

ANOVA: Explanatory variable with 2 levels:

The CLASS that identifies the categorical explanatory variable in this case is breastcancerper100th_NV01 has two levels and the values of this variable are 0 and 1.

Of the 213 observations 198 were included in the analysis.

PROC ANOVA displays an analysis of variance table for the response variable from the model statement. In this case, the response (dependent) variable is urbanrate.

The calculated F Value in the output is 24.86 and the significance probability (P value) associated with the F Value is 0.0001, less than the P Value cut point of 0.05.

The means table shows that countries with a percentage equal or lower than 50 percent of cases of breast cancer in women, indicated by the value 0, live in urban areas with a means of 52.42 and that those countries with a percentage equal or greater than 51 percent of cases of breast cancer in women, indicated by the value 1, live in urban areas with a means of 72.39.

Since the P Value is less than 0.05, in this case 0.0001, we can reject the null hypothesis and state that these means are not statistically equal.

In other words, there is an association between the presence of breast cancer in women who live in urban areas.

By normal scientific standards this is an adequate certainty to reject the null hypothesis and say that there is an association.

The P Value is less than 0.05 and we are confident to say that there is significant association between the presence of breast cancer in women who live in urban areas.

```
LIBNAME mydata "/courses/dl406ae5ba27fe300 " access=readonly;
DATA new; set mydata.gapminder;
LABEL breastcancerper100th="2002 breast cancer new cases per 100,000 female - Number of new cases of breast cancer i
      urbanrate="2008 urban population (% of total)(calculated using World Bank population estimates and urban ratic
      breastcancerper100th_NV01="When the rate of breast cancer per 100 women per country is equal or greater than 5
/*Set appropriate missing data as needed*/
IF breastcancerper100th=0 THEN breastcancerper100th_NV01=.;
IF breastcancerper100th GE 51 THEN breastcancerper100th_NV01=1;
IF breastcancerper100th LE 50 THEN breastcancerper100th_NV01=0;
PROC SORT; by COUNTRY;
PROC ANOVA; CLASS breastcancerper100th_NV01;
MODEL urbanrate=breastcancerper100th_NV01;
MEANS breastcancerper100th_NV01;
RUN;
```

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
breastcancerper100th_NV01	2	0 1

Number of Observations Read	213
Number of Observations Used	198

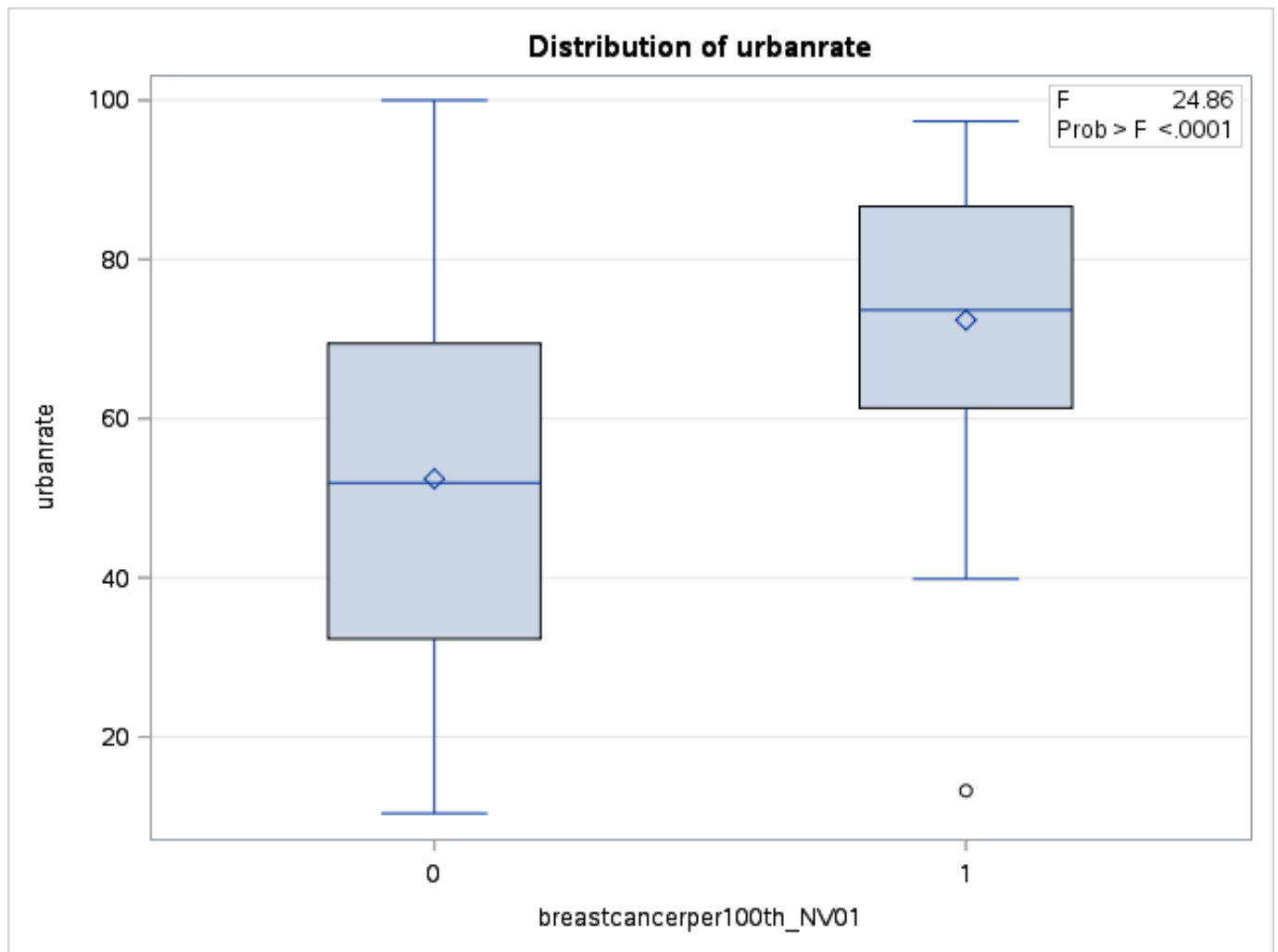
The ANOVA Procedure

Dependent Variable: urbanrate 2008 urban population (% of total)(calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects)

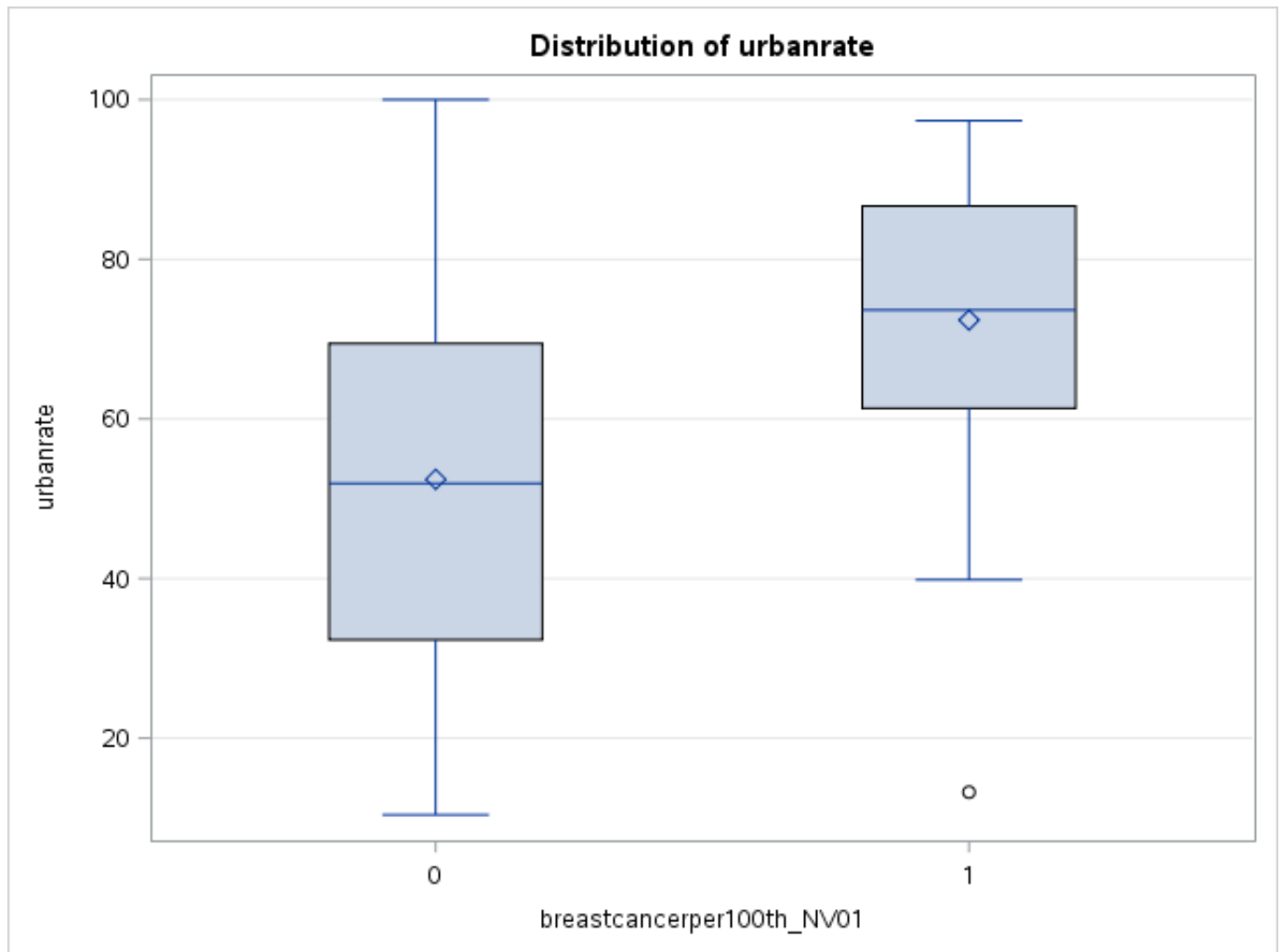
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	12485.2595	12485.2595	24.86	<.0001
Error	196	98423.5746	502.1611		
Corrected Total	197	110908.8342			

R-Square	Coeff Var	Root MSE	urbanrate Mean
0.112572	39.76459	22.40895	56.35404

Source	DF	Anova SS	Mean Square	F Value	Pr > F
breastcancerper100th	1	12485.25953	12485.25953	24.86	<.0001



The ANOVA Procedure



Level of breastcancerper100th_NV01	N	urbanrate	
		Mean	Std Dev
0	159	52.4212579	23.3585345
1	39	72.3876923	17.9292731

ANOVA: Explanatory variables with more than 2 levels (3 in this case):

Based on the PROC ANOVA results for the greater than 2-level explanatory variable.

The F Statistics is 20.52 with an associated P Value of less than 0.0001. We can safely reject the null hypothesis and that there is an association between the presence of breast cancer in women who live in urban areas.

Since the F-test and the p-value do not provide insight into why the null hypothesis can be rejected, because there are multiple levels to the categorical explanatory variable, at this point we can principally conclude that none of the means are equal. We can only eye ball each mean in the means table and take a guess as to which pairs are significantly different from one another, for example: countries with a percentage equal or lower than 33 percent of cases of breast cancer in women, indicated by the value 1, live in urban areas with a means of 49.50 and that those countries with a percentage equal or greater than 35 percent of cases of breast cancer in women, indicated by the value 3, live in urban areas with a means of 70.55.

```
LIBNAME mydata "/courses/dl406ae5ba27fe300 " access=readonly;
DATA new; set mydata.gapminder;
LABEL breastcancerper100th="2002 breast cancer new cases per 100,000 female - Number of new cases of breast cancer i
      urbanrate="2008 urban population (% of total)(calculated using World Bank population estimates and urban ratic
      breastcancerper100th_NV01="When the rate of breast cancer per 100 women per country is equal or greater than 5
/*Set appropriate missing data as needed*/
IF breastcancerper100th=0 THEN breastcancerper100th_NV01=.;
IF breastcancerper100th LE 33 THEN breastcancerper100th_NV01=1;
IF breastcancerper100th GT 34 THEN breastcancerper100th_NV01=2;
IF breastcancerper100th GE 35 THEN breastcancerper100th_NV01=3;
PROC SORT; by COUNTRY;
PROC ANOVA; CLASS breastcancerper100th_NV01;
MODEL urbanrate=breastcancerper100th_NV01;
MEANS breastcancerper100th_NV01;
RUN;
```

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
breastcancerper100th_NV01	3	1 2 3

Number of Observations Read	213
Number of Observations Used	201

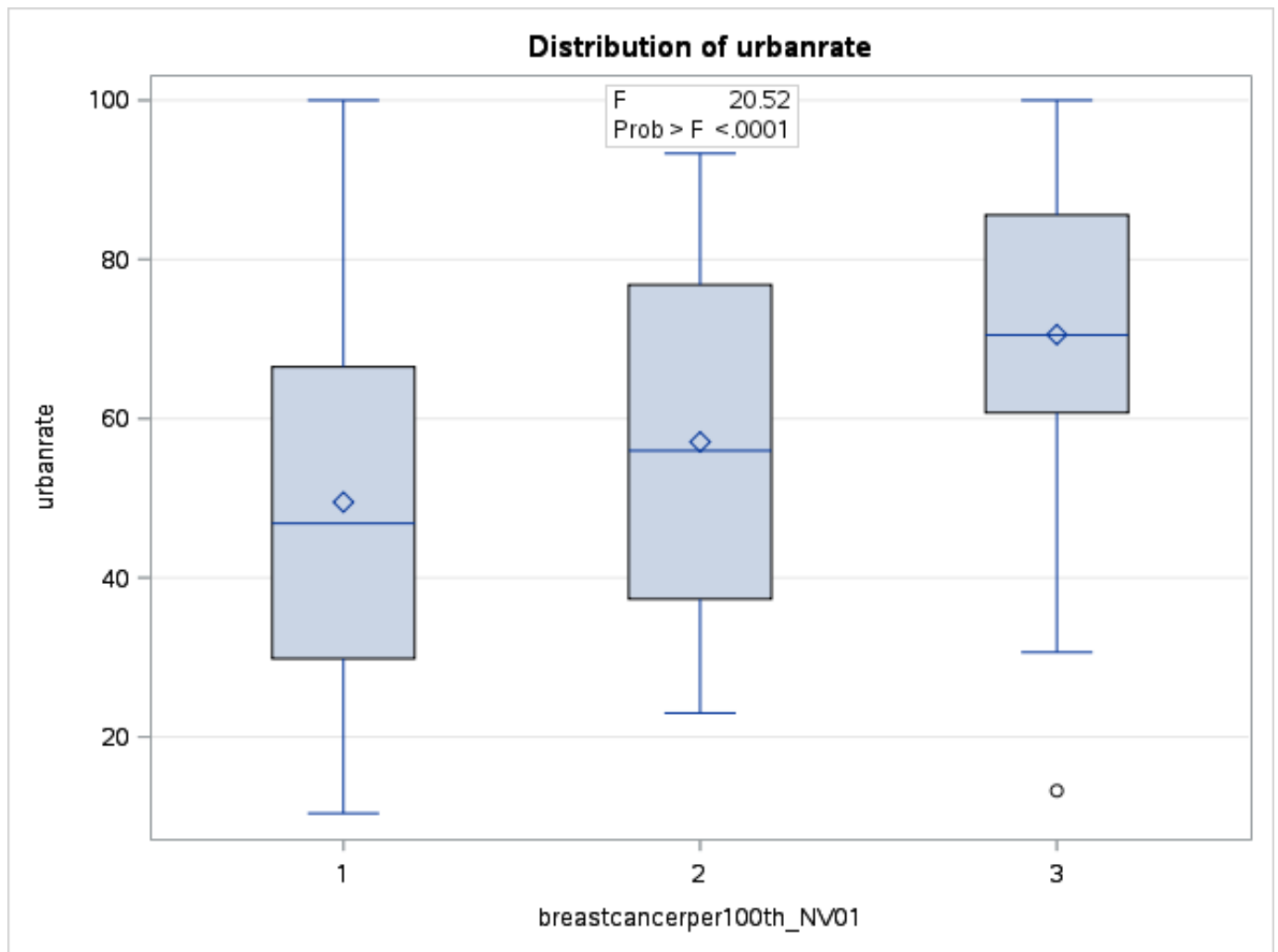
The ANOVA Procedure

Dependent Variable: urbanrate 2008 urban population (% of total)(calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects)

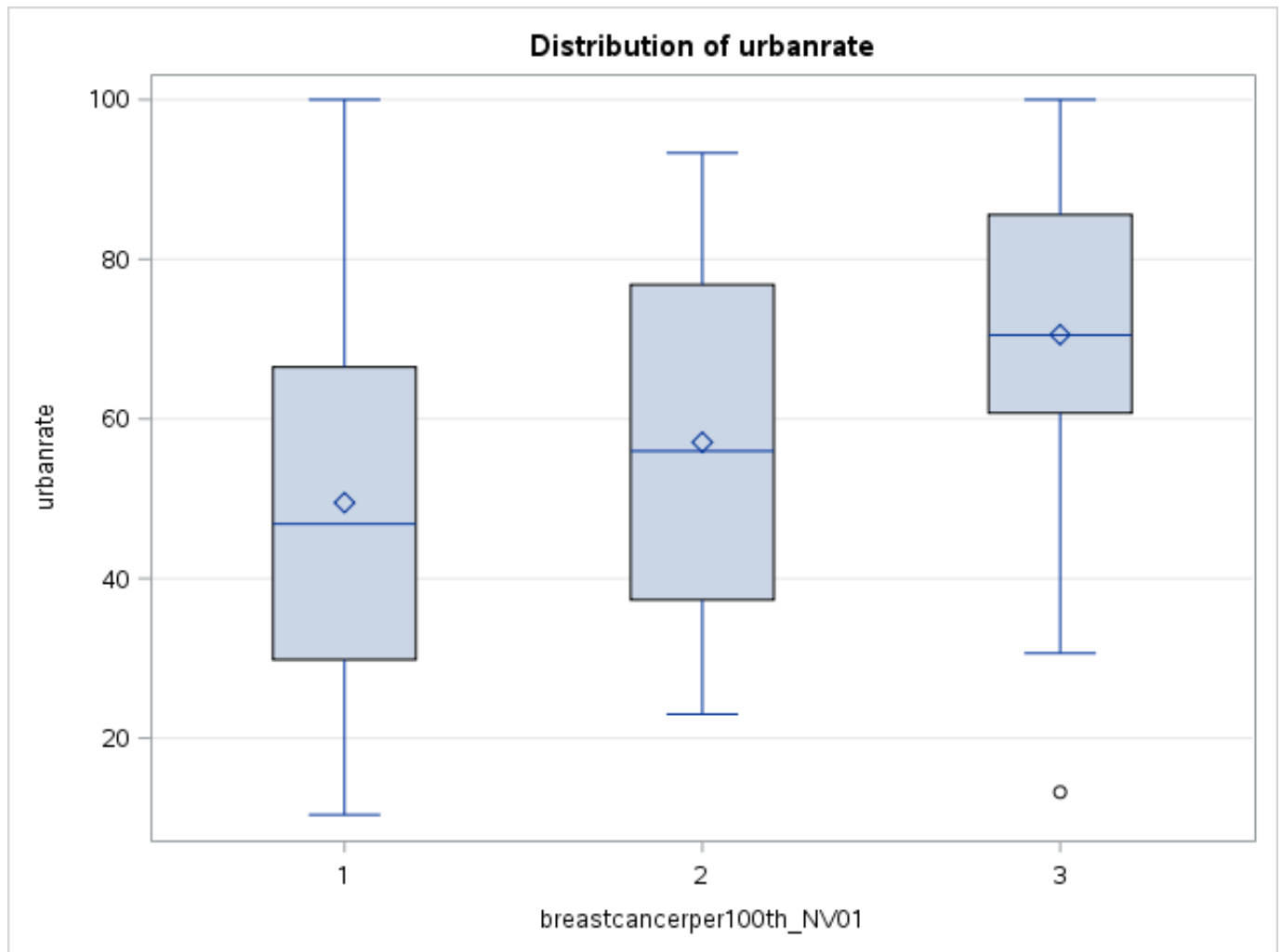
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	19453.6800	9726.8400	20.52	<.0001
Error	198	93872.1744	474.1019		
Corrected Total	200	113325.8544			

R-Square	Coeff Var	Root MSE	urbanrate Mean
0.171661	38.49561	21.77388	56.56199

Source	DF	Anova SS	Mean Square	F Value	Pr > F
breastcancerper100th	2	19453.67996	9726.83998	20.52	<.0001



The ANOVA Procedure



Level of breastcancerper100th_NV01	N	urbanrate	
		Mean	Std Dev
1	131	49.4984733	23.3538740
2	4	57.0650000	28.9525192
3	66	70.5515152	17.7395617

DUNCAN TEST on the ANOVA results: Explanatory variables with more than 2 levels (3 in this case):

In the case that the explanatory variable has more than two groups, a significant ANOVA does not tell which groups are different from the others.

To determine which groups are different from the others and to prevent excessive Type 1 error, a Duncan post hoc test was performed.

In order to examine the association between the presence of breast cancer in women and the fact that these live in urban areas,

The Proc ANOVA results with Duncan Post Hoc Test show that the top of the results looks the same as in the original test where the F value or F statistic is 20.52 and it is significant at the $P < 0.0001$ level.

However, the table displaying the results of the paired comparisons conducted by the Duncan Multiple Range Test show that: group 2 (countries with greater than 33 percent of women with breast cancer) and group 3 (countries with greater than 35 percent of women with breast cancer) are not significantly different because they both have the capital letter A and groups 1 and 2 are also not significantly different because each has a B.

However, group 1, which indicates the range of countries with less and equal than 33 percent of women with breast cancer, has a significant difference based on the urban living than group 3.

```
LIBNAME mydata "/courses/dl406ae5ba27fe300 " access=readonly;
DATA new; set mydata.gapminder;
LABEL breastcancerper100th="2002 breast cancer new cases per 100,000 female - Number of new cases of breast cancer i
      urbanrate="2008 urban population (% of total)(calculated using World Bank population estimates and urban ratic
      breastcancerper100th_NV01="When the rate of breast cancer per 100 women per country is equal or greater than 5
/*Set appropriate missing data as needed*/
IF breastcancerper100th=0 THEN breastcancerper100th_NV01=.;
IF breastcancerper100th LE 33 THEN breastcancerper100th_NV01=1;
IF breastcancerper100th GT 34 THEN breastcancerper100th_NV01=2;
IF breastcancerper100th GE 35 THEN breastcancerper100th_NV01=3;
PROC SORT; by COUNTRY;
PROC ANOVA; CLASS breastcancerper100th_NV01;
MODEL urbanrate=breastcancerper100th_NV01;
MEANS breastcancerper100th_NV01/DUNCAN;
RUN;
```

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
breastcancerper100th_NV01	3	1 2 3

Number of Observations Read	213
Number of Observations Used	201

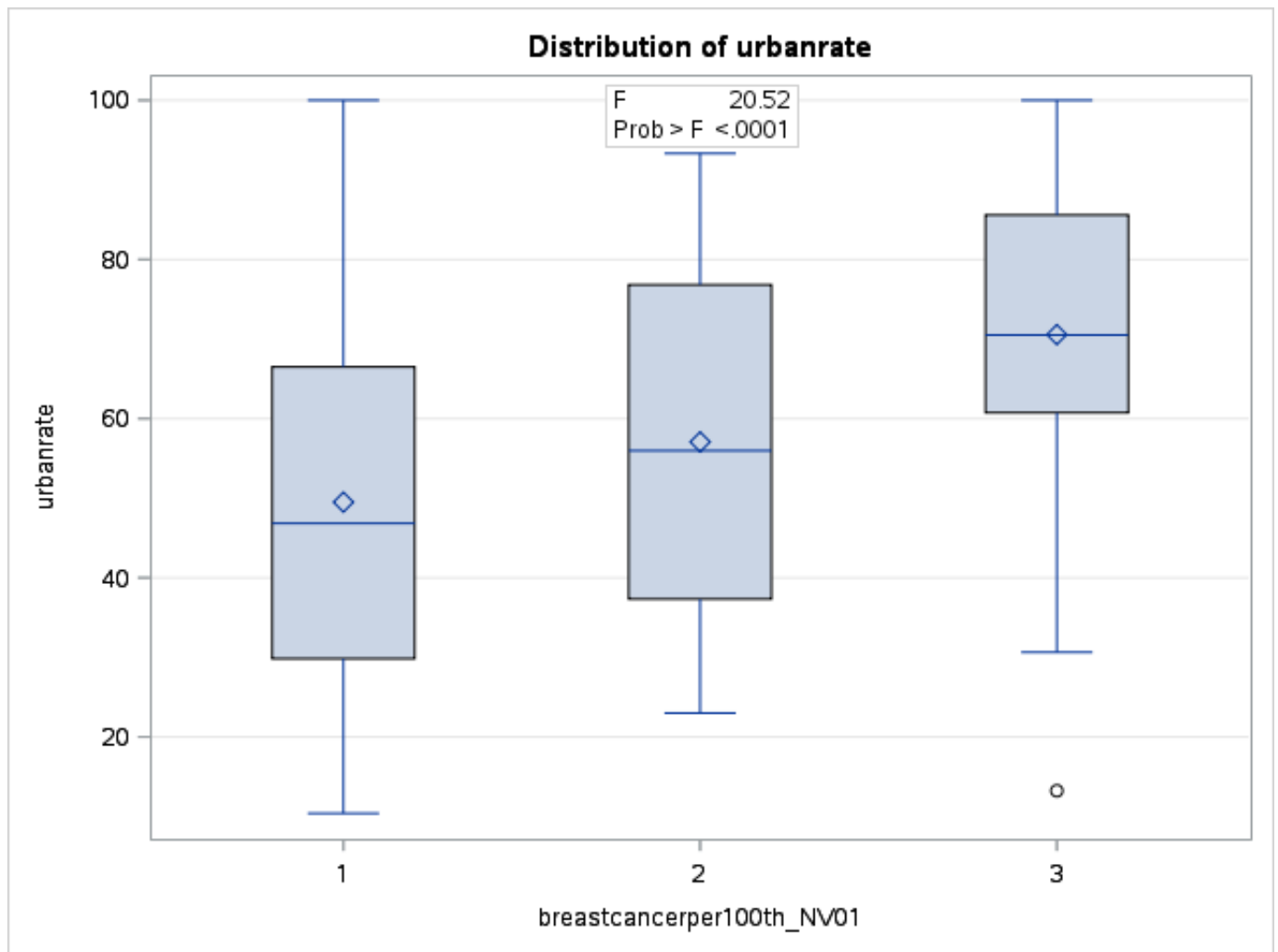
The ANOVA Procedure

Dependent Variable: urbanrate 2008 urban population (% of total)(calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects)

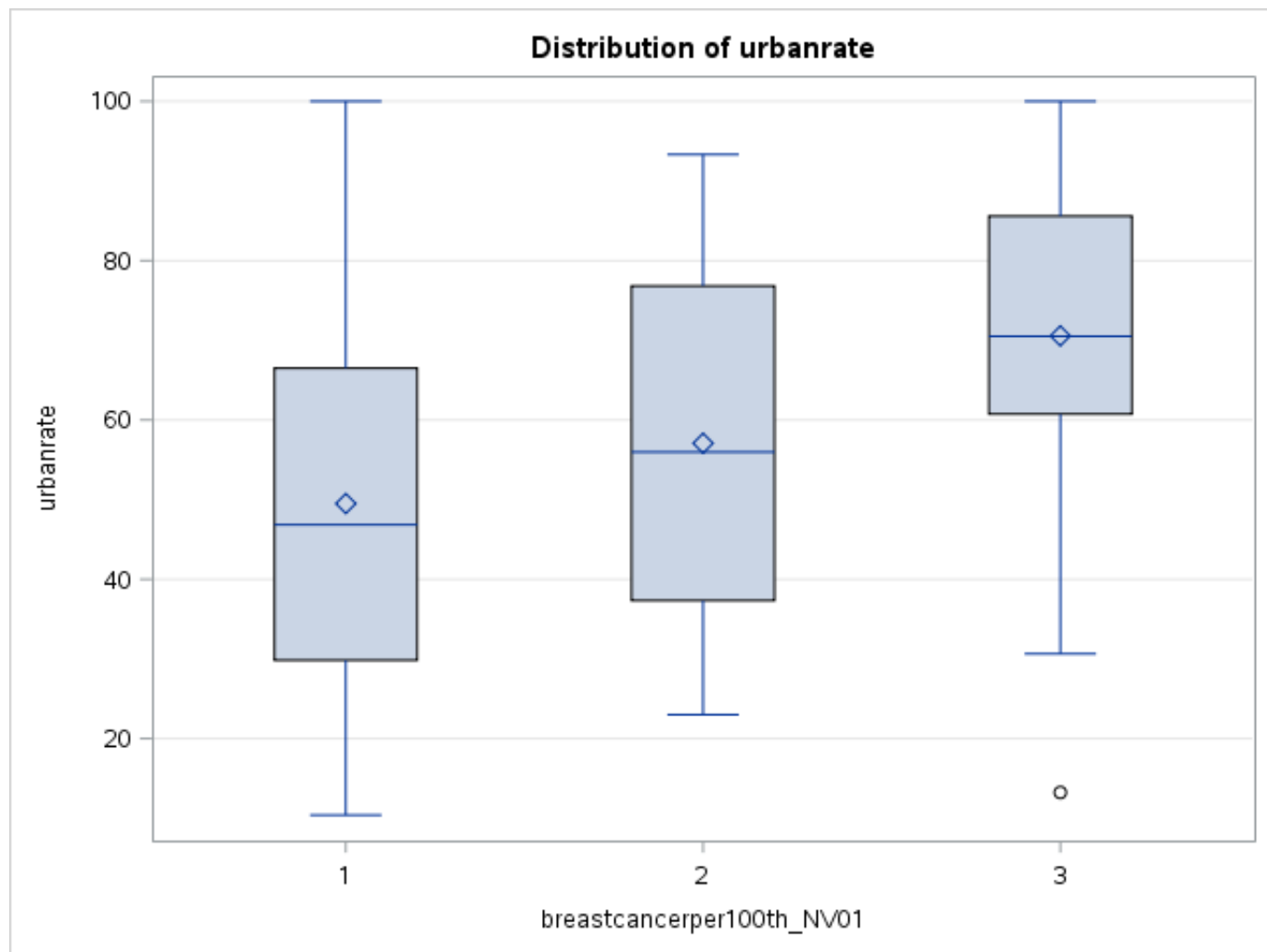
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	19453.6800	9726.8400	20.52	<.0001
Error	198	93872.1744	474.1019		
Corrected Total	200	113325.8544			

R-Square	Coeff Var	Root MSE	urbanrate Mean
0.171661	38.49561	21.77388	56.56199

Source	DF	Anova SS	Mean Square	F Value	Pr > F
breastcancerper100th	2	19453.67996	9726.83998	20.52	<.0001



The ANOVA Procedure



The ANOVA Procedure
Duncan's Multiple Range Test for urbanrate

Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	198
Error Mean Square	474.1019
Harmonic Mean of Cell Sizes	10.99767

Note: Cell sizes are not equal.

Number of Means	2	3
Critical Range	18.31	19.27

Means with the same letter are not significantly different.				
Duncan Grouping		Mean	N	breastcancerper100th_NV01
	A	70.552	66	3
	A			
B	A	57.065	4	2
B				
B		49.498	131	1

