ANOVA and Regression Analysis Connection

We can use qualitative method to check the difference among mean of different populations.

H0:  µ1  = µ2 = … = µc

H1:  At least one population mean is different

There are two methods for finding relationship between independent and dependent variables.

1. **Qualitative Method**: Use following steps:
* Draw two bracket lines that separate observations from the rest of the space. These lines represent upper and lower limits.
* Find values for each group.
* Draw a straight line that has least distance from all values.

If the line is almost parallel to the ground, i.e., having slope almost equal to zero, then we conclude that there is not enough evidence to reject H0: µ1  = µ2 = … = µc.

Sample mean

**2) Quantitative Method (ANOVA by Regression):**

Following example shows a regression with multiple responses that is an extension of simple regression with no replications for each independent variable.

**Page 1/3**

|  |  |  |
| --- | --- | --- |
| **Supplier**  | **Tensile Strength** |  |
| 1 | 18.5 |
| 1 | 24.0 |
| 1 | 17.2 |
| 1 | 19.9 |
| 1 | 18.0 |
| 2 | 26.3 |
| 2 | 25.3 |
| 2 | 24.0 |
| 2 | 21.2 |
| 2 | 24.5 |
| 3 | 20.6 |
| 3 | 25.2 |
| 3 | 20.8 |
| 3 | 24.7 |
| 3 | 22.9 |
| 4 | 25.4 |
| 4 | 19.9 |
| 4 | 22.6 |
| 4 | 17.5 |
| 4 | 20.4 |

In order to check equality of all populations’ expected values, we perform the following hypothesis.

H0:  β1  = 0

H1:  β1   ≠  0

|  |
| --- |
| ANOVA for Regression Table |
|  | *df* | *SS* | *MS* | *F* |
| Regression | 1 | 3.0625 | 3.0625 | **0.349496** |
| Residual | 18 | 157.727 | 8.762611111 |  |
| Total | 19 | 160.7895 |
|  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* |
| Intercept | 21.07 | 1.621352316 | 12.99532482 |
| Supplier  | 0.35 | 0.592034158 | **0.591182105** |

**Page 2/3**

Computed TSlope

TSlope = (0.35 – 0) / 0.592034158= 0.591182105

At significance level α = 0. 05, the critical value is

T n-2, 0.025 = T 20-2, 0.025 = T 18, 0.025 = 2.101

We conclude that there is no significant evidence to reject H0, (population slope is almost zero).

This conclusion that there is no relationship can be tested using F-computed and Critical value of F-table:

Computed F1, 18, 0.05  = **0.349496**

The critical value of F statistic is:

Critical F1, 18, 0.05  = 4.41

The computed F is much less than the value of table (notice that for F in ANOVA table of regression larger F is better, it should be say, 5-time larger). Therefore, we conclude that there is no relationship, i.e., **there is no dependency**.

We need to have at least three points to perform simple linear regression analysis. This is because degree of freedom for mean square error is n-2. The relationship between t distribution in regression and F distribution in ANOVA table is (t α/2)2 = Fα

 t2 n, α/2= F1, n, α

(a) Checking this result using empirical data:

(0.591182105) 2 = 0.349496, with α = 0.05

(b) Checking this result using statistical tables (the theoretical relation between T, and F tables):

2.1012 = 4.41 with α = 0.05

**Page 3/3**