I won't have to change it later.

Again, I'll use a dynamic seed, let the computer select it.

Press Compute, and that column, Staff 1-- I'll just get rid

of the 1-- has been added.

And I'm only going to use the first 11 that do not repeat.

It looks like the first 11 there are clean.

OK, here are the 33 random numbers

we generated for resident students.

We needed 28, which means that we're supposed to end here.

And we've scoured the first 28 looking for any repeats,

and in this case, there are none,

so we can discard the 5 extra values that I requested.

And there are the 28 resident students

that will be included in the study.

Now your 28 and my 28 will be different

because they're selected at random.

Let's look at the non-resident students.

Selected 66 values.

I only need the first 61.

And in the first 61, I go looking for repeats,

and there actually are repeats.

12,731 was elected twice, so I can discard the second one,

and my first 61 values-- that was 61.

Add one more and discard the rest like this.

Now again, your 61 will be different than my 61

because we selected them randomly.

Finally, let's take a look at the staff.

Here are the 16 values I selected.

I only need 11 of them.

So as I look at those first 11 staff members,

there are no repeats, so I can get rid of the extra 5

I asked for.

And there are the 11 staff members

that will be included in the study.

So we used the first 28 non-repeating values

that we found for the resident, the first 61

non-repeating values for the non-residents,

and the first 11 non-repeating values for the staff.

You could have done this using Microsoft Excel,

some other software package.

You also could have done this with a TI calculator.

Just keep going until you get your 28 unique residents, 61

unique non-residents, and so on.

An advantage of stratified sampling

over simple random sampling is that the researcher

can determine characteristics within each strata.

This allows an analysis to be performed on each stratum

to see if any significant differences among them exist.

For example, we could analyze the data obtained in example 1

to see if there's a difference in the opinions of students

versus staff.