Polytechnic University

MA 2222 | Final | May 7, 2002

Print Name:

Signature:

ID #:

Directions: You have two hours to answer the following questions. You must show all your work as neatly and clearly as possible and indicate the final answer clearly. You may use a calculator.

Problem	Possible	Points
1	20	
2	15	
3	10	
4	20	
5	15	
6	20	
Extra Credit	5	
Total	105	

(1) (20 points) Circle the correct statement, (i)– (v), for each question.

(a) In hypothesis testing, what happens to the rejection region when α, the level of significance, is reduced? (Assume this is the only change.)

(i) The rejection region is reduced in size.

(ii) The rejection region is increased in size.

(iii) The rejection region is unchanged.

(iv) The answer depends on the value of β.

(v) The answer depends on the form of the alternative hypothesis.

(b) In hypothesis testing, if β is the probability of committing a Type II error, then $1 - \beta$ (the power of the test) is:

(i) the probability of rejecting H_0 when H_a is true.

(ii) the probability of failing to reject H_0 when H_a is true.

(iii) the probability of rejecting H_0 when H_0 is true.

(iv) the probability of failing to reject H_0 when H_0 is true.

(v) the probability of rejecting H_0.

(c) Two samples, each of size 50, were taken to compare the mean absorbencies of two brands of paper towels. If the null hypothesis H_0 is $\mu_1 = \mu_2$, and the alternative hypothesis H_a is $\mu_1 > \mu_2$, and the computed statistic equals 2.05, which of the following is the correct p-value?

(i) 0.0101.

(ii) 0.0202.

(iii) 0.0404.

(iv) another value.

(d) The confidence interval $(\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}})$ for a population mean with unknown population variance and small sample size is obtained from a population having

(i) a t distribution.

(ii) a normal distribution.

(iii) a binomial distribution.

(2) (15 points, Homework from sections 7.3 and 7.4) Two computers are often compared by running a collection of different benchmark programs and recording the difference in CPU times required to complete the same program. Six benchmark programs, run on two computers, produced the following CPU times, in minutes. Assume the running time for each machine is normally distributed.

Computer\ Program	1	2	3	4	5	6
1	1.12	1.73	1.04	1.86	1.47	2.10
2	1.15	1.72	1.10	1.87	1.46	2.15

(a) Find a 90% confidence interval for **the standard deviation** of the running time for **machine number 1**.

(b) Should you treat the running times from machine 1 and machine 2 as independent or dependent variables? Why?

(c) Does the data provide sufficient evidence at significance level $\alpha = 0.1$ to indicate a difference in **mean CPU time** required for the two computers to complete a job?

(3) (10 points, Homework for sections 7.5 and 7.6) To obtain an estimate of the proportion, p, of New York City residents who feel that the quality of life in NYC has become worse in the past few years, a telephone poll by Time/CNN in 1990 revealed that 686 out of 1009 New York City residents said that the quality of life in NYC had become worse.

(a) Give a 90% confidence level for p based on the poll mentioned above.

(b) To update the above estimate, how large a sample is required to be 98% confident that the maximum error of the estimate of p is 2.5%?

(4) (20 points, Worksheet 4) For a sample of 12 automobiles whose horsepower is between 290 and 390, the following measurements give the time in seconds for the car to go from 0 to 60 mph.

6.2 7.9 6.6 6.4 5.2 4.9 5.5 7.0 6.6 5.4 5.3 5.1

(a) The interval $(x_{(1)}, x_{(7)})$ could be used as a confidence interval for $\pi_{0.3}$, the 30th percentile. Find it and give its confidence coefficient.

(b) Use the Wilcoxon signed rank test to test $H_0 : m = 6.5$ vs. $H_a : m \neq 6.5$ at significance level $\alpha = 0.05$, where m stands for median of the population. State the test statistic, p-value, and your conclusion clearly.

(Note: For the Wilcoxon test statistic, W, based on a sample of size n, we know that $E(W) = 0$ and $Var(W) = \dfrac{n(n+1)(2n+1)}{6}$.

(5) (15 points, Homework, Section 8.5) Two hundred university students were randomly selected and asked if they favor a revision of the university's student government constitution. The following results were obtained:

	Class Standing			
Opinion	Freshman	Sophomore	Junior	Senior
Favor	21	24	25	19
Against	41	35	21	14

Does the data suggest that opinion on this issue is dependent on class standing? Test at significance level $\alpha = 0.05$. State the hypothesis H_0 and H_a, carry out the appropriate test, and state your conclusion clearly.

(**Note:** A goodness-of-fit test between observed and expected frequencies is based on the quantity

$$Q = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i},$$

where Q is a value of a random variable whose sampling distribution is approximately chi-square distribution with degrees of freedom $k - p - 1$, where p is the number of independent parameters estimated on the basis of the sample data. The symbols o_i and e_i represent the observed and expected frequencies, respectively, for the ith cell.)

(6) (20 points, Worksheet 5) A researcher is studying the relation between the first-year maintenance cost Y for a mini-van and the number of miles X the mini-van has been driven during the first year after purchase. The data collected is the following:

Van	Y: Maintenance costs ($)	X: Miles driven
1	652	16500
2	422	8000
3	724	14200
4	746	18400
5	571	9300
6	644	13900
7	548	11000
8	553	13400
9	792	17200
10	739	16500
11	742	18400
12	763	18700
13	698	17700
14	568	10100
15	663	16300

Note: The Minitab output of the regression analysis is on the next page.

(a) Fit a least-squares regression line to these data.

(b) Test (using the t-test) whether or not a linear association exists between the maintenance cost and the number of miles driven. Use $\alpha = 0.05$. State the hypothesis, test statistic, rejection region, and conclusion.

(c) Give a 95% confidence interval on the rate at which the maintenance cost changes with respect to the miles driven.

(d) The ANOVA table is not complete in the print-out. What are the values for **MSR** and **MSE**?

(e) By how much, relatively, is the total variation in maintenance cost reduced when the miles driven is introduced into the analysis?

(f) Suppose a company bought 20 such new minivans. If each is driven 19000 miles in the first year, about 19 of these cars should have their maintenance cost in the interval (__________, __________).

????

(7) (Extra Credit, 5 points) As we have seen, the Wilcoxon signed-rank test in Minitab gives different test statistic from the one given in the book. The test statistic in Minitab (W_M) equals the sum of all the positive ranks, while the test statistic in the book equals the sum of all the positive ranks minus the sum of all the negative ranks. Derive the method to standardize the Minitab signed rank test statistic, i.e., find $E(W_M)$ and $Var(W_M)$, and explain why it is approximately normal.

A table for some well-known distributions. (Note: In the following table $q = 1 - p$.)

Name	p.d.f	m.g.f	mean	variance
Bernoulli(p)	$f(x) = p^x q^{1-x}, \quad x = 1, 2$	$M(t) = q + pe^t$	p	pq
Binomial(n, p)	$f(x) = \binom{n}{x} p^x q^{n-x}$ $x = 0, 1, 2, \ldots, n$	$M(t) = (q + pe^t)^n$	np	npq
Geometric(p)	$f(x) = q^{x-1} p$ $x = 1, 2, \ldots$	$M(t) = \frac{pe^t}{1 - qe^t}$	$\frac{1}{p}$	$\frac{q}{p^2}$
Negative Binomial (r, p)	$f(x) = \binom{x-1}{r-1} p^r q^{x-r}$ $x = r, r+1, r+2, \ldots$	$M(t) = \frac{(pe^t)^r}{(1 - qe^t)^r}$	$\frac{r}{p}$	$\frac{rq}{p^2}$
Poisson(λ)	$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}$ $x = 0, 1, 2, \ldots$	$M(t) = e^{\lambda(e^t - 1)}$	λ	λ
Exponential(θ)	$f(x) = \frac{1}{\theta} e^{-x/\theta}$ $0 \leq x < \infty$	$M(t) = \frac{1}{1 - \theta t}$	θ	θ^2
Gamma (α, θ)	$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$ $0 \leq x < \infty$	$M(t) = \frac{1}{(1 - \theta t)^\alpha}$	$\alpha\theta$	$\alpha\theta^2$
Chi-Square(r)	$f(x) = \frac{1}{\Gamma(\frac{r}{2}) 2^{\frac{r}{2}}} x^{r/2-1} e^{-x/2}$ $0 \leq x < \infty$	$M(t) = \frac{1}{(1 - 2t)^{r/2}}$	r	$2r$
Normal $N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$ $-\infty < x < \infty$	$M(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$	μ	σ^2