# Unit 1 Practice Test

Data Analysis – Part I

Name: **Solutions**

A 1. We collect these data from 50 male students. Which variable is categorical?
   (A) eye color   B) head circumference   C) hours of homework last week
   D) number of cigarettes smoked daily   E) number of TV sets at home

D 2. Which of those variables is most likely to be bimodal?   *(smokers, nonsmokers)*

B 3. Which of those variables is most likely to follow a Normal model?   *(natural processes)*

B 4. The mean number of hours worked for the 30 males was 6, and for the 20 females was 9.
   The overall mean number of hours worked ... $\dfrac{(30)(6)+(20)(9)}{50}$
   A) is 6.5   (B) is 7.2   C) is 7.5   D) is none of these.   E) cannot be determined.

D 5. We might choose to display data with a stemplot rather than a boxplot because a stemplot
   ✓ I.   reveals the shape of the distribution.
   ✗ II. is better for large data sets.   *actually, harder to produce*
   ✓ III. displays the actual data.
   A) I only   B) II only   C) III only   (D) I and III   E) I, II, and III

A 6. Which is true of the data whose distribution is shown?
   ✓ I.   The distribution is skewed to the right.
   ✗ II. The mean is probably smaller than the median.
   ✗ III. We should summarize with mean and standard deviation.   *(median & IQR for skew)*
   (A) I only   B) II only   C) I and II   D) II and III   E) I, II, and III

mean $\overline{x}$
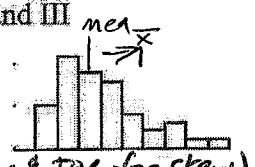
B or C 7. The standard deviation of the data displayed in
   this dotplot is most likely to be ...
   choose one (we may accept both as correct)
   A) 5.   (B) 8.   (C) 12.   D) 18.   E) 20.

E 8. Suppose that a Normal model describes the
   acidity (pH) of rainwater, and that water tested
   after last week's storm had a z–score of 1.8.
   This means that the acidity of that rain ...
   A) had a pH of 1.8.
   B) varied with a standard deviation of 1.8
   C) had a pH 1.8 higher than average
      rainfall.
   D) had a pH 1.8 times that of average
      rainwater.   *(z = # std devs)*
   (E) had a pH 1.8 standard deviations
      higher than that of average rainwater.

around 10
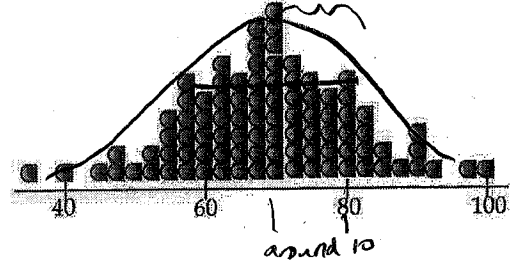
B 9. The ages of people attending the opening
   show of a new movie are summarized in
   the ogive shown. Estimate the IQR of the
   ages.
   A) 5   (B) 13   C) 21   D) 30   E) 37

Cumulative %   *Q3 75*   *25 Q1*   *Q1 30*   *Q3 44*
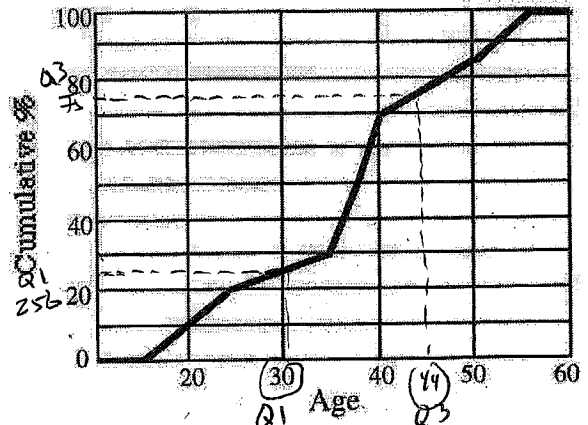
$IQR = Q_3 - Q_1 = 44 - 30 = 14$

E 10. Environmental researchers have collected rain acidity data for several decades. They want
   to see if there is any evidence that attempts to reduce industrial pollution have produced a
   trend toward less acidic rainfall. They should display their data in a(n)...
   A) contingency table   B) bar graph   C) boxplot   D) histogram   (E) timeplot

acidity / years

11. **Paying for purchases** One day a store tracked the
way shoppers paid for their purchases. Their data
are summarized in the table.

	Cash	Check	Charge	Total
Male	18	10	12	40
Female	18	12	30	60
Total	36	22	42	100

   a. What percent of the men paid cash? $\frac{18}{40} = 45\%$
   b. What is the conditional relative frequency
       distribution of payment method for women? (%)

   cash   check   charge
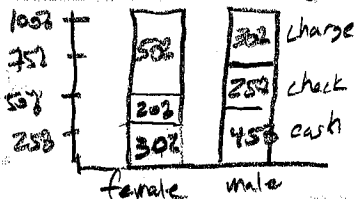   18   12   30   / 60
   30%   20%   50%

   c. If you wanted to show the association between gender and method of payment visually,
       what kind of graph would you make? (Just name it.)

   segmented bar graphs

   d. Is there evidence of an association between gender and method of payment? Explain
       briefly.

   for men:
   cash check charge
    18   10   12   /40
   45%   25%   30%

   The percentage using charge is quite
   different (50% female vs. 30% male),
   check is closer (20% vs. 25%) and
   cash slightly more male (45% vs 30%)
   There is an association between gender
   and method of payment

12. **Repair bills** An automobile service shop reported the summary statistics
    shown for repair bills (in $) for their customers last month.

Min	27
Q1	88
Median	132
Q3	308
Max	1442
Mean	284
SD	140

   a. Were any of the bills outliers? Show how you made your decision.

   $IQR = Q3 - Q1 = 308 - 88 = 220$

   upper fence: $Q3 + 1.5 IQR = 308 + 1.5(220) = 638$

   lower fence: $Q1 - 1.5 IQR = 88 - 1.5(220) = -242$

   min is 27 (not < -242) so no low outliers but
   max is 1442 (is > 638) so there is at least one outlier (1442).

   b. After checking out a problem with your car the service manager gives you an estimate of
       only $90." Is he right to imply that your bill will be unusually low? Explain briefly.

   Q1 is $88 so 25% of bills are $88 or less, $90 does not seem
   to be unusually low.
   — or — $z = \frac{90 - 284}{140} = -1.38$ $90 is only 1.4 std devs below the mean
   which is not very unusual. (note: can't use
                                 normalcdf --
13. **Salary conversions** You learn that your company is sending you and several other employees   we don't
    to staff a new office in China. While there everyone will earn the equivalent of their current   know it
    salary, converted to Chinese currency at the rate of 8 yuans per dollar. In addition, everyone   is normal
    will earn a weekly foreign living allowance of 200 yuans. For example, since you are earning   shaped)
    $1000 per week, your weekly salary in China will be 1000 x 8 + 200 = 8200 yuans.   x8 + 200

   a. Shown are some summary statistics
       describing the current salaries of this group
       being sent overseas. Fill in the table to show
       what these statistics will be for the salaries
       you all will earn while in China.

Statistic	In the US	In China
Minimum salary	$400	3400 yuan
Standard deviation	$250	2000 yuan
Median	$750	6200 yuan
IQR	$300	2400 yuan

   b. Among this group of employees going to
       China, your US salary has a z-score of +1.20.
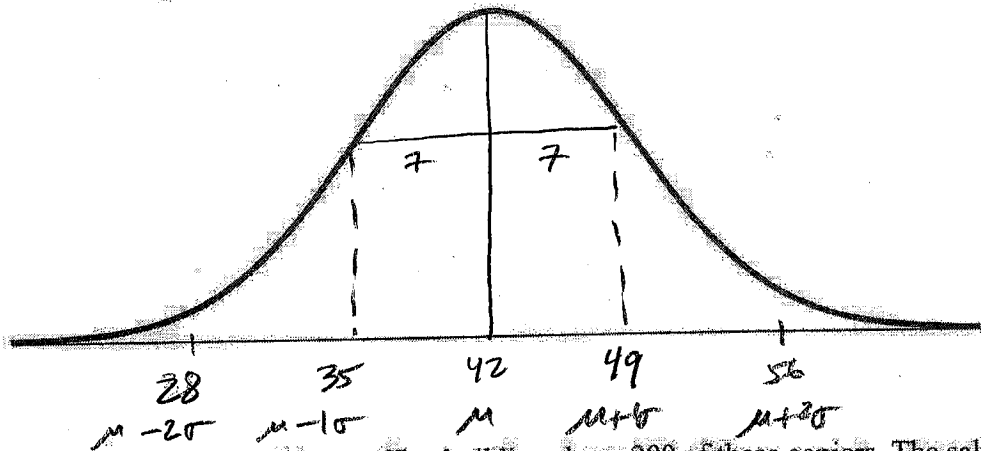   What will your new z-score be, based on everyone's China salary?

   +1.20

(a value) min: $400(8) + 200 = 3400$
(a spread) std dev: $250(8) = 2000$
(a value) median: $750(8) + 200 = 6200$
(a spread) IQR: $300(8) = 2400$

14. **Copy machines** A manufacturer claims that lifespans for their copy machines (in months) can be described by a Normal model $N(42,7)$. Show your work.

a. Draw and clearly label the model.
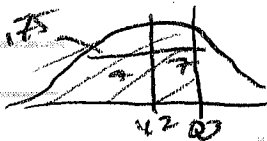

$$7 \qquad 7$$

$$28 \qquad 35 \qquad 42 \qquad 49 \qquad 56$$

$$\mu - 2\sigma \qquad \mu - 1\sigma \qquad \mu \qquad \mu + \sigma \qquad \mu + 2\sigma$$

b. A company with a several large office buildings buys 200 of these copiers. The salesman tells the boss "190 (95%) of your new copiers will last between __28__ and __56__ months." Comment on this claim.

(too strong)

± 2 std devs

The statement is likely correct but is worded too strongly.
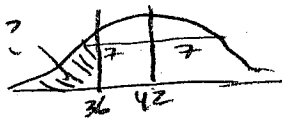
c. What is the 3rd quartile of copier lifespans?

$Q3 = invNorm(.75, 0, 1) = 46.7$
mean SD
area to left

__46.7 months__

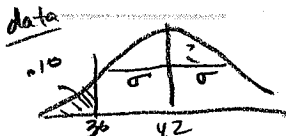d. What percent of the copiers are expected to fail before 36 months?

$normalcdf(-999, 36, 42, 7) = .196$
lower upper mean SD

__19.6%__

e. The manufacturer wants to reduce the 36-month failure rate to only 10%. Assuming the mean lifespan will stay the same, what standard deviation must they achieve?

data   z-score

$z = invNorm(area to left = 0.10, mean = 0, SD = 1) = -1.28155$

$z = \dfrac{x - \mu}{\sigma}$

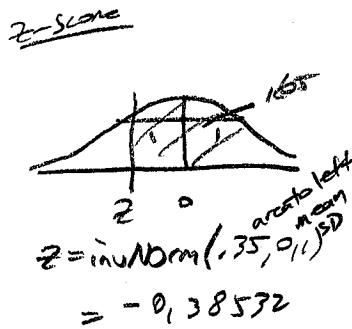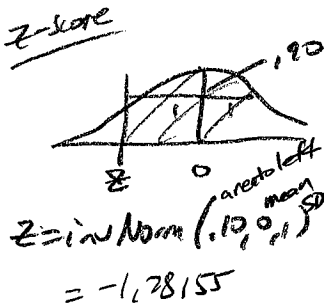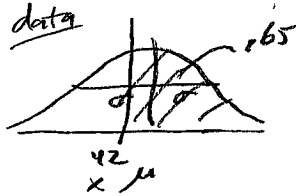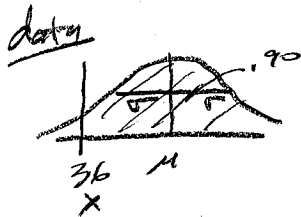$-1.28155 = \dfrac{36 - 42}{\sigma}$

$\sigma = $ __4.68 months__

f. Briefly explain what that change in standard deviation means in this context.

The machine wear out times will group more closely to the mean wear out time of 42 months.

g. A competing manufacturer says that not only will 90% of their copiers last at least 36 months, 65% will last at least 42 months. What Normal model parameters is that manufacturer claiming? Show your work.

$N($ __44.6__ , __6.7__ $)$ months

work on back →

data

.90

36   μ
X

data

.65

42 μ
X

z-score

.90

Z   0
area to left
mean

$$z = \text{invNorm}\left(.10, 0, 1\right)_{\text{SD}}^{\text{mean}}$$

$$= -1.28155$$

z-score

.65

Z   0
area to left
mean

$$z = \text{invNorm}\left(.35, 0, 1\right)_{\text{SD}}^{\text{mean}}$$

$$= -0.38532$$

$$z = \frac{x - \mu}{\sigma}$$

$$-1.28155 = \frac{36 - \mu}{\sigma}$$

$$\mu - 1.2815\sigma = 36$$

$$z = \frac{x - \mu}{\sigma}$$

$$-0.38532 = \frac{42 - \mu}{\sigma}$$

$$\mu - 0.38532\sigma = 42$$

System:

$$\begin{cases} \mu - 1.28155\sigma = 36 \\ \mu - 0.38532\sigma = 42 \end{cases}$$

Solve w/ matrices or substitution

matrices

$$\begin{bmatrix} 1 & -1.28155 & | & 36 \\ 1 & -0.38532 & | & 42 \end{bmatrix}$$

rref

$$\begin{bmatrix} \mu & \sigma & | & \text{RHS} \\ 1 & 0 & | & 44.6 \\ 0 & 1 & | & 6.7 \end{bmatrix}$$

μ = 44.6 months
σ = 6.7 months

Substitution

$$\mu = 36 + 1.2815\sigma$$

$$\mu = 42 + 0.38532\sigma$$

so

$$36 + 1.28155\sigma = 42 + 0.38532\sigma$$

$$0.89623\sigma = 6$$

$$\sigma = \frac{6}{.89623} = 6.7 \text{ months}$$

back substitute: $\mu - 1.28155(6.7) = 36$

$$\mu = 44.6 \text{ months}$$