FACULTY OF INDUSTRIAL SCIENCES & TECHNOLOGY
EXAMINER ANSWER SCRIPT (FINAL EXAM)
SEMESTER: I SESSION: 2018/2019

Course: APPLIED STATISTICS **Course Code: BUM2413**
Prepared By: MOHD KHAIRUL BAZLI BIN MOHD AZIZ

	QUESTION 1 (19 MARKS)
i	The weekly weight of chicks
ii	$n_1 = 20, s_1 = 2.5, v_1 = 19, n_2 = 17, s_2 = 1.8, v_2 = 16$ **Step 1: Hypothesis** $H_0 : \sigma_1^2 = \sigma_2^2 \quad (Claim)$ $H_1 : \sigma_1^2 \neq \sigma_2^2$ **Step 2:** $f_{0.05,19,16} = 2.2880$ $f_{0.05,16,19} = 2.2149$ A 90% CI for $\frac{\sigma_1^2}{\sigma_2^2} = \left(\frac{2.5^2}{1.8^2}\left(\frac{1}{2.2880}\right), \frac{2.5^2}{1.8^2}(2.2149)\right)$ $= (0.8431, 4.2726)$ **Step 3: Decision** since $0.8431 < (\sigma_0^2 = 1) < 4.2726$, do not reject H_0. **Step 4: Conclusion** At $\alpha = 0.1$, there is insufficient evidence to reject the claim. The assumption of equal population variances is validated.
iii	$n_1 = 20, s_1 = 2.5, v_1 = 19, n_2 = 17, s_2 = 1.8, v_2 = 16$ **Step 1: Hypothesis** $H_0 : \mu_1 - \mu_2 \leq 0$ $H_1 : \mu_1 - \mu_2 > 0 \quad (Claim)$ **Step 2: Test statistic** $s_p = \sqrt{\frac{19(2.5^2) + 16(1.8^2)}{20 + 17 - 2}} = 2.2077$ $t_{test} = \frac{(62 - 51) - 0}{2.2077\left(\sqrt{\frac{1}{20} + \frac{1}{17}}\right)} = 15.1040$ **Step 3: critical value** critical value : $t_{0.1,35} = 1.3062$

	Step 4: Decision Since $(t_{test} = 15.1040) > (t_{0.1,35} = 1.3062)$, reject H_0. **Step 5: Conclusion** At $\alpha = 0.1$, there is sufficient evidence to accept the claim. The egg production is higher with the 20-hour per day as compared to a 24-hour per day.
	QUESTION 2 (10 MARKS)
i	Ratio of two population variances/variances
ii	**Step 1: Hypothesis** $H_0 : \sigma_X^2 \geq \sigma_Y^2$ $H_1 : \sigma_X^2 < \sigma_Y^2$ $(claim: \text{risk stock X is less than stock Y})$ **Step 2:** $P\text{-value} = 0.0229$ **Step 3: Decision** since $(P\text{-value} = 0.0229) < (\alpha = 0.05)$, reject H_0. **Step 4: Conclusion** At $\alpha = 0.1$, there is sufficient evidence to support the claim. Stock X provides less risk as compared to Stock Y.
iii	Stock X
iv	$H_0 : \sigma_X^2 \geq \sigma_Y^2$ $H_1 : \sigma_X^2 < \sigma_Y^2$ $(claim: \text{risk stock X is less than stock Y})$ since $(P\text{ - value} = 0.0229) < (\alpha = 0.05)$, reject H_0. **No error** since reject H_0 when H_0 is not true.
	QUESTION 3 (15 MARKS)
	H_0 : There is no interaction effect gender and level of protein consumed H_1 : There is an interaction effect gender and level of protein consumed P-value = 0.1555 Since $(P - value = 0.1555) > (\alpha = 0.05)$, do not reject H_0.

	At $\alpha = 0.05$, there is no interaction effect between gender and level of protein consumed on teenager's performance during a physical fitness test. Since there is NO interaction effect, proceed with test of row effect and column Effect. Row effect H_0 : There is no effect of gender H_1 : There is an effect of gender *P*-value = 0.0004 Since $(P-value = 0.0004) < (\alpha = 0.05)$, reject H_0. At $\alpha = 0.05$, there is an effect of gender on teenager's performance during a physical fitness test. Column effect H_0 : There is no effect of amount of protein intake H_1 : There is an effect of amount of protein intake *P*-value = 0.0088 Since $(P-value = 0.0088) < (\alpha = 0.05)$, reject H_0. At $\alpha = 0.05$, there is an effect amount of protein intake on teenager's performance during a physical fitness test.
QUESTION 4 (12 MARKS)	
a	The objective of linear regression model is to model the **relationship** between the variables./ achieve parsimony model
b. i	Independent variable: Pressure Dependent variable: Yield
ii	$\sum x = 7720, \sum x^2 = 3797800, \sum y = 868, \sum y^2 = 57452, \sum xy = 422860$ $\bar{x} = 482.5$, $\bar{y} = 54.25$, $\hat{y} = 27.4230 + 0.0556x$, $se(\hat{\beta}_1) = 0.0997$ If $x = 241.25$, $y = 27.4230 + 0.0556(241.25) = 40.8365$ The pressure of 241.25 bars will produce 40.8365 liters of peanut oil.
iii	$H_0 : \beta_1 = 0$ or the slope is zero (no linear relationship between x and y) $H_1 : \beta_1 \neq 0$ or the slope is not zero (there is linear relationship) $t_{test} = \dfrac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \dfrac{0.0556 - 0}{0.0997} = 0.5577$ $\alpha = 0.02$, $t_{0.01,14} = 2.6245$, $-t_{0.01,14} = -2.6245$ Since $t_{0.01,14} = 2.6245 < t_{test} = 0.5577 < -t_{0.01,14} = -2.6245$, Do not reject H_0

	At $\alpha = 0.02$, there exist significant relationship between pressure and the peanut oil yield.
QUESTION 5 (20 MARKS)	
i	Coefficient of determination: Adjusted R^2: 0.9208 92.08% of the variation in the yield can be predicted by the pressure, temperature and particle sizes.
ii	Coefficient of x3= -16.0650: when pressure (x1) and temperature (x2) are held constants, the estimated yield will decrease by 16.0650 for each unit of particle size.
iii	$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ $H_1 : \beta_j \neq 0$ for at least one j=1,2,3 p-value = 1.8476-E07 or $f_{test} = 59.1169$, $f_{0.05,3,12} = 3.4903$ Since $(\text{p-value} = 1.8476 - \text{E07}) < (\alpha = 0.05)$, Reject H_0 or Since $f_{test} = 59.1169 >$ $f_{0.05,3,12} = 3.4903$, Reject H_0 At $\alpha = 0.02$, at least one of the independent variables is related to the dependent variable.
iv	(see table below)

Predictor	P-value	r^2	Adjusted r^2	Regression Model
x_1	0.05860	0.0217	-0.0482	$\hat{y} = 27.4230 + 0.0556x_1$
x_2	0.1375	0.1506	0.0899	$\hat{y} = 37.3214 + 0.2821x_2$
x_3	9.48984E-06	0.7644	0.7475	$\hat{y} = 97.0632 - 16.0650x_3$
x_1, x_2	0.2926	0.1723	0.0449	$\hat{y} = 10.5159 + 0.0556x_1 + 0.2821x_2$
x_1, x_3	4.4343E-05	0.7861	0.7532	$\hat{y} = 70.2576 + 0.0556x_1 - 16.0650x_3$
x_2, x_3	1.1069E-07	0.9149	0.9018	$\hat{y} = 80.1346 + 0.2821x_2 - 16.0650x_3$
x_1, x_2, x_3	1.84761E-07	0.9366	0.9208	$\hat{y} = 53.3290 + 0.0556x_1 + 0.2821x_2 - 16.0650x_3$

The best multiple linear regression model is $\hat{y} = 97.0632 - 16.0650x_3$ with value P-value = 9.48984E-06 and r^2 =0.7644

Pressure and temperature are not significant variables since P-values > $\alpha = 0.05$.

v	$x_1 = 400,\ x_2 = 60,\ x_3 = 2.28$ $\hat{y} = 97.0632 - 16.0650(2.28) = 60.4350$ liter

QUESTION 6 (13 MARKS)	
i	Any acceptable answer, Students prefer amusement park.
ii	H_0: There is no preference on the type of field trip H_1: There is a preference on the type of field trip (Claim) **(B1)** **Test Statistic:**

x_i	O_i	P_i	$E_i = nP_i$	$\frac{(O_i - E_i)^2}{E_i}$
Sporting event	35	0.25	100(0.25)=25	$\frac{(35-25)^2}{25} = 4$
Play at the local college	5	0.25	25	16
Science museum visit	10	0.25	25	9
Play at amusements parks	50	0.25	25	25
				$\chi^2_{test} = 54$

Critical Value: $\chi^2_{0.025,3} = 9.3484$

Decision: Compare: Since $\left(\chi^2_{test} = 54\right) > \left(\chi^2_{0.025,3} = 9.3484\right)$, reject H_0

(M1A1)

Conclusion:

At 2.5% significance level, the is a preference on the type of field trip.

iii	χ^2_{test} small, Observed and expected values closed to each other, decision do not reject null hypothesis
iv	No.

QUESTION 7 (11 MARKS)

H_0: Families and singles have the same distribution of cars. (claim)

H_1: families and singles do not have the same distribution of cars.

	Sport	Sedan	Hatchback	Truck	Van/SUV
Family	5	15	35	17	28
Single	45	65	37	46	7

5	15	35	17	28	100
45	65	37	46	7	200
50	80	72	63	35	300

O_{ij}	$E_{ij} = \dfrac{n_{i.} \times n_{.j}}{n_{..}}$	$\dfrac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$
$O_{11} = 5$	$E_{11} = \dfrac{100 \times 50}{300} = 16.6667$	$\dfrac{(5 - 16.6667)^2}{16.6667} = 8.1667$
$O_{12} = 15$	$E_{12} = \dfrac{100 \times 80}{300} = 26.6667$	$\dfrac{(15 - 26.6667)^2}{26.667} = 5.1042$
$O_{13} = 35$	$E_{13} = \dfrac{100 \times 72}{300} = 24$	$\dfrac{(35 - 24)^2}{24} = 5.0417$
$O_{14} = 17$	$E_{14} = \dfrac{100 \times 63}{300} = 21$	$\dfrac{(17 - 21)^2}{21} = 0.7619$
$O_{15} = 28$	$E_{21} = \dfrac{100 \times 35}{300} = 11.6667$	$\dfrac{(28 - 11.6667)^2}{11.6667} = 22.8665$
$O_{21} = 45$	$E_{21} = \dfrac{200 \times 50}{300} = 33.3333$	$\dfrac{(45 - 33.3333)^2}{33.3333} = 4.0834$
$O_{22} = 65$	$E_{21} = \dfrac{200 \times 80}{300} = 53.3333$	$\dfrac{(65 - 53.3333)^2}{53.3333} = 2.5521$
$O_{23} = 37$	$E_{23} = \dfrac{200 \times 72}{300} = 48$	$\dfrac{(37 - 48)^2}{48} = 2.5208$
$O_{24} = 46$	$E_{24} = \dfrac{200 \times 63}{300} = 42$	$\dfrac{(46 - 42)^2}{42} = 0.3810$
$O_{25} = 7$	$E_{21} = \dfrac{200 \times 35}{300} = 23.3333$	$\dfrac{(7 - 23.3333)^2}{23.3333} = 11.4333$
$\chi^2_{test} = \sum_{i=1}^{r} \sum_{j=1}^{c} \dfrac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = 62.9116$		

$$
\begin{aligned}
\chi^2_{critical} &= \chi^2_{\alpha,(r-1)(c-1)} \\
&= \chi^2_{0.005,(2-1)(5-1)} \\
&= \chi^2_{0.005,4} \\
&= 14.8603
\end{aligned}
$$

	Since $\left(\chi^2_{test} = 62.9116\right) > \left(\chi^2_{critical} = 14.8603\right)$, then, we reject H_0. (M1A1) Hence, there is sufficient evidence to conclude that families and singles do not have the same distribution of cars at $\alpha = 0.005$. (A1)