Stat 160 Final Exam
Fall 2004
December 2004
Form A
There are 35 problems. Select the best answer to each problem.

For the next 4 questions consider the following problem. A study is being conducted on whether entering college students gain weight during the freshman year. Below are the "Before" and "After" weights for a random sample of 4 students.

Before	137	152	106	120
After	140	160	108	123

1. What test statistic would you compute for this design?

a) The p-value.

b) Wilcoxon test statistic for two independent samples.

c) Hodges-Lehmann test statistic

d) X Wilcoxon signed-rank test statistic.

2. For the given problem compute the point estimate of the effect ($\hat{\Delta}$).

a) 10

b) 1.5

c) 5

d) X 3

3. For the above design, the pairs are Independent, but within each pair there is

a) X Dependency

b) Shift in location

c) Mostly positive differences

d) Walsh averages

4. Consider the following hypothesis H_0: $\Delta = 0$ versus H_A: $\Delta \neq 0$. A confidence interval based on the paired differences is $(2.0, 8.0)$. What conclusion do you draw based on the CI?

a) Reject H_0 and conclude students do not gain weight during freshman year.

b) Accept H_0 and conclude students do not gain weight during freashman year.

c) Do not have enough evidence to make a conclusion.

d) X Reject H_0 and conclude that students gained weight during freshman year.

5. Suppose I look at the clock. What is the probability that the second hand is between 15 and 30?

a) 15

b) -0.25

c) X 0.25

d) 0.50

6. From a group of 5 boys and 3 girls in a Math club, 2 will be chosen to represent the school for a competition. The faculty decided to pick the two at random. What is the probability that they would be of opposite gender?

a) 0.3571

b) 0.1071

c) 0.2679

d) X 0.5357

7. Which of the following graphical methods is best for outlier identification?

a) Least Squares Fit

b) X Box plots

c) Tree diagrams

d) Histograms

For the next three questions: A large company would be willing to build a graduate school at a certain place in Chicago if the average household income of the families living there is at least \$50,000 and the percentage of people who finished college is about 60%. They sampled 100 households randomly and interviewed the head. The mean income of the sample was \$53,000 with a standard deviation of \$15000. And out of the 100, only 53 had college degrees.

8. Compute a 95% confidence interval for the true mean income of all households who lived in that place.

a) $(\$41,500, \$64,500)$

b) $(\$52,706, \$53,294)$

c) $(\$38,000, \$68,000)$

d) X $(\$50,060, \$55,940)$

9. Compute a 95% confidence interval for the true proportion of all heads of households who are college graduates.

a) $(0.2251, 0.8349)$

b) $(0.4801, 0.6799)$

c) X $(0.4322, 0.6278)$

d) $(0.3275, 0.7325)$

10. Based on the confidence intervals, would you advise the company to build the school there?

a) No, a p-value is needed before we can make conclusions.

b) No, since the confidence interval for the mean includes \$50,000 although the confidence interval for the proportion includes 0.6.

c) Yes, since the confidence interval for the mean includes \$50,000 and the confidence interval for the proportion includes 0.6.

d) X Yes, since the confidence interval for the mean is greater than \$50,000 and the confidence interval for the proportion includes 0.6.

11. Suppose the length of songs of a certain rock and roll band are normally distributed with a mean of 3.5 minutes and a standard deviation of 40 seconds.

```
Rweb:> # CUMULATIVE NORMAL DISTRIBUTION
Rweb:> pnorm(3, 3.5, .6667)
[1] 0.2266274
Rweb:> # NORMAL PERCENTAGE POINT
Rweb:> qnorm(.6667, 3.5, 3)
[1] 4.792182
Rweb:> # CUMULATIVE NORMAL DISTRIBUTION
Rweb:> pnorm(3, 3.5, 40)
[1] 0.4950134
```

Using the Rweb output, what is the probability that a randomly selected song is at least 3 minutes.

a) 0.5049866

b) 0.2266274

c) X 0.7733726

d) 4.792182

12. Based on the following stem-leaf plot, determine the Range, IQR and Median.

```
Stem  Leaves
11    079
12    357889
13    1122334466689
14    0379
15    0
```

a) 10, 40, 133

b) X 40, 10, 133

c) 10, 40, 138

d) 40, 10, 138

The next two questions refer to the two samples: X: 16 18 7 7 10 and Y: 15 10 12 7

13. What is the expected value of the Wilcoxon test statistic under H_0?

a) 10.5

b) X 10

c) 9

d) 20

14. What is the value of the Wilcoxon test statistic for these data?

a) 20

b) 9

c) X 9.5

d) 10

The next 3 question refer to the following situation. Suppose a drug company is interested in knowing the effect that their new drug has on lowering cholesterol. So they obtain a sample of 10 people and randomally assign them to one of two groups, either placebo or active compound. Here are the reductions in cholesterol for the Placebo and active compound groups:
Placebo: 3 4 2 −3 −6
Active compound: 9 5 8 5 3

15. Suppose in fact that the active compound has no effect on lowering cholesteral, what can be said if I reject the null hypothesis?

a) X I have committed a type I error.

b) The people should have eaten more walnuts then their cholesterol would have been lower and then we would not have made an error.

c) As there is an outlier in the data I should have used least squares.

d) I have committed a type II error.

16. What is the expected value under H_0 of the Wilcoxon test statistic for these data?

a) 3

b) 22

c) 25

d) X 12.5

17. Suppose that $(-2, 9)$ is a 95% confidence interval for the difference in the treatments. What can be said?

a) Since zero is not in the interval we should reject the null hypothesis.

b) Since zero is in the interval we should reject the null hypothesis.

c) Since zero is not in the interval we should not reject the null hypothesis.

d) X Since zero is in the interval we should not reject the null hypothesis.

18. A study was performed to investigate the relationship between the carborator jetting size x and the time Y of a Camero for a quarter-mile run. The data are:

Jet Size	76	68	70	72	74	76
Time	15.08	14.60	14.50	14.53	14.79	15.02

The simple linear model $Y = a + bx + e$ was fit. The results are:

	Estimate	Std. Error
(Intercept)	9.89375	1.39254
x	0.06688	0.01915

The confidence interval for estimated slope b is

a) X $(.0293, .1044)$

b) $(.06688, .01915)$

c) $(.1044, .0293)$

d) $(7.164, 12.623)$

The next two questions refer to the following. We wish to test whether or not there is a differnce between the ages of all men (Y) and all women (X) that have ever taken Stat 160. We will use the results from last semester's survey as our sample data. Below are the data as well as summary statistics and p-values. Recall only one of the p-values is correct for this test. (m1 is for males and m2 for females)

```
Ages for male students:
20 33 20 22 18 18 20 19 18 19 21 21 21 20 23 19 21 20 21 21 21 20 20 24 21
20 23 21 22 20 20 21 22 19 44 31 19 20 22 23 22 20 18 22 20 19 22 20 20 19
39 18

Ages for female students:
22 21 21 19 22 18 19 21 21 20 20 20 19 21 22 21 18 19 22 18 33 20 21 20 19
22 20 23 21 21 20 19 21 21 44 21 23 20 32 20 45 21 20 20

Rweb:> summary(variables)
      male
 Min.   :18.00
 1st Qu.:20.00
 Median :20.00
 Mean   :21.67
 3rd Qu.:22.00
 Max.   :44.00
Rweb:> ou<-outer(male,male,"+")/2
Rweb:> hl<-median(ou[row(ou) <= col(ou)])
Rweb:> # HODGES-LEHMANN ESTIMATE for male
Rweb:> hl
[1] 20.5
Rweb:> # STANDARD DEVIATION of male
Rweb:> var(male)^.5
[1] 4.829327
Rweb:>

    female
 Min.   :18.00
 1st Qu.:20.00
 Median :21.00
 Mean   :22.07
 3rd Qu.:21.25
 Max.   :45.00
Rweb:> ou<-outer(female,female,"+")/2
Rweb:> hl<-median(ou[row(ou) <= col(ou)])
Rweb:> # HODGES-LEHMANN ESTIMATE for female
Rweb:> hl
[1] 20.5
Rweb:> # STANDARD DEVIATION of female
Rweb:> var(female)^.5
[1] 5.695358

Alternate Hypothesis:
m1 < m2            m1 < > m2          m1 > m2
p-value 0.35       p-value 0.7        p-value 0.651
```

19. What is the correct p-value and the conclusion?

- **a)** p-value =0.7, so reject the null hypothesis
- **b)** p-value = 0.651, so reject the null hypothesis
- **c)** X p-value = 0.7, so do not reject the null hypothesis
- **d)** p-value = 0.35, so do not reject the null hypothesis

20. What is an estimate of the Δ (delta), the difference in mean ages (mean(Y) - mean(X))?

a) -0.866031

b) X -0.40

c) 0

d) -1

21. Weight limits of luggages transported by a certain airline is to be modified. A random sample of 20 luggages were weighted. Give a 95% bootstrap CI for the true median weight of all luggages the airline carries. `Sorted Medians`

```
---------------------------

46.0 47.0 48.0 48.0 49.0 49.5 50.5 50.5 50.5 50.5
50.5 50.5 51.5 51.5 51.5 53.0 53.0 53.0 53.0 53.0
54.0 54.0 54.0 54.5 54.5 54.5 54.5 55.0 55.0 55.0
55.5 55.5 55.5 55.5 55.5 55.5 55.5 55.5 56.0 56.0
56.0 56.0 56.0 56.0 56.0 56.0 56.0 56.0 56.0 56.5
56.5 56.5 56.5 56.5 56.5 56.5 56.5 56.5 57.0 57.0
57.0 57.0 57.0 57.0 57.0 57.0 57.0 57.0 57.0 57.0
57.0 57.0 57.0 57.0 57.0 57.0 57.0 57.0 57.0 57.0
57.0 57.0 57.0 57.0 57.0 57.0 57.0 57.0 57.0 57.0
57.0 57.0 57.0 57.0 57.0 57.0 59.0 60.0 61.0 62.0
```

a) X $(48, 60)$

b) $(49, 57)$

c) $(48, 59)$

d) $(47, 61)$

For next 3 questions consider the following problem. The following data gives information on the average saturated fat(in grams) consumed per day and the cholestrol level (in milligrams per hundered millimeters) for eight males.

Fat consumption (x)	Cholestrol level (Y)
55	180
65	210
50	195
34	165
43	170
58	204
69	235
36	150

The class code gives the LS fit as follows

	Estimate	Std. Error	t-value	$Pr(>\|t\|)$
(Intercept)	85.9598	16.5148	5.205	0.002005
x	2.0032	0.3137	6.386	0.000694
R-Squared 0.87				

22. Which is prediction equation based on the LS output?.

a) $\hat{Y} = 16.5148 + 2.0032 * x$

b) X $\hat{Y} = 85.9598 + 2.0032 * x$

c) $\hat{Y} = 2.0032 + 85.9598 * x$

d) $\hat{Y} = 16.5148 + 0.3137 * x$

23. Predict the cholestrol level for the fat consumption of 72 gms.

a) 160.75

b) 6191.11

c) X 230.19

d) 18.77

24. Interpret the R^2 value.

a) The probability that there is a linear realtionship between the two variables is 87

b) 87% of the variation in fat consumption is being accounted for by cholestrol level.

c) X 87% of the variation in cholestrol level is being accounted for by fat consumption.

d) For every one gram increase of fat consumption, the cholestrol level goes up by 2.0032.

25. For a perfect positive linear relationship between two variables, what would be the value of the r, (correlation coefficient)?.

a) X 1.0

b) 0.97

c) 0.0

d) -1

26. A dietitian wants to test three different diets to find out if the mean weight loss for each of these diets is the same. She randomly selected 21 overweight persons, randomly divided them into three groups, and put each group on one of the three diets. The data given is the weight(in pounds) lost by these persons after being on these diets for 2 months:

Diet I	15	8	11	7	22	12	8
Diet II	11	16	9	13	24	17	19
Diet III	9	17	11	8	11	6	14

What kind of design are we talking about?

a) X Completely randomized design.

b) Regressional design

c) Randomized pair design

d) Uncontroled design

27. Consider the folllowing stem-leaf plot. How would you describe its shape?

```
0 : 1223444444
0 : 556666667777777888999
1 : 0011111123333344
1 : 5555566788889999
```

```
2 : 011222333334
2 : 56666789999
3 : 0114
3 : 668
4 : 02
4 : 58
5 : 02
```

a) X Asymmetric and Right skewed

b) Symmetric and Right skewed

c) Asymmetric and Left skewed

d) Symmetric and Left Skewed

For the next two problems: Suppose that we are investigating the safety of a dangerous intersection. Past police records indicate a mean of 5 accidents per month at this intersection. Suppose the number of accidents is distributed according to a Poisson distribution. And we have the following RWEB output:

```
Rweb:> # CUMULATIVE POISSON DISTRIBUTION
 Rweb:> ppois(3, 5)
 [1] 0.2650259
 Rweb:> # POISSON PROBABILIY
 Rweb:> dpois(3, 5)
 [1] 0.1403739
 Rweb:> # CUMULATIVE POISSON DISTRIBUTION
 Rweb:> ppois(4, 5)
 [1] 0.4404933
 Rweb:> # POISSON PROBABILIY
 Rweb:> dpois(4, 5)
 [1] 0.1754674
```

28. What is the variance of the number of accidents per month in the intersection?

a) 25

b) X 5

c) 10

d) 2.2361

29. What is the probability that fewer than 4 accidents in the intersection will be observed this month?

a) 0.1404

b) 0.4405

c) X 0.265

d) 0.8594

30. Which of the following statements is true about the central limit theorem?

a) X For most distributions, as the sample size increase, the distribution of the sample mean will become more and more like a normal distribution.

b) The distribution of the sample mean will become uniform with mean $\sigma/\sqrt{n}$ and standard deviation μ.

c) For most distributions, as the sample size increases, the distribution of the sample mean will become more and more like a uniform distribution.

d) If the distribution of $X_1, X_2, \ldots, X_n$ is χ^2, then the distribtion of the sample standard deviation will become more and more u-shaped.

For the next two questions: The "house" rolls two 20-sided dice and the "player" rolls one 20-sided die. If the player rolls a number on his die between the two numbers the house rolled, then the player wins. Otherwise, the house wins (including ties). What are the probabilities of the player winning? We use resampling to answer this problem. For each trial, let the first and the last number be the numbers rolled by the "house" and the middle number be the one for the "player".

```
Trial 1
13 20 18

Trial 2
10 7 5

Trial 3
7 13 19

Trial 4
1 9 2

Trial 5
1 5 16

Trial 6
6 6 8

Trial 7
15 3 8

Trial 8
1 11 18

Trial 9
16 8 4

Trial 10
14 12 18

Trial 11
8 6 5

Trial 12
19 2 17

Trial 13
2 7 18

Trial 14
7 15 20

Trial 15
18 14 6
```

31. What is the estimate of the probability that the player wins and its error?

a) estimate = 0.8, error = 0.2066

b) estimate = 0.6667, error = 0.2434

c) estimate = 0.7333, error = 0.2284

d) X estimate = 0.6, error = 0.253

32. Which is the correct sampling code for the above experiment?

a) X
```
Minimum Value = 1
Maximum Value = 20
Number to be Drawn = 3
With Replacement
```

b)
```
Minimum Value = 1
Maximum Value = 20
Number to be Drawn = 20
With Replacement
```

c)
```
Minimum Value = 1
Maximum Value = 3
Number to be Drawn = 20
With Replacement
```

d)
```
Minimum Value = 1
Maximum Value = 20
Number to be Drawn = 3
Without Replacement
```

For the next two problems: Test for impurities commonly found in drinking water from private wells showed that 30% of all wells in a particular country have impurity A. Twenty wells are selected at random and we have the following RWEB output:

```
Rweb:> # CUMULATIVE BINOMIAL DISTRIBUTION
 Rweb:> pbinom(5, 20, 0.3)
 [1] 0.4163708
 Rweb:> # BINOMIAL PROBABILITY
 Rweb:> dbinom(5, 20, 0.3)
 [1] 0.1788631
 Rweb:> # CUMULATIVE BINOMIAL DISTRIBUTION
 Rweb:> pbinom(4, 20, 0.3)
 [1] 0.2375078
 Rweb:> # BINOMIAL PROBABILITY
 Rweb:> dbinom(4, 20, 0.3)
 [1] 0.1304210
```

33. What is the probability that at least 5 of the wells will have impurities?

a) 0.4164

b) 0.1787

c) 0.8696

d) X 0.7625

34. What is the probability that 4 or 5 of the wells will have impurities?

a) X 0.3093

b) 0.4164

c) 0.5468

d) 0.6539

35. The probability that a students passes a Philosophy course is 0.75, while the probability of passing a Statistics course is 0.45. Assuming that the courses are independent of one another, what is the probability that a student enrolled in both courses will fail on both?

a) X 0.1375

b) 0.3375

c) 0.4125

d) 0.1125