

E 370 – Fall 2004  
In-Lab Examination One  
Statement of Academic Integrity:

“I swear that I have neither given nor received assistance on this exam and that I will not discuss this exam until all sections have completed it, that is until 14:00 on Saturday, October 12, 2004.”

I have read and agree with the above statement.

Signature: \_\_\_\_\_

Instructions

1. Write your name on every page of your exam.
2. Answer the questions in the spaces provided on the exam.
3. Do not provide computer output unless it is **required**. Whenever you print out anything, you **MUST TYPE your Name and 5-digit Team Number on your work before you print it.**
4. **You must show your work and explanation to receive full credit. You must write any EXCEL functions used, with arguments, as well as the numerical output. For example Include 4 decimal places in your answer.**
5. You may only use EXCEL for this exam.
6. Only the exam, pencils, erasers, and the tool cards may be on your desk. Put all the rest of your belongings along the wall or at the front of the room.
7. Remember, a student is to avoid even the appearance of cheating. Keep your eyes on your exam or on your computer screen. Any questionable behavior on your part is sufficient reason for me to confiscate your exam and ask you to leave the room.
8. The value of each question is given by each question. **Budget your time accordingly.**
9. When you are finished, you may turn in all pages of the exam and leave the room as soon as you are able to without disturbing your classmates.
10. Stay calm and do your best.
11. **DATA INSTRUCTION: Open Internet Explorer. Type the URL your instructor gives you to get YOUR data to use on the exam. Save it to the desktop as a TEMP file. Close Internet Explorer.**

Name \_\_\_\_\_

1. The Wood County sheriff classifies crimes by age (in years) of the criminal and whether the crime is violent or nonviolent. As shown below:

Type of Crime	Age (in years)			Total
	Under 20	20 to 40	Over 40	
Violent	27	46	19	92
Nonviolent	17	34	17	68
Total	44	80	36	160

a. complete the contingency table given above. (6 points)

If a case is selected at random, what is the probability that:

b. It involved a violent crime? (3 points)

$$=92/160=0.575$$

c. It involved a violent crime or an offender less than 20 years old? (3 points)

$$=(44+92-27)/160=0.6813$$

d. It was committed by offender above 20 given that it involved a violent crime? (3 points)

$$=(46+19)/92=.7065$$

e. It involved nonviolent crime by someone between 20 and 40 years old? (3 points)

$$=34/160=.2125$$

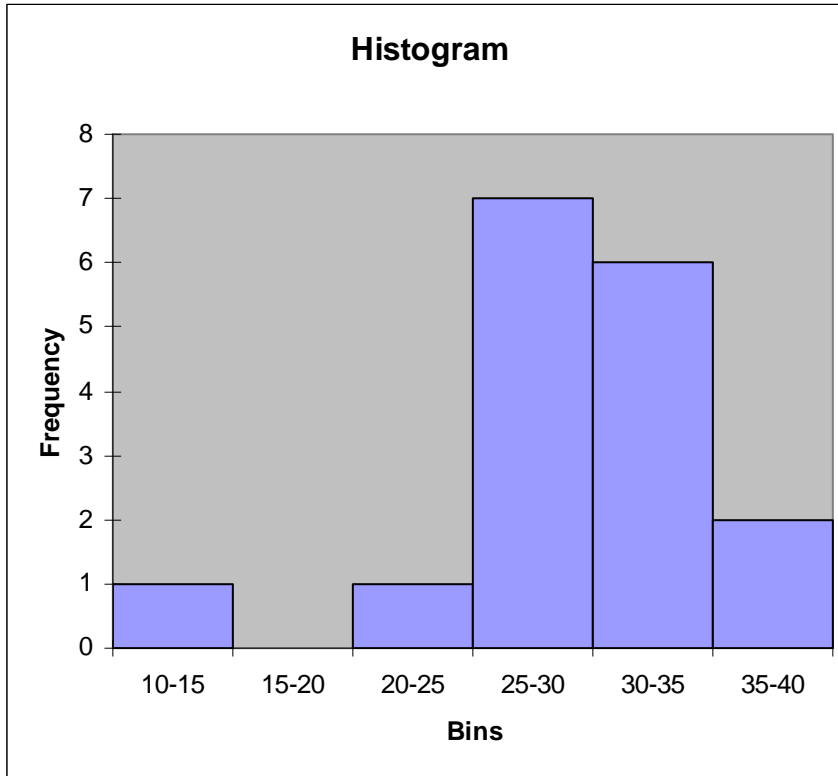
f. Based on this information, do you think age has any influence on type of crime? Justify your answer. (2 points)

$$(44/160)*(92/160) \text{ vs } 27/160$$
$$1581 \neq 1688 \Rightarrow \text{Dependent}$$

Name \_\_\_\_\_

2. Open the 'Fund' worksheet in your data set. The data is a sample of domestic general stock funds' returns in dollars.

a. Create a frequency distribution table and histogram with 6 classes using excel. (**Print out your results with your name and 5-digit team number typed and turn it in with your exam**) (5 points)



<i>Bins</i>	<i>Frequency</i>
10-15	1
15-20	0
20-25	1
25-30	7
30-35	6
35-40	2

b. Describe the shape of the distribution according to your graph. (5 points)

Data appears to have a significant left-side outlier, so left skewed.

c. Calculate all relevant descriptive statistics for the stock fund returns. Be sure to include inter-quartile range. Then use these statistics to **describe the data set and its distribution**. (5 points)

		location	value
<u>1-Yr Return</u>	$=18*0.25$		27.75
	$=18*0.75$		32.75
			$=32.75-27.75 =$
Mean	29.3471	<b>IQR</b>	<b>5</b>
Standard Error	1.48428	Q3-Q1	
Median	29.9		
Mode	30.5		
Standard Deviation	6.11986	<b>CV</b>	$=(6.1199/29.3471)*100$ <b>20.85</b>
Sample Variance	37.4526		
Kurtosis	4.59699		
Skewness	-0.271	$=3*(29.3471-29.9)/6.1199$	
Range	26.1		
Minimum	11		
Maximum	37.1		
Sum	498.9		
Count	17		

This slightly skewed data set has a mean of 29.35 with a median of 29.9, accounting for the slight negative skew. The variability within the data can be best understood by looking at the Coefficient of Variation, which says that the standard deviation is about 21% of the size of the mean, which is relatively small. The slight negative skew suggests that there was at least one fund that with returns much smaller than the bulk of the data. The IQR of 5 compared to the range of 26.1 bears this out, and it eliminates the impact of the minimum value of 11, which is very different from the other returns.

d. What would be the best measure of center for the stock fund returns? Why? (5 points)

Since the mean and the median are not equal, and the mean is less than the median, then the best measure of center for data that are skewed is the median.

Name \_\_\_\_\_

1. Open the “MLB” worksheet in your data set. This is a sample of data from Major League Baseball about number of wins (WINS), earned run average (E.R.A.) and runs score (Runs Scored). (20 points)

- a. Calculate the covariance between WINS and E.R.A., and between WINS and Runs Scored. Describe the relationships between WINS and E.R.A., and between WINS and Runs Scored. (5 points)

	<i>Wins</i>	<i>E.R.A.</i>	<i>Runs Scored</i>
Wins	88.99489796		
E.R.A.	-3.03410714	0.284890179	
Runs Scored	344.6836735	10.51678571	5583.316327

- b. Wins and ERA have a negative relationship. When ERA is low, number of wins is high. Runs and Wins have a positive relationship. When Runs are High, so are wins.
- c. Between E.R.A. and Runs Scored, which one has the stronger relationship with WINS? Why? (Justify your answer) (5 points)

	<i>Wins</i>	<i>E.R.A.</i>	<i>Runs Scored</i>
Wins	1		
E.R.A.	-0.60257293	1	
Runs Scored	0.488981064	0.263692738	1

The relationship between Wins and ERA is stronger because  $|-0.603| > |0.489|$ . Since these are correlation coefficients, then the larger the absolute value the stronger the relationship.

- d. If you want to predict team wins by using only runs scored, what would be your model for prediction (equation)? (5 points)
- i.  $Wins = 33.296 + 0.062 * Runs$
- e. Among these three variables, which one has the highest variation? (5 points) Using the Coefficient of Variation, ERA has the highest variation at about 12.4%.

11.87074058	12.3885	9.862048
Wins	ERA	Runs

Name \_\_\_\_\_

4. The following table is a report on costs for minor repairs during last month from a randomly selected auto shop. (20 points)

Costs for minor repair (\$)	Cumulative Frequency
0-100	10
100-200	28
200-300	88
300-400	158
400-500	200
500-600	228

(a) Calculate average costs and standard deviation for minor repair. (7 points)

RELATIVE FREQ	MID CLASS	MID*FREQ	MID-MEAN	(MID-MEAN)^2	FREQ*(MID-MEAN)^2
10	50	500	-287.719	82782.39	827823.9
18	150	2700	-187.719	35238.53	634293.6
60	250	15000	-87.7193	7694.675	461680.5
70	350	24500	12.2807	150.8156	10557.09
42	450	18900	112.2807	12606.96	529492.2
28	550	15400	212.2807	45063.1	1261767
		77000		183536.5	3725614
	MEAN	337.7193		VARIANCE	16412.4
				STD. DEV	128.1109

$$\text{MEAN} = [(10*50)+(18*150)+(60*250)+(70*350)+(42*450)+(28*550)]/228$$

$$= \mathbf{337.7193}$$

$$\text{VARIANCE} = [10*(50-337.72)^2+18*(150-337.72)^2+\dots+28*(550-337.72)^2]/(228-1)$$

$$= \mathbf{16412.4}$$

$$\text{STDARD DEVIATION} = \text{SQRT}(16412.4) = \mathbf{128.1109}$$

(b) What would be the Modal and Median classes for this data? WHY?(7 points)

MODAL CLASS = **[300-400]** because it is the most frequent class

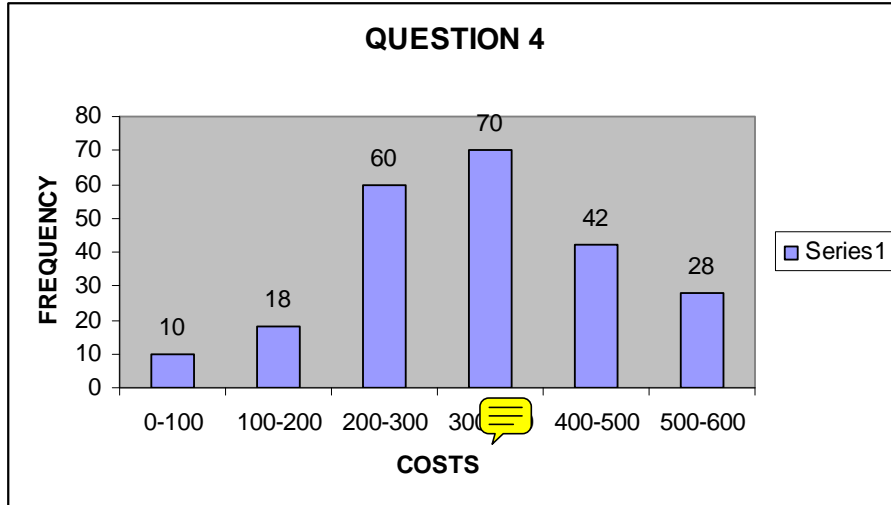
MEDIAN CLASS = **[300-400]** because the median (mid value) falls within this class

(c) What would the sign of the skewness coefficient be for this data and why? (6 points) (two answers are accepted)

Negative because  $337.72 < 350$  (midpoint of median class)

Zero (positive) because  $337.72$  is also in the class of **[300-400]** which is the same as median class.

Name \_\_\_\_\_



5. A vendor at the baseball stadium in Bloomington is going to determine whether to sell ice cream or soft drinks at tomorrow's IU vs. Purdue game. The vendor estimates the following profits that will be made given tomorrow's possible weather conditions when he invests \$100. (20 points)

Probability	Weather	Sell Soft Drinks	Sell Ice Cream
.25	Cool	\$50	\$20
.60	Warm	\$60	\$80
.15	Hot	\$70	\$90

- (a) Compute the expected profit and standard deviation for selling only soft drinks and those for selling only ice cream. (12 points)

5a  $E(\text{softdrinks}) = 0.25 \cdot 50 + 0.6 \cdot 60 + 0.15 \cdot 70 = 59$   
 $V(\text{softdrinks}) = 0.25 \cdot (50 - 59)^2 + 0.6 \cdot (60 - 59)^2 + 0.15 \cdot (70 - 59)^2 = 39$   
 $S(\text{soft drinks}) = \text{SQRT}(39.0225) = 6.2450$

$E(\text{icecreams}) = 0.25 \cdot 20 + 0.6 \cdot 80 + 0.15 \cdot 90 = 66.5$   
 $V(\text{icecreams}) = 0.25 \cdot (20 - 66.5)^2 + 0.6 \cdot (80 - 66.5)^2 + 0.15 \cdot (90 - 66.5)^2 = 732.75$   
 $S(\text{ice creams}) = \text{SQRT}(732.75) = 27.0694$

- (b) Compute the covariance of selling soft drinks and selling ice cream. (8 points)

5b  $\text{COV}(\text{soft drinks}, \text{ice creams})$   
 $= 0.25 \cdot (20 - 66.5)(50 - 59) + 0.6 \cdot (80 - 66.5)(60 - 59)$   
 $+ 0.15 \cdot (90 - 66.5)(70 - 59)$   
 $= 151.5$

- (d) **(BONUS QUESTION)** Now, the vendor is going to sell both soft drinks and ice cream by the proportion of 7:3. Compute the expected profit and standard deviation for selling both items. (5 extra credit points)

$E(\text{profit}) = 0.7 \cdot 59 + 0.3 \cdot 66.5 = 61.25$   
 $V(\text{profit}) = (0.7^2) \cdot 39 + (0.3^2) \cdot 732.75 + 2 \cdot 0.7 \cdot 0.3 \cdot 151.5 = 148.6875$   
 $S(\text{profit}) = \text{SQRT}(148.6875) = 12.1937$