CENTRE FOR MATHEMATICAL SCIENCES
EXAMINER ANSWER SCRIPT (FINAL EXAM)
SEMESTER: II SESSION: 2019/2020

Course: APPLIED STATISTICS
Course Code: BUM2413

QUESTION 1	
i)	$s_X = 0.1114$ $s_Y = 0.1949$ $f_{1-\alpha/2, v_Y, v_X} = f_{0.98,5,5} = \frac{1}{f_{0.02,5,5}} = \frac{1}{7.9529} = 0.1257$ $f_{\alpha/2, v_Y, v_X} = f_{0.02,5,5} = 7.9529$ **A 96% confidence interval for** σ_X^2 / σ_Y^2 $= \left(\frac{s_X^2}{s_Y^2} f_{1-\alpha/2, v_Y, v_X}, \frac{s_X^2}{s_Y^2} f_{\alpha/2, v_Y, v_X} \right)$ $= \left(\frac{(0.1114)^2}{(0.1949)^2}(0.1257), \frac{(0.1114)^2}{(0.1949)^2}(7.9529) \right)$ $= (0.0411, 2.5982)$
ii)	**Step 1:** $H_0 : \sigma_X = \sigma_Y$ (*Claim*: similar variability) $H_1 : \sigma_X \neq \sigma_Y$ **Step 2:** The confidence interval is $(0.0411, 2.5982)$ **Step 3:** Since $\sigma_0^2 = 1$ is in the interval $(0.0411, 2.5982)$, Do not reject H_0 **Step 5:** At $\alpha = 0.04$, there is enough evidence to support the claim that prototype X and Y has similar variability of capacity.
iii)	$\bar{x}_X = 2.9083$ $\bar{x}_Y = 2.85$ **Step 1:** $H_0 : \mu_Y - \mu_X \leq 0$ $H_1 : \mu_Y - \mu_X > 0$ (Claim: wind turbine prototype Y has higher capacity compared to prototype X) **OR** $H_0 : \mu_X - \mu_Y \geq 0$

	$H_1 : \mu_X - \mu_Y < 0$ (Claim: wind turbine prototype Y has higher capacity compared to prototype X) **Step 2:** $s_p = \sqrt{\frac{5(0.1114)^2 + 5(0.1949)^2}{10}} = 0.1587$ $t_{test} = \left(\frac{(\bar{x}_Y - \bar{x}_X) - \mu_0}{s_p \sqrt{\frac{1}{n_Y} + \frac{1}{n_X}}} \right) = \left(\frac{(2.85 - 2.9083) - 0}{(0.1587)\sqrt{\frac{1}{6} + \frac{1}{6}}} \right) = -0.6363$ **OR** $t_{test} = \left(\frac{(\bar{x}_Y - \bar{x}_X) - \mu_0}{s_p \sqrt{\frac{1}{n_Y} + \frac{1}{n_X}}} \right) = \left(\frac{(2.9083 - 2.85) - 0}{(0.1587)\sqrt{\frac{1}{6} + \frac{1}{6}}} \right) = 0.6363$ **Step 3:** $t_{\alpha,\nu} = t_{0.03,10} = 2.1202$, right-tailed test OR $t_{\alpha,\nu} = t_{0.03,10} = -2.1202$ **Step 4:** Since $(t_{test} = -0.6363) < (t_{0.03,10} = 2.1202)$, Do not reject H_0. **OR** Since $(t_{test} = 0.6363) > (t_{0.03,10} = -2.1202)$, Do not reject H_0. **Step 5:** At $\alpha = 0.03$, there is sufficient evidence to reject the claim that the wind turbine prototype Y has higher capacity compared to prototype X.
	QUESTION 2
i)	**Step 1:** $D = B - C$ $H_0 : \mu_D \le 40$ (Claim: Biodiesel produces more UHC by at most 40ppm) $H_1 : \mu_D > 40$ **Step 2:** $P - value = 0.0000$ or $1.8398E - 04$ **Step 3:** Since $(P - value = 0.0000) < (\alpha = 0.1)$, Reject H_0 **Step 4:** At $\alpha = 0.1$, there is enough evidence to reject the claim that biodiesel produces more UHC emission by at most 40ppm compared to conventional diesel.
ii)	No
iii)	State of nature: H_0 is true Statistical conclusion: H_0 is rejected So this is Type I error

QUESTION 3	
i)	P=583.1667 Q= 2 R= 4 S= 35 T=12.8426 U= 22.704 V= 2.7278
ii)	ii) Row effect H_0 : There is no effect on type of paints H_1 : There is an effect on type of paints P-value = 0.0000 @ 0.000001 Since $(P-value=0.0000)<(\alpha=0.05)$, reject H_0. At $\alpha=0.05$, there is an effect type of paints on corrosion. Column effect H_0 : There is no effect on type of steel-alloys panels H_1 : There is an effect on type of steel-alloys panels P-value = 0.0000 Since $(P-value=0.0000)<(\alpha=0.05)$, reject H_0. At $\alpha=0.05$, there is an effect type of steel-alloys panel on corrosion.
QUESTION 4	
i)	Control: Intensity of the light Respond: Time to fall asleep.
ii)	Mean, $\bar{x}=\frac{\sum_{n=1}^{12}285}{12}=23.75\ watt$ Mean, $\bar{y}=\frac{\sum_{n=1}^{12}308}{12}=25.6667\ minutes$
iii)	$H_0:\beta_1=0$ The slope is zero(no linear relationship between x and y) $H_1:\beta_1\neq 0$ The slope is not zero(there is linear relationship) $\hat{\beta}_1=\frac{S_{xy}}{S_{xx}}=\frac{1375}{1792.25}=0.7671$ $t_{test}=\frac{\hat{\beta}_1-\beta_1}{se(\hat{\beta}_1)}=\frac{0.7671-0}{\sqrt{0.004664}}=11.1893$ $\alpha=0.06, t_{0.03,10}=2.1202, -t_{0.03,10}=-2.1202$ Since $t_{test}=11.1893>t_{0.03,10}=2.1202$, Reject H_0. At $\alpha=0.06$, there exist significant relationship between the intensity of the light and the time people to fall asleep. OR

$H_0: \beta_1 = 0$ The slope is zero(no linear relationship between x and y)
$H_1: \beta_1 \neq 0$ The slope is not zero(there is linear relationship)

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{1375}{1792.25} = 0.7671$$
$$t_{test} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{0.7671 - 0}{\sqrt{0.0039}} = 12.2850$$

$$\alpha = 0.06, t_{0.03,10} = 2.1202, -t_{0.03,10} = -2.1202$$

Since $t_{test} = 12.250 > t_{0.03,10} = 2.1202$, Reject H_0.

At $\alpha = 0.06$, there exist significant relationship between the intensity of the light and the time people to fall asleep.

iv) No, any rational answer.
Yes, any rational answer.

QUESTION 5

i) 10

ii) 0.8309. 83.09% variation in the carbon emission can be predicted by the fuel consumption, the average of fuel efficiency, and the average of distance travel.

iii) When the fuel consumption (x_1) and the fuel efficiency (x_2) are held constant, the estimated of carbon emission is increase by 0.0156 million ton for every 1 unit of average of distance travel (x_3).

iv) Based on the *P*-value in the coefficients table from **Figure 3**:

$x_1 : P - value(0.0147) < \alpha(0.01)$, it is not significant predictor
$x_2 : P - value(0.6707) > \alpha(0.01)$, it is not a significant predictor
$x_3 : P - value(0.3249) > \alpha(0.01)$, it is not significant predictor
Hence, the is no significant predictors.

v) Complete the table:

lictor(s)	***P*-value**	r^2	**Adjusted r^2**	**Regression equation**
$, x_2, x_3$	0.003	0.8873	0.8309	$\hat{y} = 14.1432 + 0.0023x_1 - 0.7267x_2 + 0.0156x_3$

The best model: x1
$\hat{y} = 10.7771 + 0.0026x_1$
Reason: This model has the lowest significant *P*-value and the highest coefficient of determination.

vi) When $x_1 = 2889,\ x_2 = 7.47$ and $x_3 = 106.4$.
$\hat{y} = 10.7771 + 0.0026x_1 = 18.2885$ million ton.

QUESTION 6

(a) We should combine the adjacent

(b)	We assume $m = 0$
(c) i)	H_0: The fat content follows a Normal distribution H_1: The fat content does not follows a Normal distribution

Fat Content	**Frequency,** (O_i)	**Probability,** P_i	**Expected,** $E_i = nP_i$
$26 \le X < 28$	7	0.043	$E_1 = 175(0.043)$ $= 7.5250$
$28 \le X < 30$	22	0.108	18.9
$30 \le X < 32$	36	0.214	37.45
$32 \le X < 34$	45	0.270	47.25
$34 \le X < 36$	33	0.214	37.45
$36 \le X < 38$	28	0.108	18.9
$38 \le X < 40$	4	0.043	7.5250

No need to combine

$$\chi^2_{test} = \sum_{i=1}^{7} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(7-7.525)^2}{7.525} + \frac{(22-18.9)^2}{18.9} + \frac{(36-37.45)^2}{37.45} + \frac{(45-47.25)^2}{47.25} + \frac{(33-37.45)^2}{37.45} + \frac{(28-18.9)^2}{18.9} + \frac{(4-7.525)^2}{7.525}$$

$$= 7.2629$$

Since the value of mean and variance are given, hence $m = 0$.

$\chi^2_{critical} = \chi^2_{\alpha, k-m-1} = \chi^2_{0.025, 7-0-1} = \chi^2_{0.025,6} = 14.4494$

Since $\left(\chi^2_{test} = 7.2629\right) < \left(\chi^2_{0.025,6} = 14.4494\right)$, then we do not reject H_0

At $\alpha = 0.025$, there is sufficient evidence to conclude that the fat content follows a Normal distribution.

ii)	It's a good fit because we do not reject null hypothesis

QUESTION 7

i)	Number of accidents / Number of drivers / type of drivers
ii)	Ratio-level / Ratio-level / Nominal
iii)	H_0: All three groups of driver have the same distribution of number of accidents H_1: All three groups of driver have the different distribution of number of accidents **OR**

H_0 : All proportions are the same

H_1 : At least one proportion is different from the others

OR

$H_0 : \pi_1 = \pi_2 = \pi_3$

$H_1 : \pi_i \neq \pi_j$ for at least one $i \neq j$ where $i, j = 1, 2, 3$

Number of Accidents	**Drivers**			
	Group 1	**Group 2**	**Group 3**	$x_{i.}$
0	209	280	240	729
1	291	220	260	771
$\boldsymbol{x_{.j}}$	**500**	**500**	**500**	$x_{..}$= **1500**

O_{ij}	$E_{ij} = \dfrac{n_{i.} \times n_{.j}}{n_{..}}$	$\dfrac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$
$O_{11} = 209$	$E_{11} = \dfrac{500 \times 729}{1500} = 243$	$\dfrac{(209-243)^2}{243} = 4.7572$
$O_{12} = 280$	$E_{12} = \dfrac{500 \times 729}{1500} = 243$	$\dfrac{(280-243)^2}{243} = 5.6337$
$O_{13} = 240$	$E_{13} = \dfrac{500 \times 729}{1500} = 243$	$\dfrac{(240-243)^2}{243} = 0.0370$
$O_{21} = 291$	$E_{21} = \dfrac{500 \times 771}{1500} = 257$	$\dfrac{(291-257)^2}{257} = 4.4981$
$O_{22} = 220$	$E_{22} = \dfrac{500 \times 771}{1500} = 257$	$\dfrac{(220-257)^2}{257} = 5.3268$
$O_{23} = 260$	$E_{23} = \dfrac{500 \times 771}{1500} = 257$	$\dfrac{(260-257)^2}{257} = 0.0350$
$\chi^2_{test} = \sum_{i=1}^{r} \sum_{j=1}^{c} \dfrac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = 20.2878$		

$$\begin{aligned}
\chi^2_{critical} &= \chi^2_{\alpha,(r-1)(c-1)} \\
&= \chi^2_{0.005,(2-1)(3-1)} \\
&= \chi^2_{0.005,2} \\
&= 10.5966
\end{aligned}$$

Since $\left(\chi^2_{test} = 15.5668\right) > \left(\chi^2_{critical} = 10.5966\right)$, then, we reject H_0.

Hence, all three groups of driver have the different distribution of number of accidents at α = 0.005