

Introduction to Art of Stat 2022-23

Introduction

www.artofstat.com

Table of Contents

	Page
1. Welcome	2
2. <u>Illustration:</u> Describe a Sample of Categorical Data.....	4
3. <u>Illustration:</u> Describe a Sample of Continuous Data	11

Before you begin:

Right click to download icu_540.xlsx from the BIOSTATS 540 course website:

RIGHT click to download **icu_540.xlsx**:

https://people.umass.edu/biep540w/datasets/icu_540.xlsx

Place this excel file where you can access it easily; I recommend your desktop

Launch Excel. At the bottom you will see 3 tabs:

tab	Name	Description
1	Sheet 1	7 columns of data
2	Sheet 2	1 column of data
3	Coding manual	Data dictionary

1. Welcome

www.artofstat.com is a nice “go to” for the newcomer to biostatistics who might not be quite ready to jump into a statistical software package, much less doing one’s own programming! This site was developed and continues to be maintained by authors Agresti A, Franklin C, and Klingenberg B who authored an excellent introductory statistics text:

Agresti A, Franklin C, and Klingenberg B.
Statistics: The Art and Science of Learning From Data, 4th Edition
2017
Pearson
ISBN 13: 9780321997838

How to Launch.

From your favorite browser, launch www.artofstat.com

From the welcome page, at RIGHT, click anywhere inside the picture under the title **Online Web Apps**

The screenshot shows the homepage of the Art of Stat website. The navigation bar at the top includes links for Home, Web Apps, Mobile Apps, Datasets, R Code, and More. The main heading is 'THE ART & SCIENCE OF LEARNING FROM DATA'. Below this, there are three main sections: Textbook, Mobile Apps for your Phone, and Online Web Apps. The Textbook section features the cover of the 5th edition of 'STATISTICS: The Art and Science of Learning from Data' by Agresti, Franklin, and Klingenberg. The Mobile Apps section shows icons for 'Art of Stat Explore Data', 'Art of Stat Distributions', and 'Art of Stat Inference', along with download links for the App Store and Google Play. The Online Web Apps section displays a screenshot of an interactive app interface for 'Inference for a Population Mean'. A red arrow points from the 'Online Web Apps' title to the app interface screenshot. At the bottom of the Online Web Apps section, there is a link to 'Go to Web Apps'.

Home Web Apps Mobile Apps THE ART & SCIENCE OF LEARNING FROM DATA Datasets R Code More

Introductory Statistics Textbook & Apps

Textbook

Statistics: The Art and Science of Learning from Data, 5th edition. By Agresti, Franklin, Klingenberg. An introductory statistics textbook for a one or two-semester course.

Go to Publisher's Website >

ofstat.com/web-apps

Mobile Apps for your Phone

Art of Stat Explore Data Art of Stat Distributions Art of Stat Inference

Download on the App Store GET IT ON Google Play

New mobile apps for exploring and working with a variety of distributions and carrying out inference with your own data, right on your phone. (Explore Data will be released soon.) Download it for [iOS](#) or [Android](#).

Online Web Apps

Online interactive apps for illustrating & learning statistics. Explore statistical concepts (Coverage, Central Limit Theorem) or carry out data analysis (descriptive & inferential) right in your browser.

Go to Web Apps >

Look around

Clicking on **Online Web Apps** brings you to a listing of several apps, organized into categories. Trust me. You will find them to be very handy in BIOSTATS 540!

Exploratory Data Analysis

Explore Categorical Data

Explore Quantitative Data

Time Series Plots

Good for producing summary statistics and some basic graphs (BIOSTATS 540 Units 1, 2)

Association, Correlation & Regression

Association Between Two Categorical Variables

Scatterplots & Correlation

Guess the Correlation

Good for simple linear regression (BIOSTATS 540 Unit 12)

Distributions: Explore Shapes & Find Probabilities

Binomial Distribution

Normal Distribution

t Distribution

We'll use these a lot for confidence intervals and p-values (BIOSTATS 540 Units 8, 9, 10)

Sampling Distributions and the Central Limit Theorem

Sampling Distribution of the Sample Proportion

Sampling Distribution of the Sample Mean (Continuous Population)

Sampling Distribution of the Sample Mean (Discrete Population)

I find these handy for illustrating key ideas. (BIOSTATS 540 Units 1, 5, 7)

Confidence Intervals & Significance Tests (One Sample)

Inference for a Proportion

Inference for a Mean

Bootstrap for One Sample

Good for one sample estimation and hypothesis testing. (BIOSTATS 540 Units 8,9)

Confidence Intervals & Tests Comparing Two Groups

Compare Two Proportions

Compare Two Means

Bootstrap for Two Samples

Good for two sample estimation and testing and chi square tests. (BIOSTATS 540 Units 8, 10,11)

2. Illustration

Describe a Sample of Categorical Data

Example #1 – Question #1 of the Unit 2 (Data Visualization) Homework

Data Entry Method: Direct entry of frequency table (“by hand”)

The World Health Organization (WHO) records the annual number of confirmed cases of human Avian Influenza A/(H5N1). Following are a subset of their data:

Year:	2003	2004	2005	2006	2007	2008
# cases:	4	32	43	79	59	26



Pssst! You will be entering this data into ArtofStat. Stay tuned.

In this illustration, there are two (2) variables.

year: This is a nominal variable with 6 levels = 2003, 2004, 2005, 2006, 2007, 2008

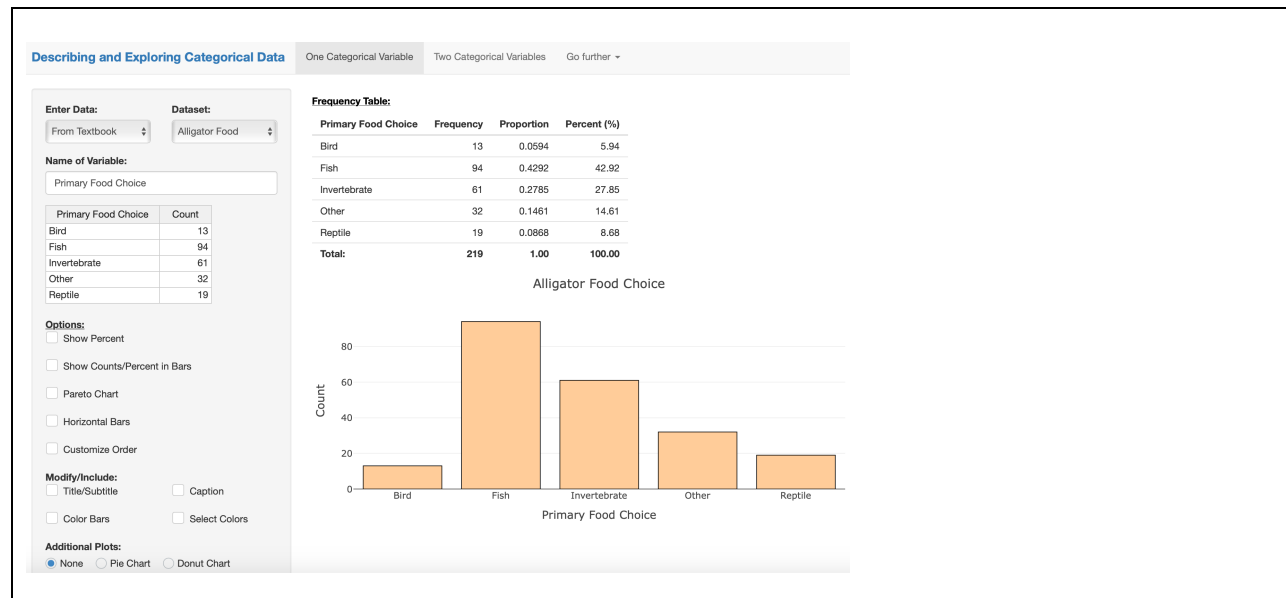
cases: This is the associated number of confirmed cases of influenza in that year)

GOAL: Produce a description of the distribution of **cases** over time (**year**). The description is to include a **frequency/relative frequency table** and a **bar graph**.

Step 1 – Launch Art of Stat

www.artofstat.com > **Online Web Apps**
 > **EXPLORATORY DATA ANALYSIS**
 > **EXPLORE CATEGORICAL DATA** (at top choose tab “One Categorical Variable”)

You should see:



Key: On the horizontal (X-axis) you see the possible outcomes (“food choice”). On the vertical (Y-axis) Art of Stat is showing you the frequency (“how often”) /relative frequency (“proportion of observations”) table and associated bar graph for Artof Stat’s own dataset titled “ALLIGATOR FOOD”

Step 2 – Happily, we can instead use own data! Here is where we enter the data on the previous page.

Next, at the upper left, from the dialogue boxes and drop-down menus, make the following selections/inputs:

Enter Data: **Frequency Table**

Tip – ArtofStat lets you enter EITHER a frequency table or individual data

Number of Categories: **6**

Name of Variable: **Year**

You should now see an empty Art of Stat analysis, waiting for you to enter your data. Fill in cells at LEFT!

Enter Data:
Frequency Table

Number of Categories:
6

Name of Variable:
Year

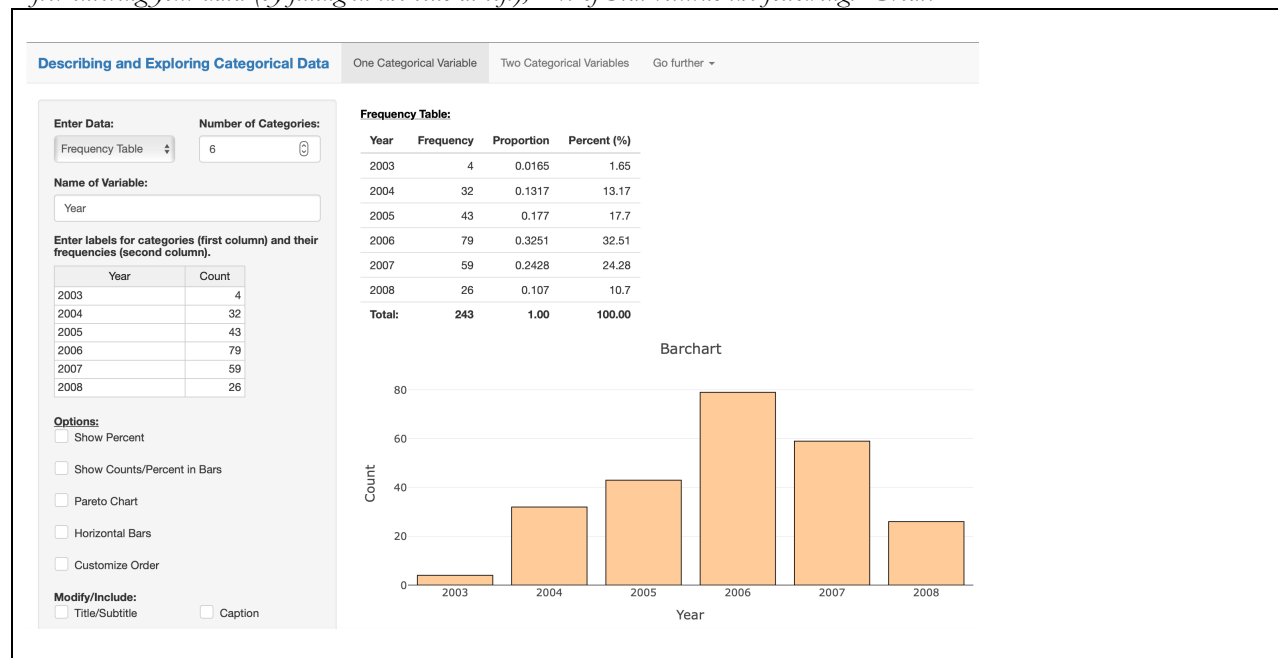
Enter labels for categories (first column) and their frequencies (second column).

Year	Count

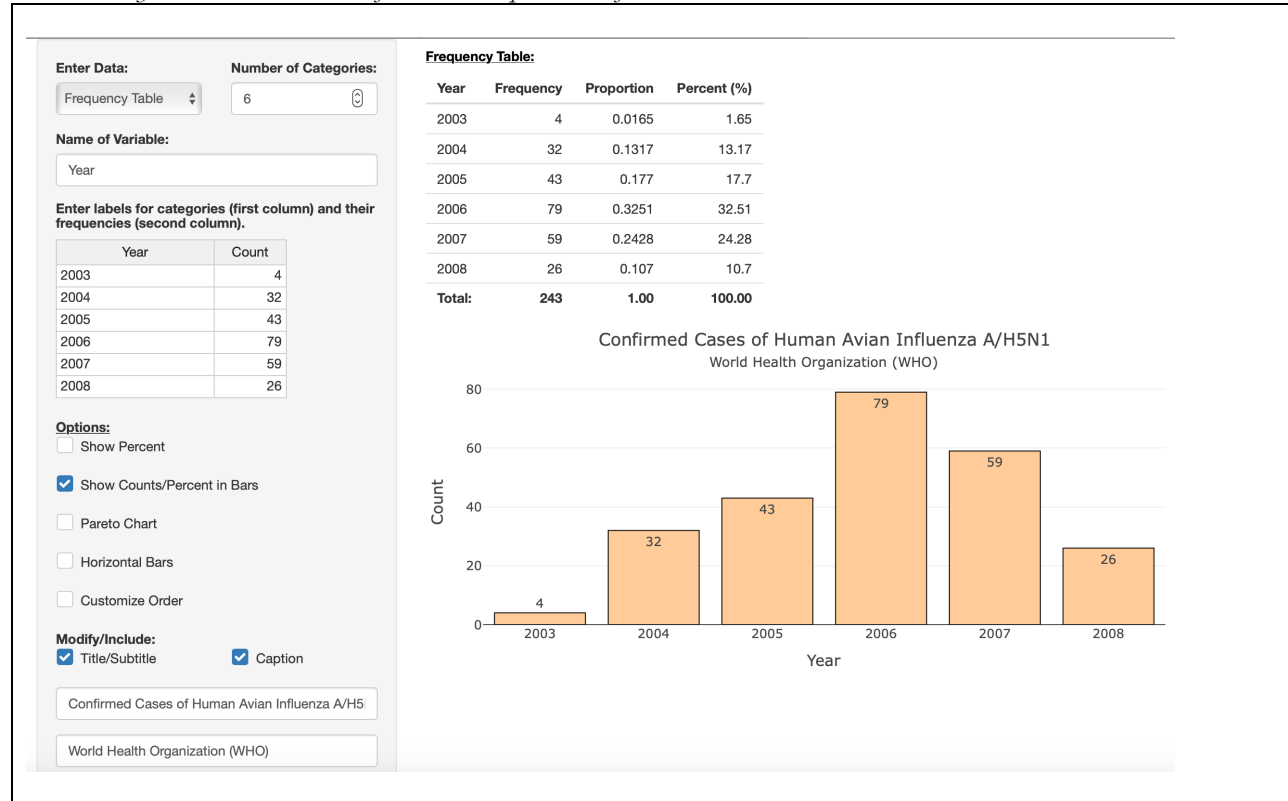
Frequency Table:

Year	Frequency	Proportion	Percent (%)
Total:	NA	1.00	100.00

After entering your data (by filling in the cells at left), Art of Stat returns the following. Great!



There's more you can do! Have some fun with the options at left



Example #2 – Unit 1 (Summarizing Data)**Data Entry Method: From Excel – Paste of individual observations**

This time we will enter data as individual observations and we will see how to do this as a “cut and paste” from an excel file. Handy!

As you will soon learn, this was a study of n=25 consecutive patients entering the general medical/surgical intensive care unit at a large urban hospital. One of the variables measured was ICU_TYPE, which is a nominal variable indicating the type of ICU admission. Recall how it was coded:

- 1 = medical
- 2 = surgical
- 3 = cardiac
- 4 = other

GOAL: Produce a description of the distribution of the types of ICU admissions (icu_type) in this sample.

How to get excel data into Art of Stat:

The 25 observations of icu_type are contained in the 2nd sheet of the datafile icu_540.xlsx.

Step 1 – In Excel

Open icu_540.xlsx.

At the bottom, click on the 2nd sheet. It is labelled “sheet 2”

You should now see the following:

	A
1	icu_type
2	medical
3	surgical
4	medical
5	medical
6	other
7	medical
8	medical
9	other
10	medical
11	medical
12	surgical
13	surgical
14	cardiac
15	cardiac
16	cardiac
17	medical
18	medical
19	surgical
20	cardiac
21	medical
22	surgical
23	medical
24	cardiac
25	medical
26	surgical

Step 2 – COPY/PASTE: The Copy Part

In Excel

Be sure to begin with row 2!!!

1st: Highlight (click and drag) to select cells in rows 2 through 26 (you don't want row 1)

2nd: EDIT > COPY

3rd: Minimize Excel. But do not exit Excel (you may want to come back to this later)

Step 3 – COPY/PASTE: The Paste Part

In Art of Stat

www.artofstat.com > **Online Web Apps**
 > **EXPLORATORY DATA ANALYSIS**
 > **EXPLORE CATEGORICAL DATA**

1st: In “ENTER DATA” drop down box, choose: Individual Observations

2nd: In “NAME OF VARIABLE” box, type: Type of ICU

3rd: Place your cursor in the box titled “Enter categorical observations,...”

4th: EDIT > PASTE

Enter Data: Individual Observations

Name of Variable: Type of ICU

Enter categorical observations, separated by a space. Or, simply copy and paste values from a column in your spreadsheet:

Bird Invertebrate Invertebrate Invertebrate Reptile Other Bird Bird Invertebrate Bird Other Invertebrate

Frequency Table:

Type of ICU	Frequency	Proportion	Percent (%)
-	-	-	-

Art of Stat returns the following:



Again, have some fun with the options at left



Key: On the horizontal (X-axis) you see the possible outcomes ("medical", "surgical", "other", "cardiac"). On the vertical (Y-axis) you now see frequencies (e.g., there are 12 occurrences of admissions to a medical ICU) /relative frequency (e.g. in this sample of 25, 48% of the admissions are to a medical ICU)

3. Illustration

Describe a Sample of Continuous Data

Example – Unit 1 (Summarizing Data)

Data Entry Method: From Excel – Paste of individual observations

Consider again the same study of $n=25$ consecutive patients entering the general medical/surgical intensive care unit at a large urban hospital. There is one continuous variable in this dataset: **age**.

GOAL: Produce a description of the distribution of age.

How to get excel data into Art of Stat:

The 25 observations of age are contained in the 1st sheet of the datafile icu_540.xlsx.

Step 1 – In Excel

Open icu_540.xlsx.

At the bottom, click on the 1st sheet. It is labelled “sheet 1”

You should now see the following:

A	B	C	D	E	F	G
id	age	type_adm	icu_type	sbp	icu_los	vit_stat
1	15.00	1.00	1.00	100.00	4.00	0.00
2	31.00	1.00	2.00	120.00	1.00	0.00
3	75.00	0.00	1.00	140.00	13.00	1.00
4	52.00	0.00	1.00	110.00	1.00	0.00
5	84.00	0.00	4.00	80.00	6.00	0.00
6	19.00	1.00	1.00	130.00	2.00	0.00
7	79.00	0.00	1.00	90.00	7.00	0.00
8	74.00	1.00	4.00	60.00	1.00	1.00
9	78.00	0.00	1.00	90.00	28.00	0.00
10	76.00	1.00	1.00	130.00	7.00	0.00
11	29.00	1.00	2.00	90.00	13.00	0.00
12	39.00	0.00	2.00	130.00	1.00	0.00
13	53.00	1.00	3.00	250.00	11.00	0.00
14	76.00	1.00	3.00	80.00	3.00	1.00
15	56.00	1.00	3.00	105.00	5.00	1.00
16	85.00	1.00	1.00	145.00	4.00	0.00
17	65.00	1.00	1.00	70.00	10.00	0.00
18	53.00	0.00	2.00	130.00	2.00	0.00
19	75.00	0.00	3.00	80.00	34.00	1.00
20	77.00	0.00	1.00	130.00	20.00	0.00
21	52.00	0.00	2.00	210.00	3.00	0.00
22	19.00	0.00	1.00	80.00	1.00	1.00
23	34.00	0.00	3.00	90.00	3.00	0.00
24	56.00	0.00	1.00	185.00	3.00	1.00
25	71.00	0.00	2.00	140.00	1.00	1.00

Step 2 – COPY/PASTE: The Copy Part
In Excel

Be sure to begin with row 2!!!

1st: Highlight (click and drag) to select cells in rows 2 through 26 of column B (age)

2nd: EDIT > COPY

3rd: Minimize Excel. But do not exit Excel (you may want to come back to this later)

Step 3 – COPY/PASTE: The Paste Part
In Art of Stat

www.artofstat.com > **Online Web Apps**
> **EXPLORATORY DATA ANALYSIS**
> **EXPLORE QUANTITATIVE DATA** (at top choose tab "Single Group")

1st: In "ENTER DATA" drop down box, choose: **Your Own**

2nd: Under "DO YOU HAVE": choose **Individual Observations**

3rd: In "NAME OF VARIABLE" box, type: **age**

4th: Place your cursor in the box titled **"Enter observations,..."**

5th: EDIT > PASTE

The screenshot shows the 'Enter Data' section of the Art of Stat web application. It includes a dropdown menu for 'Enter Data' set to 'Your Own', radio buttons for 'Do you have:' with 'Individual Observations' selected, a text box for 'Name of Variable:' containing 'age', and a large text area for 'Enter observations (separated by space) or copy and paste from spreadsheet:'. The text area contains the following data: 0 70 140 200 180 210 150 100 130 140 180 190 160 290 50 220 180 200 210. A red arrow points to the first '0' in the first row of data.

Art of Stat returns the following:



Playing around with options at left, I now have the following. You might make different selections; Go for it!

