Name:_________________________________

Check one: Section 1 (Mon.-Wed.. 10:30-noon):_____

Section 2 (Mon,.-Wed. 1:30-3:00):_____

Statistics 431 Midterm
March 1, 2004 6-8pm
Solutions

- **This exam is closed book.**
- **You may use a calculator.**
- **You must write the exam using pen (not pencil).**
- **A table of formulas and statistical tables have been provided for you at the end of this exam.**
- **Show all your work.**
- **Be sure to state the null hypothesis, test statistic, conclusion, etc for any statistical test that you use.**
- **There are blank pages at the end of the exam if you need more room.**

Question	Total Points	Points Received
1	20	
2	11	
3	10	
4	6	
5	11	
6	3	
Bonus		
Total	61	

20 pages (including 2 blank pages at end
in case you need extra space for answers)
+ packet of statistical tables
+ 3 pages of formulas.

1) The following is a list of some of the statistical methods that you have learned about so far in this course:

- One sample t/Z-test of a population mean.
- One sample Z-test of a population proportion.
- Two sample t/Z-test of population means.
- Two sample Z-test of population proportions.
- Paired or Matched pairs t/Z-test.
- One-way ANOVA
- Two-way ANOVA
- Randomized Block Design

For each of the situations described below, state the technique (from the list above) that you believe is appropriate. **If none are appropriate, state "none of the above".**

No calculations are required.

a) [2 points] A company is investigating how long it takes its drivers to deliver goods from its factory to a nearby port for export. The company recorded the delivery times for 48 recent deliveries on the standard pre-specified driving route. The company then devised a new route and recorded the delivery times for 25 deliveries on this new route. The company wants to decided if the new route is quicker on average than the standard route.

2 sample t/Z-test

b) [2 points] In an experiment to determine if the best type of golf driver, 10 golfers are asked to do the following: hit 5 golf balls with each of 4 drivers (so that they each hit 20 balls in total). The total straight-line distance traveled by the ball is recorded for each hit (farther is better). Are some drivers better than others?

Randomized block design

c) [2 points] A manufacturer claims that its cars achieve an average of at least 35 miles per gallon in highway driving. A consumer interest group tests this claim by driving a random selection of the cars in highway conditions and measuring their fuel efficiency (e.g. miles per gallon observed in the test drives)

1 sample t/Z test.

d) [2 points] A new drug is being compared with a standard drug for treating a particular illness. In the clinical trials, a group of 200 patients was randomly split into two groups, with one group being given the standard drug and one group being given the new drug. Altogether, 83 out of the 100 patients given the new drug improved their condition, while only 72 out of the 100 patients given the standard drug improved their condition. Is this enough evidence to say that the new drug is better than the standard drug?

2 sample Z test of proportions

e) **[2 points]** A retail company is interested in finding out whether the time taken for it to obtain approval for a credit card number by phone varies from one day of the week to another. The manager collects data on the average time (measured in seconds) taken to obtain approval for a random sample of 20 cases on Mondays, 15 cases on Wednesdays, and 25 cases on Fridays.

1-way anova

f) **[2 points]** A garage sells tires of types A, B, and C. This year's and last year's sales of the three types of tires are shown in the table below. The company wants to determine if there is evidence of a change in the preferences for the three types of tire between the two years (AND they want to do only a single hypothesis test to get the answer).

		Type of Tire		
		A	**B**	**C**
Sales	**This Year**	113	82	26
	Last Year	211	143	61

None of the above

g) **[2 points]** Two independently operated laboratories provide historical dating services using radioactive carbon dating methods. A researcher suspects that one laboratory tends to provide older datings than the other laboratory. To investigate, 18 samples of old material are split in half. One half is sent to lab A for dating, and the other half is sent to lab B for dating. The labs are asked to submit their answers in years (rounded to the nearest decade).

Paired t/Z-test

h) **[2 points]** Suppose a legal agreement has been reached that says if 10% or more of the building tiles are cracked one year after installation then the construction company that originally installed the tiles must pay for the building repair costs. The owner of the building surveys a randomly selected set of 1250 tiles and finds that 98 of them are cracked. Should the construction company be required to pay for repair costs?

1 sample Z test of proportion

i) **[2 points]** Four drugs are being studied to determine their effectiveness in treating headaches. All four drugs require a series of injections. There are two possible schedule for the injectons: (1) one injection a week for 4 weeks and (2) one injection every other day for a total of 4 injections. Researchers want to determine if (a) there is a difference between the two schedules of injections and (b) whether there is a difference between the four drug mixtures and (c) whether different drugs require different injection schedules. The conduct an experiment in which 40 patients were treated. 5 patients were randomly assigned to each combination of injection schedule and drug. The patients were asked to report the frequency, duration, and severity of his or her headache during the 30 days following the last injection. An index ranging from 0 to 100 was constructed for each patient, where 0 indicates no headache pain and 100 specifies the worst headache pain.

2-way anova

j) **[2 points]** A federal agency responsible for enforcing laws governing weights and measures routinely inspects packages to determine whether the weight of the contents is at least as great as that advertised on the package. A random sample of 18 containers whose packaging state that the contents weight 8 ounces was selected. The contents of each package was then weighed carefully. The auditors are trying to determine if they can conclude that the containers are mislabled, on average.

1 sample t/Z-test

2) An audit of a federal assistance program implemented after a major regional disaster discovered that out of 85 randomly selected applications processed during the first two weeks after the disaster, 17 contained errors due to either applicant fraud or processing mistakes. However, out of 132 randomly selected applications processed later than the first two weeks after the disaster, only 16 contained errors.

a) [6 points] Test whether or not we can conclude that errors in the assistance applications are more likely in the initial aftermath of the disaster. Use a 1-sided test with $\alpha = .05$. Report the Pvalue of your test. Check Any necessary assumptions.

$H_0 : p_{\text{first 2 weeks}} = p_{\text{after first 2 weeks}}$ $\quad H_0 : p_{\text{first 2 weeks}} > p_{\text{after first 2 weeks}}$

where p is the percent of applications that contained errors.

$$\hat{p}_{pooled} = \frac{\text{\#successes in group 1 + \#successes in group 2}}{n_1 + n_2} = \frac{17+16}{85+132} = 0.152$$

Test Statistic:

$$\frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p}_{pooled}(1-\hat{p}_{pooled})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{.2-.121}{\sqrt{.152(1-.152)\left(\frac{1}{85}+\frac{1}{132}\right)}} = 1.58 \sim Z$$

Pvalue=P(Z>1.58)=1-.9429=.0571

Since the Pvalue is greater than .05 we do not reject the null hypothesis. Not enough evidence to conclude that there are more errors in the first 2 weeks than afterwards.

Note:

$n_1 \hat{p}_1 = 85(.2) = 17 \geq 10$

$n_1(1-\hat{p}_1) = 85(.8) = 68 \geq 10$

$n_2 \hat{p}_2 = 132(.121) = 15.97 \geq 10$

and $n_2(1-\hat{p}_2) = 132(.879) = 116.028 \geq 10$

so normal approximation to binomial ok.

b) **[3 points]** Using the results from this sample, estimate how large a sample of applications processed during the first two weeks after the disaster would be needed to estimate the percent of applications processed during the first two weeks that contain errors to within 3 percentage points (with 95% confidence). [You can ignore the applications processed later than the first two weeks for this problem.]

$$n = \left(\frac{\left(Z_{\alpha/2}^2\right)\hat{p}(1-\hat{p})}{e^2} \right) = \left(\frac{\left(1.96^2\right).14(1-.14)}{.03^2} \right) = 513.9 \approx 514$$

c) **[2 points]** Now let's consider all applications made after the disaster (both before 2 weeks, and after two weeks after the disaster). Suppose you wanted to estimate the proportion of applications that ask for over $10,000 in assistance. Estimate how large a sample size you need to estimate this proportion to within 2 percentage points (with 95% confidence).

Note that we have no info on the proportion of applications that ask for over $10,000 in assistance, therefore we use p=.5

$$n = \left(\frac{\left(Z_{\alpha/2}^2\right)\hat{p}(1-\hat{p})}{e^2} \right) = \left(\frac{\left(1.96^2\right).5(1-.5)}{.02^2} \right) = 2401$$

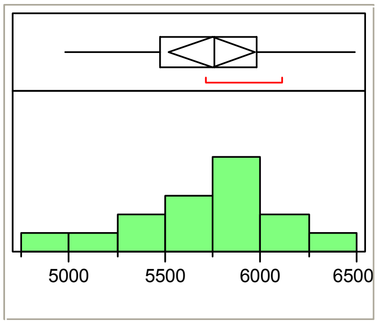
3) advertising t-test problem

1. A restaurant located in an office building decides to adopt a new strategy for attracting customers to the restaurant. Every week it advertises in the city newspaper. To measure how well the advertising is working, the restaurant owner recorded the weekly gross sales for the 15 weeks after the campaign began and the weekly gross sales for the 24 weeks immediately prior to the campaign. These data are analyzed below.

After

Distributions

Sales



Quantiles

100.0%	maximum	6489.0
99.5%		6489.0
97.5%		6489.0
90.0%		6303.0
75.0%	quartile	5982.0
50.0%	median	5757.0
25.0%	quartile	5477.0
10.0%		5024.6
2.5%		4979.0
0.5%		4979.0
0.0%	minimum	4979.0

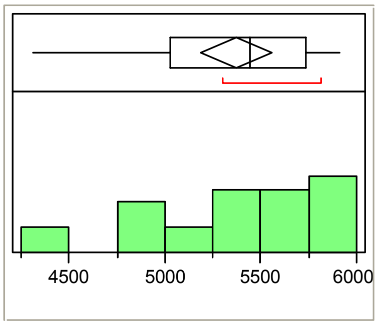
Moments

Mean	5746.067
Std Dev	409.010
Std Err Mean	105.606
upper 95% Mean	5972.568
lower 95% Mean	5519.565
N	15.000

Before

Distributions

Sales



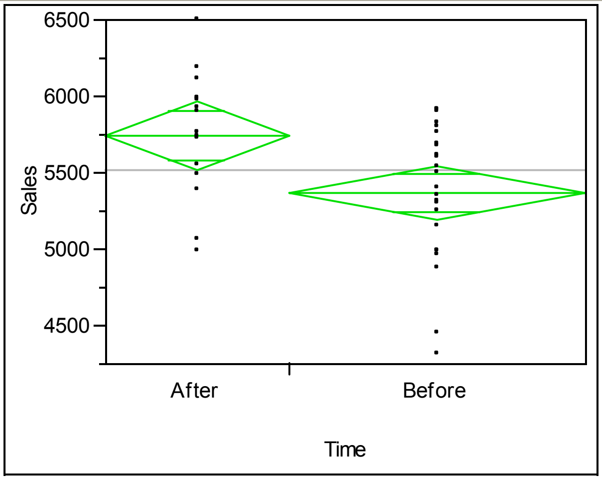
Quantiles

100.0%	maximum	5911.0
99.5%		5911.0
97.5%		5911.0
90.0%		5902.5
75.0%	quartile	5736.8
50.0%	median	5442.5
25.0%	quartile	5026.8
10.0%		4657.5
2.5%		4310.0
0.5%		4310.0
0.0%	minimum	4310.0

Moments

Mean	5372.125
Std Dev	441.330
Std Err Mean	90.086
upper 95% Mean	5558.481
lower 95% Mean	5185.769
N	24.000

Oneway Analysis of Sales By Time



Missing Rows	9

Oneway Anova

Summary of Fit

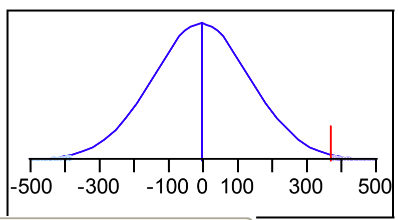
Rsquare	0.159106
Adj Rsquare	0.136379
Root Mean Square Error	429.3872
Mean of Response	5515.949
Observations (or Sum Wgts)	39

t Test

After-Before

Assuming equal variances

Difference	373.942	t Ratio	2.645902
Std Err Dif	141.329	DF	37
Upper CL Dif	660.301	Prob > |t|	0.0119
Lower CL Dif	87.583	Prob > t	0.0059
Confidence	0.95	Prob < t	0.9941



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Time	1	1290760.3	1290760	7.0008	0.0119
Error	37	6821813.6	184373		
C. Total	38	8112573.9			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
After	15	5746.07	110.87	5521.4	5970.7
Before	24	5372.13	87.65	5194.5	5549.7

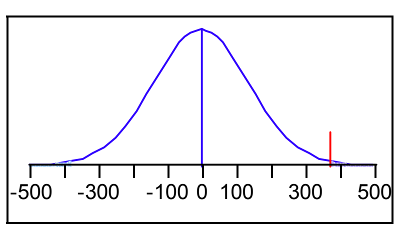
Std Error uses a pooled estimate of error variance

t Test

After-Before

Assuming unequal variances

Difference	373.942	t Ratio	2.693916
Std Err Dif	138.810	DF	31.60234
Upper CL Dif	656.827	Prob > |t|	0.0112
Lower CL Dif	91.056	Prob > t	0.0056
Confidence	0.95	Prob < t	0.9944



a) [4 points]Can we conclude at the 5% significance level that the advertising campaign is successful?

$H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 > \mu_2$ (1=with ads, 2=without ads)

Since $\frac{s_1^2}{s_2^2} = \frac{441.33^2}{409.01^2} = 1.16 < 3$ and $\frac{n_1}{n_2} = \frac{24}{15} = 1.6 < 2$ we can use the equal-variance 2-sample t-test.

From JMP the test statistic is 2.646 ~t(37) and the Pvalue =.0059

Since the Pvalue is less than .05 we reject the null hypothesis and conclude that the ads are effective in increasing sales.

b) [2 points] What assumptions are necessary for the test in (a)?

We need to assume the data is normal in each group and that the variances are equal in the two groups (this assumption seems reasonable based on our rules of thumb).

We also need to assume the data is independent.

Give Full marks even if they don't say the independence part (just normality and equal variance ok).

c) [4 points] Assume that the profit is 20% of the gross. If the ads costs $50 per week, can the restauranteur conclude that the ads are profitable? (i.e. Do the ads generate enough profit to cover the cost of the ads, hopefully more?) Justify your answer with statistics.

If the ads cost $50/week, the gross would have to increase by $250 per week in order to cover the cost of the ads.

$H_0: \mu_1 - \mu_2 = 250$ vs $H_1: \mu_1 - \mu_2 > 250$ (1=with ads, 2=without ads)

Since $\frac{s_1^2}{s_2^2} = \frac{441.33^2}{409.01^2} = 1.16 < 3$ and $\frac{n_1}{n_2} = \frac{24}{15} = 1.6 < 2$ we can use the equal-variance 2-sample t-test.

$$s_p^2 = \frac{(24-1)441.33^2 + (15-1)409.01^2}{(24-1)+(15-1)} = 184373.2002$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(5746.067 - 5372.125) - (250)}{\sqrt{184373.20\left(\frac{1}{24} + \frac{1}{15}\right)}} = 0.877 \sim t(37)$$

P-value (using t(35) since t(37) not on table)=.187
P-value (using t(40))=.187

Since the Pvalue is greater than .05 we do not reject the null hypothesis. There is not enough evidence to conclude that the ads are profitable (e.g. not enough evidence to say they generate for then $250 in extra sales).

Note: Average weekly gross before ads: 5372.13
Average weekly gross after ads: 5746.07
Difference: 373.94
Standard Error of Difference: 141.329

So Although the estimate of the difference in gross is $373.94 (so increase in profit is .2*373.94=74.79, enough to cover the $50 for the ads), the standard error of this estimate is so large that in a test we cannot be sure that the difference is larger than $250.

Another strategy is to construct a 95% confidence interval. If the interval is totally above $250 then we can conclude (at the 5% level) that the increase in gross is above $250 and so the ads are profitable.

4) Suppose you want to design a study to determine how much time your employees spend each day reading and sending email.

a) [3 points] Suppose you survey 25 people. You want to test the null hypotheis: Ho: the average number of minutes spend on email is 30 against the alternative Ha: the average number of minutes spent on email is greater than 30, using α =.05. Suppose you estimate the sd (σ) in the number of minutes spent on email to be 15 minutes. How much power do you have to reject the null hypothesis if the true mean is 35 minutes?

$$\text{Power}=1-\Phi\left(Z_\alpha+\frac{\mu_0-\mu'}{\sigma/\sqrt{n}}\right)=1-\Phi\left(1.645+\frac{30-35}{15/\sqrt{25}}\right)=1-\Phi(-0.02)=1-.4920=.508$$

b) [3 points] How large a sample size would you need to have 80% power to detect a difference of 5 minutes?

Note: Power=.80, so β =.20

$$n=\left(\frac{\sigma\left(Z_\alpha+Z_\beta\right)}{\mu_0-\mu'}\right)^2=\left(\frac{15(1.645+.845)}{30-35}\right)^2=55.8\approx 56$$

5) An experiment was conducted to study productivity of software engineers. Management was interested in the time it takes to code a software module. Two factors that may affect the coding time are the size of the module and whether or not the programmer has access to a library of previously coded submodules. Module size is studied at two levels, large and small. Access to a library of submodules is coded as either available (yes) or not (no). Data from this experiment is analyzed below.

a) [4 points] Fill in the missing values on the ANOVA table.

A=______________1_______

B=______________2572.361________

C=_____________3.88________

D=_____________P>.10________

*****ignore the "true" values on the attached JMP output. The above answers will be given full marks. There was a slight typo in my JMP dataset. The data was actually unbalanced (not the same sample size in each cell) and this is the one case where the sum of squares for JMP does not add up. (The JMP tests are still reasonable, just the sum of squares as given don't add up). We'll stick to balanced designs in the future to avoid this problem.**

(an approximate Pvalue using the appropriate table is fine.)

b) [7 points] Interpret the results from this analysis. Explain any hypothesis test you use (state null hypothesis, test statistic, etc). Give a complete answer.

Ho: no interaction between size and library
Ha: there is a interaction between size and library

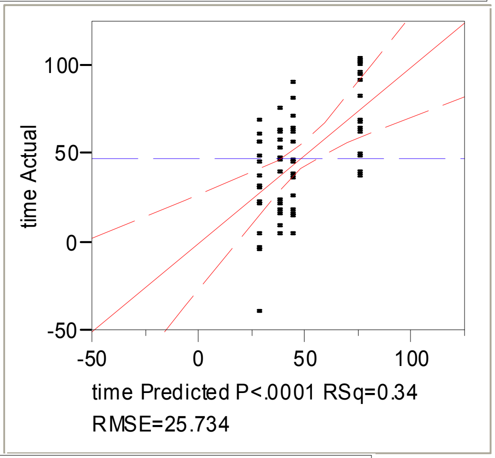
Test statistic: $F(1,56)=15.8026$, Pvalue=.0002

Since the Pvalue is less than .05 we reject the null hypothesis and conclude that there is an interaction between size and library.

From the interaction plot we see that having access to a library reduces the time substantially for large jobs, but there is no difference in time to complete the job for small jobs (e.g. roughly the same completion time whether or not a library is available).

Tests of the main effects of size and library are not appropriate since the interaction was significant.

Actual by Predicted Plot



Summary of Fit

RSquare	0.339659
RSquare Adj	0.304284
Root Mean Square Error	25.73418
Mean of Response	47.93333
Observations (or Sum Wgts)	60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	19075.848	6358.62	9.6016
Error	56	37085.885	662.25	Prob > F
C. Total	59	56161.733		<.0001

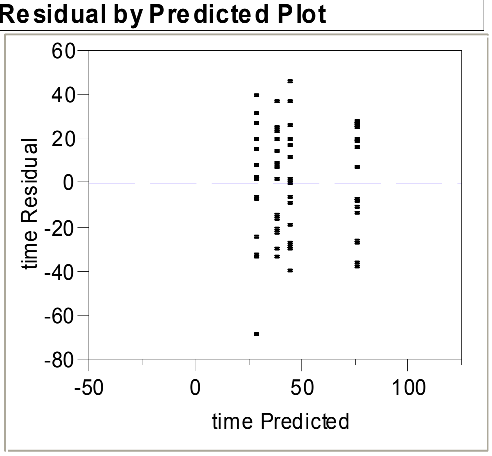
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	47.309673	3.325974	14.22	<.0001
size[large]	5.488244	3.325974	1.65	0.1045
library[no]	10.043006	3.325974	3.02	0.0038
size[large]*library[no]	13.221577	3.325974	3.98	0.0002

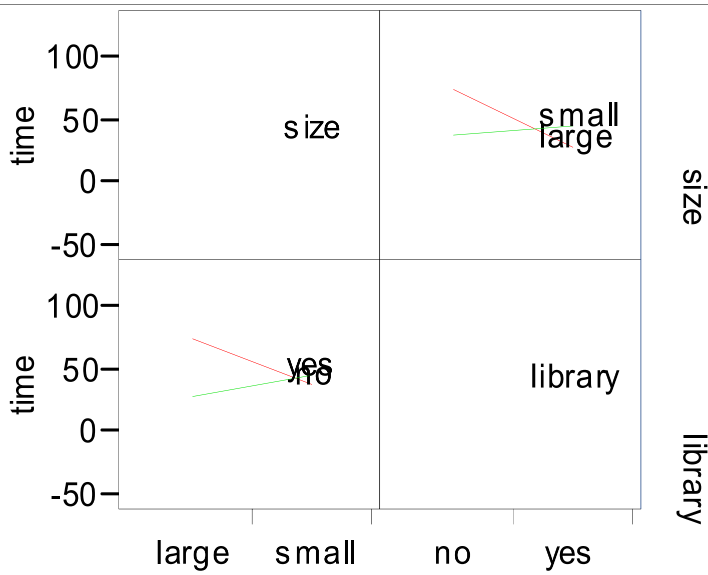
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
size	1	1	1803.224	2.7229	0.1045
library	1	1	6038.240	9.1178	0.0038
size*library	1	1	10465.247	15.8026	0.0002

Residual by Predicted Plot



Interaction Profiles

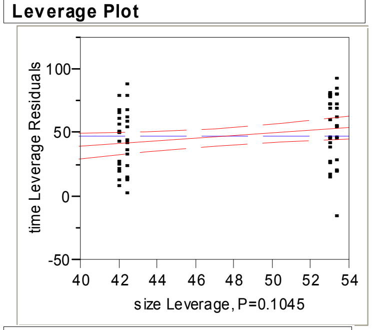


Starting at Left
Top line=large
Bottom line=small

Starting at Left
Top line=no
Bottom line=yes

size

Leverage Plot

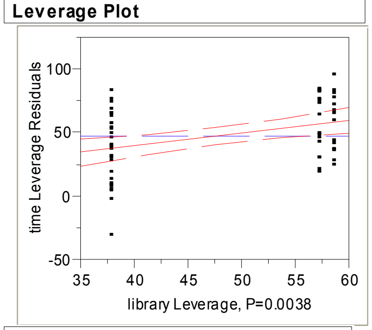


Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
large	52.797917	4.6244016	53.5484
small	41.821429	4.7815606	41.9310

library

Leverage Plot

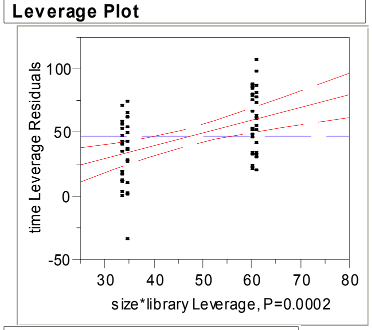


Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
no	57.352679	4.7088725	58.6000
yes	37.266667	4.6983967	37.2667

size*library

Leverage Plot



Least Squares Means Table

Level	Least Sq Mean	Std Error
large,no	76.062500	6.4335447
large,yes	29.533333	6.6445363
small,no	38.642857	6.8777485
small,yes	45.000000	6.6445363

7) [**3 points]** You will receive 3 points for answering the following questions.

We will analyze these data later in the course.

Your answers to these questions will in NO WAY affect your grade for this exam.

a) What is your gender? Male Female (circle one)

b) About how many hours did you study for this exam? ______________________

c) Estimate your score (out of 61) on this exam______________________________
You will get **2 bonus** points if your answer is within +/- 2 points of your actual score.
Do not include these bonus points as part of your estimate.