Stat 21 - Class 8

Determining Which Method to Use - Solutions

1. A consumer magazine plans to poll car owners to see if they are happy enough with their vehicles that they would purchase the same model again. They randomly selected 450 owners of American-made cars and 450 owners of Japanese models and ask the owners whether or not they are happy with their vehicle (yes/no).

Suppose we want to make a statistical statement quantifying the difference in owner satisfaction of American versus Japanese cars.

Population: All American or Japanese car owners who read this magazine

Parameter of interest (if there is one): $p_J - p_A$ where p_J is the proportion of satisfied owners of Japanese cars and p_A is the proportion of satisfied owners of American cars

Type of variable(s): Categorical, binary (satisfaction)

Sample size: $n_1 = 450 \quad n_2 = 450$

Observational units: Different owners of cars

Method: inference for the difference in two proportions OR chi-squared test for homogeneity

2. Data was collected on the average high temperatures in the months of January and July of 2020 for a dozen different cities.

Suppose we want to determine if there is a significant difference in the mean temperatures between January and July.

Population: The temperature difference in all cities (of what size? in what country? etc. not specified)

Parameter of interest (if there is one): $\mu_{jan} - \mu_{july}$ where μ_{jan} = mean temperature across all cities in January and μ_{july} = mean temperature across all cities in July

Type of variable(s): Quantitative and continuous (temperature)

Sample size: $n = 12$

Observational units: Each different city

Method: Inference for a paired difference in means

3. The Masterfoods Company claims that yellow candies make up 20% of its milk chocolate M&M's, red another 20%, orange, blue, and green 10% each. The rest are brown. You purchase a bag of plain M&M's and count the number of yellow, red, orange, blue, green, and brown candies.

Suppose we want to determine if your sample is consistent with the company's stated proportions.

Population: All bags of candy produced by Masterfoods Company

Parameter of interest (if there is one): No parameters

Type of variable(s): Categorical, 6 levels corresponding to colors

Sample size: $n =$ however many M&Ms are in the bag you bought

Observational units: The individual M&M candy pieces

Method: Chi-squared goodness-of-fit test

4. Hepatitis C causes bout 10,000 deaths each year in the US but often lies undetected for years after infection. A study at a large medical center randomly surveyed visitors over a particular time period. The participants were asked whether or not they had any tattoos and from where (all tattoos from a parlor, at least one tattoo from somewhere besides a parlor, or no tattoos) and were tested to determine if they had hepatitis C or not.

Suppose we want to test if a hepatitis C infection is related to whether and how people have tattoos.

Population: All potential/past visitors of this medical center

Parameter of interest (if there is one): No parameters

Type of variable(s): Two categorical variables, the first with three levels and the second with two levels

Sample size: $n =$ however many participants the medical center surveys

Observational units: Different people

Method: Chi-squared test of independence

5. Some people think that a full moon elicits unusual behavior in people including increased illegal activities. Suppose we collect data on the number of arrests made in a small town during the weeks of six full moods and six other randomly selected weeks during the same year. For each week we note whether or not there was a full moon and we count the number of arrests for violent crime, property damage, drug offenses, and other offenses.

Suppose we want to test whether or not there is evidence of a difference in the types of illegal activity that take place during a full moon.

Population: All people residing in or visiting through this small town

Parameter of interest (if there is one): Here there are actually many parameters: $p_{vc,full} - p_{vc,not}$, $p_{pd,full} - p_{pd,not}$, $p_{do,full} - p_{do,not}$, $p_{oo,full} - p_{oo,not}$, where vc stands for violent crime, pd stands for property damage, do stands for drug offenses, and oo stands for other offenses and $p_{vc,full}$, for example, is the proportion of total arrests for violent crime during a full moon.

Type of variable(s): Categorical variable with 4 levels

Sample size: $n =$ however many arrests were made during these 12 weeks in total

Observational units: Different people

Method: Chi-squared test for homogeneity