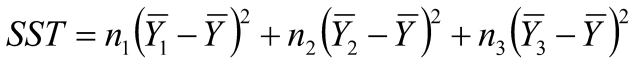
**ANOVA Test**

Three students, Linda, Tuan, and Javier, are given five laboratory rats each for a nutritional experiment. Each rat's weight is recorded in grams. Linda feeds her rats **Formula A**, Tuan feeds his rats **Formula B**, and Javier feeds his rats **Formula C**. At the end of a specified time period, each rat is weighed again, and the net gain in grams is recorded. Using a significance level of 5%, test the hypothesis that the three formulas produce the same mean weight gain. (10 marks)

|  |  |  |
| --- | --- | --- |
| Linda’s rats | Tuan’s rats | Javier’s rats |
| 43.5 | 47.0 | 51.2 |
| 39.4 | 40.5 | 40.9 |
| 41.3 | 38.9 | 37.9 |
| 46.0 | 46.3 | 45.0 |
| 38.2 | 44.2 | 48.6 |

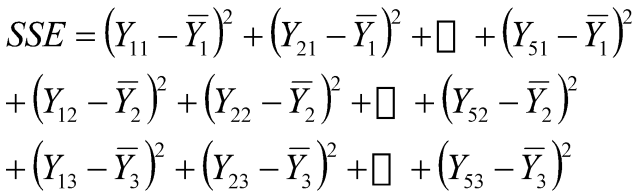
There are three samples, we use ANOVA F-test. Assume the weight of Linda’s rats Y1, the weight of Tuan’s rats Y2, the weight of Javier’s rats Y3.

=43.26, =41.68, =43.38, =44.72



SST = 5\*(41.68-43.26)2+5\*(43.38-43.26)2+5\*(44.72-43.26)2

SST = 5\*(2.4964+0.0144+2.1316) = 23.212



SSE = (43.5-41.68)2+(39.4-41.68)2+(41.3-41.68)2+(46.0-41.68)2+(38.2-41.68)2

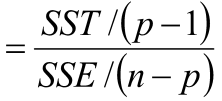
+(47.0-43.38)2+(40.5-43.38)2+(38.9-43.38)2+(46.3-43.38)2+(44.2-43.38)2

+(51.2-44.72)2+(40.9-44.72)2+(37.9-44.72)2+(45.0-44.72)2+(48.6-44.72)2

SSE = 39.428+50.668+118.228 = 208.324

*p-1* = 2

*n-p* = 12

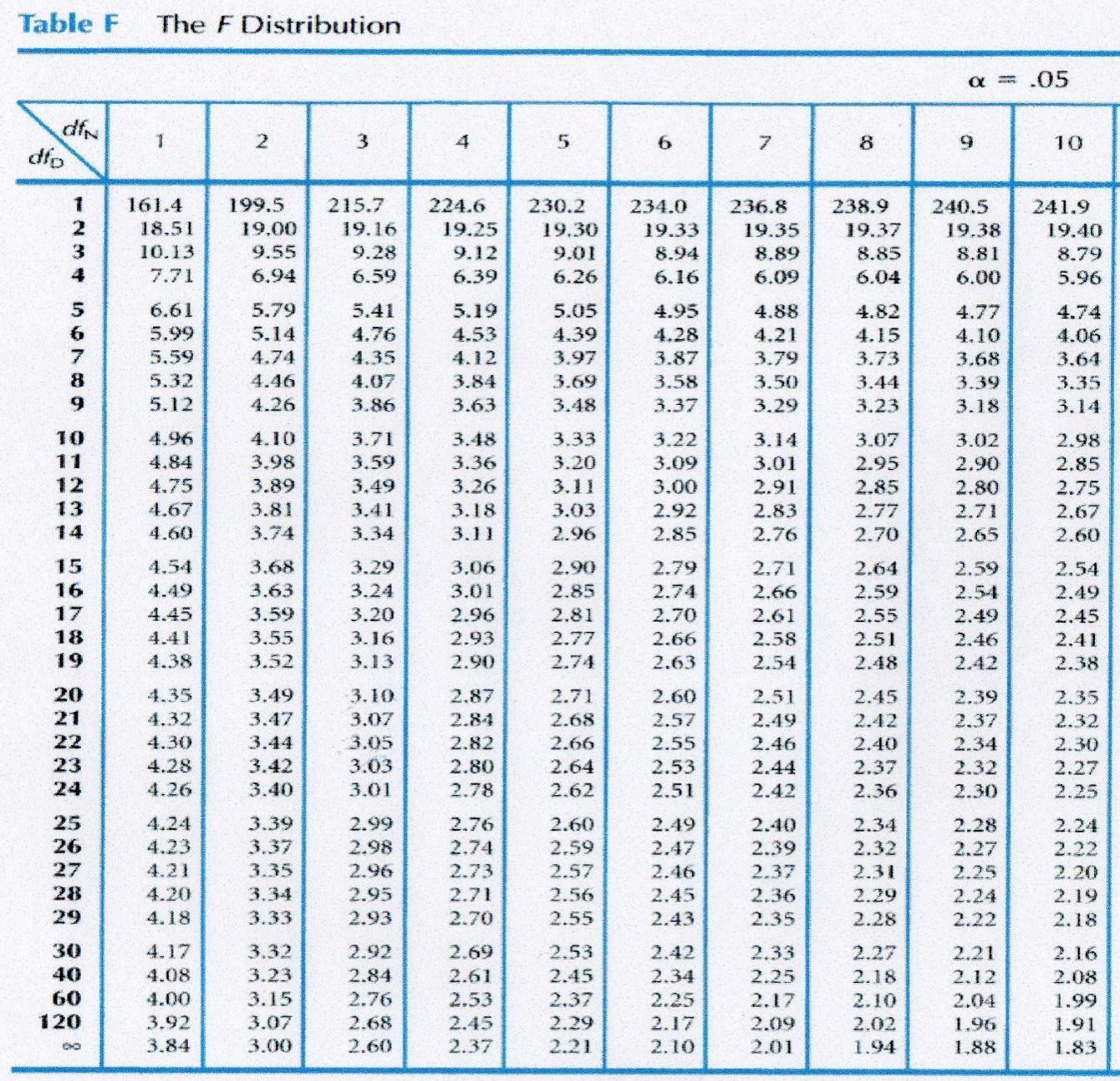
F

F = (23.212/2) / (208.324/12)

Fcrit = 3.89

DfN = p-1 = 2

DfD = n-p = 12



F = 11.606/17.36 =0.67 < 3.89

Accept the hypothesis that the three formulas produce the same mean weight gain.

UK Biobank database has recruited more than 5 million participants from different cities of UK. It collected the participant’s lab test results, lifestyle and disease history from questionnaire and clinical records. In Table 1, we showed the selected information (including disease, weight, age and smoking status) of 30 white British participants.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Participant ID | Disease | Weight (kg) | Age | Smoking |
| 7 | Hypertension | 52 | 86 | Smoker |
| 9 | Hypertension | 59 | 54 | Smoker |
| 12 | Hypertension | 79 | 64 | Smoker |
| 14 | Hypertension | 63 | 23 | Non-Smoker |
| 17 | Hypertension | 50 | 75 | Non-Smoker |
| 21 | Hypertension | 85 | 73 | Non-Smoker |
| 24 | Hypertension | 75 | 28 | Smoker |
| 25 | Hypertension | 60 | 23 | Smoker |
| 28 | Hypertension | 61 | 23 | Non-Smoker |
| 30 | Hypertension | 83 | 31 | Smoker |
| 2 | Lung cancer | 54 | 74 | Smoker |
| 4 | Lung cancer | 51 | 26 | Smoker |
| 5 | Lung cancer | 49 | 89 | Non-Smoker |
| 6 | Lung cancer | 46 | 54 | Smoker |
| 8 | Lung cancer | 40 | 46 | Smoker |
| 11 | Lung cancer | 40 | 64 | Smoker |
| 18 | Lung cancer | 56 | 25 | Smoker |
| 20 | Lung cancer | 44 | 26 | Smoker |
| 22 | Lung cancer | 58 | 73 | Non-Smoker |
| 26 | Lung cancer | 56 | 46 | Smoker |
| 1 | Type 2 Diabetes | 70 | 69 | Non-Smoker |
| 3 | Type 2 Diabetes | 68 | 31 | Non-Smoker |
| 10 | Type 2 Diabetes | 65 | 31 | Non-Smoker |
| 13 | Type 2 Diabetes | 84 | 22 | Smoker |
| 15 | Type 2 Diabetes | 74 | 41 | Smoker |
| 16 | Type 2 Diabetes | 52 | 89 | Smoker |
| 19 | Type 2 Diabetes | 88 | 60 | Non-Smoker |
| 23 | Type 2 Diabetes | 81 | 45 | Non-Smoker |
| 27 | Type 2 Diabetes | 81 | 53 | Smoker |
| 29 | Type 2 Diabetes | 64 | 70 | Non-Smoker |

**Table 1.** The information of thirty participants of UK Biobank

**Population, mean, std**

**Question 6 (5 marks).** According to the information shown in **Table 1**, calculate the mean weight and its 95% confidence interval ofthe participants with Type 2 Diabetes (4 marks)? Is it sample mean or population mean (1 marks)?

Answer:

Mean weight:

Sample standard deviation is

11.03

5% confidence interval: [72.7-2\*3.49, 72.7+2\*3.49]=[65.72,79.68]

This is a sample mean

**Correlation**

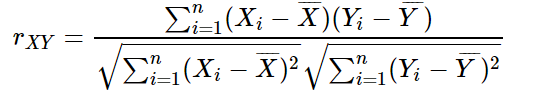
**Question 7.** Calculate Pearson correlation, Spearman correlation between age and weight for the smokers with lung cancer. **(6 marks)**

Collect data based on the targeted group (the smokers with lung cancer)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Participant ID** | **Disease** | **Weight (kg)** | **Age** | **Smoking** |
| 2 | Lung cancer | 54 | 74 | Smoker |
| 4 | Lung cancer | 51 | 26 | Smoker |
| 6 | Lung cancer | 46 | 54 | Smoker |
| 8 | Lung cancer | 40 | 46 | Smoker |
| 11 | Lung cancer | 40 | 64 | Smoker |
| 18 | Lung cancer | 56 | 25 | Smoker |
| 20 | Lung cancer | 44 | 26 | Smoker |
| 26 | Lung cancer | 56 | 46 | Smoker |

Answer:

Pearson correlation:



(age)= (74+26+54+46+64+25+26+46)/8 = 45.125

(weight) = (54+51+46+40+40+56+44+56)/8 = 48.375

*rXY* =( (74-45.125)\*(54-48.375)+ (26-45.125)\*(51-48.375)+……

+(26-45.125)\*(44-48.375)+ (46-45.125)\*(56-48.375) ) / sqrt( (74-45.125)2+(26-45.125)2)+ …… +(26-45.125)2+(46-45.125)2 ) \* sqrt( (54-48.375)2+(51-48.375)2)+ …… +(44-48.375)2+(56-48.375)2 )

*rXY* = -137.375 / 17.885 \* 49.060 = -0.157

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Participant ID** | **Disease** | **Weight (kg)** | **Age** | **Smoking** |
| 2 | Lung cancer | 54 | 74 | Smoker |
| 4 | Lung cancer | 51 | 26 | Smoker |
| 6 | Lung cancer | 46 | 54 | Smoker |
| 8 | Lung cancer | 40 | 46 | Smoker |
| 11 | Lung cancer | 40 | 64 | Smoker |
| 18 | Lung cancer | 56 | 25 | Smoker |
| 20 | Lung cancer | 44 | 26 | Smoker |
| 26 | Lung cancer | 56 | 46 | Smoker |

**Spearman correlation:**

Step1: calculate rank for each variable and rank difference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Participant ID** | **Disease** | **Weight (kg)** | **Age** | **Smoking** | **rank dif.** |
| 18 | Lung cancer | 7.5 | 1 | Smoker | 6.5 |
| 20 | Lung cancer | 3 | 2.5 | Smoker | 0.5 |
| 4 | Lung cancer | 5 | 2.5 | Smoker | 2.5 |
| 8 | Lung cancer | 1.5 | 4.5 | Smoker | -3 |
| 26 | Lung cancer | 7.5 | 4.5 | Smoker | 3 |
| 6 | Lung cancer | 4 | 6 | Smoker | -2 |
| 11 | Lung cancer | 1.5 | 7 | Smoker | -5.5 |
| 2 | Lung cancer | 6 | 8 | Smoker | -2 |

rs = 1-(6\*(6.52+0.52+…+(-5.5)2 +(-2)2)) / 8\*(82-1)

=1- 6\*105 / 8 \* 63

=-0.25

Spearman correlation: -0.25

2.Are the weights of smokers and non-smokers with hypertension significantly different each other (=0.05, tcrit=±2.306)? Please state (A) what is the null hypothesis and alternative hypothesis? (B) What kind of t-test would you like to use? (C) One tailed or two tailed tests?

Answer:

(A) H0 (null hypothesis): The weights between smokers and non-smokers with hypertension are the same

H1 (Alternative hypothesis): The weights between smokers and non-smokers with hypertension are different from each other

(B) We will select the independent sample T Test

(C) We will select two-tailed T test

Full procedure for independent t-test

1. Identify

Pop1 (variable ): The weights of smokers with hypertension

Pop2 (variable ): The weights of non-smokers with hypertension

2. State the null and research hypotheses

H0 (null hypothesis): The average weights between smokers and non-smokers with hypertension are the same

H1 (Alternative hypothesis): The average weights between smokers and non-smokers with hypertension are different from each other

3. Determine characteristics of comparison distribution (distribution of differences between means)

Population: *μx = μy* (i.e., no difference between means)

*dfTotal* = *dfx* + *dfy* = 5 + 3 = 8

=

= 99.5+80.59=180.09

. Determine critical value (cutoffs)

In Behavioral Sciences, we use = .05 (5%)

Our hypothesis is *nondirectional*, so our hypothesis test is *two-tailed*.

*tcrit*= ±2.306

5. Calculate a test statistic

6. Make a decision

Because，we fail to reject the null hypothesis and make the conclusion that the average weights between smokers and non-smokers with hypertension are the same.

3. Use Mann-Whitney U test to evaluate whether the weights of the participants with Hypertension and Lung cancer are significantly different (U=23).

Answer：

Step 1: Rank all scores together, ignoring which group they belong to

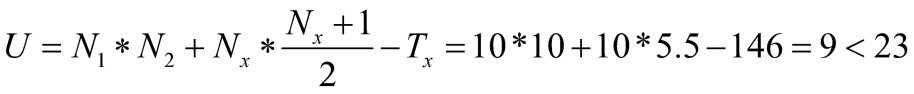
Step 2: Add up the ranks for Hypertension and Lung cancer

|  |  |  |  |
| --- | --- | --- | --- |
| **Hypertension** | **Rank** | **Lung cancer** | **Rank** |
| 52 | 8 | 54 | 9 |
| 59 | 13 | 51 | 7 |
| 79 | 18 | 49 | 5 |
| 63 | 16 | 46 | 4 |
| 50 | 6 | 40 | 1.5 |
| 85 | 20 | 40 | 1.5 |
| 75 | 17 | 56 | 10.5 |
| 60 | 14 | 44 | 3 |
| 61 | 15 | 58 | 12 |
| 83 | 19 | 56 | 10.5 |
| **Sum of Ranks** | **146** | **Sum of Ranks** | **64** |

Step 3: Select the larger one, and call it TX. In this case TX is the rank total for Hypertension, which is 138 ().

Step 4: Calculate N1, N2 and NX.

N1= N2= Nx=10



We can conclude that the difference between the weights of the participants with Hypertension and Lung cancer is significant