

STAT212 Assignment 10

Mustafa Sadiq (ms3035)

May 3, 2021

Answer to Question 1

Let u_1, u_2, u_3, u_4 and u_5 denote the mean survival (in days) for each cancer type Stomach, Bronchus, Colon, Ovary and Breast respectively.

$$H_0 : u_1 = u_2 = u_3 = u_4 = u_5 \text{ (mean survival are equal)}$$

$$H_A : \text{Not all the means are equal.}$$

We now find the test statistic:

$$k = 5$$

$$n = 13 + 17 + 17 + 6 + 11 = 64$$

The overall mean \bar{x} :

$$\bar{x} = \frac{\sum x_i}{n} = \frac{35752}{64} = 558.63$$

Region	Size, n_j	Mean, \bar{x}_j	$\bar{x}_j - \bar{x}$	$(\bar{x}_j - \bar{x})^2$	$n_j(\bar{x}_j - \bar{x})^2$
Stomach	13	286.00	-272.63	74324.39	966217.08
Bronchus	17	211.59	-347.04	120434.52	2047386.77
Colon	17	457.41	-101.21	10244.12	174150.02
Ovary	6	884.33	325.71	106085.92	636515.51
Breast	11	1395.91	837.28	701044.65	7711491.14
					11535760.52

From the final column of the table, we see that

$$SSTR = 11535760.52$$

Therefore,

$$MSTR = \frac{SSTR}{k-1} = \frac{11535760.52}{5-1} = 2883940.13$$

To compute MSE, we first compute the sample variance of each sample and construct the following table:

Region	Size, n_j	Variance, \bar{s}_j^2	$n_j - 1$	$(n_j - 1)\bar{s}_j^2$
Stomach	13	119930.33	12	1439164.00
Bronchus	17	44040.63	16	704650.12
Colon	17	182473.01	16	2919568.12
Ovary	6	1206875.47	5	6034377.33
Breast	11	1535038.49	10	15350384.91
				26448144.48

From the final column, we see that

$$SSE = 26448144.48$$

Therefore,

$$MSE = \frac{SSE}{n - k} = \frac{26448144.48}{64 - 5} = 448273.64$$

Finally, we determine F,

$$F = \frac{MSTR}{MSE} = \frac{2883940.13}{448273.64} = 6.43$$

We are to perform the test at the 10% significance level; so $\alpha = 0.1$.

$$df = (k - 1, n - k) = (5 - 1, 64 - 5) = (4, 59)$$

The P-value is:

$$P = 0.00023$$

The rejection region is $P \leq 0.1$. Since the P-value is less than the specified significance level of 0.1, we reject the null hypothesis. There are statistically significant differences in the mean survival times for individuals depending on which organ is afflicted by cancer.

Answer to Question 2

The following table gives the means and sizes for the sample data:

Region	Stomach	Bronchus	Colon	Ovary	Breast
j	1	2	3	4	5
\bar{x}_j	286.00	211.59	457.41	884.33	1395.91
n_j	13	17	17	6	11

We then obtain the test statistic for $u_i - u_j$:

$$\left| \frac{u_i - u_j}{\sqrt{s \times (1/n_i) + (1/n_j)}} \right|$$

where $s = MSE$. Do so for all possible pairs of means with $i < j$.

In Answer to Question 1, we found $MSE = 448273.64$ for the data. Now, we are ready to obtain the required test statistics. All ten are displayed in the following table:

$u_i - u_j$	Stomach (1)	Bronchus (2)	Colon (3)	Ovary (4)
Bronchus (2)	74.41			
Colon (3)	-171.41	-245.82		
Ovary (4)	-598.33	-672.75	-426.92	
Breast (5)	-1109.91	-1184.32	-938.50	-511.58

$(1/n_i) + (1/n_j)$	Stomach (1)	Bronchus (2)	Colon (3)	Ovary (4)
Bronchus (2)	0.136			
Colon (3)	0.136	0.118		
Ovary (4)	0.244	0.225	0.225	
Breast (5)	0.168	0.150	0.150	0.258

Test Statistic	Stomach (1)	Bronchus (2)	Colon (3)	Ovary (4)
Bronchus (2)	0.30			
Colon (3)	0.69	1.07		
Ovary (4)	1.81	2.12	1.34	
Breast (5)	4.05	4.57	3.62	1.51

Each entry in the table is the test statistic for the difference between the mean labeled by the column and the mean labeled by the row.

$$k = 5$$

$$n = 64$$

$$\text{degree of freedom} = n - k = 64 - 5 = 59$$

We perform these tests so that the family wise error rate is controlled at 10%. The individuals level of confidence is $\frac{0.1}{\binom{5}{2}} = 0.01$.

The cutoff value is

$$2.662$$

We will reject the null hypothesis if the test statistic is greater than 2.662. Referring to the table above, we see that we can declare the means u_1 and u_5 different, the means u_2 and u_5 different and the means u_3 and u_5 different; all other pairs of means are not declared different.

To summarize, the following cancer types have statistically significant differences in survival times:

- Stomach (1) and Breast (5)
- Bronchus (2) and Breast (5)
- Colon (3) and Breast (5)

Answer to Question 3

The overall standard deviation (i.e. the standard deviation using all observations) of survival times is:

$$\sigma = 776.48$$

The standard deviation within each organ cancer type are:

$$\sigma_{Stomach} = 346.31$$

$$\sigma_{Bronchus} = 209.86$$

$$\sigma_{Colon} = 427.17$$

$$\sigma_{Ovary} = 1098.58$$

$$\sigma_{Breast} = 1238.97$$

Answer to Question 4

The standard deviation within each group is not the same. Stomach, Bronchus and Colon have a lower standard deviation while Ovary and Breast have a higher standard deviation.