1. (10 Marks) A recent study at a university concluded that the distribution for the number of soft drinks consumed by students per week on campus is uniformly distributed. A random sample of n = 300 students produced the following data:

x = drinks	0	1	2	3	4	5	Total
number of students	55	60	50	45	50	40	300
e_i	50	50	50	50	50	50	300

Test this claim at 1% significance level.

1. H_0: The dist. for the number of soft drinks consumed by students per week on campus is uniformly distributed. ② pts

H_A: The dist. is NOT uniform.

2. $e_i = \frac{300}{6} = 50$ for all categories. ① pt

$$\chi_c^2 = \sum_{i=1}^{K} \frac{(o_i - e_i)^2}{e_i} = \frac{(55-50)^2}{50} + \frac{(60-50)^2}{50} + \frac{(50-50)^2}{50} + \frac{(45-50)^2}{50} + \frac{(50-50)^2}{50} + \frac{(40-50)^2}{50}$$

$$= 0.5 + 2.0 + 0 + 0.5 + 0 + 2.0$$

$$= 5$$

③ pts

3. $\chi^2_{\alpha, K-1} = \chi^2_{.01, 5} = 15.0863$ ① pt

Reject H_0 if $\chi_c^2 > \chi^2_{\alpha, K-1}$

$5 \not> 15.0863$

$\therefore$ Do not reject H_0 ② pts

4. Conclusion: Based on the sample data, the dist. for the number of soft drinks consumed by students per week on campus is UNIFORMLY distributed. ① pt

2. (12 Marks) A study was recently conducted in which people were asked to indicate which new medium was their preferred choice for national news. The following data were observed:

	radio	television	newspaper	total
under 21	30 / 25.19	50 / 33.05	5 / 26.76	85
21-40	20 / 22.22	25 / 29.17	30 / 23.61	75
41 and over	30 / 32.59	30 / 42.78	50 / 34.63	110
total	80	105	85	270

1 pt 1 pt 1 pt

Given this data, we wish to test whether the preferred news source is independent of age, use 5% level of significance.

1. H_0: The preferred news source is indep. of age.
   H_A: " " " " " not indep. of age. } 1 pt

2. $$\chi^2_c = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

$$= \frac{(30-25.19)^2}{25.19} + \frac{(50-33.05)^2}{33.05} + \frac{(5-26.76)^2}{26.76} + \frac{(20-22.22)^2}{22.22} + \frac{(25-29.17)^2}{29.17}$$

$$+ \frac{(30-23.61)^2}{23.61} + \frac{(30-32.59)^2}{32.59} + \frac{(30-42.78)^2}{42.78} + \frac{(50-34.63)^2}{34.63}$$

$$= 0.9185 + 8.6930 + 17.6942 + 0.2218 + 0.5961$$
$$+ 1.7294 + 0.2058 + 3.8179 + 6.8217$$

$$= 40.6957.$$

4 pts

3. $\chi^2_{\alpha,(r-1)(c-1)} = \chi^2_{.05,4} = 9.4877$ → } 1 pt

   Reject H_0 if $\chi^2_c > \chi^2_{\alpha,(r-1)(c-1)}$

   $40.6957 > 9.4877$ ✓

   ∴ Reject H_0 } → 2 pts

4. Conclusion: The preferred news source is NOT indep. of the age. } 1 pt

3. (18 Marks) A nation job placement company is interested in developing a model that might be used to explain the variation in starting salaries (in thousand dollars) for college graduates based on the college GPA. The following data were collected through a random sample of the clients with which this company has been associated.

GPA (X)	3.2	3.4	2.9	3.6	2.8	2.5	3.0	3.6	2.9	3.5
Starting salary (Y)	35.0	29.5	30.0	36.4	31.5	29.0	33.2	37.6	32.0	36.0

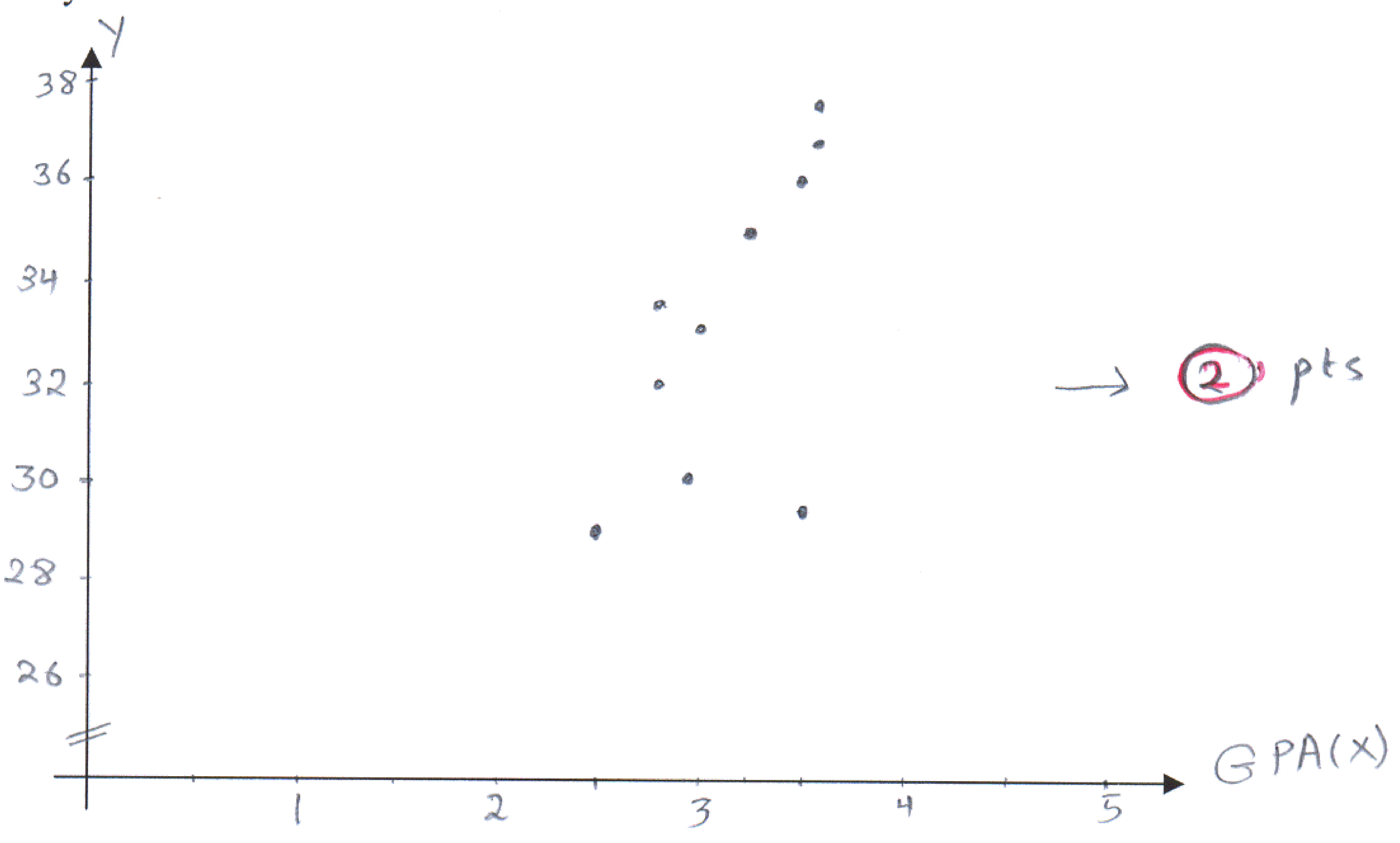
You may use the following results:

$\sum x = 31.4$, $\sum y = 330.20$, $\sum x^2 = 99.88$, $\sum y^2 = 10989.46$, $\sum xy = 1044.8$,

$SSE = 36.7873$

3

a) Plot the scatter and comment on the relation between the GPA and the starting salary.



→ 2 pts

Comment There is a positive linear relationship between the GPA(X) and the Starting Salary Y. } 1 pt

5

b)

I. Find the correlation coefficient between the two variables,

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = \frac{(10)(1044.8) - (31.4)(330.20)}{\sqrt{[(10)(99.88) - (31.4)^2][(10)(10989.46) - (330.20)^2]}}$$

$$= \frac{79.720}{105.2391} = 0.7575. \quad \Big\} \text{ 2 pts}$$

II. Test the hypothesis that there is no linear relation between the two variables.

1. $H_0: \rho = 0$ vs $H_A: \rho \neq 0$ → 1 pt

2. $t_c = \dfrac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \dfrac{.7575}{\sqrt{\frac{1-(.7575)^2}{10-2}}} = 3.2819$ → 1 pt

3. $t_{\alpha/2, n-2} = t_{.025, 8} = 2.3060$, Reject H_0 if $|t_c| > t_{\alpha/2, n-2}$

$\Rightarrow 3.2819 > 2.3060 \Rightarrow$ Reject H_0 } 1 pt

4. Conclusion: There is a significant linear relation between X & Y.

4

c) Find the equation for predicting the salary using the GPA. What is your prediction for the salary if GPA is 3.9?

I. $b_1 = \dfrac{\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} = \dfrac{1044.8 - \frac{(31.4)(330.2)}{10}}{99.88 - \frac{(31.4)^2}{10}} = \dfrac{7.972}{1.284} = 6.2087.$ } 1

$b_0 = \bar{y} - b_1\bar{x} = 33.02 - (6.2087)(3.14) = 13.5246$ } 1 pt

$\therefore \hat{y} = 13.5246 + 6.2087x.$ } → 1 pt

II. $\hat{y}(3.9) = 13.5246 + 6.2087(3.9)$
$= 37.7385$ } 1 pt

4

d) Test the hypothesis that the value of the slope is more than 5

1. $H_0: \beta_1 \leq 5$ vs $H_A: \beta_1 > 5$ } → 1 pt

2. $t_c = \dfrac{b_1 - \beta_{10}}{s_{b_1}}$, ⟹ $S_\varepsilon = \sqrt{\dfrac{SSE}{n-2}} = \sqrt{\dfrac{36.7873}{8}} = 2.1444.$

$s_{b_1} = \dfrac{S_\varepsilon}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}} = \dfrac{2.1444}{\sqrt{99.88 - \frac{(31.4)^2}{10}}}$

$= 1.8924.$

$\therefore t_c = \dfrac{6.2087 - 5}{1.8924}$

$= 0.6387$ } → 2 pts

3. $t_{\alpha, n-2} = t_{.05, 8} = 1.8595$

Reject H_0 if $t_c > t_{\alpha, n-2}$

$0.6387 \not> 1.8595$

$\therefore$ Do NOT reject H_0. } → 1 pt

4. Conclusion: Based on the sample data, the slope parameter is not more than 5.

2

e) Find a 95% confidence interval to estimate the mean salary if the GPA is 3.9.

$1 - \alpha = .95 \Rightarrow \alpha = .05 \quad \therefore \quad t_{\alpha/2, n-2} = t_{.025, 8} = 2.3060$

A 95% C.I. for the mean salary (Y) given $X_p = 3.9$ is:

$37.7385 \pm (2.3060)(2.1444)\sqrt{\dfrac{1}{10} + \dfrac{(3.9 - 3.14)^2}{99.88 - \frac{(31.4)^2}{10}}}$

37.7385 ± 3.6661

$[34.0724, 41.4046]$ } 2 pts

4. (10 Marks) An electric Company wants to estimate the relationship between the daily summer temperature in degrees Fahrenheit and the amount of electricity used by its customers in millions of kilowatts. A random sample of temperature and resulting amount of electricity used were collected and analyzed using Minitab as shown in the next page, use the Minitab output to answer the following questions:

2

a) Find the correlation coefficient between the temperature and the amount of electricity used and comment on your finding.

$r = +\sqrt{R\text{-}Sq} = \sqrt{0.88} = 0.9381$ } 1 pt

Comment: There is a strong positive linear relationship between the daily summer temperature in F° and the amount of electricity used by its customers. } 1 pt

5 pts

b) Find a 95% confidence interval for the slope and use your finding to test the significance of the regression model. Give your conclusion.

I. $1-\alpha = .95 \Rightarrow \alpha = .05$

$t_{\alpha/2, n-2} = t_{.025, 13} = 2.1604$ 1 pt, $S_{b_1} = 0.01971$ } 1 pt

A 95% C.I. for β_1 is:

$b_1 \pm t_{\alpha/2, n-2} \cdot S_{b_1}$

$0.1928 \pm (2.1604)(0.01971)$ } 1 pt

$[0.1497, 0.2354]$

II. Testing: $H_0: \beta_1 = 0$ vs $H_A: \beta_1 \neq 0$ } 1 pt

By using the C.I. for β_1, then at 95% confidence level, the zero value $\notin$ C.I., which means that } 1 pt the regression model is significant.

3

c) What are the assumptions that you need to answer par **b**?

As the error assumptions.

1. Errors are normally distributed
2. " have a mean equal to zero
3. " " a constant variance
4. " are indep.
5. X & Y are linearly related.

} 3 pts

Regression Analysis: Elect versus Temp

The regression equation is
Elect = 6.28 + 0.193 Temp

Predictor	Coef	SE Coef	T	P
Constant	6.284	1.826	3.44	0.004
Temp	0.19280	0.01971	9.78	0.000

S = 0.6873 R-Sq = 88.0% R-Sq(adj) = 87.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	45.215	45.215	95.72	0.000
Residual Error	13	6.141	0.472		
Total	14	51.356			

Unusual Observations

Obs	Temp	Elect	Fit	SE Fit	Residual	St Resid
5	84	24.200	22.479	0.240	1.721	2.67R

R denotes an observation with a large standardized residual