

## Final Exam Preparation

### *Narrative: Grades*

Suppose a class of 100 students took their statistics final and their grades are shown in the table below.

A	B	C	D	F
25	28	34	10	3

1. {Grades narrative} Choose one student at random. What is the probability that he/she received a B or a C?
2. {Grades narrative} What is the probability that a student selected at random passed the final (where a D is considered to be a passing grade)?
3. In what sense is a coincidence not an improbable event?
4. Suppose that 20% of the individuals in a particular population own a cell phone. Suppose you took a random sample of size 2,500 from this population and found that 17.6% of them owned a cell phone. The two-sided p-value here is 0.006. Do you consider this to be a reasonable value given the size of this sample? Is it plausible that the true population fraction really is 20%? Give as precise an answer as you can.

### Narrative: Politics

Suppose a population contains 60% Republicans and 40% Democrats.

5. {Politics narrative} Suppose you take a random sample of 10 people from this population. Are you certain that you would get 6 Republicans and 4 Democrats in your sample? Explain your answer. How surprised would you be if your sample contained 10 Republicans?

6. Explain (in words that a non-statistics student would understand) what is meant by a '95% confidence interval.'

7. Which would be wider, a 90% confidence interval or a 95% confidence interval? (Assume both of them were calculated using the same sample data.) Explain your answer.

8. Explain the difference between the statistical meaning of the word 'significant results' and the regular conversational meaning of the word 'significant results.'

Narrative: Shaking hands

Suppose the chances of picking up a cold from someone by shaking hands with them is 0.01 (assuming you don't know whether they have a cold or not), and that each encounter you have is independent of other encounters.

9. {Shaking hands narrative}. Suppose you shake hands with 5 people in a given day. What is the probability that you don't pick up a cold from any of these people?

Narrative: Cell phone use

How do men and women compare when it comes to talking on the cell phone? Suppose you take a random sample of 100 male cell phone owners and a random sample of 100 female cell phone owners. The average number of minutes for the women per month was 280 with a standard deviation of 42; the average number of minutes for the men per month was 270 with a standard deviation of 30.

10. {Cell phone use narrative} Based on these sample results, a 95% confidence interval for the difference in average time spent on the cell phone for females versus males is (-0.3, 20.3). What does this mean? Give as precise an answer as you possibly can.

11. Suppose you suspect that a majority of the people in your neighborhood would like to start a neighborhood watch program, but you want to conduct a hypothesis test to find out. What are your null and alternative hypotheses?

12. Large samples make it easier to detect real relationships or differences in the population than small samples. Explain how this is taken into account in the formula for the test statistic for testing a population mean.

Narrative: Tiger Woods

Washington Post columnist Sally Jenkins described an anecdote about golfer Tiger Woods' ability to detect subtle differences in golf equipment. Tiger Woods was sent six golf clubs to test. The six clubs looked identical, but one was heavier than the rest by two grams (about the weight of a dollar bill). Tiger Woods swung each club and quickly declared, "This one's heavier." He was right.

Suppose this basic experiment is carried out for a random sample of 25 professional golfers and five of them guess the correct club. A statistician carries out a significance test of the null hypothesis that these golfers were just guessing versus the alternative hypothesis that a greater proportion of professional golfers than expected under random chance can recognize the heaviest of six clubs.

15. {Narrative: Tiger Woods} The p-value here is about 0.2. What does that mean?

16. In testing hypotheses, what should we do if the consequences of rejecting the null hypothesis are very serious?

17. What should you conclude if a hypothesis test gives a p-value of 0.50?

18. What should you conclude if a hypothesis test gives a p-value of 0.03?

Narrative: Blood clot risk

Suppose you hear on the news that sitting on an airplane for five hours or more without getting up and walking around increases your chances of developing blood clots by a factor of 10.

19. {Blood clot risk narrative} The baseline risk is missing. Explain what that means in terms of what conclusions you can make about the results.

20. {Blood clot risk narrative} Does this increased risk apply to anyone and everyone? Explain why or why not.

21. Suppose a researcher wanted to find out which of two chemicals works best to kill poison ivy. He conducted a controlled experiment involving 200 poison ivy plants, randomly assigned half of them to each treatment (chemical #1 or chemical #2). Five days after applying the chemicals, he recorded the data below. The p-value from a chi-square test here is 0.36. What do you conclude?

	DIED	LIVED	TOTAL
CHEMICAL #1	78	22	100
CHEMICAL #2	89	11	100
TOTAL	167	33	200

#### MULTIPLE-CHOICE

Narrative: Genetic defect

Suppose a population contains 100,000 individuals, of which 500 are carriers of a certain genetic defect.

22. {Genetic defect narrative} Which of the following is an appropriate way to express this result in terms of odds?

- a. The odds of carrying the genetic defect are 5/995.
  - b. The odds of carrying the genetic defect are 500 out of 100,000.
  - c. The odds of carrying the genetic defect are 500 to 1.
  - d. None of the above
23. Which of the following statements is false?
- a. Sample results will always be very close to their respective population values.
  - b. Sample results vary from one sample to the next.
  - c. The key to interpreting statistical results is to understand what kind of dissimilarity we should expect to see in various samples from the same population.
  - d. None of the above statements are false.
24. Which of the following statements is not true regarding a 95% confidence interval for the mean of a population?
- a. In 95% of all samples, the sample mean will fall within 2 standard errors of the true population mean.
  - b. 95% of the population values will lie within 2 standard errors of the sample mean.
  - c. In 95% of all samples, the true population mean will be within 2 standard errors of the sample mean.
  - d. If you add and subtract two standard errors to/from the sample mean, in 95% of all cases you will have captured the true population mean.
25. Suppose a confidence interval for the difference in mean weight loss for two different weight loss programs (Program 1 – Program 2) is entirely above zero. What does this mean?
- a. We can't say with any confidence that there is a difference in mean weight loss for the populations of people on these two programs.
  - b. We can say with confidence that there is a difference in mean weight loss for the populations of people on these two programs; further, we can say that the average weight loss on Program 1 is higher.
  - c. We can say with confidence that there is a difference in mean weight loss for the populations of people on these two programs; further, we can say that the average weight loss on Program 2 is higher.

d. None of the above.

26. Which of the following conclusions do you draw if the p-value is smaller than the level of significance?

a. Reject the null hypothesis.

b. Accept the alternative hypothesis.

c. The true population value is significantly different from the value in the null hypothesis.

d. All of the above.

27. Suppose the p-value for your hypothesis test for a difference of two means was .001. Which of the following is an appropriate conclusion?

a. You reject the null hypothesis.

b. If there really was no difference in the population means, we would see results as extreme as these only .1% of the time.

c. You conclude that there is a statistically significant difference between the population means.

d. All of the above.

28. What do we call a hypothesis test when the alternative hypothesis includes a possible difference in either direction?

a. A two-direction test.

b. A two-sided test.

c. A test that is not possible to conduct.

d. None of the above.

29. Which of the following statements is false?

a. If the sample size is large enough, almost any null hypothesis can be rejected.

b. There is almost always a slight relationship between two variables, or a difference between two groups, and if you collect enough data, you will find it.

- c. Larger samples sizes always produce more meaningful results than small sample sizes.
- d. None of the above statements are false.

16. Which of these is an example of Simpson's paradox?

- a. Teachers' salaries and sales of alcoholic beverages have risen together over time, but paying teachers more does not cause higher alcohol sales.
- b. Alaska Air has a lower percent of late flights than America West at every airport, but America West has a lower percent when we combine all airports.
- c. The percent of surgery patients given Anesthetic A who die is higher than the percent for Anesthetic B, but this is because A is used in more serious surgeries.
- d. States in which a smaller percent of students take the SAT exam have higher median scores on the SAT.