

# Test bank: means and proportions

[Literature](#), [Russian Literature](#)



**Group Assignment**

According to the output, three variables (section, bed, and pool) are insignificant because the p-value of them is larger than 0.05. The relationship between the selling price and variables should be:

$$Y = -49.59 + 4.04X_1 + 32.97X_2 + 11.09X_3 + 29.15X_4 + 22.52X_5 + 12.92X_6 - 25.66X_7 + 1.59X_8$$

X<sub>1</sub> = lot size X<sub>2</sub> = number of bathrooms X<sub>3</sub> = number of other rooms X<sub>4</sub> = number of stories X<sub>5</sub> = number of fireplaces X<sub>6</sub> = car garages X<sub>7</sub> = whether or not the lot is fenced X<sub>8</sub> = age

Q4: Based on the results of the estimation in step 1, answer the following questions:

- a. How do you interpret the intercept, the coefficient of lot size, beds, and other variables? The intercept (-49.58856) is the expected mean value of Price Y when all variables X = 0. And the coefficient means that if all other variables are fixed when lot size change 1 unit, the price Y changes 4.05389 units. So do the others.
- b. What does the sign of the coefficient tell you? The Fence has a negative sign on the coefficient, which means that as the Section, Fence increase, the Price would decrease. And the others (Lot size, Bath, Other, Stories, Fireplaces, Cars, Age) have a positive sign on the coefficient, which means that as these variables increase, the Price

would increase. These three variables: Section, Bed, Pool are insignificant.

- c. Based on the relationship you have estimated, how do you determine if a variable such as a lot size has a significant (statistically significant) impact on the sales price? Base on the P-value. Assume the confidence is 95%, P-value should be less than 0.05. Otherwise, the variables are insignificant.
- d. Are the signs and the magnitude of the coefficients consistent with intuition? Why/why not? There is a little bit of a difference. Section of the town, pool, and a number of bedrooms should be significant variables in the real world but they have no influence on Price according to the result.
- e. Does the model fit the data well - what criteria do you use to assess goodness of fit? Yes, the model fits the data well because the R-square is 0.8705, which is close to 1. We set the confidence level 95% and only 3 variables are insignificant.

Based on the results of the estimation in step 2, answer the following questions: a. What do the correlations between variables reveal?

We assume:  $|r| < 0.4$ , weakly correlated;  $0.4 < |r| < 0.7$ , significantly correlated;  $0.7 < |r| < 1$ , strongly correlated. As the picture shows, section and Age are highly correlated, and is weakly correlated with other variables; Lot size and other variables are weakly correlated; Bed and Bath, Stories are significantly correlated, and is weakly correlated with other variables; Bath and Other, Stories are significantly correlated and b. Assign labels to the

factors and provide a justification for these labels –justification should be based on results in step 2. Base on the rotated factor pattern, I will assign a bath, other, stories, fireplaces, and cars to Factor 1 because the absolute value of these factors is all very big. The variables section and age will be assigned to Factor 2 and the absolute value of them are bigger than 0.8. Lot size will be assigned to Factor 3 and Fence will be assigned to Factor 4. The variable pool will not be assigned to any of the factors because its ratio is very small in all of the factors. c. Based on the results how many factors would you consider in summarizing the information in the explanatory variables - why?

What percentage of the total information in the explanatory variables is summarized in the factors that you have decided to retain? Only 4 factors will be considered in summarizing the information in the variables because there are only 4 eigen value > 1, which means these 4 factors will do a good job. According to the estimates, 64.89% of the total information in the explanatory variables in the factors will be retained.

Based on the results of the estimation in step 3, answer the following questions:

- a. How do you interpret the intercept and the coefficient of the factors?  
First, the P-value for this model is less than .5 so this model is significant. Then we can figure out the P-value for each factor is significant, which means Factor 1, 2, 3, and 4 do have some effects on Price.

$$\text{Price} = 110.70 + 48.5 * F1 + 33.49 * F2 + 57.20 * F3 - 8.30 * F4$$

It means every unit increased on factor1, the price of this house would be increased 48.5 units as well. If the value of factor2 (section and age) increase one unit, the section of this house closer to uptown by one unit, the price of the house would be increased by 33.49 units. If the value of factor3 (lot size) increases one unit, the price of this house would be increased by 57.0 units. If the value of factor4 (fence) increases one unit then the price of this house would decrease by 8.30 units. b. Are the signs and the magnitude of the coefficients of the factors consistent with intuition and the labels you have assigned in 6c - why/why not? Due to the structure and lot size have a positive effect on the price, factor 1 and 3 are consistent with intuition, On the other hand, the price should be decreased with the increasing age of the house, the factor 2 and 4 are not consistent with intuition, and the increasing offense should cause a higher price.

Does this model fit the data better than the model in step 1 - why/why not? No, because the R square and Adj R square are both smaller in this model than in model 1. It should be due after factor reduction after abandoning some variables which may cause a bigger error in regression. Suppose a seller wanted to spend \$10,000 on home improvement before selling the house. Assume that the lot size cannot be changed. a. If \$10,000 were barely enough to make one update - add a bedroom, add a pool, etc. where would you recommend that the money be spent? Why? Spend on the bath. Because the coefficient of bath is the largest, which means if the seller spends money on bath, he can sell the house at the highest price. b. If \$10,000 were enough to make two updates - add a bedroom and a pool, etc., which two updates would you recommend, and why? I will recommend

spending on bath and stories because the two have the largest coefficients of all the variables.