

Contents

The following files are found in the **Instructor Files** folder. Exercises can be used for homework or in-class assignments. Corresponding Excel files can be found in the Data folder and are noted in the descriptions below. Student versions of the same files can be found in the **Student Files** folder. We’ve also included descriptions of the files found in the **Student Resources** folder below the lesson information. Tipsheets in that folder are meant to provide supplementary information for students interested in further pursuing data and story ideas presented in exercises. Additional tipsheets can be found in the IRE [tipsheet database](#).

Introduction	
0b_Welcome	Introduction to NICAR Courses
0c_InstructorNotes	Detailed information about this NICAR course, tips for teaching Excel and additional resources
Excel Basics	
1a_IntroExcel	Teaching guide - Introduction of basic concepts, exercise on creating spreadsheet. Skills covered: entering data, cursors, formatting, sum, average, median, change, percent change
1b_Reviewing_the_basics_&_sorting	Teaching guide - Review of basic concepts and sorting. Citybudget.xls
1c_Exercise1	Practice using change, percent change and sorting to examine population of major urban areas throughout the world. Urbanpop.xls

1d_Exercise2	Practice using sum, change and percent change to look at university tuition. Instructors can download their schools' information or they can use our sample data: UniversityOfMissouri.xls
1e_Exercise3	Practice using sum, change, percent change and percent of total to examine university student fees. UniversityFees.xls
Rates and Ratios	
2a_Rates_and_Ratios	Teaching guide - Introduction to rates and ratios. CA_Discipline.xls and RI_Student_Teacher.xls
2b_Exercise4	Practice with ratios using data that reporters in Chicago used following a breaking news story. Transit.xls
2c_Exercise5	Practice using rates to analyze crime data. Crime2013.xls
2d_Exercise6	Practice using rates to analyze school capacity. Capacity.xls
Filtering and PivotTables	
3a_Filtering_and_PivotTables	Teaching guide - Introduction to filtering, grouping and PivotTables. WorldBank.xls and MLB2014.xls
3b_Exercise7	Practice filtering and PivotTables by analyzing lottery ticket winnings. Lotterywinners.xls
3c_Exercise8	Practice filtering and PivotTables by analyzing campus incident data. CampusIncidents.xls
3d_Exercise9	Practice sum, sort, change, percent change, filtering and PivotTables with university employee salary data. UniversitySalaries.xls
Importing	
4a_Importing	Teaching guide - How to get data from text, PDF and the Web into Excel. Debtcsv.txt, Debtfix.txt and Madoff.pdf

Student Resources	
Formulas	A cheat sheet of Excel formulas covered in the course
3920	IRE Tipsheet - Why does college cost so much? (Exercises 2 and 3)
CampusCoverage2014	IRE Tipsheet - Student loans, debt and aid (Exercise 2)
Measuring_Up	IRE Tipsheet - Finding and using data to track your campus' performance (Exercises 2 and 3)
DelvingintoCrime	From the 2013 IRE Conference blog, highlights of FBI data and its flaws by reporters Ben Poston of the Los Angeles Times and Steve Thompson of The Dallas Morning News. (Exercise 5)
3373	IRE Tipsheet - Poston and Thompson offer "a cookbook" for investigating police crime reporting and statistics. (Exercise 5)
3313	IRE Tipsheet - Hidden crimes: UCR data, and what's not there. (Exercise 5)
4234	IRE Tipsheet – Campus coverage: Sexual assault and other crimes on campus (Exercise 8)
T1656	IRE Tipsheet - Untold Stories on Campus (and how to find them) (Exercise 8)

Practice: Finding trends in lottery data

Most states have a lottery system. Reporters can request data on lottery winnings or ticket sales to mine for stories and practice data analysis skills.

In this exercise you'll work through lottery winnings for a county in New York. Each record, or row, represents one winning ticket of \$1,000 or more. Please keep in mind that this is practice data and should not be used for a story. If you're interested in looking at lottery data contact your state's [gaming commission](#). IRE members can listen to "[Getting lucky: Uncovering those gaming the system](#)" from the 2014 IRE Conference in San Francisco for more tips on investigating gaming, located on our website [ire.org](#).

***Teaching tip:** We recommend using this as an in-class exercise. Break the class up into groups and have them work through the data and analysis as a team. After 20-30 minutes ask each team to pitch a story idea based on their analysis. Have them explain the steps they took in Excel to come up with the idea. If there isn't time in class or you prefer having them work on their own, we've included some questions as a guide.*

GETTING STARTED

Open up the file Lotterywinners.xlsx. Before you go any further, make sure you make a copy of the data and do a four corners check. Look at the data and see what columns you have.

You should have 7,174 records of lottery winnings. Take a look at what the data include:

Selling Retailer	Name of the store that sold the winning ticket
Business Address	Address of the store that sold the winning ticket
City	City of the store that sold the winning ticket
Date Claimed	The date the winnings were claimed (not necessarily the "win" date); dates range from November 1996 through 2007
Game	Name of the game played
PrizeAmount	Amount of the prize (note: \$1,000 is the smallest amount included in this dataset)
Primary Claimant's Name	Name of the person who claimed the winning ticket
Claimant's City	City of the person who claimed the winning ticket
State	State of the person who claimed the winning ticket

QUESTIONS

It's time to test your skills. Work through the questions below. Answers can be found on the next page.

- 1) What was the single highest amount won? What was the game? Who claimed the ticket?
- 2) Which game had the most total winnings? (Find the answer for total times won and for total amount.)
- 3) Which retailer sold the most winning tickets?
- 4) Why can't you say that this retailer is the luckiest store?
- 5) Did anyone win more than once? If so, who won the most times and for how much in total winnings?

Rates and ratios in Excel

Use this as a guide to introduce rates and ratios to your students then follow up with an in-class exercise. We recommend Exercise 5 – crime rates. From there, have your students practice with Exercise 4 on their own either in class or for homework.

INTRODUCTION

In the first two lessons we covered a lot of ground. We learned how to navigate Excel's interface and enter formulas. We learned how to sort data. Additionally, we heard about some best practices for working with spreadsheets, such as saving an original copy of our data files and performing a four-corners check.

With many of our formulas, we used Excel to compare one number to another. As journalists, it's our responsibility to make sense out of the numbers for our audience. Numbers standing alone are meaningless. So it's up to us to compare our numbers to others and provide context.

In this lesson, we will continue our exploration of how we can compare numbers by learning how to calculate rates and ratios.

NUMBER COMPARISONS

Before we dive into rates and ratios, let's review the comparisons that we've already made.

1. New number \rightarrow Old number. This is the amount, or raw, change. In Excel we subtracted the old number from the new: $= \text{New} - \text{old}$.
2. Change \rightarrow Old number. This is the percent change. In Excel we divided the raw, or amount change, by the old number: $= \text{Change} / \text{Old}$. The easy way to remember this is N-O-O or (New – Old, divided by Old).
3. Individual amount \rightarrow Total amount. This is the percent of total. It's the numeric equivalent of a pie chart. In Excel we divided the individual amount by the total amount. $= \text{Individual} / \text{Total}$.
4. All numbers to themselves. These are the averages – the mean and median. We don't typically think of the averages as number comparisons, but they really are. Each one compares a set of numbers and generates one number that best speaks for the data set. Statisticians call these calculations central tendencies. In Excel we calculate the mean with the $=\text{AVERAGE}()$ function and the median with the $=\text{MEDIAN}()$ function.

Teaching tip: Take this as an opportunity to review the number comparison concepts that you covered in the earlier lessons. Jot down the comparisons on a whiteboard.

Now we are going to introduce two new number comparisons.

1. Individual → Total population. This is the **rate**. It tells us how frequently something occurs. In Excel we will divide the individual number by the total population. Then we will multiply that by a standard number to express the rate in terms of a consistent number of people. = Individual number/population * x.
2. Item 1 → Item 2. This generates a **ratio**, which is a number that expresses the proportional relationship between two things. In Excel we will divide the first item by the one we comparing it to. = Item 1/Item 2.

RATES

For journalists, learning how to compute rates is vital. That's because rates help us get closer to the truth and may even help us uncover stories that otherwise might get buried.

Here's an example, some journalists have called Chicago the most dangerous city in the United States because FBI Uniform Crime Report data say it had the greatest number of murders of any U.S. metropolitan area. One news account even called it the "murder capital" in a blog post headline. Other big cities, such as New York and Los Angeles have spots near Chicago at the top of the list.

However, focusing on those big numbers is misleading because it fails to account for population. Chicago, L.A. and New York also have the three largest metro area populations. We'd expect them to have the greatest number of homicides, or most anything else. It's more newsworthy when a small city has a high number of murders, or a big one has a small number.

When journalists took population into account, they found that Flint, Mich., was the most dangerous metro area in the United States in 2012, when it came to murders. Detroit came in second place.

In another example, The New York Times reported that that small towns like Rehoboth Beach, Del., had become gay enclaves. The Times' report relied on 2010 census data that went beyond raw numbers and took population into account.

CALCULATING RATES

To calculate a rate we take our number and divide it by the population. This gives us the per capita or per person rate. (In Latin, per capita means per head.)

Unfortunately, per capita numbers are often fractions. When we calculate them in spreadsheets we will often get numbers with many decimal places. So, we usually then multiply the per capita

rate by a standard number to generate a number that's more meaningful, that reflects a defined number in the population. The FBI uses 100,000 as its standard multiplier in the Uniform Crime Reports. So we will see crime rates per 100,000 people. Likewise, the incidence and prevalence of diseases is often reported per 100,000 people.

If we are unsure what the multiplier should be, there are a few ways we might be able to dig up that information. First, contact the agency that created the data. Second, look for academic or government studies using the data. If those avenues fail, we could create our own multiplier by examining the populations and picking an increment that makes sense. For example, with school enrollment data whose populations is in the hundreds pick 100.

Teaching tip: Note that the population used in the rates does not need to be the number of people. It could be some other unit, such as households or vehicles.

Now we are ready to calculate rates. Make a copy of the CA_discipline.xlsx spreadsheet and open it. The spreadsheet contains information about California public school discipline summarized by county for the 2012-2013 school year.

County	Enrollment	Suspended	Expelled
Riverside	457537	25326	984
San Bernardino	445015	26590	846
Fresno	209936	14066	699
San Diego	541115	22091	684
Los Angeles	1644601	58253	578
Kern	190425	16127	509
Orange	524396	16629	387
Stanislaus	112554	9816	296
San Joaquin	149544	13902	285
Sacramento	254000	18001	228
Santa Clara	288260	9127	226
Kings	31295	2123	191
Alameda	231631	10079	185
Ventura	149534	6249	164
Butte	33483	2881	152
Contra Costa	180925	11668	140
Sonoma	74531	4368	133
San Mateo	98610	3946	131
Tulare	105398	6085	127
Madera	32807	2977	112
Santa Cruz	41719	743	97
Monterey	77692	4904	96
Sutter	23582	1473	85
San Luis Obispo	36663	1975	84
Yuba	15435	1422	78
Solano	68937	6512	76
Yavapai	45660	4460	68

Note that we have four columns in the spreadsheet. Column A lists county, B the enrollment, C the number of students suspended and D the number expelled.

Before we calculate rates, let's sort our sheet to answer two questions: Which county had the greatest number of students who were suspended? Which county had the greatest number of students who were expelled?

Select the data table as we did in the last lesson, then sort from largest to smallest.