Practice test 1 – KEY

Chapter 1 Conceptual Questions

1. It is a way of *learning* from data. It is how we collect and organize measurements. It is how we collect clues and adjudicate the evidence.
2. Measurement is the assigning of numbers or codes according to prior set rules.
3. GIGO = "Garbage in, garbage out"
4. An observation is the individual unit upon which measurements are made.
   A variable is a measurement that can take on different values.
   A value is a realized measurement.
5. Nominal measurement or qualitative measurement
6. Continuous data.
7. Observations in rows, variables in columns, and values in cells.
8. objective
9. A statistical population is the set of all possible values for a variable.
10. A sample is a subset of the population.
11. The sampling fraction
12. (c)
13. (b)
14. equal
15. This sample is not a random because it excludes people toward the end of the alphabet. Simple random samples are selected so that everyone in the population has an equal chance of selection and every conceivable sample of size *n* is equally likely to comprise the sample.
16. Statistics serves many different scientific disciplines, helping to make its research more objective, precise, and accurate.
17. Sampling *with* replacement means that a once a unit is sampled from the population they are tossed back in and may be sampled again. Sampling *without* replacement means that the sample unit is ineligible for a second selection.

Chapter 1 Exercise

1.8 Cerebellar toxicity
(A) The population is leukemia and lymphoma patients that are candidates to receive this form of chemotherapy.
(B) The sample is the 54 patients that comprise the data.
(C) AGE = quantitative; SEX = categorical; GENERIC= categorical; DIAG = categorical; STAGE = categorical TOX = categorical; DOSE = quantitative; SCR = quantitative; WEIGHT = quantitative
(D) 49
(E) 1 = leukemia

Chapter 2 Conceptual Review Questions

1. *n* (one per observation)

2. Start with between 4 and 12 stem values. Then experiment with different stem multipliers and splitting schemes to see which stemplot best shows the shape of the distribution.
3. You split stem values if it gives you a better picture of the distribution's shape.
4. The depth of an observation is its location on the ordered array..
5. Its purpose is to help viewer decipher the magnitude of values.
6. positive
7. spread
8. median
9. $(n + 1) / 2$
10. "Potential outlier"
11. frequency = count; relative frequency = frequency count expressed as proportion of total; cumulative relative frequency = relative frequency up to and including the current value
12. A distribution is the frequency of occurrence for each value and each range of values in a batch of number.
13. Stem-and-leaf plot
14. You evaluate its shape, location, and spread, and whether there are any potential outliers.
15. So that frequencies are accurately depicted by the height of each set of values.
16. Truncate extra digits.

Chapter 2 Exercises

2.4 *Outpatient wait time*.

(A) The stemplot (below) reveals a positive skew (right tail) and no obvious outliers (shape). The median is 31 (central location). Data that range from 6 minutes to 77 minutes (spread).

```
0|6
1|033669
2|23489
3|1235
4|259
5|15
6|3
7|227
x10
```

(B) Frequency table, 10 day class intervals

MINUTES	Freq.	%	Cumulative %
0 - 9	1	4%	4%
10 - 19	6	24%	28%
20 - 29	5	20%	48%
30 - 39	4	16%	64%
40 - 49	3	12%	76%
50 - 59	2	8%	84%
60 - 69	1	4%	88%
70 - 79	3	12%	100%
TOTAL	25	100%	--

(C) 28% are less than 20 minutes in length.

(D) 18 / 25 = .72 or 72% are more than 20 minutes.

2.10 ***Health Insurance.*** There are several acceptable and unacceptable ways to create a stemplot for these data (below). The acceptable plots show a positive skew. Acceptable plots #2 show a high outlier. The median is 13 (central location). Data vary from 8 to 25 (spread).

<u>Unacceptable plot</u> [multiplier of x10 with non-split stem values; too few stem values to show shape]

```
0|89
1|00000000011111111233333334444555666667777888899
2|15
x10
```

<u>Acceptable plot #1</u> [multiplier of x10 with split stem values]

```
0|89
1|00000000011111111233333334444
1|555666667777888899
2|1
2|5
(x10)
```

<u>Acceptable plot #2</u> [multiplier of x1]

```
08|5
09|9
10|145566889
11|00457889
12|7
13|4556789
14|0224
15|099
16|16678
17|0239
18|2458
19|12
20|
21|4
22|
23|
24|
25|1
(x1)
```

Chapter 3 Conceptual Review Questions

1. Mean, median, mode
2. Standard deviation (or variance) and IQR (or quartiles)
3. This suggests a positive skew or high outlier.
4. <u>median</u>
5. <u>upper inside value</u>.
6. Hinge-spread (IQR), whisker-to-whisker spread, and range.
7. (1) Expected value of observation selected at random from the sample. (2) Expected value of an observation selected at random from the population. (3) Expected value of the population mean.

8. Variances carry units squared, so it is difficult to interpret.
9. IQR.
10. Q0 (min), Q1 (25th percentile), Q2 (median, Q3 (75th percentile), Q4 (maximum)
11. bottom
12. No, the fences do *not* appear on the plot. They are used to determine if there are any outside values.
13. *n* - 1
14. μ represents the population mean; $\bar{x}$ represents the sample mean. σ = *population* standard deviation; *s* = *sample* standard deviation.
15. Group A has lower values on the average and has less spread than Group B.
16. Both groups

Chapter 3 Practice Exercises

3.10 Air pollution at two sites.

(A) Data, site 1: 22, 24, 28, 32, 36, 38, 42, 68

Step by step mean and standard deviation demonstrating deviations and squared deviations:

i	X	deviation	squared deviation
1	22	-14.2500	203.0625
2	24	-12.2500	150.0625
3	28	-8.2500	68.0625
4	32	-4.2500	18.0625
5	36	-0.2500	0.0625
6	38	1.7500	3.0625
7	42	5.7500	33.0625
8	68	31.7500	1008.0625
Sums	290	0.0000	1483.5000

xbar = 290 / 8 **= 36.25** = 36.3 [reporting with three significant digits]
s^2 **=** 1483.5 / (8 – 1) = 211.929
s = sqrt(211.929) = 14.5578 = 14.6

(B) Data, site **2:** 32, 33, 34, 36, 36, 38, 39, 40
Calculated with TI-30XIIS:
xbar **= 36.0**
***s* = 2.9**

(C) The means barely differ (36.3 vs. 36**). However, site 1 has much** greater variability than site 2 (SD, 14.6 vs. 2.9)**.**

(D) Side-by-side boxplot

<u>Site 1</u> {Data: 22, 24, 28, 32, 36, 38, 42, 68}
5-point summary: 22, 26, 30, 40, 68
IQR = 40 – 26 = 14
$\text{Fence}_{\text{Upper}}$ = 40 + 1.5(14) = 61 → 68 is outside; 42 is inside
$\text{Fence}_{\text{Lower}}$ = 26 – 1.5(14) = 5 → none outside; 22 is inside

<u>Site 2</u> {Data: 32, 33, 34, 36, 36, 38, 39, 40}
5-point summary: 32, 33.5, 36, 38.5, 40
IQR = 38.5 – 33.5 = 5
$\text{Fence}_{\text{Upper}}$ = 38.5 + 1.5(5) = 46 → none outside; 40 is inside
$\text{Fence}_{\text{Lower}}$ = 33.5 – 1.5(5) = 26 → none outside; 32 is inside