

Example 2

A company pays production workers \$630 per week. The union claims that these workers are paid below the industry average for their work. A sample of 15 workers from other sites gives a mean wage of \$670/week with a standard deviation of \$58/week. Is the union's claim justified?

Solution:

Step 1: $H_0: \mu \leq \$630$ (industry weekly average is not significantly different to \$630)

$H_a: \mu > \$630$ (The industry weekly average is greater than \$630)

Step 2: Test Statistic - As we don't know the population variance, and the sample size is < 30 , we shall use the t test.

Step 3: Significance level - We will use $\alpha = 0.10$ (as we want to be liberal rather than conservative)

Step 4: Decision rule - From 't' table, $t_{(0.1, 14df)} = 1.345$



$H_0/H_a: \text{pop_mean / industry average} == 630$

$H_1/H_a: \text{pop_mean / industry average} > 630$

630 : hypothesized mean

samples : $n=15$, mean = 670, $sd=58$ degrees of freedom $=(n-1)=14$

sample_mean - (sample_mean or hyp value)

t score = -----

sample_SD / sqrt(n)

$$\frac{670 - 630}{58 / \sqrt{15}} = 2.67$$

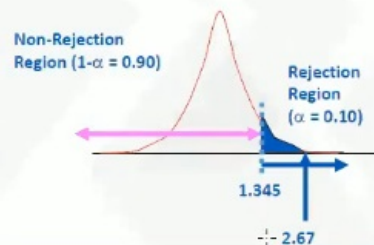
Example 2 (contd.)

Step 5: Calculate test statistic;

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{SE} = \frac{\bar{x} - \mu_{\bar{x}}}{s / \sqrt{n}} \quad t = \frac{670 - 630}{58 / \sqrt{15}} = 2.67$$

Step 6: Make a decision - As 2.67 is > 1.345 , we will reject the H_0 .

Step 7: Conclusion - "Production workers at the company earn an average of \$40 per week less than the industry standard ($t = 2.67$, $df = 14$, $p < 0.1$)"



Insurance Claim amount - cont

gender -binary categorical

s_gold : Claim_Amount of gold customers
s_silver : Claim_Amount of silver customers
s_basic : Claim_Amount of basic customers

Ho : s_plat == s_gold == s_silver == s_basic
Ha : s_plat != s_gold != s_silver != s_basic

ANOVA/F Test var Between groups
F Score = $\frac{\text{var Between groups}}{\text{var within the group}}$

F Score is high, we reject NULL
p value as output

Chi Square Test

- Check if one cat var influences other cat variable
(Observed - Expected)²

chi square score = $\sum \left(\frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \right)$

if chi_square is more and p value is low
then : there is an influence

1. Card usage has been improved significantly from last year usage which is 50. (Hint: Comparing card usage of post campaign of 1 month with last year hypothesized value 50)
2. The last campaign was successful in terms usage of credit card. (Hint: Comparing means for card usage of pre & post usage of campaign)
3. Is there any difference between males & females in terms of credit card usage? (Hint: Comparing means of card usage for males & females)
4. Is there any difference between segments of customers in terms of credit card usage? (Hint: Comparing means of card usage of different segment customers)
5. Is there any relation between region & Segment? (Hint: Finding the relationship between categorical variables region and Segment)
6. Is the relationship between card usage in the latest month and pre usage of campaign? (Hint: find the correlation between latest_mon_usage and pre_usage)

Scipy : Sci Python math, sci, eng, linear programming, img processing stats from scipy import stats
1. One sample T test : stats.ttest_1samp() Inputs: a pop_mean output: t score and p value
2. two sample ttest : stats.ttest_rel() Inputs: a(samp1) b(samp2) output: t score and p value
3. Independent sample ttest : stats.ttest_ind() Inputs: a(cont val for cat1) b(cont val for cat2) output: t score and p value
4. ANOVA or f test : stats.f() or stats.f_oneway() Inputs: a(cont val for cat1) b(cont val for cat2) c(cont val for cat3) output: f score and p value
5. Chi square : stats.chi2_contingency Inputs: cross tab output: x2 square score p value df(degree of freedom) expected value matrix
One sample T-Test last year usage = 50 Present month usage=cust.Latest_mon_usage.mean() = 63.17
Ho: Last month usage ==50 Ha: Last month usage >50

In []:

In []: