Practice Final Statistics 0800 Fall 2013

Answer questions 1 through 6 after reading the article **Down syndrome link**: "Women who have trouble metabolizing the vitamin folic acid are at a higher risk of having children with Down Syndrome, a discovery by government researchers that raises the question of whether folic acid supplements might fight the syndrome. Mothers with a genetic abnormality that hinders how the body processes folic acid were 2.6 times more likely to have a child with Down syndrome than mothers without that genetic defect, concludes the study published yesterday in the American Journal of Clinical Nutrition. But the discovery by Food and Drug administration researchers is only one piece in the complicated puzzle of Down syndrome, cautioned FDA Commissioner Jane Henney. That's because millions of women appear to have this genetic abnormality, yet the risk of having a child with Down syndrome actually is small—one in 600 births. So something else has to help trigger the devastating condition.

1. This was (Circle one): (a) a sample survey (b) an experiment (c) an observational study

2. Identify each of these Critical Components, or write "?" if it cannot be determined from the article. Whenever possible, quote a key word or phrase directly from the article. Be brief!

   (a) source _____

   (b) researchers _____

   (c) individuals studied _____

   (d) measurements/questions: what are the two variables of interest here? _____
   _____

   (e) setting in which measurements were taken _____

   (f) What groups are compared?

      i. _____
      ii. _____

   (g) magnitude of claimed differences _____

3. The number of mothers studied is not mentioned, but we can assume it was (a) less than 100 (b) between 100 and 1000 (c) more than 1000

4. Which one of the following most warrants further research? (a) possible confounding variables (b) possible interacting variables (c) possible causation in the reverse direction (d) the placebo effect

5. Ability to process folic acid plays the role of (a) explanatory variable (b) response variable (c) confounding variable (d) interacting variable

6. Assume the researchers' data consist of proportions of Down syndrome births in mothers with and without the genetic abnormality. What would be the best picture to display the data? (a)a single piechart (b) bargraph (c) multiple boxplots (d) scatterplot (e) histogram

7. A scale is used in a doctor's office to measure patients' weights.

    (a) If the patients are always weighed while wearing heavy coats and shoes, the measurements are (i) invalid (ii) unreliable

    (b) If the doctor is mainly concerned with changes in weight, and the patients are sometimes weighed with coats and sometimes without, then the measurements are (i) invalid (ii) unreliable

8. Consider this stemplot for durations in minutes of 30 recent films:

```
 7 7
 8
 9 333
10 1666
11 00002558
12 000055
13 000055
14 0
15
16 0
```

    (a) What is the value of the median?

    (b) The shape of the stemplot is (i) skewed left/low outliers (ii) fairly symmetric (iii) skewed right/high outliers

    (c) Considering the shape of the stemplot, the mean should be (i) less than the median (ii) about equal to the median (iii) greater than the median [No calculations are necessary.]

    (d) What is the value of the first quartile?

9. Suppose test-taking times follow a normal curve with mean 100 minutes and standard deviation 8 minutes.

    (a) What proportion of students are finished in more than 110 minutes?

    (b) The shortest 5% of times are less than how many minutes?

10. Age y of husbands was regressed on age x of wives, resulting in the equation $y = 3.6 + .97x$

    (a) Predict the age of the husband of a 55 year old wife.

    (b) According to the equation, older wives tend to have husbands who are (i) younger (ii) older (iii) can't say.

11. A study found that the closeness x of a child's relationship with his or her father was related to higher academic performance y.

For each of the possible explanations (a) through (g), circle the "reason for a relationship" that applies; each number 1 through 5 should be used exactly once.

(a) A child's doing well academically causes his or her father to form a closer relationship.

1   2   3   4   5

(b) Data from previous years had more children enjoying a good relationship with their fathers, and academic scores were higher at that time

1   2   3   4   5

(c) Having a good relationship with dad can lead to higher academic performance.

1   2   3   4   5

(d) The study just happened to include quite a few good students who had close relationships with their fathers.

1   2   3   4   5

(e) Gender plays a role in a child's relationship with his or her father and is also tied in with academic performance.

1   2   3   4   5

1. explanatory variable does contribute to changes in response variable

2. response variable is causing change in explanatory variable

3. confounding variables may exist

4. both variables are changing over time

5. the association may be nothing more than coincidence

12. Consider this two-way table classifying subjects in a study according to whether or not they were smokers, and whether or not they consumed enough alcohol to be considered "drinkers":

Observed	Smoker	non-Smoker	Total
Drinker	180	180	360
non-Drinker	90	150	240
Total	270	330	600

Expected	Smoker	non-Smoker
Drinker		
non-Drinker		

Compare	Smoker	non-Smoker
Drinker		
non-Drinker		

(a) State the null and alternative hypotheses to test for a relationship between smoking and drinking.

(b) Compute the numbers expected if there were no relationship.

(c) Compare the observed and expected numbers by computing each $\frac{(observed-expected)^2}{expected}$ rounded to the nearest tenth (one decimal place).

(d) Compute the chi-squared statistic.

(e) What does the chi-squared statistic tell you about the P-value?

(f) Is there statistical evidence of a relationship between smoking and drinking?

13. A certain brand of tire is supposed to last a mean of 22,000 miles, but a consumer advocacy group suspects they are not as durable as claimed. A random sample of 100 tires was found to last a mean of 21,800 miles, with standard deviation 1200 miles. Does this provide convincing evidence against the claim that the tires average 22,000 miles?

(a) State the null and alternative hypotheses.

(b) Calculate the test statistic.

(c) Find the P-value.

(d) Draw your conclusion; choose one of the following: (i) The tires may average 22,000 miles as claimed (ii) There is convincing evidence that the tires are not as durable as claimed (iii) The results are borderline.

14. In the U.S., the current probability of death by heart attack is .33; the probability of death by cancer is .20.

    (a) What is the probability of death by heart attack or cancer?

    (b) What is the probability of death by some other cause?

15. The probability of a Labor Department employee in 1995 being black was .34. The probability of being laid off was .05. The probability of being black and laid off was .03.

    (a) Were lay-offs independent of race?

    (b) What was the probability of being black or being laid off?

16. A family has 4 children.

    (a) What is the probability of 2 boys followed by 2 girls?

    (b) Which is most likely? (i) 4 boys (ii) 2 boys followed by 2 girls (iii) a total of 2 boys and 2 girls (iv) all the same

17. For an insurance company, the probability of gaining $500 from a customer in any given year is .9; the probability of having to pay $1500 is .1. What is the company's expected gain from a customer in any given year?

18. Pick a card from an ordinary deck of 52. Which of these is most likely? (a) getting a red queen (b) getting a queen (c) getting the queen of hearts

19. Viewers of the presidential debates tended not to change their minds about their preferred candidate: they would focus on his good points, and overlook his faults. This is an example of (a) anchoring (b) availability heuristic (c) representativeness (d) forgotten base rates (e) optimism (f) over-confidence (g) conservatism

20. A clothing store makes up price tags for sweaters with $60 printed in black; these are crossed off with a red pen and $25 is written underneath. The store is taking advantage of (a) anchoring (b) availability heuristic (c) representativeness heuristic (d) forgotten base rates (e) optimism (f) over-confidence (g) conservatism

21. In our survey, more students thought they were better than average drivers than worse than average. This is because of (a) anchoring (b) availability heuristic (c) representativeness heuristic (d) forgotten base rates (e) optimism (f) over-confidence (g) conservatism

22. Researchers found that people believed accidents and diseases to be equally probable causes of death in the U.S., whereas in fact death by disease is 16 times more likely than death by accident. The misperception is due to (a) anchoring (b) availability heuristic (c) representativeness heuristic (d) forgotten base rates (e) optimism (f) over-confidence (g) conservatism

23. Suppose a woman has a lump on her breast; the probability that it is malignant is 0.01. The probability of testing positive, given that the lump is malignant, is 0.8. The probability of testing positive when the lump is not malignant is 0.1.

   (a) What is the probability that the lump is malignant and the test is positive?

   (b) What is the probability that the lump is not malignant and the test is positive?

   (c) What is the overall probability of testing positive?

25. Assume the proportion of women in a population is .5. Consider the behavior of sample proportion of women in repeated samples of size 81.

   (a) What should be the mean (center) of all those sample proportions?

   (b) What should be the standard error (spread) of sample proportions?

   (c) What should be the shape of the frequency curve for all sample proportions?

   (d) 68% of the sample proportions should fall between _____and _____

   (e) 95% of the sample proportions should fall between _____and _____

   (f) 99.7% of the sample proportions should fall between _____and _____

26. Suppose a sample of 81 students registering for a particular course has 52 women.

   (a) What is the sample proportion of women in that course?

   (b) Calculate a 95% confidence interval for the proportion of women out of all students who would register for that course over many years' time.

27. I had been going under the assumption that men were just as likely as women to register for Stats 800, but maybe more women tend to register for the course because they major in areas for which it is a requirement.

   (a) State the null and alternative hypotheses to be tested.

   (b) Out of 81 students registering for the course this semester, 52 were women. Calculate the test statistic. (Remember to assume the null hypothesis to be true when you calculate the standard error.)

   (c) What is the P-value?

   (d) Draw your conclusion; choose one of the following:

      i. Overall, men are just as likely as women to register for this course; there just happened to be a higher proportion of women in this sample of 82.

      ii. The proportion of women is significantly higher than .5; they are more likely to register for this course than men are.

28. A random sample of 200 men had a mean height of 68.2 inches, with standard deviation 2.7 inches.

   (a) Give a 95% confidence interval for the mean height of all men in the population from which the sample was taken.

   (b) Interpret your confidence interval.

29. (5 pts.) Under what circumstances is it best to combine results for several studies?

   (a) When the conditions of the studies, including the individuals involved, are quite similar.

   (b) When the conditions of the studies, including the individuals involved, are quite different.

(c) Both of the above.

30. (5 pts.) Under what circumstances is it best to look at results for several studies separately?

    (a) When the conditions of the studies, including the individuals involved, are quite similar.

    (b) When the conditions of the studies, including the individuals involved, are quite different.

    (c) Both of the above.