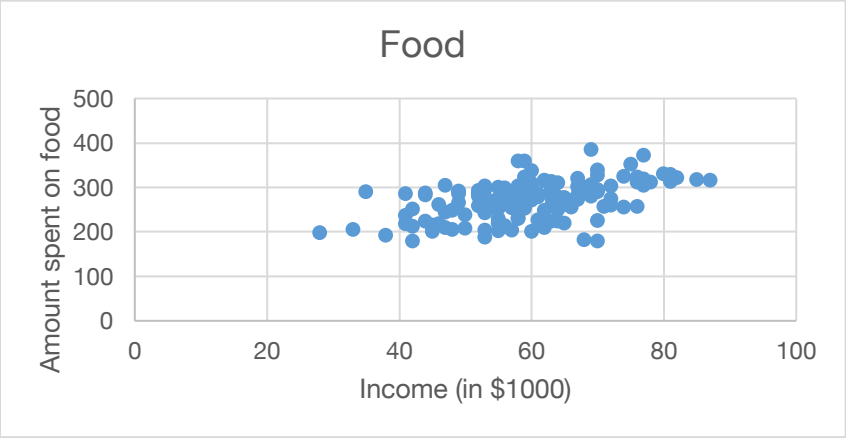
Homework two

Ugochukwu Achara
MSBA 620

1. $Y=\beta_0 + \beta_1 X_1 + \varepsilon$
   Y = amount spent on food
   X = Household income



-
- Scatter plot shows a moderate linear relationship with no outliers or leverage points.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.495853374
R Square	0.245870569
Adjusted R Square	0.2407751
Standard Error	36.93932351
Observations	150

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	65841.58033	65841.58033	48.25278347	1.1047E-10
Residual	148	201948.016	1364.513622		
Total	149	267789.5963			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*MarrErr*
Intercept	153.8985946	17.01994055	9.04225218	7.94473E-16	120.2651072	187.532082	33.63348737
Income	1.95820496	0.281901224	6.946422351	1.1047E-10	1.401133611	2.515276309	0.557071349

- At 95% confidence level, Anova table = $1.1 \times 10^{-10} < \alpha = 0.05$. Therefore, we conclude that we have a statistically significant model. With $R^2 = 0.245$, 25% of the variation in the income* is explained by the model. The variable coefficient for food is significant.

a. $\bar{x} = 270.26$ +/- 6.83
b. Est of $\beta_1 = 1.96$ +/- 0.56

c. The prediction interval for the amount of money spent each week on food for a household whose annual income is $40,000 is $232 +/- $74.

Prediction interval	
Estimate	232.2268
Margin of error	74.02609
LCL	158.2007
UCL	306.2529

2. $Y=\beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \varepsilon$

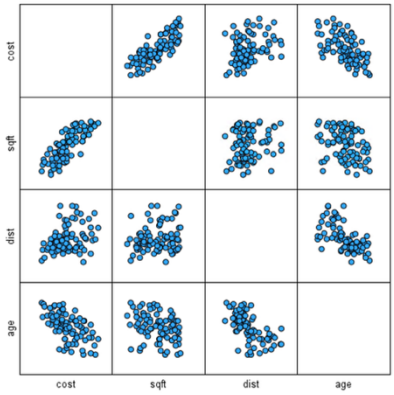
Y = Cost of apartments

X_1 = Square footage of apartment

X_2 = Distance from downtown

X_3 = Age of apartment

Graph



Correlations

		cost	sqft	dist	age
cost	Pearson Correlation	1	.801**	.311**	-.579**
	Sig. (2-tailed)		<.001	.004	<.001
	N	84	84	84	84
sqft	Pearson Correlation	.801**	1	.146	-.332**
	Sig. (2-tailed)	<.001		.186	.002
	N	84	84	84	84
dist	Pearson Correlation	.311**	.146	1	-.619**
	Sig. (2-tailed)	.004	.186		<.001
	N	84	84	84	84
age	Pearson Correlation	-.579**	-.332**	-.619**	1
	Sig. (2-tailed)	<.001	.002	<.001	
	N	84	84	84	84

**. Correlation is significant at the 0.01 level (2-tailed).

a. Yes, the data can be appropriately modeled with a linear model as evidenced by the graph and correlation matrix below. Both show that there is a linear relationship between cost and square footage (positive), cost and distance (positive) and cost and age (negative).

b. At 95% confidence level, ANOVA table shows that the model is statistically significant. With $R^2 = 0.752$, 75% of the variation in cost is explained by the model.

Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	age, sqft, dist[b]	.	Enter

a. Dependent Variable: cost

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.867[a]	.752	.743	7.392

a. Predictors: (Constant), age, sqft, dist

b. Dependent Variable: cost

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13243.200	3	4414.400	80.785	<.001[b]
	Residual	4371.503	80	54.644		
	Total	17614.702	83			

a. Dependent Variable: cost

b. Predictors: (Constant), age, sqft, dist

c.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	42.677	6.345		6.726	<.001	30.050	55.305
	sqft	.044	.004	.684	11.537	<.001	.036	.051
	dist	-.130	.867	-.011	-.150	.881	-1.856	1.596
	age	-.505	.105	-.358	-4.806	<.001	-.714	-.296

a. Dependent Variable: cost

$\beta_1X_1 = 0.044 +/- 0.0075$
$\beta_2X_2 = -0.130 +/- 1.726$
$\beta_3X_3 = -0.505 +/- 0.209$

β_1X_1 (square footage) and β_3X_3 (age) are statistically significant while β_2X_2 (distance) is not statistically significant.

d. No since distance is not statistically significant, the best model would be a reduced model with square footage and age of apartment as they are statistically significant.

SUMMARY OUTPUT
full model

Regression Statistics	
Multiple R	0.867079284
R Square	0.751826484
Adjusted R Square	0.742519977
Standard Error	7.392143313
Observations	84

ANOVA

	df	SS	MS	F	Significance F
Regression	3	13243.19976	4414.39992	80.78503531	3.86845E-24
Residual	80	4371.502621	54.64378276		
Total	83	17614.70238			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	MarrErr
Intercept	42.67718059	6.34526255	6.725833683	2.3744E-09	30.04970569	55.30465549	12.6274749
sqft	0.043839849	0.003799806	11.53739173	1.06521E-18	0.036277994	0.051401704	0.007561855
dist	-0.129945693	0.867269083	-0.149833189	0.881273331	-1.855866172	1.595974787	1.725920479
age	-0.504970589	0.105078843	-4.805635235	7.10211E-06	-0.71408415	-0.295857028	0.209113561

SUMMARY OUTPUT
Small model

Regression Statistics	
Multiple R	0.867039123
R Square	0.75175684
Adjusted R Square	0.74562738
Standard Error	7.347401803
Observations	84

ANOVA

	df	SS	MS	F	Significance F
Regression	2	13241.97301	6620.986504	122.6464894	3.10835E-25
Residual	81	4372.729373	53.98431325		
Total	83	17614.70238			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Marrerr
Intercept	42.02570087	4.593388184	9.149172503	4.02876E-14	32.88629922	51.16510252	9.139401648
sqft	0.043885831	0.003764469	11.65790736	5.24067E-19	0.036395719	0.051375944	0.007490113
age	-0.495349855	0.082675195	-5.991517218	5.48312E-08	-0.65984756	-0.33085215	0.164497705

Full Model		Reduced Model	
SSE (Residuals)	df(F)	SSE (Residuals)	df(R)
4371.502621	80	4372.729373	81
F-statistic	0.022449985		
p-value	0.881273331		

Since p-value = 0.88 > α = 0.05, we do not reject the null hypothesis of our F-test and conclude that there is no difference between the full model and the reduced model. Therefore, we would use the reduced model as any additional variable does not improve our model.

3.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.700[a]	.489	.475	2.8930

a. Predictors: (Constant), Weight

b. Dependent Variable: Height

Correlations

		Height	Weight
Height	Pearson Correlation	1	.700**
	Sig. (2-tailed)		<.001
	N	38	38
Weight	Pearson Correlation	.700**	1
	Sig. (2-tailed)	<.001	
	N	38	38

**. Correlation is significant at the 0.01 level (2-tailed).

a. I would not include both height and weight in my model as the data from both are highly correlated. Height and weight have a correlation coefficient (r) = 0.7 and coefficient of determination (r^2) = 0.489, 49% of variation in weight is explained by model. With a correlation coefficient of 0.7, the model indicates that taller people tend to weigh more with 49% of the variation in weight explained by the model with height and vice versa.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	111.276	55.867		1.992	.054	-2.141	224.692
	Brain	2.061	.547	.662	3.770	<.001	.951	3.170
	Height	-2.730	.993	-.482	-2.749	.009	-4.746	-.714

a. Dependent Variable: PIQ

b.

$Y=\beta_0 + \beta_1X_1 + \beta_2X_2 + \varepsilon$
X_1 = Brain size
X_2 = Height
$\beta_1X_1 = 2.061 +/- 0.547$
$\beta_2X_2 = -2.730 +/- 0.993$

Interpretation

$\beta_1X_1 = 2.061 +/- 0.547$
Holding other variables constant, for every 10,000-count increase in brain size, there is a 2.061 increase in PIQ.

$\beta_2X_2 = -2.730 +/- 0.993$
Holding other variables constant, a one-inch increase in height, leads to a 2.730 decrease in PIQ.