

Below you will find a series of examples. For each example, you should provide the following information:

- What units are being measured (or observed)?
- What variables are being measured? Are they categorical? binary? quantitative?
- Which is the response and which is the explanatory variable (if appropriate)?
- Which type of procedure would you follow to address the issue in the example?
 - (1) Binomial test (or z-test) of a proportion
 - (2) One sample z-interval for a proportion
 - (3) One sample t-test of a mean
 - (4) One sample t-interval for a mean
 - (5) Two sample t-test of means
 - (6) Two sample t-interval for means
 - (7) Two sample permutation test mean/median
 - (8) z-interval for OR
 - (9) t-test for linear regression slope
 - (10) t-interval for linear regression slope
 - (11) Permutation test for two categorical variables
 - (12) Chi-square test
 - (13) Analysis of Variance (ANOVA)
 - (14) none of the above are appropriate
- If the choice of procedures is unclear from the information given, state what additional information you would need.
- If you choose a hypothesis test, state the null and alternative hypotheses, and define the relevant population parameter(s).
- If you choose a confidence interval, define the relevant population parameter(s).
- Check any needed conditions applicable to the test.
- See pages 315-317 of your text for a summary of the tests.

Examples

1. A campus administrator wants to know if some campus groups are more likely to consume alcohol than others. He takes a random sample of 1500 students and classifies them as high risk or low risk drinkers, and whether they belong to a sorority, fraternity, or neither.
2. A researcher wants to determine whether people with “positive attitudes” tend to live longer than those without positive attitudes. He collects data on those who were classified with a positive attitude and those who were not, and he records how long each person lived.
3. A financial investor is interested in whether the number of houses purchased in a particular city is related to the current interest rate. Every day for 3 months, she records the number of houses purchased in the city and also the current interest rate published by the Federal Reserve.

Solutions

1. Chi-square test (12)

observational units = students

variable 1 (explanatory variable) - whether or not the student belongs to a sorority, fraternity, or neither (categorical)

variable 2 (response variable) - whether the student is a high risk or low risk drinker (categorical, binary)

H_0 : no association between Greek membership and alcohol risk level

H_a : an association between Greek membership and alcohol risk level

Note: since variable 1 has 3 categories, we can't use a two-sample procedure

check: make sure that each *expected* cell has at least one observation, and that 80% of the *expected* cells are greater than 5.

2. Two independent samples t-test of means (5) (as long as the people were randomly selected)

observational units = people

variable 1 (explanatory) - whether or not the individual has a positive attitude (categorical, binary)

variable 2 (response) - number of years lived (quantitative)

μ_1 = mean lifetime of those with a positive attitude in the population

μ_2 = mean lifetime of those without a positive attitude in the population

H_0 : $\mu_1 = \mu_2$ (no difference in mean lifetime between the two populations)

H_a : $\mu_1 > \mu_2$ (those with positive attitudes tend to live longer on average)

check: independent random samples with at least 15-30 people each. If sample sizes are small or particularly not normal, the researchers may prefer a permutation test (7).

3. No applicable method (14)

We can't use regression here because we don't have **independent** observations (the interest rates from day to day depend on each other.) We would use time series techniques to address this problem.

4. ANOVA (13) (as long as we can justify the selection of cars as random)

observational units = cars

variable 1 (explanatory) - color of car

variable 2 (response) - time to travel between two points on the highway

μ_r = mean time for "racy colored" cars in the population

μ_l = mean time for "light" colored cars in the population

μ_o = mean time for "other" colored cars in the population

H_0 : $\mu_r = \mu_l = \mu_o$ (no difference in mean time across colors in the populations)

H_a : not H_0 (at least one of the colors shows a different mean time than the others)

check: independent random samples, normally distributed travel time, equal variance for different colors (or, large independent random samples that are balanced)

5. One-sample z-interval for a proportion (2)
 - observational units = students
 - variable (response) - whether the student considers parking a problem (categorical, binary)
 - p = proportion of all students at this college who find parking a problem
 - (we want a CI for p)
 - check: first check the binomial conditions. Then make sure that $np \geq 10$ and $n(1 - p) \geq 10$. With a sample size of 200, we would need $p < 0.05$ or $p > .95$ to violate the conditions. Because this is unlikely, it seems that our conditions will be met.

6. One sample t-interval for a mean (4)
 - observational units = lizards
 - variable (response) - flight speed of the lizard
 - μ = true average flight speed for this species of lizard
 - (we want a CI for μ)
 - check: randomly selected lizards, at least 15-30 or from a normally distributed population

7. t-test for linear regression slope (9)
 - observational units = couples
 - variable 1 (explanatory) - length of time the couple co-habitated before marriage (quantitative)
 - variable 2 (response) - number of years the couple's marriage lasted (quantitative)
 - $H_0 : \beta_1 = 0$ (no linear assoc btwn length of co-habitation and length of marriage in the pop)
 - $H_a : \beta_1 \neq 0$ (a linear assoc btwn length of co-habitation and length of marriage in the pop)
 - check: randomly selected observations, a sample size of at least 15-30 or from a population that is normally distributed

8. Two-sample t-interval for means (6)
 - observational units = customers
 - variable 1 (explanatory) - gender (categorical, binary)
 - variable 2 (response) - total amount of bill (quantitative)
 - μ_m = average amount spent by men
 - μ_w = average amount spent by women
 - (we want a CI for $\mu_m - \mu_w$)
 - check: randomly selected observations, sample size of at least 30 each

9. z-interval for OR (8)
 - observational units = individuals involved in bicycle accidents
 - variable 1 (response) - whether their accident was fatal or not (categorical, binary)
 - variable 2 (explanatory) - BAC above the legal limit or not (categorical, binary)
 - (we find a CI for the OR)
 - Our OR represents the multiplicative change in odds of fatalities for a BAC above the legal limit versus a BAC not above the legal limit. If the CI overlaps 1, we will not be able to claim a significant change in odds of fatality for BAC above the legal limit vs. below the legal limit.
 - check: no conditions to check except that we have a random sample