NAME: Answer Key

Directions: Be sure to show all of your work. An answer alone will not receive any credit. You must show a formula or how you arrived at your answer. Partial credit will be given on all problems.

[10] 1. True/False questions.

T (F) [2]a. Too many intervals in a Histogram or too many stems in a stem/leaf plot stretch out the graphical summary of the data, making it difficult to discern features of the spread of the distribution.

10/10

(T) F [2]b. Statistics and parameters are numerical quantities used to summarize ANY sample and ANY population, respectively.

T (F) [2]c. Mutually exclusive events are independent events.

(T) F [2]d. Using fences in boxplots to detect outliers is a robust outlier detection method.

(T) F [2]e. Histograms and stem/leaf plots must have non-overlapping, equal length intervals and stems.

[15] 2. A random sample of ten gas stations across the US were contacted and asked what a gallon of regular gas was currently selling for last week. Below is the Minitab summary statistics output for this data. Use it to answer the questions that follow.

15/15

Variable	Total Count	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum	Kurtosis
Gas Prices	10	1.9650	0.0589	0.1861	1.4800	1.9150	2.0100	2.0775	2.1300	5.96

[3] a. What is the population of interest in this example?

(-2) if price of gas

popln = all gas stations in US selling regular gas. (-1)

[3] b. What would the variance of gas prices be?

$s^2 = STDEV^2 = 0.1861^2 = 0.0346$

[9] c. In which direction is this data skewed and give two different arguments that justify your answer.

[3]Skew direction: skewed negatively or to the left.

Reasons: [3]i. since mean/avg 1.9650 < median = 2.01

[3]ii. since min (1.48) further from median (2.01) than max (2.13) from median (2.01)

(iii) since Q_1 (1.915) further from median (2.01) than Q_3 (2.0775) from median

[20] 3. Below is the view data (in millions of viewers) for every show aired on the FOX network for the week of January 15th.

min fence fence max

3.3 4 4.3 4.5 5 5.1 5.1 5.5 **5.6** 5.9 6.1 6.3 6.3 6.7 15.7 36.9 37.4

$Q_1 = \frac{4.5+5}{2} = 4.75$ med. $Q_3 = \frac{6.3+6.7}{2} = 6.5$

[10] a. Obtain the five number, **BY HAND!!!!**, being sure to show your work! $n = 17$ so median is 9th value

{3.3, 4.75, 5.6, 6.5, 37.4}. 2 2 2 2 2

[5] b. Verify that there are outliers in this data set, using your five number summary information.

IQR = 6.5 − 4.75 = 1.750 ✓ Upper Fence = $Q_3 + 1.5\,IQR = 6.5 + 1.5(1.75) = 9.125$ ✓

Lower Fence = $Q_1 - 1.5\,IQR = 4.75 - 1.5(1.75) = 4.75 - 2.6250 = 2.1250$ ✓

only 3 outliers – 15.7, 36.9, 37.4 2

[5] c. The average and standard deviation are 9.63 and 10.70, respectively. Are the outliers found in b) also outliers by the "z-score" rule?

$z = \frac{15.7-9.63}{10.70} = 0.5673$ ✓ $z = \frac{36.9-9.63}{10.70} = 2.5486$ ✓ $z = \frac{37.4-9.63}{10.70} = 2.5953$ ✓

since no z-score is > 3 so no outliers. 2

20/20

[20] 4. Every Wednesday, *USA Today* reports the Nielsen TV ratings for every show aired the pervious week. One of the measurements reported is the Number of Viewers (in millions) of each show along with the network on which the show aired. Use the attached Network side-by-side boxplot of the Numbers of Viewers for the week of January 15th to answer the questions below.

[10] a. True/False questions.

T (F) [2]i. The network with the most normal distribution is Univision.

(T) F [2]ii. 50% of NBC's shows are within 1.5 million viewers of the median for NBC.

T (F) [2]iii. Since the distributions of number of viewers for CBS is skewed negatively, the average for CBS would be less than the median of CBS.

(T) F [2]iv. The network with the smallest measures of spread would be CW.

(T) F [2]v. The network with the most symmetric distribution is Univision.

20/20

[10] b. In the *USA Today* article, FOX was rated as the number two network. Give **TWO** statistical reasons, based on the boxplot, why this statement is misleading. **BE BRIEF, BUT TO THE POINT!!!**

i. Median # for Fox is less than CBS, ABC, and NBC

ii. 75% of Fox's shows were less than the Q_1's of CBS, ABC, & NBC.

iii. IQR of Fox is lower/shifted down from IQR's of CBS, ABC + NBC.

[5] 5. In a certain federal prison it is known that 2/3 of inmates are under 25 years of age. It also known that 3/5 of the inmates are males and that 5/8 of the inmates are female or 25 years of age or older. What is the probability that a prisoner selected at random from this prison is female and at least 25 years old?

$P(<25) = 2/3$ $P(M) = 3/5$ $P(F \text{ or } \geq 25) = 5/8$ $P(F \text{ and } \geq 25) = ?$

-3 ¬ indep

$P(F \text{ or } \geq 25) = \frac{5}{8} = P(F) + P(\geq 25) - P(F \text{ and } \geq 25)$

$P(F \text{ and } \geq 25) = P(F) + P(\geq 25) - P(F \text{ or } \geq 25)$

$= [1 - P(M)] + [1 - P(<25)] - 5/8 = \left[1 - \frac{3}{5}\right] + \left[1 - \frac{2}{3}\right] - \frac{5}{8}$

$= \frac{2}{5} + \frac{1}{3} - \frac{5}{8} = \frac{13}{120} = 0.1083$

5/5

[10] 6. An allergist claims that 50% of the patients she tests are allergic to some type of weed. What is the probability that exactly three of her next 4 patients are allergic to weeds? What do you need to assume?

3 of next 4 $= (A_1 A_2 A_3 A_4') \text{ or } (A_1 A_2 A_3' A_4) \text{ or } (A_1 A_2' A_3 A_4) \text{ or } (A_1' A_2 A_3 A_4)$ (E_1, E_2, E_3, E_4)

patients are indepdent (3)

A_i = patient i allergic to weed

M.E. events

so $P(A_1 A_2 A_3 A_4') = P(A_1)P(A_2)P(A_3)P(A_4') = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{16} = .0625$

each of E_1 to E_4 are identical in prob so $4 \cdot \frac{1}{16} = \frac{1}{4} = .25$ (3)

-5 E.L.

-4 ¬ .5³

10/10

[10] 7. A large industrial firm uses two local motels to provide overnight accommodations for its clients. From past experience it is known that 20% of the clients are assigned rooms at the Ramada Inn and 80% at the Sheraton. If the plumbing is faulty in 5% of the rooms at the Ramada Inn and in 4% of the rooms at the Sheraton, what is the probability a person with a room having faulty plumbing was assigned accommodations at the Ramada Inn?

$P(RI) = 0.20$ $P(S) = 0.80$ $P(FP|RI) = 0.05$ $P(FP|S) = 0.04$ (3)

$P(RI|FP) = \frac{P(RI \cap FP)}{P(FP)} = \frac{P(RI)\cdot P(FP|RI)}{P[(FP \cap RI) \cup (FP \cap S)]} = \frac{0.20(0.05)}{P(FP \cap RI) + P(FP \cap S)} = \frac{0.01}{0.01 + P(S)\cdot P(FP|S)}$

ME since $RI \cap S = \phi$

	P	P′	
R	.2(.05)=.01	.19	.20
S	.8(.04)=.032	.7680	.80
	.042	.9580	1

$= \frac{0.01}{0.01 + 0.80(0.04)} = \frac{0.01}{0.01 + 0.0320} = \frac{0.01}{0.0420} = 0.2381$

.01/.042 = 1/4

-2 ¬ F|R

-3 ¬ F ∩ R

10/10

[10] 8. In a class of 4 students (none of whom were born in leap year), what is the probability at least two students sharing birthdays? What do you need to assume?

students B-days in class are indep and all days equally likely (3)

$P(\text{at least two share b-days}) = 1 - P(\text{non share})$ (2)

$= 1 - P(BD_1 \cap BD_2 \cap BD_3 \cap BD_4)$ where $BD_1 \ldots BD_4$ are 4 different B-days.

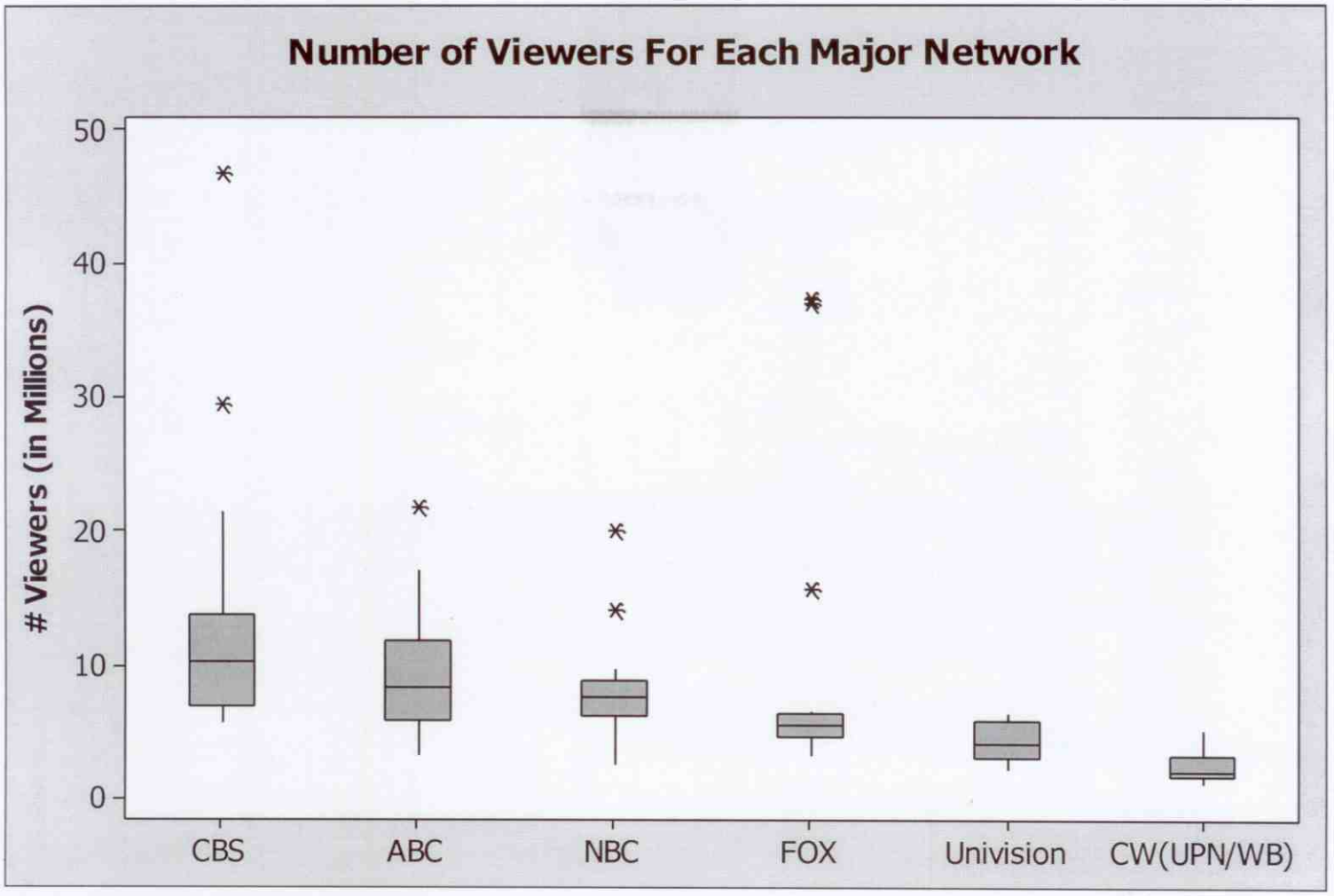
(#4)

$= 1 - P(BD_1)\cdot P(BD_2|BD_1)\cdot P(BD_3|BD_1 \cap BD_2)\cdot P(BD_4|BD_1 \cap BD_2 \cap BD_4)$

$= 1 - \frac{365}{365}\cdot\frac{364}{365}\cdot\frac{363}{365}\cdot\frac{362}{365} = 1 - 0.9836 = 0.0164$

1

10/10


Number of Viewers For Each Major Network
Viewers (in Millions)
50
40
30
20
10
0
CBS
ABC
NBC
FOX
Univision
CW(UPN/WB)