Birzeit University
Mathematics Department
First Semester 2014/2015
STAT 331 – Test No. 1
Instructor: Dr. Hani Kabajah

10/10

Name (الاسم): ________فراس توهاله________ **Student No.:** ____1121698____

Question 1: Independent simple random samples are taken to test the difference between the means of two populations whose variances are not known, but are assumed to be equal. The sample sizes are $n_1 = 32$ and $n_2 = 40$. The correct distribution to use is the

$\sigma_1 = \sigma_2$ unkn

t $df = n+n-2$

$dF = n_1 + n_2 - 2$
$= 32 + 40 - 2$
$= 72 - 2 = 70$

a. t distribution with 73 degrees of freedom
b. t distribution with 72 degrees of freedom
c. t distribution with 71 degrees of freedom
d. t distribution with 70 degrees of freedom

1

Question 2: Salary information regarding male and female employees of a large company is shown below.

	Male	Female
n Sample Size	64	36
$\bar{X}$ Sample Mean Salary (in $1,000)	44	41
σ^2 Population Variance σ^2	128	72

The 95% confidence interval for the difference between the means of the two populations is

$\alpha = 0.05$
0.025
$z = 1.96$

$\alpha = 0.05$
$\alpha/2 = 0.025$

$(44-41) \mp$

$\bar{X}_1 - \bar{X}_2 \mp z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

3 ∓ 1.96

∓ 39.2

$44 - 41 \mp z_{\frac{\alpha}{2}} \sqrt{\frac{128^2}{64} + \frac{72^2}{36}}$

$3 \mp 1.96(20)$

$3 \mp 1.96(20)$

a. 0 to 6.92
b. -2 to 2
c. -1.96 to 1.96
d. -0.92 to 6.92

Question 3: The daily production rates for a sample of workers before and after a training program are shown below.

Worker	Before	After	d
1	20	22	-2
2	25	23	2
3	27	27	0
4	23	20	3
5	22	25	-3
6	20	19	1
7	17	18	-1

$\bar{d} = 0$
$s = 2.16$

The null hypothesis to be tested is $H_0: \mu_d = 0$. The test statistic is

$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$

a. -1.96
b. 1.96
c. 0
d. 1.645

Question 4: An insurance company selected samples of clients under 18 years of age and over 18 and recorded the number of accidents they had in the previous year. The results are shown below.

Under Age of 18	Over Age of 18
$n_1 = 500$	$n_2 = 600$
Number of accidents = 180	Number of accidents = 150

$\pi_1 - \pi_2 = 0$
$\pi_1 - \pi_2 \neq 0$

We are interested in determining if the accident proportions differ between the two age groups.

The pooled proportion is

$$\frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{180+150}{600+500}$$

a. 0.305
b. 0.300
c. 0.027
d. 0.450

n obsd

Question 5: A sample of n observations is taken from a population. The appropriate chi-square distribution has

a. n degrees of freedom
b. $n - 1$ degrees of freedom
c. $n - 2$ degrees of freedom
d. $n - 3$ degrees of freedom

Question 6: The producer of a certain medicine claims that their bottling equipment is very accurate and that the standard deviation of all their filled bottles is 0.1 ml or less. A sample of 20 bottles showed a standard deviation of 0.11. The test statistic to test the claim is

a. 400
b. 22.99
c. 4.85
d. 20

$s \leq 0.1$ $n = 20$ $s = 0.11$ $\quad \frac{(n-1)s^2}{\sigma_0^2} = \frac{19(0.11)^2}{(0.1)^2}$

Question 7:

	Sample A	Sample B
s^2	12.1	5
n	11	10

We want to test the hypothesis that population A has a larger variance than population B.

The p-value is approximately

$\sigma_A^2 \leq \sigma_B^2$
$\sigma_A^2 > \sigma_B^2$

a. 0.10
b. 0.05
c. 0.025
d. 0.01

$F = \frac{s_1^2}{s_2^2} = 5.8$ $\quad df = (10)(9) = 90$

$F = \frac{s_1^2}{s_2^2} = \frac{(12.1)}{(5)} =$ $\quad F = \cancel{5.8564}$ $\quad 2.42$

$df_1 = 10$
$df = 9$

Question 8: The sampling distribution for a goodness of fit test is the

a. Poisson distribution
b. t distribution
c. normal distribution
d. chi-square distribution

Question 9: In the past, 35% of the students at ABC University were in the Business College, 35% of the students were in the Liberal Arts College, and 30% of the students were in the Education College. To see whether or not the proportions have changed, a sample of 300 students was taken. Ninety of the sample students are in the Business College, 120 are in the Liberal Arts College, and 90 are in the Education College.

	f_i	e_i	$\frac{(f_i-e_i)^2}{e_i}$
B	90	105	2.143
L	120	105	2.143.
E	90	90	0
			4.286
			4.29

The calculated value for the test statistic equals

$$\chi^2 = \sum \frac{(f_i - e_i)^2}{e_i}$$

a. 0.01
b. 0.75
c. 4.29
d. 4.38

Question 10: The table below gives beverage preferences for random samples of teens and adults.

	Teens (Tn)	Adults (A)	Total
C Coffee	50	200	250
T Tea	100	150	250
S Soft Drink	200	200	400
O Other	50	50	100
Total	400	600	1,000

We are asked to test for independence between age (i.e., adult and teen) and drink preferences.

With a .05 level of significance, the critical value for the test is

$df = (n-1)(m-1)$
$(4-1)(2-1)$
$\boxed{= 3}$
χ^2_α

a. 1.645
b. 7.815
c. 14.067
d. 15.507

(2-1)(4-1)
(3)

χ_α

25 + 16.6 + 0
40 + 10 + 16
2.5 + 1.67

e_i

100	150	250
100	150	
160	240	
40	60	

$\chi^2 =$

Birzeit University
Mathematics Department
Second Semester 2014/2015
STAT 331 – Hour Exam No. 2
Instructor: Dr. Hani Kabajah

Name (بالعربية): ____KEY________________ **Student No.:** ____________

Question 1 (10 points)

We are interested in the following model $Y = \beta_0 + \beta_1 X + \varepsilon$ and we are given the following sample:

x	1	2	3	4	5
y	4	5	4	2	1

1) Calculate b_0 and b_1 the least square estimates of β_0 and β_1
2) Calculate the coefficient of determination R^2
3) Calculate the adjusted R^2
4) Calculate the correlation coefficient
5) Using the F-test
   a) calculate the test statistic
   b) find the critical value at 10% significance
   c) Is the above model 10% significant or not? Explain your answer?

(1) 1) $b_1 = \dfrac{\sum xy - n\bar{x}\bar{y}}{(n-1)S_x^2} = \dfrac{39 - 5(3)(3.2)}{10} = -0.9$

(1) $b_0 = \bar{y} - b_1\bar{x} = 3.2 - (-0.9)(3) = 5.9$

(1) (1) 2) $R^2 = \dfrac{SSR}{SST} = \dfrac{b_1^2(n-1)S_x^2}{(n-1)S_y^2} = \dfrac{(-0.9)^2(10)}{10.8} = \dfrac{8.1}{10.8} = 0.75$

(1) 3) $adj\text{-}R^2 = 1 - (1-R^2)\dfrac{(n-1)}{(n-p-1)} = 1 - (1-0.75)\dfrac{(4)}{(3)} = 0.67$

(1) (1) 4) $r = (\text{sign } b_1)\sqrt{R^2} = (-1)\sqrt{0.75} = -0.87$

5) a)

(1)

Source of Error	df	SS	MS	F
Regression	1	8.1	8.1	9
Error	3	2.7	0.9	
Total	4	10.8		

(1) b) critical value : $F_\alpha = F_{0.10} = 5.54$ $(df_1 = 1, df_2 = 3)$

(1) c) $H_0: \beta_1 = 0$, $F = 9 > F_\alpha = 5.54 \Rightarrow$ Reject $H_0 \Rightarrow \beta_1 \neq 0 \Rightarrow$ Model is significant $(\alpha = 0.10)$

Question 2 (10 points)

Given the output on the next page answer the following questions.

1) Write the estimated regression equation (use the right variable and the right values).
2) What is the sample size?
3) What is the value of R^2?
4) What is the value of adjusted R^2?
5) How much is the model good-to-fit?
6) Is the whole model 1% significant? Explain your answer.
7) Using 1% significance level, which variable should remain in the model and which variable should be eliminated? Explain your answer.
8) Is there multicolinearity in the model? Explain your answer.
9) Is the model valid, approximately valid or not valid? Explain your answer.
10) Estimate the output variable if all input variables are equal to 1? Is the estimation good? Explain your answer.

1 1) $\hat{Y} = -0.60 + 10.21 x_1 + 89.03 x_2 + -0.14 x_3 + 18.21 x_4$

1 2) sample size: $n = 33$

1 3) $R^2 = 0.95$

1 4) adj-$R^2 = 0.94$

1 5) The model is 94% good-to-fit.

1 6) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, p-value $= 3.1 \times 10^{-17} \leq \alpha = 0.01$

$\Rightarrow$ Reject H_0 ($\alpha = 0.01$) $\Rightarrow$ The whole model is significant ($\alpha = 0.01$)

1 7) $H_0: \beta_1 = 0$, p-value $= 0.03 > \alpha = 0.01 \Rightarrow$ Don't reject $H_0 \Rightarrow \beta_1 = 0 \Rightarrow$ Eliminate X_1

$H_0: \beta_2 = 0$, p-value $= 3.7 \times 10^{-18} \leq \alpha = 0.01 \Rightarrow$ Reject $H_0 \Rightarrow \beta_2 \neq 0 \Rightarrow$ Keep X_2

$H_0: \beta_3 = 0$, p-value $= 0.94 > \alpha = 0.01 \Rightarrow$ Don't Reject $H_0 \Rightarrow \beta_3 = 0 \Rightarrow$ Eliminate X_3

$H_0: \beta_4 = 0$, p-value $= 0.0002 \leq \alpha = 0.01 \Rightarrow$ Reject $H_0 \Rightarrow \beta_4 \neq 0 \Rightarrow$ Keep X_4

$\Rightarrow$ Using 1% significance, eliminate X_1, X_3 and keep X_2, X_4

1 8) Answer 1: There is multicolinearity since some variables are not significant

Answer 2: We can't tell since we need regression output of X_i

1 9) Residuals are approximately random, normal, with zero mean, and constant variance $\Rightarrow$ The model is approximately valid

1 10) $\hat{Y}(1,1,1,1) = -0.6 + 10.21(1) + 89.03(1) - 0.14(1) + 18.21(1) = 116.71$

Answer 1: The estimation is good since adj-$R^2 = 0.94$, model significant ($\alpha = 0.01$) and approx. valid

Answer 2: We can't tell since we need to build a new model

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.972240012
R Square	0.945250642
Adjusted R Square	0.937429305
Standard Error	7.500388353
Observations	33

ANOVA

	df	SS	MS	F	Significance F
Regression	4	27195.27659	6798.819146	120.8553797	3.0947E-17
Residual	28	1575.163113	56.25582545		
Total	32	28770.4397			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 99.0%	Upper 99.0%
Intercept	-0.595522721	4.321666476	-0.137799324	0.89138521	-9.448055077	8.257009636	-12.53742138	11.34637594
x1	10.21187396	4.433207457	2.303495619	0.028887984	1.130860259	19.29288765	-2.038241711	22.46198962
x2	89.0341356	4.43681749	20.06711698	3.71533E-18	79.94572709	98.12254411	76.77404447	101.2942267
x3	-0.143193991	1.900522187	-0.075344551	0.940476327	-4.03623716	3.749849178	-5.39483557	5.108447589
x4	18.20886991	4.301978528	4.232673359	0.000224521	9.396666484	27.02107333	6.321374213	30.0963656

RESIDUAL OUTPUT

Observation	Predicted y	Residuals	Standard Residuals
1	51.79881588	6.845031664	1.204916042
2	28.25900829	5.542731735	0.975675017
3	24.2626542	-1.310357652	-0.230659409
4	42.0031061	0.155446157	0.027362849
5	78.73922116	-4.790816139	-0.843316949
6	53.17492307	-9.13966933	-1.608836121
7	39.81031926	-0.263080451	-0.04630948
8	31.98298093	4.352361967	0.766136814
9	73.28473294	-7.909123614	-1.392225834
10	11.31161348	-4.93401781	-0.868524428
11	71.41445952	-8.013749957	-1.410642982
12	79.97403871	6.382810921	1.123552329
13	25.11424191	-4.550917292	-0.801088077
14	26.88117038	-8.021337999	-1.411978689
15	18.50313677	-3.626090501	-0.638292827
16	76.56167275	-5.115432807	-0.900458515
17	80.0312567	8.716179773	1.534290175
18	70.47119751	5.728565752	1.008386975
19	56.64248983	-2.919960412	-0.513994283
20	77.98519187	0.285285324	0.050218155
21	21.64828857	2.545943774	0.448156947
22	67.69240932	4.468569187	0.786592519
23	11.07166128	-1.598388277	-0.281360813
24	86.39654969	-3.885551011	-0.683965097
25	6.908022981	1.19754088	0.210800518
26	5.149738289	-15.170072	-2.670354793
27	26.68292482	6.629432833	1.166964648
28	96.71353403	-2.486442733	-0.437683108
29	36.5321341	-1.494780405	-0.263122945
30	39.29424008	6.101699446	1.074068888
31	67.8574344	-3.588451684	-0.631667348
32	92.08530404	-0.945262675	-0.166392534
33	44.14653066	-3.906045025	-0.687572614

PROBABILITY OUTPUT

Percentile	y
1.515151515	-1.391990497
4.545454545	2.146063623
7.575757576	7.751305178
10.60606061	13.58843512
13.63636364	15.75791472
16.66666667	16.39711667
19.6969697	18.40640288
22.72727273	23.3835037
25.75757576	24.42571082
28.78787879	24.53737729
31.81818182	25.79076838
34.84848485	25.86009457
37.87878788	32.59988512
40.90909091	34.04912984
43.93939394	40.95652419
46.96969697	41.56279118
50	44.07914314
53.03030303	46.17663098
56.06060606	46.57302266
59.09090909	47.52069389
62.12121212	50.04790274
65.15151515	52.03222642
68.18181818	54.19968383
71.21212121	59.09991481
74.24242424	63.8823398
77.27272727	67.30757615
80.3030303	69.44732034
83.33333333	73.2262197
86.36363636	85.43455968
89.39393939	86.78744833
92.42424242	88.07226927
95.45454545	94.51830123
98.48484848	97.75162114