CSC591 HW4
Answer Key

Q1.
$H(T) = -(4/9 * \log_2(4/9) + 5/9 * \log_2(5/9)) = 0.991$

$H(T[a1 = T]) = -(3/4 * \log_2(3/4) + 1/4 * \log_2(1/4)) = 0.811$
$H(T[a1 = F]) = -(1/5 * \log_2(1/5) + 4/5 * \log_2(4/5)) = 0.722$
$IG(a1) = H(T) – (4/9 * H(T[a1 = T]) + 5/9 * H(T[a1 = F])) = 0.991 – 4/9 * 0.811 – 5/9 * 0.722$
$= 0.229$

$H(T[a2 = T]) = -(2/5 * \log_2(2/5) + 3/5 * \log_2(3/5)) = 0.971$
$H(T[a2 = F]) = -(2/4 * \log_2(2/4) + 2/4 * \log_2(2/4)) = 1.000$
$IG(a2) = H(T) – (5/9 * H(T[a2 = T]) + 4/9 * H(T[a2 = F])) = 0.991 – 5/9 * 0.971 – 4/9 * 1.000$
$= 0.007$

(Optional) a3 was converted to categorical data by binning using two bins. Let a3' = F if a3 <= 5, else T. The number of bins was chosen so that the resulting attribute is directly comparable to a1 and a2, which are binary.

$H(T[a3' = T]) = -(2/5 * \log_2(2/5) + 3/5 * \log_2(3/5)) = 0.971$
$H(T[a3' = F]) = -(2/4 * \log_2(2/4) + 2/4 * \log_2(2/4)) = 1.000$
$IG(a3') = H(T) – (5/9 * H(T[a3' = T]) + 4/9 * H(T[a3' = F])) = 0.991 – 5/9 * 0.971 – 4/9 * 1.000$
$= 0.007$

Higher IG is better: a1 > a2 = a3

Q2.
See MT2 solutions, Q2b.

Q3a.
(Claim changed to "median = 77" as discussed on forum)

H_0: median = 77
H_1: median ≠ 77

13 (-), 4 (+)
Test value = 4

From a sign test table with n = 17 and alpha = 0.01, C.V. = 2, so we are not able to reject the null hypothesis.

Q3b.
H0: number of colds does not decrease

H1: number of colds decreases

Subject	Before	After	Change
1	0	2	-
2	1	1	0
3	2	0	+
4	2	2	0
5	2	3	-
6	3	2	+
7	3	1	+
8	3	3	0
9	3	2	+
10	4	1	+
11	4	3	+
12	5	2	+
13	5	4	+
14	6	3	+

2 (-), 9 (+)
n = 11, T.V. = 2

From table, C.V. = 2, so the test value lies in the critical region and we reject the null hypothesis.

Q3c.
Wilcoxon signed rank test:

H_0: There is no reduction in blood pressure
H_1: There is a reduction in blood pressure

Before	After	Difference	|Difference|	Rank	Signed Rank
120	105	15	15	12	12
109	115	-6	6	4.5	-4.5
108	99	9	9	8	8
112	115	-3	3	2	-2
111	117	-6	6	4.5	-4.5
117	108	9	9	8	8
135	122	13	13	10.5	10.5
124	120	4	4	3	3
115	106	9	9	8	8
118	126	-8	8	6	-6
130	128	2	2	1	1
129	116	13	13	10.5	10.5

Sum of positives: 61
Sum of negatives: -17
Test value: |-17| = 17
C.V. from WSR table (n = 12, alpha = 0.01, one-sided): 10
T.V. > C.V., so we do not have enough evidence to reject the null hypothesis and cannot conclude that there is a decrease.

Q4a.

(i)

Prediction 1:

		Actual	
		1	**2**
Predicted	**1**	5	1
	2	4	5

Prediction 2:

		Actual	
		1	**2**
Predicted	**1**	5	2
	2	4	4

Prediction 3:

		Actual	
		1	**2**
Predicted	**1**	6	2
	2	3	4

(ii) Prediction 1: (5 + 5)/19 = 10/15
Prediction 2: (5 + 4)/19 = 9/15
Prediction 3: (6 + 4)/19 = 10/15

(iii) Prediction 1: 5/(5 + 1) = 5/6
Prediction 2: 5/(5 + 2) = 5/7
Prediction 3: 6/(6 + 2) = 6/8

(iv) Prediction 1: 5/(5 + 4) = 5/9
Prediction 2: 5/(5 + 4) = 5/9
Prediction 3: 6/(6 + 3) = 6/9

(v) Prediction 1: 2 * 5/6 * 5/9 / (5/6 + 5/9) = 0.667
Prediction 2: 2 * 5/7 * 5/9 / (5/7 + 5/9) = 0.625
Prediction 3: 2 * 5/9 * 6/9 / (5/9 + 6/9) = 0.606

In (iii) and (iv), it is assumed that “1” is the “positive” class.

Q4b.

Majority	1	1	1	1	1	2	2	2	2	2	2	2	2	1	2
Ground Truth	1	1	1	1	1	2	1	2	2	2	1	2	1	1	2

		Actual	
		1	**2**
Predicted	**1**	6	0
	2	3	6

Accuracy: 12/15

Precision: 6/6

Recall: 6/9

F-measure: 2 * 6/6 * 6/9 / (6/6 + 6/9) = 0.800