

## AP Statistics Practice

1) In hilly areas farmers often contour their fields to reduce the erosion due to water flow. This might have unintended effect of changing the yield since the rows may not be aligned in an east-west direction to take advantage of direct sunlight.

The following summary statistics show the yields in hundreds of bushels per acre of randomly selected plots of land along hillsides as well as level plots of land nearby with similar soil characteristics.

Level	Minimum	Q1	Median	Q3	Maximum	Mean	Std Dev	Number
Hilly	10.9	13.3	29.2	58.1	100	35.7	20.4	12
Level	10.9	18.5	39.3	61.5	118	48.6	22.1	12

- Draw side-by-side boxplots that compare the yields for the two types of land.
- Based on your boxplots drawn in part (a), write a few sentences comparing the distribution of corn yields for level and hilly areas.
- Suppose you wanted to determine if there was a significant difference in the corn yields between the two types of land, level and hilly. What statistical test would be most appropriate to assess this difference? Justify your choice.
- Given your choice of test in part (c), are the conditions for conduction that test met? Explain your reasoning.

2) In a recent survey of 1500 randomly selected U.S. adults, 68% of the respondents agreed with the statement, "I should exercise more than I do."

- Construct and interpret a 96% confidence interval to estimate the proportion of the U.S. adult population that would agree with this statement.
- For this study, state one source of potential bias and how it would affect the estimate of the proportion of adults who would agree with the statement, "I should exercise more than I do."

3) The administrators of a very large high school have installed new drink vending machines in the cafeteria, in the main entrance to the school, and in the lobby area outside the gym. They would like to know if a student's beverage choice depends on location. A random sample of students over a one-week period was selected and interviewed. The results are shown in the table.

	Location		
Beverage	Cafeteria	Main Entrance	Gym
Soda	2	8	41
Juice	19	5	1
Water	49	9	19

- What should the administrators conclude? Support your answer with appropriate statistical evidence?
- Is  $71/183$  a reasonable estimate of the proportion of all students in the high school who buy their beverages in the lobby of the gym? Explain your reasoning.

4) A candy manufacturer is marketing a gift box containing four cream-filled nuggets and five pieces of fudge. The manufacturing process for each candy is designed so that the mean weight of a cream-filled nugget is 3 ounces with a standard deviation of 0.2 ounces and the mean weight of a piece of fudge is 4 ounces with a standard deviation of 0.3 ounces. The boxes have a mean weight of 3 ounces with a standard deviation of 0.1 ounce.

- Assuming that the weights of the boxes, nuggets, and fudge are independent, what is the mean and standard deviation of the weight of the box of candy?
- Assuming that the weights of the boxes, nuggets, and fudge are approximately normally distributed, what is the probability that a randomly selected box of candy will weigh less than 34 ounces?
- If three boxes are selected at random, what is the probability that at least one of the boxes will weigh less than 34 ounces?
- Determine the probability that a random sample of three boxes will have a mean weight of less than 34 ounces.

5) Homing pigeons, when released many miles away from their coop, have the ability to find their way back home. Scientists speculate that pigeons use the sun as a means of navigating back to their coop. But

there are some experts who think that pigeons use some other way of determining their way back home, citing individual stories of pigeons finding their way home on sunless days.

Suppose that you had 80 randomly selected pigeons of the same age that were raised in the same coop but had never been released from beyond visual distance of the coop. You may assume that all pigeons are healthy and are physically able to fly the required distance.

Two plans are proposed for conducting the experiment.

### **Plan 1**

Number the pigeons from 1 to 80. Randomly assign 40 to one group and the remaining 40 to a second group. On a sunny day, release the pigeons one at a time from a location that is well beyond the visual range of the coop. Tally how many of these pigeons return to the coop within a day. On a cloudy day release the other 40 pigeons from the same location as the first group. Again, tally how many pigeons return within a day. Compare the proportion of pigeons that returned under the two different conditions.

### **Plan 2**

Number the pigeons from 1 to 80. Randomly assign 40 to one group and the remaining 40 to a second group. On a sunny day, from a location that is well beyond the visual range from the coop, release each numbered pigeon at a specified time and measure how long it takes for that numbered pigeon to return to the coop. Repeat this procedure for each pigeon in the group. On a cloudy day do the same for the other 40 pigeons. Compare the mean time to return to the coop for each group.

- (a) Which of the two plans described above is the more optimal plan? Justify your answer by comparing and contrasting the two plans.
- (b) For the plan you chose in part (a), state which statistical procedure you would use to test whether or not sunny or cloudy conditions affect pigeons finding their way back to their home coop. Define appropriate hypotheses for your test. (Do not do the test)
- (c) For your choice in part (b), what are the necessary conditions for conducting that test? Justify your answer.

**6)** Many people, when confronted by an embarrassing situation, will sometimes try to evade giving a direct and truthful answer when questioned about it. In order to get truthful responses to potentially embarrassing questions, researchers often use a technique called randomized response. They ask the embarrassing questions in conjunction with some other activity. For example, people will be asked to toss a coin. Then they randomly respond to either a decoy question, such as, “Did you obtain a head on the coin toss?” or a real question, such as “Have you ever stolen something from a store?” When a subject answers “yes,” the researcher does not know to which of the two questions the “yes” is referring, so the potential embarrassment of the respondent is avoided and a truthful answer can be more readily obtained.

Suppose that a large group of adults is interviewed to determine the proportion of respondents who have in some way not been completely honest on their income taxed in the past. Each person is given a die and placed in a private room out of eyesight of the researcher. The instructions are as follows: Roll the die and note the number of dots that turn up. Once the roll is completed the person is asked, “If you rolled either a 3 or 4 OR you have cheated on your income tax, answer BLUE. Otherwise answer YELLOW.”

- (a) Suppose that 20% of the people answer YELLOW. What is the approximate proportion of the adults in this survey that have cheated on their income tax?
- (b) Suppose that the results from part (a) came from a random sample of 90 adults. What is the 96% confidence interval for the true proportion of adults who have cheated on their income tax? Be sure to interpret your interval.

**7)** A certain city wants to set up a citywide program to recycle newspapers, glass, and plastics as a way to reduce the amount of landfill, preserve the environment, and ultimately save the city some money. An analyst for one of the companies interested in bidding on the recycling contract has computed that they would be able to make a profit if the mean weekly household contribution to recycling exceeded 5.5 pounds. If this company can make a profit, then it is financially feasible for them to spend the money to build a recycling plant, set up a collection system, and hire the necessary workers.

- (a) Identify appropriate hypotheses for this situation.
- (b) Describe a Type I and a Type II error in this setting.
- (c) Describe the economic consequences to the company for each of these types of errors.

**8)** Extremely smoky dwellings are common in the world’s underdeveloped countries, where open firepits in the home traditionally provide heat for cooking and warmth. This is thought to increase the risk of

developing lung cancer. For decades many countries have encouraged their citizens to install stoves with chimneys that funnel smoke away from living spaces. It was thought that these vented stoves generated substantially less soot and other airborne particulate matter than the traditional open firepits did. To determine whether the switch to vented stoves decreased lung cancer rates, a study was conducted in one such country. Data were gathered in 1990 in one region of this country by randomly selecting 1200 homes from that region that had switched to the vented stoves and 1300 homes that had not switched. Data were gathered on both men and women in the study. Between 1980 and 1996, 63 nonsmoking men whose homes used vented stoves and 96 nonsmoking men whose homes had not been converted to those stoves developed lung cancer.

- Is this an experiment or an observational study? Explain your answer.
- Was the program effective in decreasing lung cancer rates among men? Support your answer with appropriate statistical evidence.
- Can we conclude that using vented stoves rather than open firepits would reduce the incidence of lung cancer in all converted homes in this country? Justify your answer.

**9)** The manufacturer of exercise machines for fitness centers has designed two new elliptical machines that are meant to increase cardiovascular fitness. The two machines are being tested on 30 volunteers at a fitness center near the company's headquarters. The volunteers are randomly assigned to one of the machines and use it daily for two months. A measure of cardiovascular fitness is administered at the start of the experiment and again at the end. The table below contains the differences in the two scores (After – Before) for the two machines.

Machine A		Machine B
	5	3 5 9
6 1	4	2 5 7
9 7 4 1 1	3	2 4 8 9
8 7 6 3 2 0	2	1 5 9
5 4	1	0
	0	2

- Write a few sentences comparing the distributions of cardiovascular fitness gains from the two elliptical machines.
- Which machine should be chosen if the company wants to advertise it as achieving the highest overall gain in cardiovascular fitness? Explain your reasoning.
- Which machine should be chosen if the company wants to advertise it as achieving the most consistent gain in cardiovascular fitness? Explain your reasoning.
- Give one reason why the advertising claims of the company (the scope of inference) for this experiment would be limited. Explain how the company could broaden that scope of inference.

**10)** The manufacturer of a large tank, which is located next to a window, are 36 guppies of various sizes and ages. The length of each guppy is measured and recorded. A researcher drops a dividing wall down the middle of the tank, trapping half the guppies on each side. The researcher flips a coin to decide which side of the tank gets an extra portion of food each day. The other side of the tank gets a standard-sized portion of food. After four weeks the length of each guppy is again measured and recorded. The guppies that received an extra portion of food each day averaged 10% greater length than those in the other group.

- What was (were) the experimental unit(s) in this experiment?
- How was randomization used in this experiment? Explain your answer.
- Given the goal of the experiment was to compare the length of guppies given an extra portion of food daily to those who received a standard amount of food, was replication used properly in this experiment? Justify your answer.
- Explain the concept of confounding and include an example of confounding in the context of this situation.
- Can we conclude that getting an extra portion of food each day led to an increase in the lengths of the guppies? Explain your reasoning.

**11)** Lack of health insurance is a major worry for many people in this country. A recent telephone survey by a national public health foundation asked 6000 adults a series of questions about issues as: whether they and/or their children had health insurance coverage, who provided such coverage (their employer, paid it themselves, Medicaid, Medicare, or some other public agency), the sex and age of the respondent. The telephone survey was conducted in all 50 states and the District of Columbia.

The foundation wanted to know if the lack of health insurance was related to the region of the country where a person lived. If that is the case, the foundation wants to concentrate its resources in the region that had significantly fewer people insured. The table below summarizes the sex of the respondent and the U.S. Census region in which they lived for those in the sample who indicated they lacked health insurance. This amounted to 16% of the female and 19% of the male respondents in the survey.

	Northeast	Midwest	South	West	Total
Male	118	120	219	115	575
Female	93	119	150	117	476
Total	201	229	319	222	1050

- What is the probability that a randomly selected adult who lacks health insurance was a female or was from the Midwest?
- What is the probability that a randomly selected male who lacked health insurance was from the south?
- If a chi-square test for independence were conducted, what would be the expected count for the cell female and west?
- Define an appropriate pair of hypotheses that the foundation could test.
- The foundation conducts a chi-square test for independence since all the conditions appear to be met and gets a test statistic of 7.717. What decision should the foundation make about targeting its resources? Justify your answer.