**Lecture 27**

**Misc. Topics**

1. Confidence Interval for difference between two population means:

We have focused on testing hypothesis about population means but equally useful is to estimate the confidence interval

1. Test and Confidence Interval Concerning one and two population variances

. This test can be carried out using the Chi Square test given by

The test that two population variances are equal is:

. This hypothesis is tested using the F statistic given by:

which has an F distribution with and degrees of freedom respectively.

Some interesting data and exercises for regression

You can

**Ex1:** Consider the factors such as the number of megapixels, weight (oz.), and overall score ranges from 0 to 100 of sample of Canon and Nikon cameras used to explain prices. [Note Brand=1 for Canon and 0 for Nikon].

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Observation | Brand | Price\_$ | Megapixels | Weight\_oz | Score | Brand |
| 1 | Canon | 330 | 10 | 7 | 66 | 1 |
| 2 | Canon | 200 | 12 | 5 | 66 | 1 |
| 3 | Canon | 300 | 12 | 7 | 65 | 1 |
| 4 | Canon | 200 | 10 | 6 | 62 | 1 |
| 5 | Canon | 180 | 12 | 5 | 62 | 1 |
| 6 | Canon | 200 | 12 | 7 | 61 | 1 |
| 7 | Canon | 200 | 14 | 5 | 60 | 1 |
| 8 | Canon | 130 | 10 | 7 | 60 | 1 |
| 9 | Canon | 130 | 12 | 5 | 59 | 1 |
| 10 | Canon | 110 | 16 | 5 | 55 | 1 |
| 11 | Canon | 90 | 14 | 5 | 52 | 1 |
| 12 | Canon | 100 | 10 | 6 | 51 | 1 |
| 13 | Canon | 90 | 12 | 7 | 46 | 1 |
| 14 | Nikon | 270 | 16 | 5 | 65 | 0 |
| 15 | Nikon | 300 | 16 | 7 | 63 | 0 |
| 16 | Nikon | 200 | 14 | 6 | 61 | 0 |
| 17 | Nikon | 400 | 14 | 7 | 59 | 0 |
| 18 | Nikon | 120 | 14 | 5 | 57 | 0 |
| 19 | Nikon | 170 | 16 | 6 | 56 | 0 |
| 20 | Nikon | 150 | 12 | 5 | 56 | 0 |
| 21 | Nikon | 230 | 14 | 6 | 55 | 0 |
| 22 | Nikon | 180 | 12 | 6 | 53 | 0 |
| 23 | Nikon | 130 | 12 | 6 | 53 | 0 |
| 24 | Nikon | 80 | 12 | 7 | 52 | 0 |
| 25 | Nikon | 80 | 14 | 7 | 50 | 0 |
| 26 | Nikon | 100 | 12 | 4 | 46 | 0 |
| 27 | Nikon | 110 | 12 | 5 | 45 | 0 |
| 28 | Nikon | 130 | 14 | 4 | 42 | 0 |

Estimate the regression model, write down the estimated eq, interpret the coefficients. Predict price of Nikon camera of 14 megapixels with a weight of 6 oz and score of 55. Interpret Rsq. Test the hypothesis (at 5%) that the average price of Canon is significantly less than Nikon.

**Ex:** Consider the data of sales prices of 176 houses to be explained by value of land, value of improvement (all three variables in $1000) and the city area where the house is located. (CHEVAL is the base area). The estimated regression is as follows.

Interpret each coefficient. Predict the price of a house located in Cheval that has value of land and improvement as 100 and 200 (thousands of dollars). Test the hypothesis (at 5%) that average prices in the Hydepark area are significantly less than Cheval area. (data file propert.csv already shared).

SE 20.33 0.091 0.0468 32.536 22.731 28.396

**Titanic Data Example (Logistic Regression)**

You can use this data for further practice of logistic regression questions i.e. answering questions on whether a passenger survived or not (survived=1, died=0) based on passenger class  (class 1 an elite class is reference class), class 2 or class3, gender, age and fare and number of siblings. This is good data with both qualitative predictors with 3 classes (2 dummy added) as well binary predictor gender and also, quantitative predictors (e.g. age). Estimate model in R and give output and ask relevant questions e.g. marginal effect of gender on survival probability.

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 1.555476 0.445796 3.489 0.000484 \*\*\*

Pclass2 -1.308268 0.319741 -4.092 4.28e-05 \*\*\*

Pclass3 -2.525790 0.334601 -7.549 4.40e-14 \*\*\*

Age -0.044147 0.008262 -5.343 9.13e-08 \*\*\*

Siblings -0.393449 0.123221 -3.193 0.001408 \*\*

Fare 0.001750 0.002455 0.713 0.475990

SexF 2.616107 0.215413 12.145 < 2e-16 \*\*\*

Where,

Pclass2 = 1 if the passenger is in passenger class 2, otherwise 0.

Pclass3 = 1 if the passenger is in passenger class 3, otherwise 0.

Age = Age of a passenger in years.

Siblings = Number of siblings.

Fare = Fare is $.

SexF = 1 if female, otherwise 0.

1. Estimate the probability of survival for a 7-year-old female passenger, travelling in passenger class 2, having 3 siblings, and paying fare of $30.
2. Estimate the probability of survival for a 7-year-old male passenger, travelling in passenger class 2, having 3 siblings, and paying fare of $30.
3. Find the marginal effect of passenger’s gender on the probability of survival given that the passenger is 7-year-old female, travelling in passenger class 2, having 3 siblings, and paying fare of $30. Also, interpret this number.
4. Find the marginal effect of another year of passenger’s age on probability of survival for 7-year-old female passenger, travelling in passenger class 2, having 3 siblings, and paying fare of $30.
5. Predict whether a 50-year-old male passenger travelling in passenger class 3, having 2 siblings, and paying a fare of $8.05 will survive or not?
6. Find the probability that a 70-year-old male passenger travelling in passenger class 3 will not survive.
7. Find and interpret the odds ratios of age, siblings, fare, and SexF variables.

**Note:** Kindly guide students to be careful in writing interpretations in regression e.g. they do not mention proper units under study or just write in generic terms like one unit increase in income will increase consumption by 0.78 units (if units of income and consumption are given in the problem marks will be deducted.

Also, in R sq interpretation they write things like this is good or bad model without mentioning the proper meaning of the magnitude e.g in the context of above example. =0.686 should be interpreted as that 68.6% variation in consumption is explained by income. (or in case of multiple regression mention the names of independent variables which explain that percentage of variation in dependent variables).