

Math 2200

Midterm Examination 2 – March 16, 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 1-20: Multiple choice. Each problem is worth 4 points.

1. A linear model for predicting calories from sugar in breakfast cereals has the form

$$\widehat{\text{Calories}} = 89.5 + 2.50 \cdot \text{Sugar}.$$

Find the residual with respect to this model for “Frosted Sugar Bombs”, which has 120 calories and 16 grams of sugar.

- (a) -129.5
- (b) -120
- (c) -12
- (d) -12.95
- (e) -9.5
- (f) 12
- (g) 30.5
- (h) 129.5

2. You pick a random number from the standard normal model. What is the expected value?
- (a) -1
 - (b) -0.68
 - (c) -0.57
 - (d) 0
 - (e) 0.57
 - (f) 0.68
 - (g) 1
 - (h) None of the above
3. The linear regression relating mortality to calcium in the drinking water in towns in England is

$$\widehat{\text{Mortality}} = 1676.0 - 3.23 \cdot \text{Calcium},$$

where mortality is in deaths per 100,000 residents and calcium is in parts per million. The town of Norwich has 125ppm of calcium in the drinking water. Predict the mortality.

- (a) 789.50
- (b) 1133.25
- (c) 1272.25
- (d) 1353.00
- (e) 1500.00
- (f) 1673.00
- (g) 1676.00
- (h) 2079.75

4. The table below shows the oil production of the United States from 1996 to 2000 (in thousands of barrels per year). Find the equation of the regression line predicting oil production from year.

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

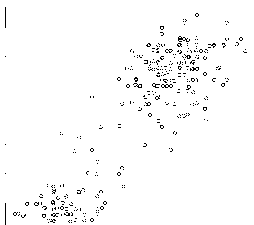
- (a) $\widehat{\text{Oil}} = 2,366,017 - 67,000.9 \cdot \text{Year}$
(b) $\widehat{\text{Oil}} = 136,124,710.4 - 67,000.9 \cdot \text{Year}$
(c) $\widehat{\text{Oil}} = 2,366,017 - 50,000.0 \cdot \text{Year}$
(d) $\widehat{\text{Oil}} = 136,124,710.4 - 50,000.0 \cdot \text{Year}$
(e) $\widehat{\text{Oil}} = -136,124,710.4 + 67000.9 \cdot \text{Year}$
(f) $\widehat{\text{Oil}} = 2028.79 - 0.000014 \cdot \text{Year}$
(g) $\widehat{\text{Oil}} = 2028.79 + 0.000014 \cdot \text{Year}$
(h) $\widehat{\text{Oil}} = 2028.79 - 2.3 \cdot \text{Year}$
5. Given the model $\log \hat{y} = 6.1 - 0.25x$, predict y for $x = 10$. (Here $\log \hat{y} = \log_{10} \hat{y}$.)
- (a) 0.6
(b) 1.3
(c) 3.6
(d) 8.6
(e) 25.6
(f) 51.6
(g) 3981.1
(h) 5431.7

6. Identify the high leverage outlier among the following table showing the relationship between female life expectancy and the average number of children women give birth to in countries in the Western Hemisphere.

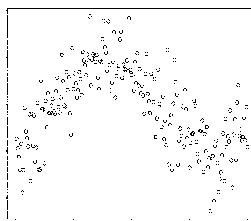
Country	Births/woman	Life exp.	Country	Births/woman	Life exp.
Argentina	2.5	77	Guatamala	4.7	68
Bahamas	2.2	77	Honduras	4.0	72
Barbados	1.8	78	Mexico	2.8	75
Canada	1.5	82	Peru	3.1	71
Costa Rica	25.0	79	Puerto Rico	3.1	71
Ecuador	3.1	71	USA	2.1	80

- (a) Argentina
 - (b) Canada
 - (c) Costa Rica
 - (d) Ecuador
 - (e) Guatamala
 - (f) Mexico
 - (g) Puerto Rico
 - (h) USA
7. Identify whether the point from Problem 5 is influential or not, and describe the best approach from the following to handling the point in devising a linear model description of the data.
- (a) Not influential, leave it in.
 - (b) Not influential, take it out.
 - (c) Influential. Create a linear regression with and without it.
 - (d) Influential. Take it out because it's a poor country.
 - (e) Influential. Take it out because it's a rich country
 - (f) Influential. Take it out because it's an error in the data.
 - (g) Influential. Take it out because of imperialist U.S. policy.
 - (h) Influential. Leave it in because of imperialist U.S. policy.

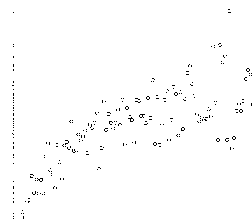
8. Which of the following 3 scatterplots may be analyzed with linear regression techniques, possibly including reexpression of variables?



I



II



III

- (a) I, II, and III.
 (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.
9. Predict the additional years of life for a 24 year old black man, according to the following table and the most appropriate reexpression.

Age	10	20	30	40	50	60	70	80	90	100
Years Left	59	50	41	32	24	17	12	7	4	3

- (a) 44.3
 (b) 44.5
 (c) 44.7
 (d) 44.9
 (e) 45.0
 (f) 45.3
 (g) 45.6
 (h) 45.9

10. Which linear model is described by the following R output?

```
Call: lm(formula = log(Pressure) ~ log(Height))
Coefficients: (Intercept)  log(Height)
              7.254          -1.001
```

- (a) $\widehat{\text{Pressure}} = 7.254 - 1.001 \cdot \text{Height}$
- (b) $\widehat{\text{Pressure}} = -1.001 + 7.254 \cdot \text{Height}$
- (c) $\sqrt{\widehat{\text{Pressure}}} = 7.254 - 1.001 \cdot \text{Height}$
- (d) $\sqrt{\widehat{\text{Pressure}}} = -1.001 + 7.254 \cdot \text{Height}$
- (e) $\log \widehat{\text{Pressure}} = 7.254 - 1.001 \cdot \text{Height}$
- (f) $\log \widehat{\text{Pressure}} = -1.001 + 7.254 \cdot \text{Height}$
- (g) $\log \widehat{\text{Pressure}} = 7.254 - 1.001 \cdot \log \text{Height}$
- (h) $\log \widehat{\text{Pressure}} = -1.001 + 7.254 \cdot \log \text{Height}$

11. You are interested in finding out the level of support among Washington University students for the proposed Metro sales tax, so you go to two cafeterias over one lunchtime, and interview every 10th person in line at each one. Which, if any, of the following sampling techniques are you using here?

- I. Stratification
- II. Clustering
- III. Simple random sample

- (a) I, II, and III.
- (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.

12. You are still interested in finding out the level of support among Washington University students for the proposed Metro sales tax. You randomly select 50 men and 50 women from the population of Washington University, and systematically go through the directory to find their e-mail addresses. You email them a survey, and follow up to those who haven't responded with an additional email reminder. Which, if any, of the following sampling techniques are you using here?

I. Stratification

II. Clustering

III. Systematic sampling

- (a) I, II, and III.
- (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.

13. Assume that birthdays are uniformly distributed throughout the year, and that no one has a leap year birthday. What is the probability that at least one of a randomly chosen group of 80 people has a July 4 birthday?

- (a) Nearly 100%
- (b) 92%
- (c) 80%
- (d) 60%
- (e) 40%
- (f) 20%
- (g) 16%
- (h) Nearly 0%

14. Lars is flying from St. Louis to Charlotte, with a connection in Atlanta. The probability that his first flight leaves on time is 0.25. If the flight is on time, the probability that his luggage makes the connecting flight in Atlanta is 0.97, but if the first flight is delayed, the probability that his luggage makes it is only 0.62. If Lars's luggage did not make it to Charlotte, what is the probability that his first flight was delayed?

- (a) 0.992
- (b) 0.974
- (c) 0.970
- (d) 0.750
- (e) 0.657
- (f) 0.613
- (g) 0.380
- (h) 0.250

15. In a group of 12 batteries, 5 are dead. You choose 2 batteries at random. What is the expected number of good batteries?

- (a) 1.432
- (b) 1.318
- (c) 1.167
- (d) 1.152
- (e) 1.000
- (f) 0.848
- (g) 0.833
- (h) 0.694

16. Given two independent trials X_1 and X_2 of a random variable X with expected value 10 and standard deviation 2, find the expected value and standard deviation of $2X_1 + X_2$.

(a) $\mu = 30, \sigma = 6.0$

(b) $\mu = 30, \sigma = 4.5$

(c) $\mu = 30, \sigma = 3.2$

(d) $\mu = 30, \sigma = 2.5$

(e) $\mu = 20, \sigma = 6.0$

(f) $\mu = 20, \sigma = 4.5$

(g) $\mu = 20, \sigma = 3.2$

(h) $\mu = 20, \sigma = 2.5$

17. You shuffle a (standard 52-card) deck of cards, and draw 5 cards, one at a time, from the top of the deck (without replacement). All 5 of them are red. What is the probability that the next two cards you draw are red?

(a) 36%

(b) 31%

(c) 28%

(d) 25%

(e) 19%

(f) 16%

(g) 8%

(h) 5%

A 2008 poll from the Harvard School of Public Health asked a sample of 1000 American adults if they thought that if we had socialized medicine in this country, whether the health care system would be better or worse than what we had now. The results were (by political party) as follows:

	Better (%)	Worse (%)	Other response (%)
Republican	17	70	13
Democrat	70	16	14
Independent	43	38	19

Assume that 35% of the sample were Democrats, 32% were Republicans, and 33% were independent.

18. Given a randomly selected person from the sample, what is the probability that she thinks socialized medicine would improve our health care system?
- (a) 70%
 - (b) 47%
 - (c) 44%
 - (d) 43%
 - (e) 42%
 - (f) 39%
 - (g) 38%
 - (h) 17%

19. Given that a randomly selected person from the sample thinks socialized medicine would worsen our health care system, what is the probability that she is a Republican?
- (a) 100%
 - (b) 70%
 - (c) 55%
 - (d) 51%
 - (e) 43%
 - (f) 32%
 - (g) 17%
 - (h) 13%
20. Suppose that a linear model predicts the price of a bushel of wheat from the yearly rainfall in inches. The standard deviation of the price of wheat is \$1.00, and the correlation is $r = -0.23$. Find the standard deviation of the residuals.
- (a) 1.23
 - (b) 1.00
 - (c) 0.97
 - (d) 0.95
 - (e) 0.88
 - (f) 0.77
 - (g) 0.48
 - (h) 0.05

Problems 21-30: True/false. Each problem is worth 2 points.

21. True/false: For each of 80 dining parties, a waiter randomly decides whether to give the party candy or not. This is a blinded experiment.
- (a) True
 - (b) False

22. True/false: Danish researchers followed a group of individuals born at a Copenhagen hospital between 1959 and 1961 for the following 40 years, and found that in this group adults who drank wine were richer and better educated than ones who did not. This is a blinded experiment.

(a) True

(b) False

23. True/false: In a box, you have a box of marbles, each of which is either red or blue, and either opaque or translucent. You draw a marble at random. This marble being red and being opaque are independent events.

	Opaque	Translucent
Red	12	28
Blue	18	42

(a) True

(b) False

24. True/false: Consider the event that randomly selected voter A favors the death penalty, and the event that randomly selectly selected voter B favors the death penalty. These events are independent.

(a) True

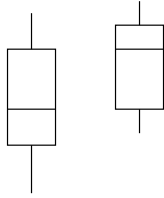
(b) False

25. True/false: Consider the event that randomly selected voter A favors the death penalty, and the event that randomly selectly selected voter B favors health care reform. These events are independent.

(a) True

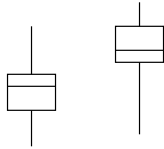
(b) False

26. True/false: The difference between the following two variables appears to be statistically significant.



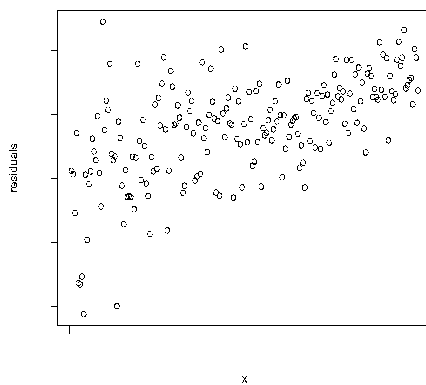
- (a) True
(b) False

27. True/false: The difference between the following two variables appears to be statistically significant.



- (a) True
(b) False

28. True/false: The following residual plot from a linear model would tend to indicate that the linear model was appropriate.



- (a) True
(b) False

29. True/false: A variable y is skewed, unimodal, and negative. This variable is a good candidate for reexpression to better fit the normal model.
- (a) True
 - (b) False
30. True/false: A variable y is skewed, bimodal, and positive. This variable is a good candidate for reexpression to better fit the normal model.
- (a) True
 - (b) False