Business Analytics **Final** Summer 2022

系級_____ 姓名 _____ 學號 _____

A. Objective questions: (70%) (2 marks each)

1. The sampling distribution for a goodness of fit test is the
   a. Poisson distribution
   b. t distribution
   c. normal distribution
   d. chi-square distribution

2. A statistical test conducted to determine whether to reject or not reject a hypothesized probability distribution for a population is known as a
   a. contingency test
   b. probability test
   c. goodness of fit test
   d. None of these alternatives is correct.

3. The following table gives beverage preferences for random samples

	Teens	Adults	Total
Coffee	50	200	250
Tea	100	150	250
Soft Drink	200	200	400
Other	50	50	100
	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
   a. 1.645
   b. 7.815
   c. 14.067
   d. 15.507

4. An important application of the chi-square distribution is
   a. making inferences about a single population variance
   b. testing for goodness of fit
   c. testing for the independence of two variables
   d. All of these alternatives are correct.

5. The degrees of freedom for a contingency table with 10 rows and 11 columns is
   a. 90
   b. 100
   c. 110
   d. 109

6. The mean square is the sum of squares divided by
   a. the total number of observations
   b. its corresponding degrees of freedom
   c. its corresponding degrees of freedom minus one
   d. None of these alternatives is correct.

7. The F ratio in a completely randomized ANOVA is the ratio of
   a. MSTR/MSE
   b. MST/MSE
   c. MSE/MSTR
   d. MSE/MST

8. In ANOVA, which of the following is not affected by whether or not the population means are equal?
   a. $\bar{x}$
   b. between-samples estimate of σ^2
   c. within-samples estimate of σ^2
   d. None of these alternatives is correct.

9. In a completely randomized design involving four treatments, the following information is provided.

	Treatment 1	Treatment 2	Treatment 3	Treatment 4
Sample Size	50	18	15	17
Sample Mean	32	38	42	48

The overall mean (the grand mean) for all treatments is
   a. 40.0
   b. 37.3
   c. 48.0
   d. 37.0

10. SSTR = 6,750 H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
    SSE = 8,000 H_a: at least one mean is different
    nT = 20
    The mean square between treatments (MSTR) equals
    a. 400
    b. 500
    c. 1,687.5
    d. 2,250

11. An ANOVA table is shown below.

Source of Variation	***Sum of Squares***	***Degrees of Freedom***	***Mean Square***	***F***
BetweenTreatments	*180*	*3*		
Within Treatments				
(Error)				
TOTAL	*480*	*18*		

The test statistic is

a. 2.25
b. 3
c. 2.67
d. 6

12. The coefficient of determination

a. cannot be negative
b. is the square root of the coefficient of correlation
c. is the same as the coefficient of correlation
d. can be negative or positive

ANS: A

13. In a simple linear regression analysis, the coefficient of correlation is 0.16. The coefficient of determination in this situation is

a. 0.4000
b. 0.0256
c. 4
d. 2.56

14. In a regression and correlation analysis if r2 = 1, then

a. SSE = SST
b. SSE = 1
c. SSR = SSE
d. SSR = SST

15. A least squares regression line

a. may be used to predict a value of y if the corresponding x value is given
b. implies a cause-effect relationship between x and y
c. can only be determined if a good linear relationship exists between x and y
d. None of these alternatives is correct.

16. A regression and correlation analysis resulted in the following information regarding a dependent variable (y) and an independent variable (x).

$\Sigma X = 90$ $\quad\quad$ $\Sigma (Y - \bar{Y})(X - \bar{X}) = 466$

$\Sigma Y = 170$ $\quad\quad$ $\Sigma (X - \bar{X})^2 = 234$

$n = 10$ $\quad\quad$ $\Sigma (Y - \bar{Y})^2 = 1434$

$SSE = 505.98$

The least squares estimate of b1 equals

a. 0.923
b. 1.991
c. -1.991
d. -0.923

17. Refer to Problem #16, the MSE is

a. 17
b. 8
c. 34
d. 42

18. In multiple regression analysis, the correlation among the independent variables is termed

a. homoscedasticity
b. linearity
c. multicollinearity
d. adjusted coefficient of determination

19. For a multiple regression model, SSR = 600 and SSE = 200. The multiple coefficient of determination is

a. 0.333
b. 0.275
c. 0.300
d. 0.75

20. A regression model between sales (Y in $1,000), unit price (X1 in dollars) and television advertisement (X2 in dollars) resulted in the following function: $\hat{Y} = 7 - 3X_1 + 5X_2$ For this model SSR = 3500, SSE = 1500, and the sample size is 18. To test for the significance of the model, the test statistic F is

a. 2.33
b. 0.70
c. 17.5
d. 1.75

21. To test for the significance of the model in Problem #21, the p-value is
    a. less than 0.01
    b. between 0.01 and 0.025
    c. between 0.025 and 0.05
    d. between 0.05 and 0.10

22. The ratio of MSE/MSR yields
    a. SST
    b. the F statistic
    c. SSR
    d. None of these alternatives is correct.

23. In a regression model of 30 observations, the following estimated regression equation was obtained: $\hat{Y}=170+34X_1-3X_2+8X_3+58X_4+3X_5$ with SSR = 1,740 and SST = 2,000. The F test for testing the significance of the above model is
    a. 32.12
    b. 6.69
    c. 4.8
    d. 58

24. Which of the following is not present in a time series?
    a. seasonality
    b. operational variations
    c. trend
    d. cycles

25. Given an actual demand of 61, forecast of 58, and an α of .3, what would the forecast for the next period be using simple exponential smoothing?
    a. 57.1
    b. 58.9
    c. 61.0
    d. 65.5

26. The time series component that reflects variability during a single year is called
    a. a trend
    b. seasonal
    c. cyclical
    d. irregular

27. For the following time series, you are given the moving average forecast.

Time Period	***Time Series Value***	***Moving Average Forecast***
1	*23*	
2	*17*	
3	*17*	
4	*26*	*19*
5	*11*	*20*
6	*23*	*18*
7	*17*	*20*

The mean squared error equals

a. 0
b. 6
c. 41
d. 164

28. Below you are given the first four values of a time series.

Time Period	***Time Series Value***
1	*18*
2	*20*
3	*25*
4	*17*

Using a 4-period moving average, the forecasted value for period 5 is

a. 2.5
b. 17
c. 20
d. 10

29. A method of smoothing a time series that can be used to identify the combined trend/cyclical component is

a. the moving average
b. the percent of trend
c. exponential smoothing
d. the trend/cyclical index

30. A qualitative forecasting method that obtains forecasts through "group consensus" is known as the

a. Autoregressive model
b. Delphi approach
c. Moving average
d. None of these alternatives is correct.

31. Parametric methods are statistical methods that
    a. require some assumptions about the population
    b. require no assumptions about the population
    c. only deal with small samples
    d. considers the sign of two matched samples

32. On a teacher evaluation form students are asked to rate their professor's performance as excellent, very good, good, and poor. This is an example of the
    a. nominal scale
    b. ratio scale
    c. ordinal scale
    d. interval scale

33. Fifteen people were given two types of cereal, Brand X and Brand Y. Two people preferred Brand X and thirteen people preferred Brand Y. We want to determine whether or not customers prefer one brand over the other. The p-value for this test is
    a. 0.0005
    b. 0.001
    c. 0.0037
    d. 0.0074

34. A nonparametric method for determining the differences between two populations based on two matched samples where only preference data is required is the
    a. Mann-Whitney-Wilcoxon test
    b. Wilcoxon signed-rank test
    c. sign test
    d. Kruskal-Wallis Test

35. The level of measurement that is simply a label for identifying an item is
    a. ordinal measurement
    b. ratio measurement
    c. nominal measurement
    d. internal measurement

B. Subjective questions (30%, 5% for each problem)

> ***Important Note:*** *Marks will be given on any crucial steps. Hypothesis statement, degree of freedom, test statistic, critical value, etc. will be considered as well. Furthermore, you should state your process clearly and neatly. NO MARKS will be given if there is only a final answer. Good luck.*

1. A clothing manufacturer purchased some newly designed sewing machines in the hopes that production would be increased. The production records of a random sample of workers are shown below.

Worker	Old Machine	New Machine
1	28	36
2	36	40
3	27	25
4	25	32
5	38	30
6	36	32
7	40	40
8	29	28
9	32	35
10	28	33
11	20	26
12	32	31
13	32	23
14	32	34
15	36	36

   Use the Wilcoxon signed-rank test to determine whether the new machines have significantly increased production. Use a .05 level of significance.

2. We want to determine the relationship between daily supply (y) and the unit price (x) for a particular item. A sample of ten days supply and associated price resulted in the following data: $\Sigma x = 66$, $\Sigma x^2 = 526$, $\Sigma y = 71$, $\Sigma y^2 = 605$, and $\Sigma xy = 557$.

   (a) Develop the least square estimated regression equation.

   (b) Compute the coefficient of determination and fully explain its meaning.

   At $\alpha = 0.05$, perform a t-test and determine if the slope is significantly different from zero.

3. In a regression analysis involving 21 observations and 4 independent variables, the following information was obtained: $R^2 = 0.80$ and $S = 5.0$. Based on the above information, fill in all the blanks in the following ANOVA.

Source	**DF**	**SS**	**MS**	**F**
Regression	_____?	_____?	_____?	_____?
Error (Residuals)	_____?	_____?	_____?	
Total	_____?	_____?		

4. The Ahmadi Corporation wants to increase the productivity of its line workers. Four different programs have been suggested to help increase productivity. Twenty employees, making up a sample, have been randomly assigned to one of the four programs and their output for a day's work has been recorded. (Note: SST = 16,350 & SSTR = 8,750)

Program A	**Program B**	**Program C**	**Program D**
150	150	185	175
130	120	220	150
120	135	190	120
180	160	180	130
145	110	175	175

a. State the null and alternative hypotheses.

b. As the statistical consultant to Ahmadi, what would you advise them? Use a .05 level of significance. Use both the critical and *p*-value approaches.

c. Use Fisher's LSD procedure and determine which population mean (if any) is different from the others. Let $\alpha = .05$.

5. The following table shows the results of a study on smoking and three illnesses. We are interested in determining if the proportions smokers in the three categories are different from each other. ($\alpha = 0.05$)

	Emphysema	Heart problem	Cancer	Total
Smoker	150	70	100	320
Non-smoker	50	130	500	680
Total	200	200	600	1000

Compute the critical values and give your conclusions by providing numerical reasoning.

6. Below you are given information on crime statistics for Middletown.

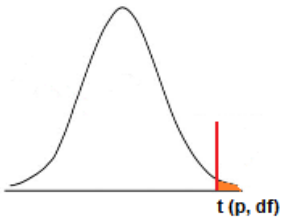
Year	Quarter	Number of Crimes Committed Y_t
1	1	10
	2	20
	3	25
	4	5
2	1	10
	2	30
	3	35
	4	25
3	1	20
	2	40
	3	35
	4	15
4	1	20
	2	50
	3	45
	4	35

The seasonal factors for these data are

Quarter	Seasonal Factor S_t
1	.589
2	1.351
3	1.335
4	.726

Deseasonalize the series. Given T = 11.5 + 1.7353×t, produce forecasts for all four quarters of year 5 by using the seasonal and trend components.

Right-sided t-distibution



Right-sided T-distribution table

df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	4.3178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
∞	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905

- Percentile of χ^2-Distribution

	P										
DF	0.995	0.975	0.2	0.1	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	.0004	.00016	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.55	10.828
2	0.01	0.0506	3.219	4.605	5.991	7.378	7.824	9.21	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.86	16.924	18.467
5	0.412	0.831	7.289	9.236	11.07	12.833	13.388	15.086	16.75	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.69	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.18	11.03	13.362	15.507	17.535	18.168	20.09	21.955	24.352	26.124
9	1.735	2.7	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.92	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.3	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.79
18	6.265	8.231	22.76	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.9	27.204	30.144	32.852	33.687	36.191	38.582	41.61	43.82
20	7.434	9.591	25.038	28.412	31.41	34.17	35.02	37.566	39.997	43.072	45.315

- Upper Critical Values of F-distribution ($\alpha = 0.05$ & $\alpha = 0.01$)

5% significance level

$F_{.05}(\nu_1, \nu_2)$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.882	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978

1% significance level

$F_{.01}(\nu_1, \nu_2)$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10
1	4052.19	4999.52	5403.34	5624.62	5763.65	5858.97	5928.33	5981.10	6022.50	6055.85
2	98.502	99.000	99.166	99.249	99.300	99.333	99.356	99.374	99.388	99.399
3	34.116	30.816	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814
9	10.561	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849

系級_____ 姓名 _____ 學號 _____

A. 選擇題

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35					

B. 計算題