HW12 - Tests of Significance

Stat 20 & 131A, Spring 2017, Prof. Sanchez

Due Apr-29

Most of the questions in this HW involve performing a hypothesis test. When performing such a test, please follow the following indications:

- The null should include:
  a. a statement about chance like "the difference is due to chance"
  b. something about the problem determining the EV like "boxes have same %".
- The alternative should be similar, like:
  a. not due to chance
  b. a statement about the problem specifying something like "too many reds"
- Conclusions should include:
  a. jargon like "reject null at 5% level" or "statistically significant"
  b. something like "doesn't seem to be chance", or "can be explained by chance".
  c. something about the problem like "possibly something else"

---

1) Each (hypothetical) data set below represents some readings on span gas. Assume the Gauss model, with errors following the normal curve. However, bias may be present. In each case, calculate a t-statistic, and determine a value or a range for the p-value to see whether the instrument is properly calibrated or not. $_{1pt}$

a. 71, 68, 79

b. 71, 68, 79, 84, 78, 85, 69

c. 71, 84

d. 71

2) Five hundred draws are made at random with replacement from a box of numbered tickets; 276 are positive. Someone tells you that 50% of the tickets in the box show positive numbers. Do you believe it? Test the hypothesis at the 5% significance level. $_{1pt}$

3) A study of the GPAs of students at UC Berkeley found that the average GPA was a 3.0 with an SD of 0.3. You wish to test whether students who go to tutoring sessions have a larger average GPA. You take a simple random sample of 10 students from a tutoring service office, and find that the average was a 3.1 with an SD of 0.4. Assume GPA's roughly follow the normal curve. Test the hypothesis at the 5% significance level. $_{1pt}$

4) Repeat the previous test but now assume that the sample statistics (average and SD) comes from a RSR of size 100 students. $_{1pt}$

5) The census data states that one city has 52% women, 48% men, an average income of \$45,000 with an SD of \$30,000, and an average weight of 150 lbs, with an SD of 10. Assume that weights are normally distributed and that incomes are not. For all problems below, determine which test to use, and what type of SD would you use. $_{1pt}$

a. You wish to test if the percent of women is correct, against the alternate that there is a lower percentage of women. You take a simple random sample of 10 people and find that 50% are women.

b. You wish to test if the income is correct, against the alternate that the average is actually larger. You take a simple random sample of 10 people, which has an average \$50,000 with an SD of \$20,000.

c. You wish to test if the weight is correct, against the alternative that the average is actually larger. You take a simple random sample of 10 people, which has an average weight of 155 lbs, with an SD of 15.

d. You want to know if a district in the city has a larger average weight. You take a simple random sample of 10 people from the district, which has an average 155 with an SD of 15. (Note this problem has the same numbers as part c.)

6) Refer to the previous question. If we conduct the above tests with samples of 100 people instead of 10, all tests are z-tests. In that case, find the SD of the box for each of the above situations. $_{1pt}$

7) You have two bags of 1000 M&M's each. You wish to know if the proportion of green M&M's is the same in both bags. In the first bag, you pull out 75 at random and counted 10 green M&M's. In the second, you pulled out 65 at random and also got 10 green M&M's. Using a significance level of 5%, test whether the proportion of green M&M's is the same in both bags. $_{1pt}$

8) Suppose you take one of your bags of M&M's and draw 80 M&M's. This is a special holiday bag of M&M's so it only has Red and Green M&M's. In that sample there were 35 green M&M's and 45 red M&M's. Test at significance level 5% whether the proportion of green M&M's is the same as the proportion of red M&M's. $_{1pt}$

9) In 1990 and 2004, the National Assessment of Educational Progress (NAEP) tested 17-year-olds on a math test. The average went up from 305 to 307, by taking a simple random sample of 1000 people in each of the two years. The SD for the 1990 data was 34, and the SD for the 2004 data was 27. Can the difference be explained by chance variation? Use 1% significance level. $_{1pt}$

10) In a class of 500, the professor randomly chooses 200 students to drink juice before the final while giving the other 300 students an indistinguishable energy drink. The group of students who drank juice got on average of 78% on their final with an SD of 16% while the students who received the energy drink scored an average of 75% with an SD of 13%. Can this difference be explained by chance? Use 5% significance level. $_{1pt}$