| First 3 letters of SURNAME | | |
|---|---|---|
| | | |

| 1) | 2) | 3) | 4) | 5) | 6) | Total. |
|---|---|---|---|---|---|---|
| | | | | | | |

# STK 121/120/161: Module Test 3 — 12 May 2022

**Time: 90 Minutes** **Total: 70 Marks**

| | |
|---|---|
| Surname | MEMO |
| Initials | |
| Student number | |
| Signature | |

- Answer the questions in the space provided.
- A list of formulae and statistical tables will be provided. **Do not write on it or make any marks on it.** Place it **in your paper** when you submit your paper.
- **Round** off your **final answers** to **four decimal places.**
- Always **write down** the **appropriate formula(e)** that you use where applicable.
- You may use the blank counter pages for your rough work or to continue with an answer if you feel the space allowed was not enough. Do not erase your rough work.

## Question 1

In a study to determine whether **Marital Status** and **Gender** are independent or not, a researcher used a sample of respondents and gathered the Observed (O) frequencies as given in the table below.

| | | Marital Status: Married | Marital Status: Single | Total |
|---|---|---|---|---|
| Gender | Female | $O_{11} = 90$<br>$E_{11} = \boldsymbol{A}$ | $O_{12} = 110$<br>$E_{12} = \boldsymbol{B}$ | ***E*** |
| | Male | $O_{21} = 110$<br>$E_{21} = \boldsymbol{C}$ | $O_{22} = 90$<br>$E_{22} = \boldsymbol{D}$ | ***F*** |
| **Total** | | ***G*** | ***H*** | ***I*** |

Use the information given above to calculate the values of $A$ to $I$ in the above table and write them down in (ii) below:

(i). What assumption is needed to calculate the values of the Expected frequencies, $A$ to $D$?

The Factors of Marital Status and Gender are Independent of each other.

(2)

(ii). The calculated values are:

$A = 100$ $B = 100$ $C = 100$ $D = 100$

$E = 200$ $F = 200$ $G = 200$ $H = 200$ $I = 400$

(11)
**[13]**

**Question 2**

In a study to determine whether **Promotion** and **Gender** are independent or not, a researcher used a sample of 1200 respondents and gathered the Observed (O) and Expected (E) frequencies as given in the table below.

| | | Promotion | |
|---|---|---|---|
| | | Yes | No |
| Gender | Female | $O_{11} = 36$<br>$E_{11} = 64.8$ | $O_{12} = 204$<br>$E_{12} = 175.2$ |
| | Male | $O_{21} = 288$<br>$E_{21} = 259.2$ | $O_{22} = 672$<br>$E_{22} = 700.8$ |

(i). Calculate the **value of the** $\chi^2$ **test statistic.**

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad \text{or} \quad \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

$$= \frac{(36-64.8)^2}{64.8} + \frac{(204-175.2)^2}{175.2} + \frac{(248-259.2)^2}{259.2} + \frac{(672-700.8)^2}{700.9}$$

$$= 12.8 + 4.734247 + 3.2 + 1.1835616$$

$$= 21.91780822$$

$$\approx 21.9178$$

(6)

(ii). Comment on the validity of the $\chi^2$ test statistic value. Motivate your answer by giving the necessary condition(s) and whether they are satisfied or not.

Yes, it is valid since all $E_{ij}$- values are greater than or equal to 5.

or

Yes, it is valid since the $E_{ij}$ - values are $\geq 5$ for all cells.

(2)

[8]

or

Yes, it is valid since all $E_{ij}$ - values are 5 or more (more than 4, exceeds 4)

or

Yes, it is valid since the expected frequencies per cell is 5 or more for all cells.

or ...... since $E_{11} \geq 5, E_{12} \geq 5, E_{21} \geq 5, E_{22} \geq 5$

**Question 3**

The manager of ***JOB SEEKERS*** **randomly** selected 180 companies and asked their human resource managers how their company planned to change their workforce over the next 12 months. Their responses could be classified into 3 categorical groups as indicated in the table below. He also categorised the companies as either private or public sector. The results are summarised in the following contingency table.

| | Employment plan | | | |
|---|---|---|---|---|
| **Company type** | Add Employees | No Change | Lay Off Employees | **Total** |
| Private sector | 37 | 19 | 16 | **72** |
| Public sector | 32 | 34 | 42 | **108** |
| **Total** | **69** | **53** | **58** | **180** |

A chi-square test for independence was conducted which yielded a test statistic value of 9.4405

a) State the **null and alternative hypotheses** for the above problem.

$H_0$: Employment Plan and Type of Company are Independent

$H_A$: Employment Plan and Type of Company are Dependent (NOT Independent) (3)

b) Determine the **degrees of freedom**.

$DF = (R-1)(C-1) = (2-1)(3-1) = (1)(2) = 2$ (1)

c) Determine the **critical value** at the 5% level of significance.

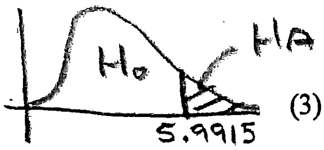
$\chi^2_{0,05;2} = 5.9915$ (2)

d) What is the **decision rule** and **what decision can be taken** at the 5% level of significance. (Motivate your answer by using among other the information above)

Reject Ho if $\chi^2$-value $\geq 5.9915$

$\chi^2 = 9.4405 > 5.9915$

$\therefore$ Reject Ho




(3)

e) What **conclusion** can be drawn based on the test results? Comment on the differences, if any, by among other making use of Row and/or Column percentages.

Employment Plan and Type of Company are Dependent. An association does exist between them.

Row %

| | EMPLOYMENT PLAN | | |
|---|---|---|---|
| Company Type | ADD | NO CHANGE | LAY OFF |
| Private | $\frac{37}{72}\times100 = 51.39$ | $\frac{19}{72}\times100 = 26.39$ | $\frac{16}{72}\times100 = 22.22$ |
| Public | $\frac{32}{108}\times100 = 29.63$ | $\frac{34}{108}\times100 = 31.48$ | $\frac{42}{108}\times100 = 38.89$ |

Private Sector are more likely to ADD employees

Public Sector are more likely to stay the same or LAY OFF employees. Better in Private sector.

(5)

[14]

**Question 3**

The manager of ***JOB SEEKERS*** **randomly** selected 180 companies and asked their human resource managers how their company planned to change their workforce over the next 12 months. Their responses could be classified into 3 categorical groups as indicated in the table below. He also categorised the companies as either private or public sector. The results are summarised in the following contingency table.

| | **Employment plan** | | | **Total** |
|---|---|---|---|---|
| **Company type** | Add Employees | No Change | Lay Off Employees | |
| Private sector | 37 | 19 | 16 | **72** |
| Public sector | 32 | 34 | 42 | **108** |
| **Total** | **69** | **53** | **58** | **180** |

A chi-square test for independence was conducted which yielded a test statistic value of 9.4405

a) State the **null and alternative hypotheses** for the above problem. (3)

b) Determine the **degrees of freedom.** (1)

c) Determine the **critical value** at the 5% level of significance. (2)

d) What is the **decision rule** and **what decision can be taken** at the 5% level of significance. (Motivate your answer by using among other the information above) (3)

e) What **conclusion** can be drawn based on the test results? Comment on the differences, if any, by among other making use of Row and/or Column percentages.

EMPLOYMENT PLAN and Company Type are not independent. An association does exist between them.

Column %

| | Employment Plan | | |
|---|---|---|---|
| Company | ADD | NO CHANGE | LAY OFF |
| Private | $\frac{37}{69} \times 100 = 53.62$ | $\frac{19}{53} \times 100 = 35.85$ | $\frac{16}{58} \times 100 = 27.59$ |
| Public | $\frac{32}{69} \times 100 = 46.38$ | $\frac{34}{53} \times 100 = 64.15$ | $\frac{42}{58} \times 100 = 72.41$ |

Private Sector is more likely to ADD employees

Public Sector is more likely to have NO CHANGE or LAY-OFF employees. (5)

[14]

Employment opportunities are better in the Private Sector

**Question 4**

The following data are the school fund fees (R1000) for a **sample of 12** model C **schools** randomly selected **in three cities** of South Africa.

| | Johannesburg | Cities<br>Cape Town | Durban |
|---|---|---|---|
| | 11 | 11 | 7 |
| | 10 | 9 | 11 |
| | 6 | 9 | 12 |
| | | 7 | 8 |
| | | 6 | |
| $\bar{x}_{j.}$ | ***A*** | ***B*** | ***C*** |
| $s_j^2$ | ***D*** | ***E*** | ***F*** |

An analysis of variance is conducted to determine whether differences exist between the mean school fund fees for the cities.

i) Calculate the values of ***A*** to ***F*** in the above table and write them down below.

$A=9.0000$ $B=8.4000$ $C=9.5000$ $D=7.0000$ $E=3.8000$ $F=5.6667$

(6)

ii) Calculate the **overall mean.**

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{n_T} = \frac{11+10+6+11+9+9+7+6+7+11+12+8}{12} = \frac{107}{12}$$

$$= 8.91666\dot{6} \approx 8.9167$$

or

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k}(n_j \bar{x}_j)}{n_T} = \frac{(3\times 9)+(5\times 8.4)+(4\times 9.5)}{3+5+4} = \frac{27+42+38}{12}$$

$$= \frac{107}{12} = 8.91666\dot{6} \approx 8.9167$$

(2)

[8]

**Question 5**

A random sample of 7 **observations** is selected **from each of 4 different** populations with normal distributions and equal variances. A Single Factor ANOVA is conducted and the following results were obtained. $K=4$ $n_j=7$ $n_T=4\times 7=28$

| Variation | SS | df. | MS | F |
|---|---|---|---|---|
| Treatments | ***A*** 780.0000525 | $k-1$ ***B*** 3 | ***C*** 260.0000175 | 9.176482 |
| Error | ***D*** 679.9992 | $n_T-K$ ***E*** 24 | 28.3333 | |
| Total | ***F*** 1459.9992 | $n_T-1$ 27 | | |

**Complete** the above **ANOVA table** by giving the **values** of ***A-F*** **below.**

$A = 780.0000$ or $780.0001$

$B = 3.0000$

$C = 260.0000$

$D = 679.9992$

$E = 24.0000$

$F = 1459.9992$ or $1459.9993$

[6]

**Question 6**
The following data are the number of Covid-19 cases reported on the first day of the school year for a **sample of 15** model C **schools** randomly selected **in three cities** of South Africa.

| | Kimberley | Bloemfontein | Pretoria |
|---|---|---|---|
| | 4 | 7 | 10 |
| | 5 | 8 | 10 |
| | 6 | 9 | 11 |
| | 6 | 7 | 11 |
| | 4 | 9 | 13 |
| $\bar{x}_j$ | 5 | 8 | 11 |
| $s_j^2$ | 1 | 1 | 1.5 |

(Cities)

An analysis of variance is conducted to determine whether differences exist between the mean number of Covid-19 cases reported for the schools in these cities.

a) Define the **response variable** for the above problem.

The number of COVID-19 cases reported on the first day of school

(1)

b) State the **null and alternative hypotheses** for the above problem.

$H_0: \mu_K = \mu_B = \mu_P$ or $\mu_1 = \mu_2 = \mu_3$

$H_A$: Not all the $\mu$'s are equal. or At least two of the $\mu$'s differ

(2)

c) Calculate the **overall mean**.

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{n_T} = \frac{4+5+6+6+4+7+8+9+7+9+10+10+11+11+13}{15}$$

$$= \frac{120}{15} = 8$$

or

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} n_j \bar{x}_j}{n_T} = \frac{5(5)+5(8)+5(11)}{5+5+5} = \frac{25+40+55}{15} = \frac{120}{15} = 8$$

or $\bar{\bar{x}} = \sum_{j=1}^{k} \bar{x}_j / k = (5+8+11)/3 = 24/3 = 8$

(2)

d) Calculate the **Sum of Squares due to Treatments** (SSTR).

$$SSTR = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2$$
$$= 5(5-8)^2 + 5(8-8)^2 + 5(11-8)^2$$
$$= 5(9) + 5(0) + 5(9)$$
$$= 45 + 0 + 45$$
$$= 90$$

(3)

e) Calculate the **degrees of freedom** for the Sum of Squares due to **Treatments**.

$k - 1 = 3 - 1 = 2$

(1)

f) Calculate the **Mean Squares** due to **Treatments** (MSTR).

$$MSTR = \frac{SSTR}{K-1} = \frac{90}{2} = 45$$

(1½)

g) Calculate the **Sum of Squares due to Errors** (SSE).

$$SSE = \sum_{j=1}^{k} (n_j - 1) s_j^2$$
$$= 4(1) + 4(1) + 4(1.5)$$
$$= 4 + 4 + 6$$
$$= 14$$

(3)

h) Calculate the **degrees of freedom** for the Sum of Squares due to **Errors**.

$$n_T - k = 15 - 3 = 12$$

(1)

i) Calculate the **Mean Squares** due to **Errors** (MSE).

$$MSE = \frac{SSE}{n_T - k} = \frac{14}{12} = 1.1666\dot{6} \approx 1.1667$$

(1½)

j) Calculate the **Test Statistic value** for the equality of the three population means.

$$F = \frac{MSTR}{MSE} = \frac{45}{1.1667} = 38.57032656 \approx 38.5703$$

or

$$F = \frac{MSTR}{MSE} = \frac{45}{1.16666} = 38.57142879 \approx 38.5714$$

(2)

k) Determine the **critical value** and what **decision can be taken** at the **1%** level of significance.

$$F_{2;12;0.01} = 6.93$$

Reject $H_0$ if $F \geq 6.93$

$H_0$ is rejected since $38.57 > 6.93$

(2)

l) What **conclusion** can be drawn based on the test results?

The mean number of Covid-19 cases in the model C shools differ (are not the same) between the three Cities.

(1)

**[21]**