Math 661-102, Spring 2012 **Name:**___________
Final Exam **Student ID:**_____________

May, 01 **Must show all work to get full credit!**
I pledge I have not violated the NJIT Honor Code______________________

1. (a)-(c) A state lottery's Pick 3 game asks player to choose a three-digit number, 000 to 999. The state chooses the winning three-digit number at random, so that each number has probability 1/1000. You win if the winning number contains the digits in your number, in any order.

(a) Your number is 377. What is your probability of winning? (5 points)

$P(\text{winning}/377) = \frac{3}{1000} = 0.003$ $\quad \# \{377, 737, 773\} = 3$

(b) Your number is 777. What is your probability of winning? (3 points)

$P(\text{winning}/777) = \frac{1}{1000} = 0.001$

(c) You pick a three-digit number whose probability of winning is the highest. Compute its probability of winning. (7 points)

(Page 247, 4.33)

A winning number = 375

$P(\text{winning}/375) = \frac{6}{1000} = 0.006$ which is greater than the ones in 1(a) & 1(b).

2. a. - c. Consider a household where the monthly bill for natural gas is normally distributed with an average of \$ 107 and a standard deviation of \$46, while the monthly bill for electricity is normally distributed with an average of \$125 with a standard deviation of \$33. The correlation between the two bills is -0.56. Let X stand for the natural gas bill and Y stand for the electricity bill. Then the total bill is X+Y is assumed to be normally distributed.

a. What is the mean total bill? $\mu_{X+Y} = \mu_X + \mu_Y = 107 + 125 = \$232$ (2 points)

b. What is the standard deviation of the total bill? (3 points)

$\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y} = \sqrt{46^2 + 33^2 + 2(-.56)(46)(33)} = \sqrt{1504.84} = 38.79226727$

c. Find the probability that the total bill will exceed \$ 300. (7 points)

(Page 273 example 4.37 and page 338, 5.78)

$P(X+Y > 300)$

$= P\left(Z > \frac{300-232}{38.79226727}\right) = P(Z > 1.75) = 1 - .9599 = 0.0401$

3. Muscular dystrophy is an incurable muscle-wasting disease. The most common and serious type, called DMD, is caused by a sex-lined recessive mutation. Specifically: women can be carriers but do not get the disease; a son of a carrier has probability 0.5 of having DMD; a daughter has .5 probability of being a carrier. As many as one-third of DMD cases, however, are due to spontaneous mutations in sons of mother who are not carriers. Mrs. Treena has a son, who has DMD.

In the absence of other information, the probability is 1/3 that the son is the victim of a spontaneous mutation and 2/3 that Mrs. Treena is a carrier. There is a screening test called the CK test that is positive with probability 0.7 if a woman is a carrier and with probability 0.1 if she is not. Mrs. Treena's CK test is positive. What is the probability that she is a carrier? (page 293, 4.131) (15 points)

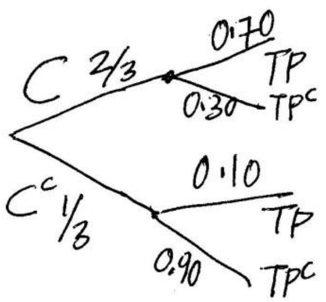
Let C: Mrs. Treena is a carrier &

TP: Ck test is positive.

Need $P(C/TP) = \dfrac{P(TP/C)P(C)}{P(TP)} = \dfrac{2/3(.7)}{1/2}$

$= \dfrac{2.8}{3} = \dfrac{14}{15}$

$= 0.9\dot{3}$



$P(TP) = \frac{2}{3}(.7)$

$+ \frac{1}{3}(.1)$

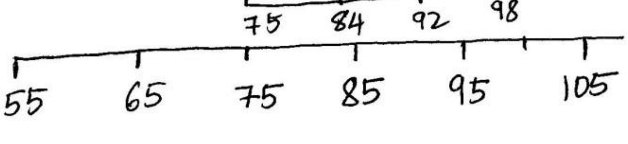
$= \dfrac{1.5}{3} = \dfrac{1}{2}$

4. Consider the scores on the first exam in an introductory statistics course for 10 students:
80 ✓ 73 ✓ 92 88 ✓ 75 ✓ 98 92
55 ✓ 80 ✓ 90.
Make a boxplot for these first-exam scores. Find the inter-quartile range and use the $1.5\times IQR$ rule to check for outliers. How low would the lowest score need to be for it to be an outlier according to this rule? (1.55, 1.56, pages 37, 39) (15 points)

55, 73, 75, 80, 80, | 88, 90, 92, 92, 98

Q_1 Q_3

median $= \dfrac{80+88}{2} = 84$



IQR = 92−75 = 17.

$Q \pm 1.5IQR$ is (49.5, 117.5). Since, all data lies inside (49.5, 117.5) there are no outliers. Any score below (<) 49.5 will be an outlier.

5. (a) – (c) Computers in some vehicles calculate various quantities related to performance. Once of these is fuel efficiency, or gas mileage, usually expressed as miles per gallon (mpg). For one vehicle equipped in this way, the mpg were recorded each time the gas tank was filled. And the computer was then reset. Here are the mpg values for a random sample of 12 of these records:

36.8, 40.1, 41.0, 42.2, 45.5, 47.7,
48.5, 39.2, 37.3, 50.7, 43.3, 44.2.

Suppose that the standard deviation is known to be σ =3.2 mpg.

(a) What is $\sigma_{\overline{x}}$ the standard deviation of $\overline{x}$? (2 points)

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.2}{\sqrt{12}} = 0.923760431$$

(b) Examine the data for skewness and other signs of non-Normality. Show your plot and numerical summaries (three of them). Do you think it is reasonable to construct a confidence interval for μ based on the Normal distribution? Explain your answer. (7 points)

45 | 2.7 = 47.7

35	1.8 4.2 2.3
40	0.1 1.0 2.2 3.3 4.2
45	0.5 2.7 3.5
50	0.7

$$\text{median} = \frac{42.2+43.3}{2} = 42.75$$

$$\overline{X} = 43.04166667$$

$$S = 4.44981273$$

Yes, because the data is close to normally distributed. Also mean ≈ median indicating a symmetric data distribution.

(c) Give a 98% confidence interval for μ, the mean mpg for this vehicle. (Page 359, 6.28) (10 points)

$$\overline{X} \pm Z^* \frac{3.2}{\sqrt{12}}$$

$$Z^* = 2.33$$

$$43.04166667 \pm (2.33)(0.923760431)$$

$$43.04166667 \pm 2.152361804$$

$$(40.88930486, 45.19402847)$$

6. a. – e. A multimedia program designed to improve dietary behavior among low-income women was evaluated by comparing women who were randomly assigned to intervention and control groups. The intervention was a 30-minute session in a computer kiosk in the Food Stamp office. One of the outcomes was the score on a knowledge test taken about two months after the program. Here is the summary of the data:

Group	n	$\bar{x}$	s
Intervention	165	5.08	1.15
Control	212	4.33	1.16

a. The test had six multiple-choice items that were scored as correct or incorrect, so the total score was an integer between 0 and 6. Do you think that these data are normally distributed? Explain why or why not. (3 points)

No, because the data are integer valued between 0 & 6.

b. Is it appropriate to use the two-sample t procedures to analyze these data? Give reasons for your answer. (3 points)

Yes, because the sample sizes are very large and for large samples or d.f. $t \simeq z$.

c. Describe appropriate null and alternative hypotheses for evaluating the intervention. Some people would prefer a two-sided alternative in this situation while others would use a one-sided significance test. Give reasons for each point of view. (2 points)

$H_0: \mu_I = \mu_C$ versus $H_1: \mu_I \neq \mu_C$; two-sided alternative also considers the possibility that the intervention group could have gone worse than control.

$H_0: \mu_I = \mu_C$ $H_1: \mu_I > \mu_C$; the research feels intervention will do better than control.

two-sided

d. Carry out the significance test at alpha = 0.05, assuming the population variances are equal. Report the test statistic with the degrees of freedom and the P-value. Write a brief summary of your conclusion. (8 points)

$H_0: \mu_I = \mu_C$ versus $H_1: \mu_I \neq \mu_C$

$$S_p^2 = \frac{164(1.15)^2 + 211(1.16)^2}{375} = \frac{500.8116}{375} = 1.33549776$$

$$t = \frac{5.08 - 4.33 - 0}{S_p\sqrt{\frac{1}{165} + \frac{1}{212}}} = \frac{.75}{S_p(0.103815159)} = 6.2514237$$

$S_p = 1.155637313$

$df = 375$

p-value $< 2(0.0005) = 0.001$

Clearly based on P-value we reject Ho and Conclude that True mean score for intervention is higher than control.

assuming population variances are equal

e. Find a 95% confidence interval for the difference between the two means. Compare the information given by the interval with the information given by the significance test. State the result of this comparison. (8 points)
(Page 455, 7.71)

$$.75 \pm t_{375, .025}\, s_p \sqrt{\frac{1}{165}+\frac{1}{212}}$$

$$.75 \pm (1.96)(1.155637313)(0.103815159)$$

$$.75 \pm 0.235146436$$

$$(.514853564,\ 0.985146436)$$

Since 0 does not belong to the C.I. $\mu_I - \mu_C$ is positive, in fact in the range 0.515 to 0.985 with 95% confidence, which is also seen in the conclusion of the significance test (two-sided).