

## Final Exam

Name: \_\_\_\_\_

Student ID: \_\_\_\_\_

**May, 9**

**Must show all work to get full credit!**

**I pledge I have not violated the NJIT Honor Code**

1. a. – d. The following displays the results of a standardized neuropsychological test administered to a sample of high school seniors. Test measures analytic skills and is scored on a scale of 0 to 100, with higher scores indicative of stronger analytic skills. Note that  $1|5 = 51$  for males and also  $6|3 = 63$  for females.

### Males

## Females

[illegible]

Compute:

- a) The median scores for females.

Compute:

a) The median scores for females.  $\frac{n}{2}$ th observation &  $\frac{n+1}{2}$ th observation (2 points)  
averaged = 31st & 32nd obs. averaged =  $62 + 63 / 2 = 62.5$

- b) The first, second and third quartiles for **males** and then the interquartile range. (9 points)

$$n_{\text{male}} = 47 \quad \text{median} = \frac{47+1}{2} \text{ observation} = 24^{\text{th}} \text{ observation}$$

position of 1<sup>st</sup> Quartile =  $\frac{85}{4} = 21.25$   
 $Q_1 = 5$   $Q_3 = 93$   $IQR = 93 - 51 = 42$

- c) Are there any outliers among the **male** scores using quartiles? Why or Why not?(4 points)

$$1.5 \text{ IQR} = \frac{3}{2}(42) = 63. \text{ The inner fence i.e.,}$$
$$Q_3 + 1.5 IQR = 93 + 63 = 156 \quad Q_1 - 1.5 IQR = 51 - 63 = -12$$

- d) Based on your answer to c) would the mean or median provide a better estimate for central location of male scores? Again, would the standard deviation or the interquartile range divide by two, provide a better estimate for spread of male scores? Give reason for your answers. (5 points)

mean and standard deviation provide better estimates for central location and spread for male scores because there are no outliers in the data.

2. a. - c. A dispensing machine is set to produce 1-pound lots of a particular compound. The machine is fairly accurate, producing mean weights of lots equal to 1.0 pounds with a standard deviation of 0.11 pounds. Thirty six lots are randomly selected.

a. What is the expected sample mean weight? (2 points)

$$\mu_{\bar{X}} = 1$$

b. What is the standard error in the weights  $\sigma_{\bar{X}}$ ? (3 points)

$$\frac{\sigma}{\sqrt{n}} = \frac{.11}{\sqrt{36}} = .11/6$$

c. Find the probability that the sample mean weight lies between 0.95 pounds and 1 pound. (7 points)

$$\begin{aligned} P(.95 \leq \bar{X} \leq 1) &= P\left(\frac{.95-1}{(.11/6)} \leq Z \leq 0\right) \\ &= P(-2.73 \leq Z \leq 0) \\ &= 0.5 - .0032 = 0.4968 \end{aligned}$$

3. We wish to estimate, using a 95% confidence interval, the mean age at which patients with hypertension are diagnosed. We randomly select 12 subjects with diagnosed hypertension and record the age at which they were diagnosed. The following data are observed in years:

32.8, 40, 41, 42, 45.5, 47, 48.5, 50, 51, 52, 54, 59.2.

We assume that age at diagnosis is approximately normally distributed (please see page 180 of the textbook). (10 points)

$$\bar{X} = 46.91666667$$

$$S = 7.159841711$$

$$46.91666667 \pm t_{.025, 11} \frac{7.159841711}{\sqrt{12}}$$

$$t_{.025, 11} = 2.201$$

$$46.91666667 \pm 4.549177062$$

$$(42.36748961, 51.46584373)$$

4. We wish to test the hypothesis that the mean weight for females who are 5' 7" is 134 pounds against the alternative that the true mean is greater than 134 pounds. Assuming  $\sigma = 15$  using 5% level of significance and with  $n = 40$ , find the power of the test if  $\mu = 140$ . (12 points)

$$|\mu_0 - \mu_1| = |134 - 140| = 6$$

$$P(Z > Z_{0.05} - \frac{6}{(15/\sqrt{40})}) = P(Z > 1.645 - \frac{2}{5}\sqrt{40})$$

$$= P(Z > -0.89)$$

$$= 0.8133$$

5. a. - c. A randomized trial is run to compare two competing medications for peripheral vascular disease. One of the outcomes is self-reported physical functioning. After taking the assigned medication for six weeks, patients provide data on their abilities to perform various physical activities, and a score is computed for each individual. The physical functioning score range from 0 to 100, with higher score indicative of better functioning. The data are:

Medication	n	$\bar{X}$	s
1	20	69.5	24.1
2	20	77.1	22.3

- a. Generate a 98% confidence interval for the difference in mean physical functioning scores between medications. (10 points)
- b. Give reason for the method used in 5. a. above use significance level  $\alpha = 0.05$ . (5 points)
- c. What <sup>additional</sup> assumptions are necessary to justify the use of the method in 5. a. above. (5 points)

a.  $\bar{X}_1 - \bar{X}_2 \pm t_{1-0.01, 38} S_p \sqrt{\frac{1}{20} + \frac{1}{20}}$

$$69.5 - 77.1 \pm (2.429)(23.21745033)$$

$$-7.6 \pm 17.83372396$$

$$S_p^2 = \frac{(24.1)^2 + (22.3)^2}{2} = 539.05$$

$$S_p = 23.21745033$$

(-25.43372396, 10.23372396) (b.)  $H_0: \sigma_1^2 = \sigma_2^2$   
 $H_1: \sigma_1^2 \neq \sigma_2^2$

$F = \frac{(24.1)^2}{(22.3)^2} = 1.16795$   $F_{19, 19, 0.025} = 2.51$   
 Since  $\frac{1}{2.51} < 1.16795 < 2.51$ , do not reject  $H_0: \sigma_1^2 = \sigma_2^2$

- c. The two data sets are random sample from Normal. The two samples are independent of one another.

6. a. -6. A nutrition expert is examining a weight-loss program to evaluate its effectiveness. Ten subjects are randomly selected for the investigation. The subjects' initial weights are recorded; they follow the program for six weeks, and are weighed again. The data are as follows:

Subject	Initial Weight	Final Weight	diff	rank diff	signedrank
1	180	165	15	10	10
2	142	138	4	3	3
3	126	138	-12	7	-7
4	138	136	2	2	2
5	175	170	5	4	4
6	205	197	8	6	6
7	116	115	1	1	1
8	142	128	14	9	9
9	157	144	13	8	8
10	136	130	6	5	5

The difference data is not assumed to be normal due to an outlier in it.

- a. Use significance level  $\alpha = 0.05$  to determine if the weight-loss program is effective. Compute the  $p$ -value of the test. (8 points)  
 b. Compute the Spearman correlation coefficient between initial weight and final weight and determine whether the correlation is significant at  $\alpha = 0.01$ . (8 points)

$$H_0 \text{ median of diff} = 0$$

$$H_1 \text{ median of diff} > 0$$

$$T = \text{sum of positive ranks} = \frac{10(10+1)}{2} - 7 = 55 - 7 = 48$$

$$Z = \frac{48 - \frac{55}{2}}{\sqrt{\frac{110(21)}{24}}} = \frac{48.0 - 27.5}{\sqrt{\frac{2310}{24}}} = \frac{20.5}{\sqrt{96.25}} = \frac{20.5}{9.81} = 2.09$$

$$p\text{-value} = P(Z > 2.09) = 0.0183 < 0.05 \text{ hence reject } H_0$$

the weight-loss program is effective

$$r_s = \frac{\text{Cov}(\text{Rank}(IW), \text{Rank}(FW))}{S.D. \text{Rank}(IW) \cdot S.D. \text{Rank}(FW)} = \frac{(68.75/9)}{(3.018461713)^2} = 0.838415$$

$$H_0 r_s = 0$$

$$H_1 r_s \neq 0$$

$$t = 0.838415 \sqrt{\frac{8}{1 - (0.838415)^2}} = 4.350937$$

$$t_{0.005, 8} = 3.355$$

Hence reject  $H_0$  i.e.  $r_s > 0$ .

7. The following data reflect ages of student at completion of eighth grade. Test if there is a significant difference in the mean age at completion of eighth grade for rural, suburban and urban students at  $\alpha = 0.025$ . The following data were collected from randomly selected students at rural, suburban and urban schools.

1=	Rural	14	14	14	14	13	13	13	12			$\bar{X}$ 13.375	S 0.744023809
2=	Suburban	14	14	14	13	13	13	13	13	12	12	13.1	0.737864787
3=	Urban	16	16	15	15	15	14	14	14	13	12	14.4	1.264911064

What additional assumptions are necessary to run the test?

(10 points)

$$\begin{array}{l}
 \text{MS} \\
 \hline
 4.62678572 \\
 .927
 \end{array}
 \begin{array}{l}
 df \\
 2 \\
 25 \\
 27
 \end{array}
 \begin{array}{l}
 SS_b = 9.25357143 \\
 SS_w = 3.875 + 4.9 + 14.4 = 23.175 \\
 SST = 32.42857143 \\
 F = 4.62678572 / 0.927
 \end{array}
 \begin{array}{l}
 \bar{X}_1 = 13.375 \\
 \bar{X}_2 = 13.1 \\
 \bar{X}_3 = 14.4 \\
 \bar{X}_{..} = 13.64285714
 \end{array}$$

$$F = 4.9911$$

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$H_1$ : at least one equality in  $H_0$  is not true.

$$F_{2,25,.025} = 4.29$$

Since  $4.991 > 4.29$  Reject  $H_0$  at  $\alpha = .025$ .  
There is significant difference in mean.

- Each of the three data sets are random sample from normal
- The three samples are independent of each other.
- The three populations have same variance