BME 1132
Probability and Biostatistics

Instructor: Ali AJDER, *Ph.D.*

Week-5

Graphical Summaries

- *Stem-and-Leaf plot*
- *Dotplot*
- *Boxplot*
- *Scatterplot*
- *Histogram*

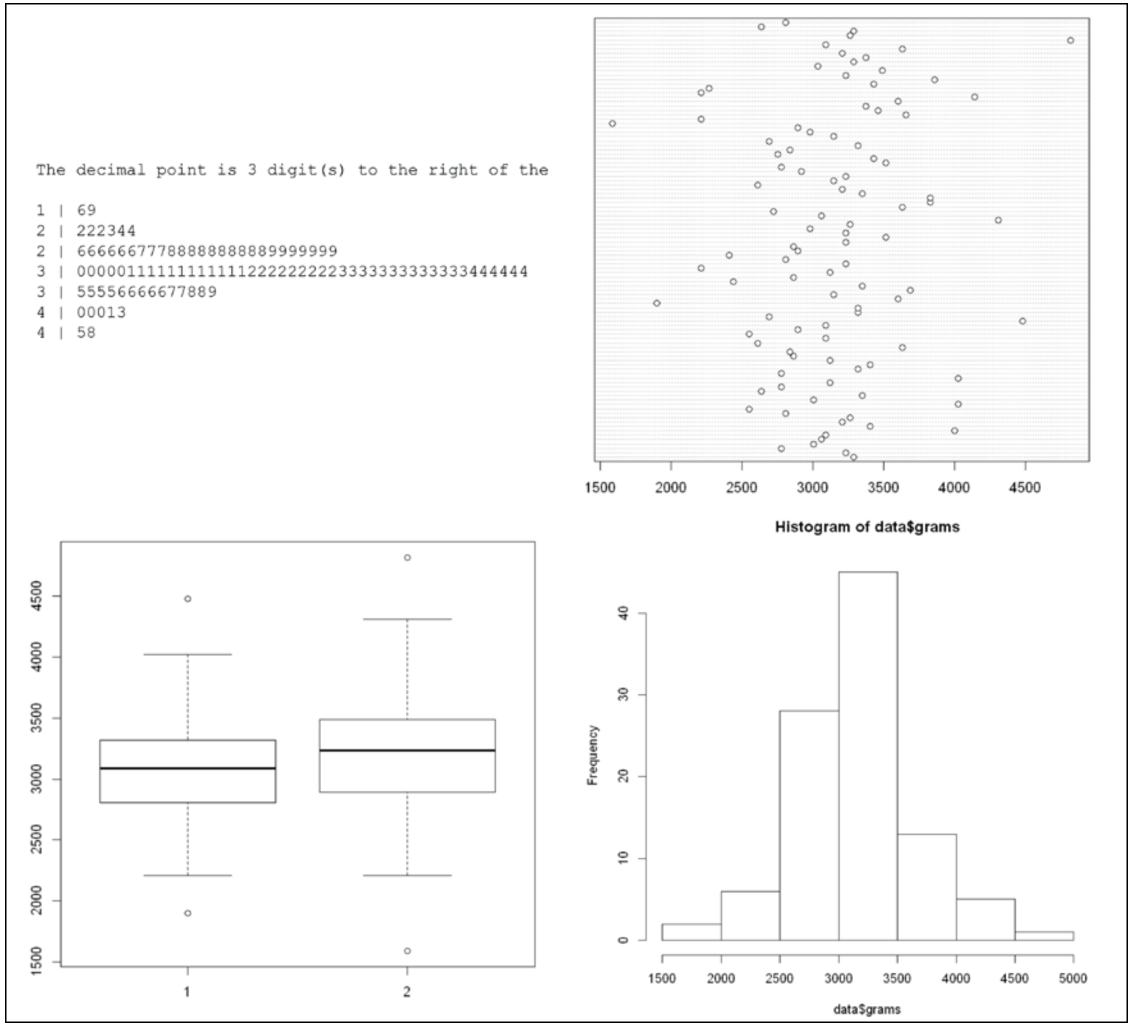
Symmetry and Skewness

The Coefficient of Variation

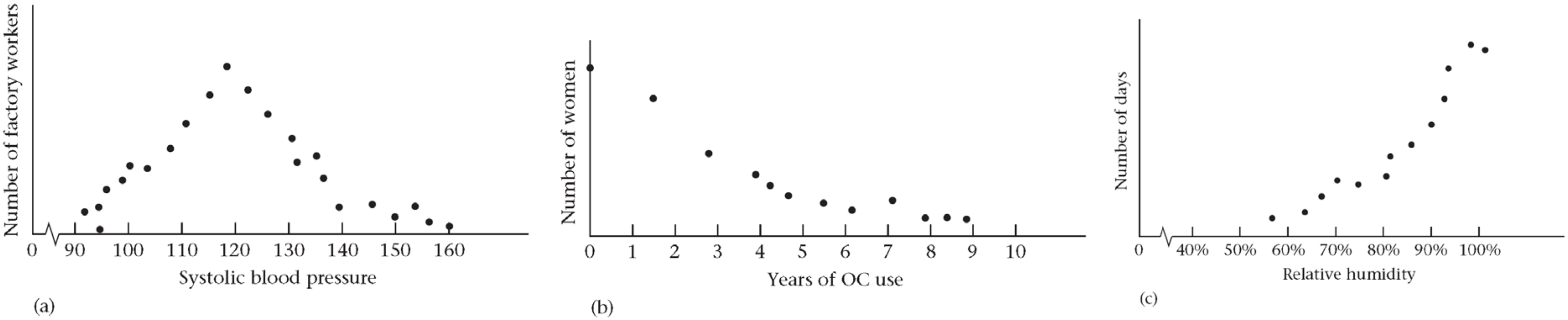
Measures of Central Tendency and Dispersion for Grouped Data

- *Mean*
- *Median*
- *Variance*
- *Standard Deviation*

Graphical Summaries



Symmetry and Skewness



Graphic displays of (a) symmetric, (b) positively skewed, and (c) negatively skewed distributions

The Coefficient of Variation

It is useful to relate the arithmetic mean and the standard deviation to each other because, for example, a standard deviation of 10 means something different conceptually if the arithmetic mean is 10 versus if it is 1000.

A special measure, the coefficient of variation, is often used for this purpose.

The **coefficient of variation** (CV) is defined by

$$100\% \times (s/\overline{x})$$

This measure remains the same regardless of what units are used because if the units change by a factor c, then both the mean and standard deviation change by the factor c; while the CV, which is the ratio between them, remains unchanged.

Example- The Coefficient of Variation

Compute the coefficient of variation for the data in Table when the birthweights are expressed in either grams or ounces.

i	x_i	i	x_i	i	x_i	i	x_i
1	3265	6	3323	11	2581	16	2759
2	3260	7	3649	12	2841	17	3248
3	3245	8	3200	13	3609	18	3314
4	3484	9	3031	14	2838	19	3101
5	4146	10	2069	15	3541	20	2834

NOTE

$$\boxed{1\ oz = 28.35\ g\ or \quad y_i = \frac{1}{28.35} x_i}$$

Example1- The Coefficient of Variation

Compute the coefficient of variation for the data in Table when the birthweights are expressed in either grams or ounces.

i	x_i	i	x_i	i	x_i	i	x_i
1	3265	6	3323	11	2581	16	2759
2	3260	7	3649	12	2841	17	3248
3	3245	8	3200	13	3609	18	3314
4	3484	9	3031	14	2838	19	3101
5	4146	10	2069	15	3541	20	2834

Solution: $CV = 100\% \times (s/\bar{x}) = 100\% \times (445.3\,\text{g}/3166.9\,\text{g}) = 14.1\%$

If the data were expressed in ounces, then

$$CV = 100\% \times (15.7 \text{ oz}/111.71 \text{ oz}) = 14.1\%$$

NOTE

$$\boxed{1\ oz = 28.35\ g\ or \quad y_i = \frac{1}{28.35} x_i}$$

Grouped Data

Sometimes the sample size is too large to display all the raw data.
Also, data are frequently collected in grouped form because the required degree of accuracy to measure a quantity exactly is often lacking due either to measurement error or to imprecise patient recall.

Consider the data set in Table, which represents the birthweights from 100 consecutive deliveries at a hospital. Suppose we wish to display these data for publication purposes.

How can we do this?

58	118	92	108	132	32	140	138	96	161
120	86	115	118	95	83	112	128	127	124
123	134	94	67	124	155	105	100	112	141
104	132	98	146	132	93	85	94	116	113
121	68	107	122	126	88	89	108	115	85
111	121	124	104	125	102	122	137	110	101
91	122	138	99	115	104	98	89	119	109
104	115	138	105	144	87	88	103	108	109
128	106	125	108	98	133	104	122	124	110
133	115	127	135	89	121	112	135	115	64

Grouped Data

How can we do this?

A **frequency distribution** is an ordered display of each value in a data set together with its **frequency**, that is, the number of times that value occurs in the data set. In addition, the percentage of sample points that take on a particular value is also typically given.

58	118	92	108	132	32	140	138	96	161
120	86	115	118	95	83	112	128	127	124
123	134	94	67	124	155	105	100	112	141
104	132	98	146	132	93	85	94	116	113
121	68	107	122	126	88	89	108	115	85
111	121	124	104	125	102	122	137	110	101
91	122	138	99	115	104	98	89	119	109
104	115	138	105	144	87	88	103	108	109
128	106	125	108	98	133	104	122	124	110
133	115	127	135	89	121	112	135	115	64

How do we develop a frequency distribution?

EXAMPLE
Creating a Frequency Distribution Table

Example- Constructing a Frequency Table

Sample of birthweights (oz) from 100 consecutive deliveries at a hospital

58	118	92	108	132	32	140	138	96	161
120	86	115	118	95	83	112	128	127	124
123	134	94	67	124	155	105	100	112	141
104	132	98	146	132	93	85	94	116	113
121	68	107	122	126	88	89	108	115	85
111	121	124	104	125	102	122	137	110	101
91	122	138	99	115	104	98	89	119	109
104	115	138	105	144	87	88	103	108	109
128	106	125	108	98	133	104	122	124	110
133	115	127	135	89	121	112	135	115	64

- What is the typical birthweight on each delivery?
- What is the largest birthweight on any delivery?
- What is the lowest birthweight on any delivery?
- Around what value did the birthweight tend to cluster?

Steps of a Frequency Distribution

1. Number of classes (k)
   - “2 to the k rule”
   - $2^k > n$

2. Class width
   - $i \geq \frac{H-L}{k}$

3. Class limits

4. Tally the observation into the classes

5. Count the number of items in each class

Example- Constructing a Frequency Table

58	118	92	108	132	32	140	138	96	161
120	86	115	118	95	83	112	128	127	124
123	134	94	67	124	155	105	100	112	141
104	132	98	146	132	93	85	94	116	113

Group interval	Frequency
$y_1 \le x < y_2$	f_1
$y_2 \le x < y_3$	f_2
.	.
.	.
.	.
$y_i \le x < y_{i+1}$	f_i
.	.
.	.
.	.
$y_k \le x < y_{k+1}$	f_k

The FREQ Procedure

Group_interval	Frequency	Percent	Cumulative Frequency	Cumulative Percent
$29.5 \le x < 69.5$	5	5.00	5	5.00
$69.5 \le x < 89.5$	10	10.00	15	15.00
$89.5 \le x < 99.5$	11	11.00	26	26.00
$99.5 \le x < 109.5$	19	19.00	45	45.00
$109.5 \le x < 119.5$	17	17.00	62	62.00
$119.5 \le x < 129.5$	20	20.00	82	82.00
$129.5 \le x < 139.5$	12	12.00	94	94.00
$139.5 \le x < 169.5$	6	6.00	100	100.00

Questions?