SOLUTION TO A3-II

QUESTION 1: SELLING AND RENTING HOMES.

(a) [1 mark] One can argue for either a negative (more likely) or positive linear association. There may also be a non-linear association.

Since people have to live somewhere we have a "zero-sum game" (or at least roughly so) between house sales and rentals. When one of these variables goes up the other one is likely to go down. Thus we anticipate a negative correlation.

For a particular real-estate agent, sales and rentals may be positively correlated if there has been a sweeping demographic shift in (or out) of your city, or if the agent's degree of effort/efficiency as a sales-person has changed (improved/worsened). Under these circumstances when sales go up rentals will tend to also go up and visa versa.

Note that X and Y being correlated does not necessarily imply a linear correlation.

1mark for anything remotely reasonable

(b). [3 marks]

Use general formula (substituting Y for X if choosing RV Y)

E[X] = sum (P[X=x]* X)

Var[Y] = sum((X-E[X])^2 P[X=x])

If choice is X

E[X] = 1*0.15 + 3*0.59+ 6*0.26 = 3.48 sales/year

Var[X] = (1-3.33)^2 * 0.15 + (3-3.33)^2 * 0.59+ (6-3.33)^2 * 0.26 = 2.71 (sales/year)^2

If choice is Y

E[Y] = 1*0.15 + 4*0.25+ 10*0.60 = 7.15 sales/year

Var[Y] = (1-7.1)^2 * 0.15 + (3-7.1)^2 * 0.59+ (6-7.1)^2 * 0.26 = 13.03 (sales/year)^2

EV

0.5 mark for expression, 0.5 mark for answer

Var

0.5 mark for expression, 1 mark for answer

0.5 marks for units

c) [6 marks]

Let AGI denote annual gross income

AGI = $15000*X + $2500*Y

(i)

E[AGI] = $15000*E[X] + $2500*E[Y]

= $15000*4 + $2500*8 = $80,000

0.5 mark for expression

0.5 mark for answer

(ii)

Var[AGI] = ($15000)^2*Var[X] + ($2500)^2*Var[Y] +2* ($15000)*($2500)*Cov[X,Y]

Where : Cov[X,Y] = rho*sigma.x*sigma.y

= -0.4*2*5 = - 4 (sales-rentals/yr^2)

Var[AGI] = ($15000)^2*(2^2) + ($2500)^2*(5^2) +2* ($15000)*($2500)*(-4)

= $900^{e}6 + 156.25^{e}6 – 300^{e}6$ (where $1^{e}6 = 1x10^6$)

= $756.25^{e}6$ ($)^2

1.5 mark for terms 1 and 2 (that would apply in the case of independence)

1.5 mark for the correct calculation of the third term (owing to correlation)

0.5 mark for the correct answer

(iii) Coefficient of variation

Nu(AGI) = sqrt(Var[AGI])/E[AGI] = 0.344 (no units)

0.5 mark for logic, 0.5 mark for answer

0.5 marks for units within part (c.)

(d) [2 marks] Area=0.75 for Q3. This is an inverse question.

A => Z-score => X

(Z=0.675)

- ⇨ (Q3.AGI -80000)/ sqrt(756.25e6) =0.675
- ⇨ Solve for Q3.AGI

Q3.AGI = Z*sigma+80000

= 0.675*27500 + 80000

= $98600.00 [98700..00 if Z=0.68; 98400 if Z=0.67]

1 mark for logic, 1 mark for answer

(e) [Bonus 3 marks] Based on the hint (and lecture notes) we know that

Covar[X,Y] = E[(X-mu.x)(Y-mu.y)]

We have had lots of practice with determining expected values (EV). We have generally exploited tabular approach to determining EV's. Below we follow this same practice taking the expected value of the quantity: (X-mu.x)(Y-mu.y) --- as per hint. The tabulation below starts by re-organizing our original probability model so that it accommodates the EV calculation. Note that the mu.x and mu.y values are those calculated in the solution above (part b).

X (sale/yr)	Y (rent/yr)	Pxy	X -mu.x	Y -mu.y	(X-mu.x)(Y-mu.y)Pxy
1	1	0	-2.48	-6.15	0.000
3	1	0.03	-0.48	-6.15	0.089
6	1	0.12	2.52	-6.15	-1.860
1	4	0.02	-2.48	-3.15	0.156
3	4	0.12	-0.48	-3.15	0.181
6	4	0.11	2.52	-3.15	-0.873
1	10	0.13	-2.48	2.85	-0.919
3	10	0.44	-0.48	2.85	-0.602
6	10	0.03	2.52	2.85	0.215
					-3.612

The summation of the last column obtains the expected cross-deviation. That is the covariance. Thus the answer is -3.612 (sale*rentals/year).

2 marks for laying out a tabular calculation as shown above or for indicating the individual elements involved in the calculation, and 1 mark for the correct answer. Since this is a BONUS question, please be strict in marking (not generous).

QUESTION 2 [10 MARKS]:

Plan	Setup. Identify the 5W's and the H for the data available, and summarize the question being asked. [1 mark for attending to some portion of these details. 0.5 marks for providing a summary, but must at least address who/what/when for full marks]	The monthly returns (in dollars) of 2500 stocks that are classified as Growth or value stocks are provided for the time period from January 1975 until June 1997 (amounting to 22 years and 5 months of data). Who [monthly stock returns]; what [month, calendar year, type]; when [see above]; where [presumably a north American market]; why [to examine which type of stock might give a better return]; and how [stock returns for those stocks falling into each of the two categories must have been averaged each month]. The case study invites us to use the data to judge which of the two categories (growth or value) would make for a better choice in the long-run. Because data are provided as a time-series we will want to be aware of autonomous trends, and how the two types are associated with one and another.
Do	Mechanics. Select the graphical displays that help summarize the stock-returns from each category; how the stock-returns from the two categories compare to one and another; and which stock category might be the better investment. Histogram is a good choice for plotting the returns from each stock type (box and whisker or stem and leaf would work too). Be sure to use a common scale, and to clearly label histograms for both types. Describe the shape of the histograms. [4 marks for both histograms or suitable counterparts; subtract	See power-point slides

	0.5mark for each of: (1) incomplete labeling (including units); (2) and unshared scale.] Provide summary statistics for each stock type. The standard 5 number summary, supplemented by xbar, s, and coefficient of variation seem like useful summaries. These summaries can be obtained for both Value and Growth stocks. Since we are trying to figure out which would be preferable, and since these stocks follow similar trends, it seems sensible to summarize the difference (Value – Growth) as calculated for each month. [1.5 marks for summary statistics; must include 5 number summary for full marks; subtract 0.25 marks for neglecting units] A scatter plot is a useful way for examining how the returns for these two stock types are associated with each other. [2 marks for scatter plot; subtract 0.5 marks for poor labeling and neglecting units] Time-series plots are also useful displays, since the data are provided as a time series, and since we anticipate that the Value and Growth returns might follow similar	

	time trends. [bonus 1 mark, provided properly labeled] Histogram of the matched (by month) differences between the Value and Growth returns. [bonus 1 mark, provided properly labeled]	
Report	Conclusion: Report what you have learned from the data and any conclusions you can make. [2.5 marks for reasonable summary of analysis. Should provide some descriptions of the distribution of the returns (for each type) as well as a description of the scatter plot for full marks. Allocate 1 mark to the aforementioned descriptions, with the remaining 1.5 marks allocated to an otherwise reasonable summary.]	See power-point slides