53.1201/6-001 and 53.1201/6-002   Introduction to Stat Analysis, Final Exam

Wednesday, April 11, 2007 – 9:00 a.m., 3 hours duration
Instructors: A. Basilevsky, B. Bector

Special Instructions:

1. Necessary statistical tables and formulae are provided.
2. Show all your work on the right-hand side in the examination booklet provided.
3. Non-programmed hand calculators only may be used.
4. Write the name of your instructor on the front of your examination booklet.

NOTE: THERE ARE THREE PARTS TO THIS EXAM

Value

PART I: – Answer all ten questions.

Each is worth 1 mark.

Questions 1, 2, 3, refer to the following data: 9, 2, 3, 5, 4, 7.

1. The median of these data is:

   (a) 30   (b) 5   (c) 6   (d) 4.5   (e) 5.5

2. The mean of these data is:

   (a) 30   (b) 5   (c) 6   (d) 4.5   (e) 5.5

3. The standard deviation of these data is:

   (a) 2.6077   (b) 2.3085   (c) 6.8000   (d) 5.6667   (e) 5.5

Questions 4, 5, 6 refer to events A, B with the following probabilities:
$P(A) = .60$, $P(B) = .40$, $P(A \cap B) = .20$

4. Events A, B are:

   (a) mutually exclusive   (b) dependent   (c) independent
   (d) Binomial probabilities   (e) none of the above

5. The probability of event A occurring, given that event B has occurred, is:

   (a) 2/3   (b) .40   (c) 1   (d) .50   (e) 0

6. The probability of event A occurring <u>or</u> event B occurring, is:

   (a) .80   (b) .40   (c) 1   (d) .50   (e) 0

/...to page 2

Value

7. A random variable x has the following probability distribution:

x	0	1	2	3	4
$p(x)$	.30	.30	.10	.10	.20

The mean of x is (a) 2.50   (b) 2.12   (c) .38   (d) .40   (e) 1.60

8. Refer to the probability distribution of Question 7. The variance of x is:

(a) 2.24   (b) 1.50   (c) 5.40   (d) 10.10   (e) 6.80

9. The $P(\text{Type I error}) = \alpha$ is

(a) a calculation error for the critical value.

(b) the acceptance region of the null hypothesis.

(c) the probability of rejecting a true null hypothesis.

(d) the probability of accepting a true null hypothesis.

(e) the probability of rejecting a true alternative hypothesis.

10. When testing a null hypothesis H_0 against an alternative hypothesis H_a

(a) we are determining which hypothesis is true.

(b) we are deciding which hypothesis to accept/reject.

(c) we are attempting to estimate a parameter.

(d) we are trying to determine the Type I error.

(e) we are trying to determine the type II error.

PART II:
- **Answer any 10 of the following 12 questions.**

- **Each is worth 6 marks.**

11. A random sample of $n = 620$ adults indicates that 340 of those sampled think Winnipeg needs a new housing development. Assuming that 48% of the adult population of Winnipeg believe that the city needs a new housing development, what is the probability of obtaining 320 individuals or more who are in favour? $P[Z \geq 1.7604] = 0.0392$

12. An average of 2.9 emergency cases arrive at a hospital by ambulance, per hour. Using the Poisson distribution, compute the probability that 3 emergencies or more per hour arrive at the hospital. 0.5540

13. In a study of a tropical fish population, there were a total of 9 females and 14 males living in the vicinity of a trap. Assuming that each of the fish is equally likely to be trapped, what is the probability of getting 4 female and 8 male fish if 12 are trapped? 0.2798

Value

14. A drug that provides relief from headaches was tried on a random sample of 8 patients. These patients had a mean time until relief of 26.2 minutes with a standard deviation of 3.74.

$(22.236, 30.164)$

(a) Construct a 98% confidence interval for the true mean time until relief.

(b) State the necessary assumption(s) and provide an interpretation of your result.

15. According to advertisements, a strain of canola planted in soil prepared with a specified fertilizer treatment has a mean yield of 500 bushels per acre. Fifty farmers who belong to a cooperative plant the canola. Each uses a 40-acre plot and records the mean yield per acre. The mean and variance for the sample of 50 farms are $\bar{x} = 485$ and $s^2 = 10,045$. Use the p-value for this test to determine whether the data provide sufficient evidence to indicate that the mean yield for the canola is difference from that advertised at the 0.10 significance level. p-value $= 0.2892$ Do not reject H_0

16. The following measurements (in ounces) were obtained on the weights of six laboratory mice picked at random: 12, 8, 7, 12, 14 and 13. At the 5 percent level of significance, is the population variance of the weights of such mice greater than 2.25? State the assumption(s) you make about the population distribution. $\chi^2_c = 17.78$ Reject H_0

17. A population has a normal distribution with variance 225. Find how large a sample must be drawn in order to be 92 percent confident that the sample mean will not differ from the population mean by more than 2 units.

$z_{\frac{\alpha}{2}} = 1.75$
$n = 173$

18. Suppose a factory manufactures a new type of electric bulbs. An engineer employed by the factory believes that she has found a new manufacturing technique that will increase the mean life of the bulb, which is currently believed to be 450 hours. Suppose the factory management adopts the decision rule on the basis of a sample of bulbs $n = 16$ to reject the null hypothesis if $\bar{x} > 475$ hours. The standard deviation of bulb life is known to be 40.0 hours. Assume the sample came from a normal population.

(a) State the Type I and Type II errors in this context.

(b) Determine α, the probability of a Type I error. 0.0062

(c) Determine β, the probability of a Type II error when in fact $\mu = 490.0$. 0.0668

Value

19. A large supermarket chain is interested in determining whether a difference exists between the mean shelf life in days of Brand A bread and Brand B bread. Random samples of 10 freshly baked loaves of each brand were tested, with the results shown below:

Brand A	Brand B
$\bar{x} = 4.1$	$\bar{y} = 5.2$
$s_1 = 1.2$	$s_2 = 1.4$

$t_c = -1.886$, Do not reject Ho

Is there sufficient evidence to conclude that a difference does exist between the mean shelf lives of Brand A and Brand B bread? Test at the $\alpha = 0.05$ level of significance. State the assumptions required for the test.

20. A sample of six persons admitted driving under the influence of alcohol. The court asked all to attend a series of counselling sessions. The following table gives the number of times these six persons drove under the influence of alcohol one month before and one month after attending the counselling sessions.

Before	11	8	13	6	4	12
After	7	5	8	3	4	8

$t_c = 4.505$ Reject Ho

(a) test at the 1% significance level if attending the counselling sessions decreases the mean number of times all persons drive under the influence of alcohol.

(b) State the assumption(s) required for the test in (a).

21. A sample of 400 items produced by Supplier A contained 70 defective items. A sample of 300 items produced by Supplier B contained 40 defective items. Test, at the 5% level of significance, whether there is a difference between the proportion of defective items for the two suppliers.

$z_c = 1.50$ Do not reject Ho

22. The Supreme Beverage Company has developed "LiteCol", a new type of diet cola. A marketing study is conducted to evaluate the new product. A random sample of 300 persons was recruited for a taste test. Each person in the sample tasted LiteCol and four other brands of diet cola and were then asked to state the preferred brand. To avoid bias, brand labels were replaced with letters A, B, C, D or E and the order of tasting was randomized for each person. Sample results are presented below.

$x_c^2 = 9.168$ Do not reject Ho

Brand	A	B	C	D	E
Number of persons preferring brand	50	65	45	70	70

Test, at the 5% level of significance, whether each brand of cola is preferred by an equal percentage of the population.

Value

PART III:
 – Answer all three questions.   – Each is worth 10 marks.

23. *The* Wall Street Journal (June, 1996) reported on a study that compared cancer risks for three classes of hypertension drugs. For 4 years the study tracked 750 elderly patients who were being treated for hypertension with either beta blockers, ACE inhibitors or calcium channel blockers (mainly short-acting forms), with the following results:

Class of Hypertension Drugs

Cancer	Beta Blockers	ACE Inhibitors	Calcium Ch. Blockers
Yes	28	6	27
No	396	118	175

Do the data suggest that for elderly patients with hypertension, the risk of developing cancer is related to the class of drugs used? Test at the 1 percent level. State the assumption(s) required for the above test.

(handwritten: $x_c^2 = 10.5322$ Reject Ho)

24. The following table gives the hourly wages (in dollars) of fourteen randomly picked workers classified according to their occupations:

Hourly Wages of Workers
Classified According to their Occupations

Plumber	Electrician	Painter	Carpenter
17.6	18.5	16.8	17.4
18.3	18.7	16.7	16.5
17.6	17.7	16.6	16.8
	18.3	16.4	

$$\sum\sum x_{ij}^2 = 4,257.43$$

(handwritten: $F = \dfrac{2.3167}{0.1394} = 16.62$ Reject Ho)

(a) At the 1 percent level of significance, test whether the true mean wages differ for the four occupations.

(b) State the assumption(s) required for the above test.

25. In the following table, X is the temperature (in degrees Celsius) and Y is the corresponding amount (in grams) of a chemical substance extracted from one pound of mineral soil:

X	210	250	270	290	310	340
Y	2.1	5.8	8.1	10.8	12.8	14.3

$$\sum x_i = 1,670 \quad \sum y_i = 53.9 \quad \sum x_i y_i = 16,040 \quad \sum x_i^2 = 475,300 \quad \sum y_i^2 = 588.63$$

(a) Use the method of least squares to find the estimated linear regression equation. *(handwritten: $\hat{Y}_i = -18.57 + (0.098998)X_i + e_i$)*

(b) Test whether the regression is significant at $\alpha = 0.05$. *(handwritten: $t_c = 15.623$, Reject Ho)*

(c) Predict the amount of the chemical substance which can be extracted from one pound of soil at 350°. *(handwritten: $\hat{Y}_i = 16.0793 \approx 16.1$)*

Total
100