

MATH 143: Introduction to Probability and Statistics

Worksheet 9 for Thurs., Dec. 10: What procedure?

For each numbered problem, identify (if possible) the following:

- (a) the variable(s) and variable type(s) of interest.
 - (b) the type of inference procedure (1-sample z , 1-sample t , 1-proportion, 1-way ANOVA, 2-sample t , 2-way ANOVA, simple linear regression, multiple regression, etc.)
 - (c) Is an hypothesis test or a confidence interval what is called for?
 - (d) If an hypothesis test were performed using this data,
 - (i) what would be appropriate hypotheses?
 - (ii) what type of test statistic? (z ? t ? χ^2 ? Something else?)
 - (e) Is it possible to use this data to construct a confidence interval? If so,
 - (i) for what population parameter (or combination of population parameters)?
 - (ii) what type of critical value would be used?
 - (iii) what formula for the appropriate standard error?
1. We have a sample of 35 frisbies for which the weight (in ounces) and distance (in feet) of flight when thrown by a mechanical arm are recorded. We wish to know a likely range of numbers that represent how that distance of flight changes when the weight of the frisbee is increased by 1 ounce.
 2. *Sports Illustrated* magazine surveyed a random sample of 757 Division I college athletes in 36 sports. One question asked was "Have you ever received preferential treatment from a professor because of your status as an athlete?" Of the athletes polled, 225 said "Yes." What value(s) do we think likely for the true percentage of athletes who believe they have received this kind of preferential treatment?
 3. When the new euro coins were introduced throughout Europe in 2002, curious people tried all sorts of things. Two Polish mathematicians spun a Belgian euro (one side of the coin has a different design for each country) 250 times. They got 140 heads. Newspapers reported this result widely. Is it significant evidence that the coin is not balanced when spun?
 4. The corn from three different seed types is planted in an experiment involving 5 different types of fertilizers. We wish to know if certain pairings of seed type–fertilizer type grow, on average, higher than others.
 5. The amount of lead in a certain type of soil, when released by a standard extraction method, averages 86 parts per million (ppm). A new extraction method was tried, with researchers wondering if this new approach would result in a significant difference in the mean amount of extracted lead. Forty one specimens were obtained, with a mean of 83 ppm lead and a s.d. of 10 ppm.

6. Many low- and middle-income families do not save enough for their retirement. It would be to their advantage to contribute to an individual retirement account (IRA), which allows money to be invested for retirement without paying taxes on it now. Would more families contribute to an IRA if the money they invest were matched by their employer? In an experiment on this question, the tax firm *H&R Block* offered to partly match IRA contributions of families with incomes below \$40,000. In all, 1681 married taxpayers were assigned at random to the control group (no match), 1780 to a 20% match, and 1831 to a 50% match. All were offered the opportunity to open an IRA. The study found that 49 married taxpayers in the control group, 240 in the 20% group, and 456 in the 50% group opened IRAs. Is there strong evidence that people in these groups decide differently about opening IRA accounts?
7. Consider the *H&R Block* study described above. We wish to estimate the difference between the percentages of people opening an IRA when offered a 50% match vs. those offered a 20% match. (Note: We are ignoring the control group here.)
8. Consider the same *H&R Block* study. Each taxpayer who opened an IRA decided how much to contribute. Those in the control group contributed \$1549 on average, with standard deviation \$1652. Those who were offered a 20% match contributed an average of \$1723 with s.d. \$1332, and those offered a 50% match contributed an average of \$1742 with s.d. \$1174. Is this strong evidence to conclude that people with these three different types of incentives for contributing to an IRA will make different decisions about how much to contribute?
9. With the data of the previous problem, estimate the difference in the average contributions among people who open an IRA under a 50% match program and those who do so under a 20% match program.

List the standard assumptions/conditions which validate

- I. inferences on a single (population) mean.
- II. inferences on the difference of two (population) means.
- III. inferences on a single proportion.
- IV. inferences on the difference of two proportions.
- V. inferences on the value of a model parameter in simple or multiple regression
- VI. inferences about whether there is some relationship between a collection of potential explanatory quantitative variables and some quantitative response variable.
- VII. our methods for determining if there is some relationship between two categorical variables.
- VIII. our methods for determining if there is some difference in population means amongst several different populations.

What does it mean to say a certain inference procedure is *robust* against a particular assumption? Go back through the list of conditions you made for the previous problem. Circle those assumptions against which the corresponding inference procedure is robust.