

Math 2200

Midterm Examination 1 – February 9, 2010

General Instructions: You may use any calculator you like. You may have a 3x5 card, but no other notes. Only the answer on the answer card will be graded.

Problems 1-19: Multiple choice. Each problem is worth 4 points.

1. The Gallup Poll conducted a representative telephone survey of 1200 American voters during the first quarter of 2007. Among the reported results were the voter's region (Northeast, South, etc), age, party affiliation, and whether or not the person had voted in the 2006 midterm Congressional election.

How many variables in the data described above are quantitative, and how many variables are there in total?

- (a) 0 quantitative, 3 total.
- (b) 0 quantitative, 4 total.
- (c) 1 quantitative, 3 total.
- (d) 1 quantitative, 4 total.
- (e) 2 quantitative, 3 total.
- (f) 2 quantitative, 4 total.
- (g) 3 quantitative, 3 total.
- (h) 4 quantitative, 4 total.

2. The follow contingency table compares the daily forecast with the actual weather in St. Louis for a 31-day winter month.

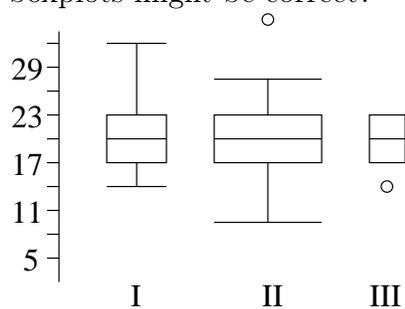
Forecast	Actual Weather	
	Snow	No snow
Snow	6	4
No snow	1	20

On what percent of days did it actually snow?

- (a) 1.92%
 - (b) 3.23%
 - (c) 16.67%
 - (d) 20.00%
 - (e) 22.58%
 - (f) 32.26%
 - (g) 67.74%
 - (h) 77.42%
3. In Question 2, what percent of the time was the forecast accurate?
- (a) 5.48%
 - (b) 7.12%
 - (c) 16.13%
 - (d) 22.58%
 - (e) 64.52%
 - (f) 80.00%
 - (g) 83.87%
 - (h) 95.00%

4. Which of the following variables is least likely to be normally distributed?
- (a) SAT scores of high school students.
 - (b) Ages of people at the Galleria mall on Sunday.
 - (c) IQs of people in St. Louis.
 - (d) Finish times in a foot race.
 - (e) Area codes of customers.
 - (f) Weights of eggs from 2-year old hens.
 - (g) Scores on a chemistry exam.
 - (h) Cholesterol levels of adult men.
5. Find the interquartile range of the numbers 1, 2, 3, 4, ..., 99, 100. That is, find the interquartile range of the natural numbers between 1 and 100.
- (a) 32
 - (b) 48
 - (c) 50
 - (d) 52
 - (e) 66
 - (f) 68
 - (g) 70
 - (h) 95

6. A student study of the effects of caffeine on memory asked volunteers to take a memory test 2 hours after drinking an espresso. The 5 number summary of the volunteers scores included: $Q1=17$, $Median=20$, $Q3=23$. Without knowing the rest of the data, which of the following boxplots might be correct?



- (a) All of the above.
 (b) I and II only.
 (c) I and III only.
 (d) II and III only.
 (e) I only.
 (f) II only.
 (g) III only.
 (h) None of the above.
7. In a normal model, find the z -score for $Q3$.
- (a) 0.00
 (b) 0.33
 (c) -0.33
 (d) 0.67
 (e) -0.67
 (f) 1
 (g) -1
 (h) 1.35

8. The following stem plot was created by R, and represents the downhill skiing times from the 2002 Winter Olympics:

```
> stem(skaters$Times..in.seconds.)
      The decimal point is at the |
    98 | 1448
    100 | 003344677888912336777889999
    102 | 23550223678
    104 | 4335
    106 | 467
    108 | 488
    110 |
    112 |
    114 | 4
```

How many skiers had a time of 101.7 seconds (to the nearest 10th of a second)?

- (a) 0
- (b) 1
- (c) 2
- (d) 3
- (e) 4
- (f) 5
- (g) 6
- (h) 7

9. The table below shows the oil production of the United States from 1996 to 2000 (in thousands of barrels per year). Find the correlation between year and oil production.

Year	Oil production
1996	2,366,017
1997	2,354,831
1998	2,281,919
1999	2,146,732
2000	2,135,062

- (a) -0.9956
(b) -0.9561
(c) -0.9166
(d) -0.8771
(e) -0.8376
(f) -0.7981
(g) -0.7586
(h) -0.7191
10. Suppose a normal model has mean $\mu = 5$ and standard deviation $\sigma = 3$. Calculate to 2 decimal places the percentage of data that you expect to be between -1 and 11 .
- (a) 5.64%
(b) 68.00%
(c) 68.27%
(d) 86.64%
(e) 95.00%
(f) 95.45%
(g) 99.73%
(h) 99.50%

11. If the scores on the SAT math test are normally distributed with a mean of 500 and standard deviation of 100, what score do you need to be in the top 20%? (Round up to the nearest 10.)

- (a) 530
- (b) 550
- (c) 570
- (d) 590
- (e) 610
- (f) 630
- (g) 650
- (h) 680

12. Suppose a normal model has mean $\mu = 20$, and 60% of the data is above 10. Calculate the standard deviation.

- (a) 39.47
- (b) -39.47
- (c) 46.74
- (d) -46.74
- (e) 50
- (f) -50
- (g) 68.27
- (h) -68.27

13. Suppose a normal model has standard deviation $\sigma = 2$, and only 10% of the data is below 0. Calculate the mean.

- (a) 0.72
- (b) -0.72
- (c) 1.28
- (d) -1.28
- (e) 1.72
- (f) -1.72
- (g) 2.56
- (h) -2.56

14. If the average Greek adult male is 177 cm, and the standard deviation for heights is 7 cm, about what percentage of Greek adult males will be 183 cm or taller? (Assume heights are normally distributed.)

- (a) $< 0.1\%$
- (b) 12.2%
- (c) 19.6%
- (d) 32.7%
- (e) 68.3%
- (f) 80.4%
- (g) 88.8%
- (h) $> 99.9\%$

15. Using the information from Question 14, convert the height of a 190 cm Greek man to a z -score.

- (a) 0.00
- (b) 0.75
- (c) 0.93
- (d) 1.06
- (e) 1.50
- (f) 1.86
- (g) 2.12
- (h) 91.00

16. Suppose that the heights of kindergarten children can be described by a normal model with a mean of 38.2 inches and standard deviation of 1.8 inches. Between what two heights do you expect to find the middle 90% of kindergarteners?

- (a) 35.0 to 45.0 inches
- (b) 36.4 to 38.2 inches
- (c) 38.2 to 40.0 inches
- (d) 36.4 to 40.0 inches
- (e) 35.2 to 38.2 inches
- (f) 38.2 to 41.2 inches
- (g) 35.2 to 41.2 inches
- (h) 33.0 to 42.0 inches

17. A study examining the health risks of smoking measured the cholesterol levels of smokers and of ex-smokers.

Smokers				Non-smokers			
225	211	209	284	250	134	300	249
258	216	196	288	213	310	175	174
250	200	209	280	160	188	321	213
225	256	243	200	257	292	200	271
351	155						

According to the median and quartile numbers for each (as calculated by your TI), which variable has a higher center? Which has a greater spread? Consider a difference of < 5 to be insignificant for this problem.

- The difference between the centers and spreads are both < 5 .
- The difference in centers is < 5 , but the spread of the smokers is significantly greater.
- The difference in centers is < 5 , but the spread of the non-smokers is significantly greater.
- Smokers have significantly higher center, but the difference between the spreads is < 5 .
- Non-smokers have significantly higher center, but the difference between the spreads is < 5 .
- Smokers have significantly higher center and significantly greater spread.
- Smokers have significantly higher center, but non-smokers have significantly greater spread.
- Non-smokers have significantly higher center, but smokers have significantly greater spread.
- Non-smokers have significantly higher center and significantly greater spread.

18. In the data from Question 17, find the distance between the endpoints of the lower and upper whiskers in the boxplot associated to smokers. (Careful, this is not the same as the distance between the ‘fences’.)
- (a) 92
 - (b) 133
 - (c) 150
 - (d) 155
 - (e) 166
 - (f) 196
 - (g) 200
 - (h) 232
19. Examine the data from Question 17 again. According to the mean and standard deviation numbers (as calculated by your TI), which variable has a higher center? Which has a greater spread? Consider a difference of < 5 to be insignificant for this problem.
- (a) The difference between the centers and spreads are both < 5 .
 - (b) The difference in centers is < 5 , but the spread of the smokers is significantly greater.
 - (c) The difference in centers is < 5 , but the spread of the non-smokers is significantly greater.
 - (d) Smokers have significantly higher center, but the difference between the spreads is < 5 .
 - (e) Non-smokers have significantly higher center, but the difference between the spreads is < 5 .
 - (f) Smokers have significantly higher center and significantly greater spread.
 - (g) Smokers have significantly higher center, but non-smokers have significantly greater spread.
 - (h) Non-smokers have significantly higher center, but smokers have significantly greater spread.
 - (i) Non-smokers have significantly higher center and significantly greater spread.

Problems 20-21: 1, 0, or -1? Each problem is worth 3 points.

20. For the population of high school seniors, you find the correlation coefficient between SAT and ACT scores. You expect it to be:

- (a) Near 1.
- (b) Near 0.
- (c) Near -1.

21. For a population of adults between the ages of 20 and 60, you find the correlation coefficient between age and shoe size. You expect it to be:

- (a) Near 1.
- (b) Near 0.
- (c) Near -1.

Problems 22-30: True/false. Each problem is worth 2 points.

In the following 4 problems, consider the following gas prices (in dollars) collected from 7 stations around St. Louis.

2.34	2.38	2.40	2.53
2.35	2.38	2.45	2.59
2.36	2.38	2.48	2.59
2.36	2.39	2.52	2.73

You may find it helpful to enter these numbers in your TI and view a histogram, boxplot, etc.

22. True/false: Gas price is a quantitative variable.

- (a) True
- (b) False

23. True/false: The gas price value of \$2.73 is an outlier, according to Tukey's boxplot heuristic.

- (a) True
- (b) False

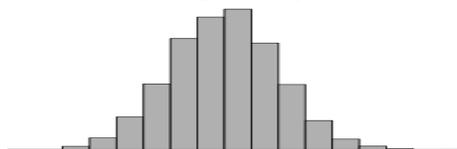
24. True/false: The gas price variable has a symmetric distribution.

- (a) True
- (b) False

25. True/false: Gas price is a unimodal variable.

- (a) True
- (b) False

26. True/false: The normal model is appropriate for the data represented by the following histogram.



- (a) True
- (b) False

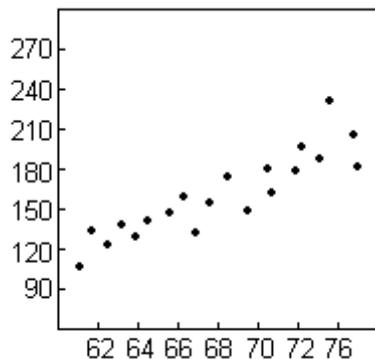
27. True/false: It is appropriate to use the normal model for a variable with a symmetric, bimodal distribution and no outliers.

- (a) True
- (b) False

28. True/false: A survey of elementary school children finds a strong positive correlation between height and reading proficiency, which is matched by a linear association in the scatterplot. We conclude that taller children have an easier time reaching their books, helping them read.

- (a) True
- (b) False

29. True/false: It is appropriate to summarize the strength of the association between weight and height shown in the below scatterplot with a correlation.



- (a) True
- (b) False

30. True/false: A college admissions office is examining whether high school SAT scores predict freshman grades. When they scatterplot SAT scores versus freshman GPA, they will put the SAT scores on the x -axis.

- (a) True
- (b) False