

Statistical Analysis Project

Doctorate of Business Administration

Graduate School of Business

Arab Academy for Science, Technology & Maritime Transport

Group 1

Member	Reg.No.
1. Saif El-Deen Ibrahim abdelslaam	25110322
2. Mohamed Ashraf Elsayed	25110327
3. Halla Nagi	25110505
4. Nourhan Nagi	25110504
5. Noor El-Shimi	25110311

Guide: Match each group member with the registration number in the same row. These are the five authors of this project.

Project scope

Ten statistical questions on 60 customer records: descriptive comparisons, mean and rank tests, categorical association, Excel correlation and multiple regression. Native SPSS tables and charts are followed by reading guides and explanations. The appendix maps all 15 lectures and states excluded methods with reasons.

Consolidated results summary

Question	Method	Key results	Conclusion
Q1	Descriptive means	Highest: Houston \$1,879.65; lowest: Jamestown \$1,281.38.	Descriptive comparison; significance is tested in Q8.
Q2	ATM descriptives	Highest mean: Houston 11.29; largest SD: San Lucas 5.82.	Houston has the highest average ATM use; San Lucas has the greatest spread.
Q3	Welch t-test	$t(56.424) = -2.153$; $p = .036$; difference = $-\$264.05$.	Reject H_0 . The interest group has a higher mean balance.
Q4	One-sample t-test; Wilcoxon check	Mean = 4.42; $t(59) = -10.122$; $p < .001$. Wilcoxon $p < .001$.	Reject the mean-equals-7 hypothesis. Average services used is below 7.
Q5	Pearson chi-square	Chi-square(3) = 0.224; $p = .974$; Cramer's V = .061.	Fail to reject H_0 . No city-debit association is detected.
Q6	Mann-Whitney U; Welch check	U = 415.000; $p = .687$. Welch mean comparison $p = .326$.	Fail to reject both null hypotheses. No balance difference is detected.
Q7	One-sample t-test	Mean = \$1,499.87; $t(59) = -6.490$; $p < .001$.	Reject the mean-equals-\$2,000 hypothesis. Mean balance is below \$2,000.
Q8	Welch ANOVA; Games-Howell	$F(3, 28.026) = 6.521$; $p = .002$. Houston-Jamestown $p = .002$.	Reject H_0 . Houston exceeds Jamestown; no other pair is significant.
Q9	Pearson correlation	$r = .705$; $p < .001$; $n = 60$; simple-model R squared = .497.	Strong positive linear association between balance and ATM use.
Q10	Multiple regression	R squared = .545; $F(2, 57) = 34.200$; $p < .001$. Predicted balance = \$1,916.93.	The model is significant; the prediction is an estimated mean, not a guaranteed balance.

Guide: Read each question's method, result and conclusion together. Tests use a .05 significance level. Q1-Q2 are descriptive. The Q3 difference is no interest minus interest. In Q6, Mann-Whitney tests distributions and Welch tests means. Q10 predicts at 15 ATM transactions and 4 other services. Failing to reject H_0 does not prove equality or independence; associations do not establish causation.

Sample and variable definitions

The analysis includes 60 complete customer records and six numeric variables. Account balance is measured in dollars; ATM transactions and other services are counts. Debit card and interest status are coded 0 = No, 1 = Yes. Cities are coded 1 = Jamestown, 2 = Houston, 3 = Dallas, 4 = San Lucas, following their order in the supplied city legend.

Analysis framework

Questions 1–8 are analyzed in SPSS; Questions 9–10 are calculated in Excel. Tests are two-sided where relevant, using $\alpha = .05$ and 95% confidence intervals. The course normality rule uses Kolmogorov–Smirnov for groups above 30 and Shapiro–Wilk below 30. Unequal variances and non-normal distributions are handled explicitly. A non-significant result is not proof of equality, and observational findings do not establish causation.

Variable Information

Variable	Position	Label	Measurement Level	Role	Column Width	Alignment	Print Format	Write Format
balance	1	Account Balance (\$)	Scale	Input	12	Right	F10.2	F10.2
atm	2	ATM transactions per month	Scale	Input	12	Right	F3	F3
services	3	Other bank services used	Scale	Input	12	Right	F3	F3
debit	4	Has a debit card	Nominal	Input	12	Right	F1	F1
interest	5	Receives interest	Nominal	Input	12	Right	F1	F1
city	6	Banking city	Nominal	Input	12	Right	F1	F1

Variables in the working file

SPSS variable definitions

Guide: Read Variable, Label, Measurement Level and format columns; value labels are shown in the next table. Explanation: All six variables are numeric; balance, ATM use and services are scale variables. Debit, interest and city are nominal codes, not numeric magnitudes.

Variable Values

	Value	Label
debit	0	No
	1	Yes
interest	0	No
	1	Yes
city	1	Jamestown
	2	Houston
	3	Dallas
	4	San Lucas

SPSS numeric codes and value labels

Guide: Match each numeric code to its value label before interpreting groups. Explanation: Debit/interest use 0 = No and 1 = Yes; cities use 1 = Jamestown, 2 = Houston, 3 = Dallas and 4 = San Lucas.

Question 1 — Average account balance by city

Draw a bar chart showing the average Account Balance for customers in each city.

Method and hypotheses

Descriptive comparison: mean account balance is plotted for each of the four numeric-coded city groups. No hypothesis test is needed for this question.

Checks and test selection

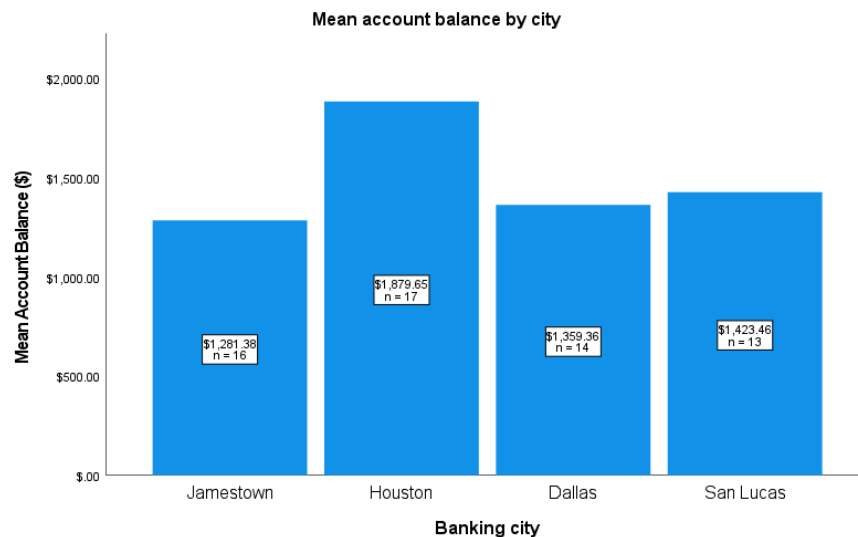
There are 60 customers: Jamestown 16, Houston 17, Dallas 14 and San Lucas 13. The vertical axis shows the mean, not the total balance.

Report

Account Balance (\$)			
Banking city	N	Mean	Std. Deviation
Jamestown	16	1281.3750	474.31675
Houston	17	1879.6471	350.87657
Dallas	14	1359.3571	683.48289
San Lucas	13	1423.4615	709.22195
Total	60	1499.8667	596.90491

SPSS mean account balance by city

Guide: Read each city's N and Mean; the outcome is balance in dollars. Explanation: Houston has the largest observed mean (\$1,879.65), followed by San Lucas (\$1,423.46), Dallas (\$1,359.36) and Jamestown (\$1,281.38). This table alone does not test significance.



SPSS bar chart of mean account balance

Guide: Compare bar heights on the dollar axis, not numeric city codes. Explanation: Houston has the highest sample mean. Use Q8's Welch/Games-Howell results to determine whether observed differences are statistically significant.

Answer and analysis

Houston has the highest average account balance (\$1,879.65), followed by San Lucas (\$1,423.46), Dallas (\$1,359.36) and Jamestown (\$1,281.38). Houston exceeds Jamestown by \$598.27 in this sample. The chart describes the sample; Question 8 tests whether the city mean differences are statistically significant.

Question 2 — ATM descriptive statistics by city

For each city, find the mean, median, mode, standard deviation, range, maximum, and minimum for number of ATM transactions per month.

Method and hypotheses

SPSS Frequencies is run separately for each city. Standard deviations are sample standard deviations. All tied modes are included in the summary below.

Checks and test selection

SPSS prints only the smallest mode when several values tie; its footnote is retained in the original table images.

Statistics ^a		
ATM transactions per month		
N	Valid	16
	Missing	0
Mean	9.38	
Median	10.00	
Mode	7 ^b	
Std. Deviation	3.442	
Range	13	
Minimum	4	
Maximum	17	

a. Banking city = Jamestown

b. Multiple modes exist. The smallest value is shown

Jamestown

Guide: Read mean, median, mode, SD and range for Jamestown. Explanation: Its 16 customers average 9.38 ATM transactions; both 7 and 10 are modes. SPSS displays only the smallest tied mode, so the table below lists both.

Statistics ^a		
ATM transactions per month		
N	Valid	17
	Missing	0
Mean	11.29	
Median	12.00	
Mode	6 ^b	
Std. Deviation	3.478	
Range	12	
Minimum	6	
Maximum	18	

a. Banking city = Houston

b. Multiple modes exist. The smallest value is shown

Houston

Guide: Read Houston's ATM summary and note the mode footnote. Explanation: The complete mode list is given in the combined table below. Mean and median describe location, while SD and range describe variation, not uncertainty in the mean.

Statistics ^a		
ATM transactions per month		
N	Valid	14
	Missing	0
Mean	10.00	
Median	9.50	
Mode	7 ^b	
Std. Deviation	4.169	
Range	15	
Minimum	2	
Maximum	17	

a. Banking city = Dallas

b. Multiple modes exist. The smallest value is shown

Dallas

Guide: Read Dallas's ATM mean, median, smallest displayed mode and spread. Explanation: SPSS's tied-mode convention can omit other modes; use the combined table below for the full list. These values describe customers in Dallas, not a time trend.

Statistics ^a		
ATM transactions per month		
N	Valid	13
	Missing	0
Mean	10.08	
Median	8.00	
Mode	6 ^b	
Std. Deviation	5.823	
Range	18	
Minimum	2	
Maximum	20	

a. Banking city = San Lucas

b. Multiple modes exist. The smallest value is shown

San Lucas

Guide: Read San Lucas's ATM mean, median, mode and spread. Explanation: Compare these with the other city summaries and the complete table below. A high range can reflect a small number of high-use customers rather than uniformly high use.

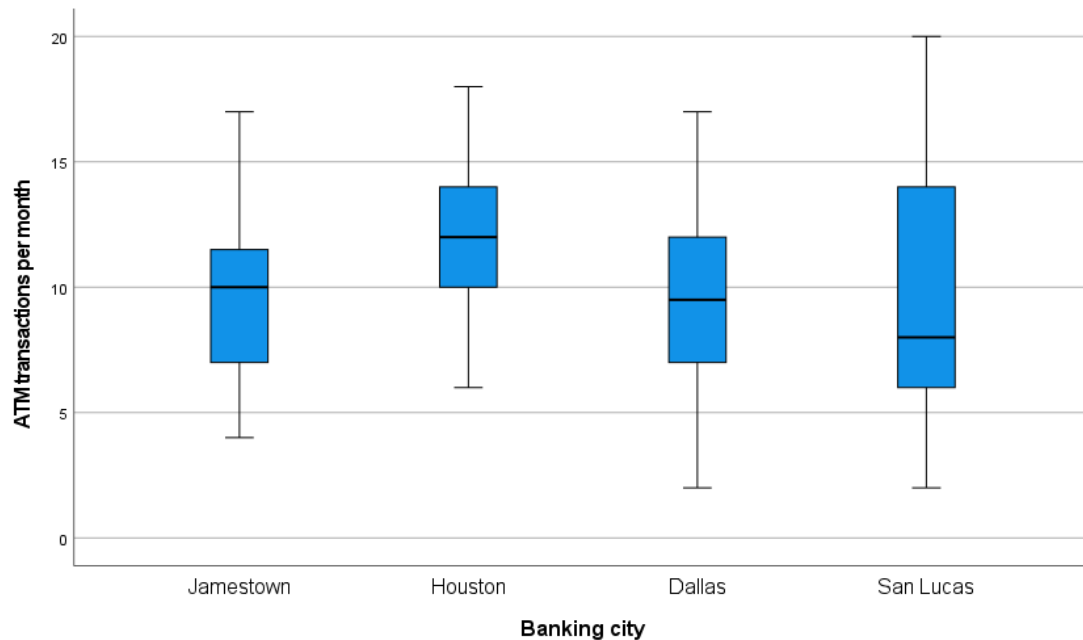
City	n	Mean	Median	All modes	SD	Range	Min	Max
Jamestown	16	9.38	10.00	7, 10	3.44	13	4	17
Houston	17	11.29	12.00	6, 10, 12, 14	3.48	12	6	18
Dallas	14	10.00	9.50	7, 8, 12	4.17	15	2	17
San Lucas	13	10.08	8.00	6, 8, 14, 20	5.82	18	2	20

Guide: Read across a city row; ATM values are monthly transaction counts. Explanation: All tied modes are listed, unlike SPSS's smallest-mode display. Mean/median describe typical use; SD/range describe customer spread. Q2 asks for descriptives, not a significance test between cities.

Answer and analysis

Houston has the highest average ATM use (11.29 transactions per month). San Lucas is the most variable city (SD = 5.82; range = 18), and its mean (10.08) is above its median (8), reflecting the influence of higher-use customers. The multiple modes show that no city has a single uniquely most frequent transaction count.

Question 2 — ATM distributions by city



SPSS ATM transaction boxplots by city

Guide: The line inside each box is the median; the box is the middle 50%, and markers identify potential outliers. Explanation: City ATM distributions differ in spread and high-use observations. This plot complements Q2; it is not a test of equal means.

Interpretation

The boxplots show medians, interquartile ranges, spread and potential outliers. They complement the mean, median, modes and ranges in the Q2 table; they do not replace those requested statistics. High-use observations help explain why some city means exceed their medians. The city labels are nominal codes, not numeric quantities.

Question 3 — Account balance and receiving interest

Test if there is a significant difference between the Account Balance and Receive interest on the account.

Method and hypotheses

Independent-samples t-test comparing customers who receive interest with those who do not. H0: the two population mean balances are equal. H1: they differ.

Checks and test selection

Normality is not rejected: no-interest group (n = 44), Kolmogorov–Smirnov p = .200; interest group (n = 16), Shapiro–Wilk p = .077. Levene's p < .001, so use the “Equal variances not assumed” (Welch) row.

Tests of Normality							
	Receives interest	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Account Balance (\$)	No	.085	44	.200 [*]	.968	44	.258
	Yes	.142	16	.200 [*]	.899	16	.077

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

SPSS normality tests

Guide: Apply the course sample-size rule separately to each interest group and compare Sig. with .05. Explanation: These normality checks are read alongside Q-Q plots. Failure to reject normality is not proof of normality; variance inequality determines the Welch row in Q3.

Group Statistics					
	Receives interest	N	Mean	Std. Deviation	Std. Error Mean
Account Balance (\$)	No	44	1429.4545	664.83442	100.22756
	Yes	16	1693.5000	282.50994	70.62749

SPSS group statistics

Guide: Read N, Mean, SD and SE for each interest group. Explanation: Interest-receiving accounts have the larger observed mean. Their mean difference is tested in the next table, rather than inferred from group summaries alone.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Account Balance (\$)	Equal variances assumed	13.269	<.001	-1.532	58	.131	-264.04545	172.30083	-608.94304	80.85213
	Equal variances not assumed			-2.153	56.424	.036	-264.04545	122.61242	-509.62692	-18.46399

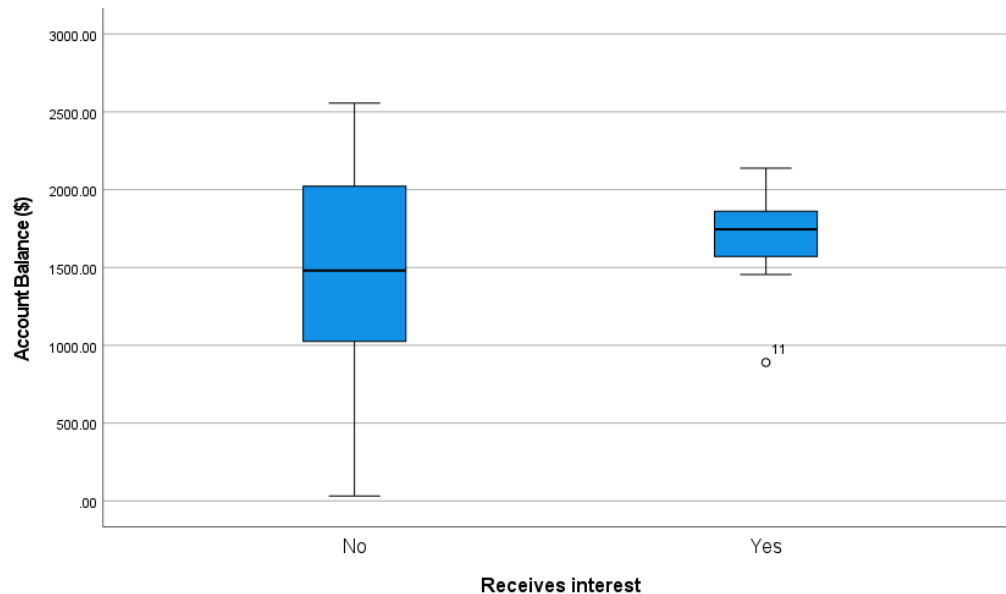
SPSS independent-samples t-test: read the unequal-variances row

Guide: Read the Equal variances not assumed row, Sig. (2-tailed) and the difference CI. Explanation: Welch $t(56.424) = -2.153$, $p = .036$; no-interest minus interest is $-\$264.05$ (95% CI $-\$509.63$ to $-\$18.46$). Reject equal means at .05.

Answer and analysis

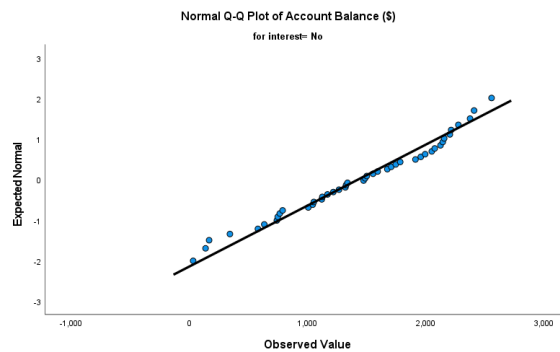
Yes. Customers receiving interest have a higher mean balance (\$1,693.50; SD = \$282.51) than customers not receiving interest (\$1,429.45; SD = \$664.83). Welch's two-sided t-test gives $t(56.424) = -2.153$, $p = .036$. Reject H_0 at .05. The difference (no interest minus interest) is -\$264.05 (95% CI: -\$509.63 to -\$18.46). Interest status is associated with balance, but this observational comparison does not show that receiving interest causes a higher balance.

Question 3 — Balance distributions by interest status



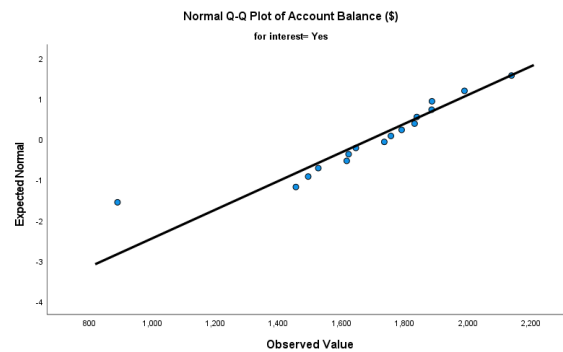
SPSS balance boxplots by interest status

Guide: Compare medians, interquartile ranges, whiskers and outliers by interest status. Explanation: The different spreads support using a variance-robust comparison. The plot does not by itself prove a difference or an interest-payment effect.



SPSS normal Q-Q plot: no interest

Guide: Compare no-interest balance points with the diagonal. Explanation: Departures show where the sample differs from a normal distribution. This visual check supplements the normality table; it cannot prove normality or establish independence.



SPSS normal Q-Q plot: receives interest

Guide: Compare interest-group balance points with the diagonal. Explanation: Interpret deviations together with group size and the normality test, not as an automatic pass/fail decision. Welch handles variance inequality, not every possible distribution problem.

Interpretation

The boxplot illustrates the different spreads that motivate the Welch row. The Q-Q plots supplement the normality test table. The visual contrast alone is not a significance test, and interest status is not a randomized treatment.

Question 4 — Other bank services compared with seven

Test the hypothesis that the average of Number of other bank services used is equal to 7.

Method and hypotheses

H0: the population mean number of other services is 7. H1: it is not 7. A two-sided one-sample t-test addresses this mean hypothesis, using a large-sample approximation ($n = 60$). A one-sample Wilcoxon test is also reported to follow the course's non-normal-data branch.

Checks and test selection

Kolmogorov–Smirnov $p = .002$ rejects normality under the course rule for $n > 30$. These are discrete counts. The t-test therefore relies on the sampling-mean approximation, not on normally distributed individual counts. Wilcoxon evaluates a location/median benchmark under its symmetry assumption; it does not test the mean directly.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Other bank services used	.150	60	.002	.969	60	.128

a. Lilliefors Significance Correction

SPSS normality tests

Guide: For all 60 service counts, read the course-selected Kolmogorov-Smirnov Sig. column.

Explanation: Discrete service counts depart from normality, motivating the Wilcoxon sensitivity check. The required hypothesis about the mean remains distinct from a rank-location hypothesis.

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Other bank services used	60	4.42	1.977	.255

SPSS one-sample statistics

Guide: Read the service-count N, Mean, SD and SE before testing against seven. Explanation: The sample mean is 4.42 services (SD 1.98), below the benchmark. Descriptive distance alone is not the inferential decision.

One-Sample Test						
	Test Value = 7					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Other bank services used	-10.122	59	<.001	-2.583	-3.09	-2.07

SPSS one-sample t-test: test value = 7

Guide: Test Value = 7; read t, df, two-sided p and the difference CI. Explanation: $t(59) = -10.122$, $p < .001$ rejects a population mean of seven. The negative difference indicates fewer services; the rank sensitivity result is explained separately.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig. ^{a,b}	Decision
1	The median of Other bank services used equals 7.	One-Sample Wilcoxon Signed Rank Test	<.001	Reject the null hypothesis.

a. The significance level is .050.

b. Asymptotic significance is displayed.

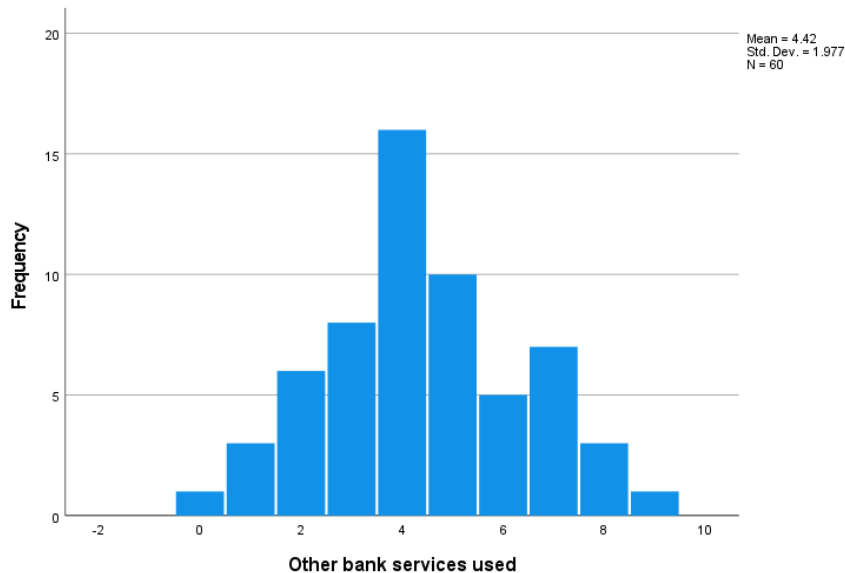
SPSS Wilcoxon sensitivity check

Guide: Read the Wilcoxon hypothesis, Sig. and decision; its benchmark is seven. Explanation: It also rejects the benchmark ($p < .001$), but tests signed-rank location under its assumptions rather than the population mean directly.

Answer and analysis

No. The sample mean is 4.42 services (SD = 1.98), significantly below 7: $t(59) = -10.122$, $p < .001$. Reject the mean-equals-7 hypothesis at .05. The estimated shortfall is 2.58 services (95% CI for mean minus 7: -3.09 to -2.07); the mean itself has a 95% CI of 3.91 to 4.93. The course-based Wilcoxon check also rejects its location/median benchmark (standardized statistic = -6.124, $p < .001$), supporting the direction of the finding. The bank could investigate barriers to using additional services rather than assume every customer needs seven.

Question 4 — Services distribution and signed ranks



SPSS histogram of other services used

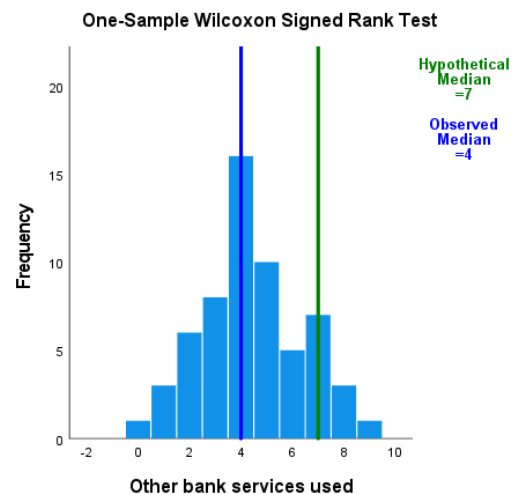
Guide: Read service-count values on the horizontal axis and customer frequencies vertically. Explanation: The histogram shows a discrete distribution concentrated below seven. It explains the normality concern but does not replace a significance test.

One-Sample Wilcoxon Signed Rank Test Summary

Total N	60
Test Statistic	27.500
Standard Error	112.345
Standardized Test Statistic	-6.124
Asymptotic Sig.(2-sided test)	<.001

SPSS one-sample Wilcoxon test statistics

Guide: Read the standardized signed-rank statistic and two-sided significance. Explanation: $Z = -6.124$, $p < .001$ supports a location below seven. Symmetry is needed for the usual median/location interpretation; this is a sensitivity check, not a mean test.



SPSS one-sample Wilcoxon signed-rank visual

Guide: Compare the observed median line (4) with the hypothetical median line (7); bars show service frequencies, not means or signed ranks. Explanation: The lower observed median illustrates the test's direction. Significance comes from the Wilcoxon summary, not the distance between the lines.

Interpretation

The histogram displays the discrete service counts. The signed-rank visual compares values with the benchmark of seven; it accompanies the Wilcoxon summary already reported. Seven is a hypothesis benchmark, not a required number of services for each customer. The mean-based t-test and location-based rank test answer different statistical hypotheses.

Question 5 — City and debit-card ownership

Test if there is a significance association between the "City where banking is done" and "whether or not the customer has a debit card".

Method and hypotheses

Pearson chi-square test of independence for a 4×2 contingency table. H0: city and debit-card ownership are independent. H1: they are associated.

Checks and test selection

Both variables are nominal. All expected cell counts exceed 5 (minimum = 5.63), so the chi-square approximation is appropriate, assuming independent customers.

Banking city * Has a debit card Crosstabulation

			Has a debit card		
			No	Yes	Total
Banking city	Jamestown	Count	9	7	16
		Expected Count	9.1	6.9	16.0
		% within Banking city	56.3%	43.8%	100.0%
		% within Has a debit card	26.5%	26.9%	26.7%
	Houston	Count	9	8	17
		Expected Count	9.6	7.4	17.0
		% within Banking city	52.9%	47.1%	100.0%
		% within Has a debit card	26.5%	30.8%	28.3%
	Dallas	Count	8	6	14
		Expected Count	7.9	6.1	14.0
		% within Banking city	57.1%	42.9%	100.0%
		% within Has a debit card	23.5%	23.1%	23.3%
	San Lucas	Count	8	5	13
		Expected Count	7.4	5.6	13.0
		% within Banking city	61.5%	38.5%	100.0%
		% within Has a debit card	23.5%	19.2%	21.7%
Total	Count	34	26	60	
	Expected Count	34.0	26.0	60.0	
	% within Banking city	56.7%	43.3%	100.0%	
	% within Has a debit card	100.0%	100.0%	100.0%	

SPSS city × debit-card crosstabulation

Guide: Compare observed with expected counts and distinguish row from column percentages.

Explanation: City-debit counts total 60 customers. All expected cells exceed five, supporting the Pearson chi-square approximation; Q5 tests association, not differences in mean balances.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	.224 ^a	3	.974
Likelihood Ratio	.225	3	.973
Linear-by-Linear Association	.108	1	.742
N of Valid Cases	60		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.63.

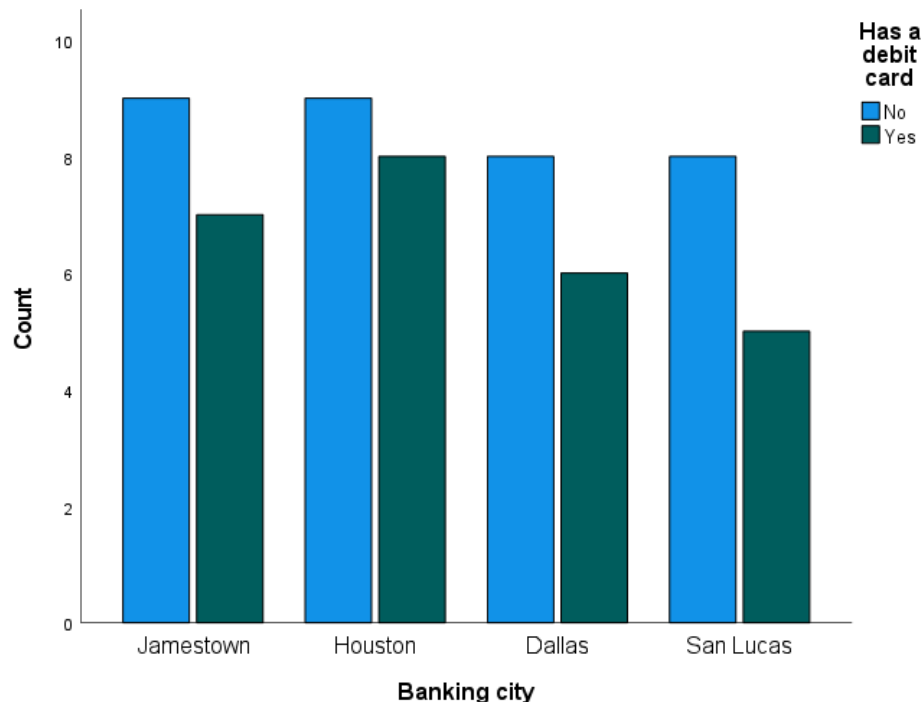
SPSS chi-square tests

Guide: Read Pearson Chi-Square, df and Asymptotic Sig. (2-sided). Explanation: chi-square(3) = .224, p = .974, so there is no detected city-debit association. Failure to reject is not evidence that association is exactly zero.

Answer and analysis

No. There is no statistically significant evidence of an association between city and debit-card ownership: $\chi^2(3, N = 60) = 0.224$, $p = .974$. Fail to reject H_0 at .05. Cramer's $V = .061$ indicates a very small observed association. Debit-card ownership ranges from 38.5% in San Lucas to 47.1% in Houston; those sample differences do not establish different ownership rates in the populations. The bank should not base city-specific debit-card policies on these data alone.

Question 5 — City and debit-card ownership: chart



SPSS grouped bar chart of customer counts

Guide: Bar height represents customer count; colors distinguish debit-card status within cities. Explanation: Similar observed proportions accompany the non-significant chi-square result. Unequal city totals mean raw bar heights should not be confused with ownership percentages.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.061	.974
	Cramer's V	.061	.974
N of Valid Cases		60	

SPSS association effect size

Guide: Read Cramer's V for the association's magnitude, alongside chi-square significance. Explanation: $V = .061$ is small in this sample. It measures categorical association, not a dollar difference or causal effect.

Interpretation

The chart shows counts, not ownership percentages. The city sample sizes differ (16, 17, 14 and 13), so compare ownership percentages as well as bar heights. Jamestown: $7/16 = 43.8\%$; Houston: $8/17 = 47.1\%$; Dallas: $6/14 = 42.9\%$; San Lucas: $5/13 = 38.5\%$. These figures are consistent with the non-significant chi-square result.

Question 6 — Account balance and debit-card ownership

Test the hypothesis that “there is no significance between the Account Balance and whether or not a customer has a debit card.”

Method and hypotheses

The course-based analysis uses a two-sided Mann–Whitney U test because one group fails normality. H0: the balance distributions are the same in the two groups. A Welch t-test is included as a sensitivity check for the narrower question about equality of mean balances.

Checks and test selection

No-card group (n = 34): Kolmogorov–Smirnov p = .044, so normality is rejected. Card group (n = 26): Shapiro–Wilk p = .987. Mann–Whitney does not require normality, but its result cannot automatically be interpreted as a test of means or medians without additional distributional assumptions.

Tests of Normality							
		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Has a debit card	Statistic	df	Sig.	Statistic	df	Sig.
Account Balance (\$)	No	.152	34	.044	.931	34	.033
	Yes	.064	26	.200*	.988	26	.987

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

SPSS normality tests

Guide: Read balance normality separately for customers with and without debit cards. Explanation: Distribution concerns motivate the Mann-Whitney analysis. Its rank hypothesis is distinguished from the Welch sensitivity test of population means.

Group Statistics					
	Has a debit card	N	Mean	Std. Deviation	Std. Error Mean
Account Balance (\$)	No	34	1435.8235	671.15068	115.10139
	Yes	26	1583.6154	483.11465	94.74658

SPSS balance group statistics

Guide: Compare debit groups' N, Mean, SD and SE. Explanation: These summarize dollar balances, while the following Mann-Whitney table uses ranks. A visible difference in group means does not alone establish statistical significance.

Test Statistics ^a	
	Account Balance (\$)
Mann-Whitney U	415.000
Wilcoxon W	1010.000
Z	-.403
Asymp. Sig. (2-tailed)	.687

a. Grouping Variable: Has a debit card

SPSS Mann–Whitney U test

Guide: Read Mann-Whitney U, Z and two-sided significance. Explanation: U = 415, Z = -.403, p = .687: no distribution/rank difference is detected. This is not automatically a test that mean balances are equal.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Account Balance (\$)	Equal variances assumed	3.928	.052	-.950	58	.346	-147.79186	155.63838	-459.33592	163.75221
	Equal variances not assumed			-.991	57.826	.326	-147.79186	149.08134	-446.22964	150.64593

SPSS t-test sensitivity check

Guide: Read the unequal-variances row to test the mean-based interpretation of Q6. Explanation: Welch $p = .326$ also fails to detect a mean difference. Neither a non-significant t-test nor rank test proves equivalence.

Answer and analysis

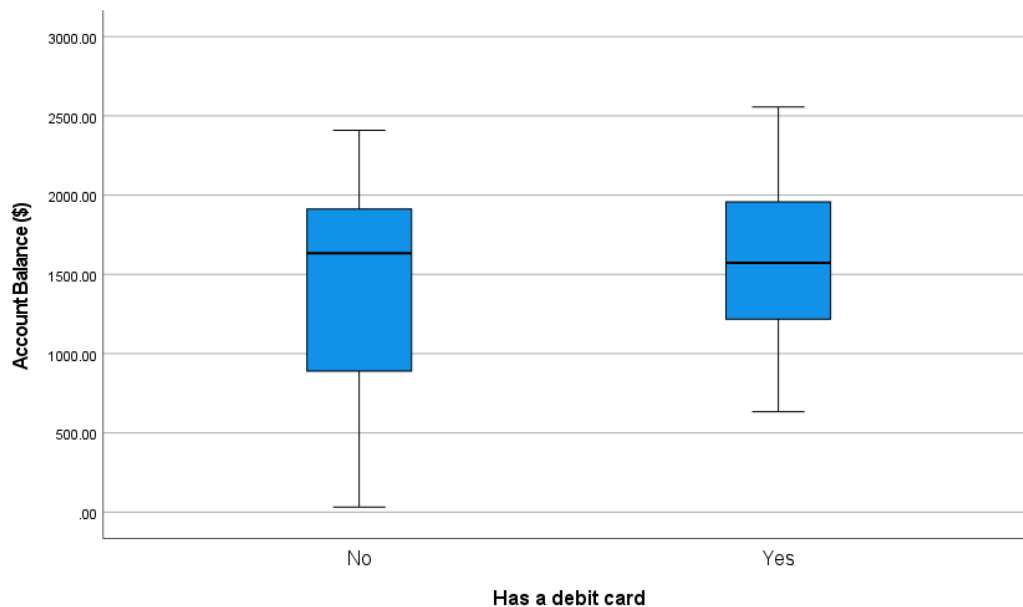
No. A significant balance difference is not detected: Mann–Whitney $U = 415.000$, $Z = -0.403$, SPSS two-sided $p = .687$. Fail to reject the distributional null hypothesis at .05. Mean balances are \$1,435.82 without a debit card and \$1,583.62 with one. The Welch mean comparison also fails to reject equality ($p = .326$); the mean difference (no card minus card) is $-\$147.79$ (95% CI: $-\$446.23$ to $\$150.65$). These results do not prove the groups are identical; debit-card ownership alone is not a reliable basis for predicting balance here.

Question 6 — Debit-card balance ranks and distribution

Ranks				
	Has a debit card	N	Mean Rank	Sum of Ranks
Account Balance (\$)	No	34	29.71	1010.00
	Yes	26	31.54	820.00
	Total	60		

SPSS Mann–Whitney ranks

Guide: Mean Rank and Sum of Ranks describe ordered balances, not dollars. Explanation: Mean ranks are 29.71 without a card and 31.54 with a card; the U test finds this separation non-significant ($p = .687$).



SPSS balance boxplots by debit-card status

Guide: Compare debit-group medians, boxes, whiskers and potential outliers. Explanation: The distributions overlap substantially and spreads differ. The plot illustrates the data; use the U and Welch tests for their respective statistical hypotheses.

Interpretation

The mean ranks are 29.71 without a debit card and 31.54 with a card. The ranks and boxplot accompany the Mann–Whitney test, as in Lecture 14. The large overlap and unequal spreads make it inappropriate to claim a clear separation of balance distributions or interpret the rank test as an automatic test of mean balances.

Question 7 — Account balance compared with \$2,000

Test the hypothesis that the average Account Balance is equal 2000 \$.

Method and hypotheses

Two-sided one-sample t-test. H0: the population mean account balance is \$2,000. H1: it differs from \$2,000.

Checks and test selection

For $n = 60$, Kolmogorov–Smirnov $p = .176$ does not reject normality under the course rule. Shapiro–Wilk $p = .063$ also does not reject normality. Independence and a representative sample are needed for population inference.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Account Balance (\$)	.103	60	.176	.963	60	.063

a. Lilliefors Significance Correction

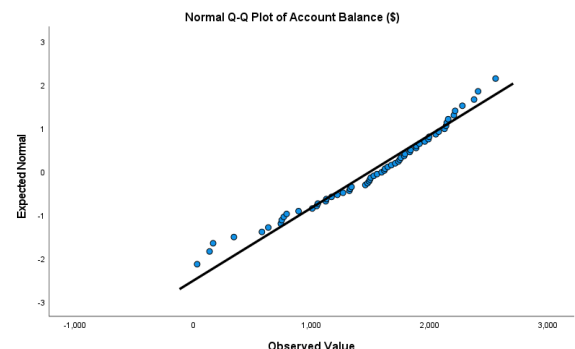
SPSS normality tests

Guide: For the 60 balance observations, read the course-selected Kolmogorov-Smirnov significance alongside Q-Q output. Explanation: The distribution check informs interpretation of the one-sample t-test. No normality test can verify the sampling independence assumption.

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Account Balance (\$)	60	1499.8667	596.90491	77.06009

SPSS one-sample statistics

Guide: Read overall N, Mean, SD and SE in dollars. Explanation: The 60-customer mean is \$1,499.87 (SD \$596.90). The next table tests the difference from \$2,000; SD is customer spread, while SE is mean precision.



SPSS normal Q-Q plot of account balance

Guide: Compare the balance quantiles with the reference line, especially the tails. Explanation: The Q-Q plot supplements the normality test for Q7. Modest visual departures do not justify ignoring sampling design or treating the t-test as a causal analysis.

One-Sample Test

Test Value = 2000						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Account Balance (\$)	-6.490	59	<.001	-500.13333	-654.3302	-345.9364

SPSS one-sample t-test: test value = \$2,000

Guide: Test Value = \$2,000; read t, df, p and Mean Difference. Explanation: $t(59) = -6.490$, $p < .001$ rejects the benchmark. The sample mean is about \$500.13 lower; the test does not predict every customer's balance.

Answer and analysis

No. The estimated mean balance is \$1,499.87 (SD = \$596.90), significantly below \$2,000: $t(59) = -6.490$, $p < .001$. Reject H_0 at .05. The estimated mean difference is -\$500.13 (95% CI: -\$654.33 to -\$345.94), and the mean balance has a 95% CI of \$1,345.67 to \$1,654.06. For planning based on customers like this sample, using \$2,000 as the expected average would overstate observed balances.

Question 8 — One-way ANOVA of account balance across cities

Test if there is a significant difference between the average Account Balance for the different cities.

Method and hypotheses

One-factor comparison of four independent city groups. H0: all four population mean balances are equal. H1: at least one differs. Use Welch one-way ANOVA and Games–Howell post-hoc comparisons because variances are unequal.

Checks and test selection

Each city has fewer than 30 customers. Shapiro–Wilk p-values are .543 (Jamestown), .839 (Houston), .257 (Dallas) and .936 (San Lucas); normality is not rejected. Levene's $p = .040$ rejects equal variances. The ordinary pooled ANOVA row is therefore not the primary test.

Tests of Normality							
	Banking city	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Account Balance (\$)	Jamestown	.158	16	.200 [*]	.953	16	.543
	Houston	.118	17	.200 [*]	.971	17	.839
	Dallas	.137	14	.200 [*]	.925	14	.257
	San Lucas	.122	13	.200 [*]	.974	13	.936

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

SPSS normality tests by city

Guide: Each city has fewer than 30 cases, so read its Shapiro-Wilk Sig. value. Explanation: These small-group checks do not reject normality, but have limited power. Levene's result, rather than normality alone, motivates Welch ANOVA.

Descriptives

Account Balance (\$)								
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
Jamestown	16	1281.3750	474.31675	118.57919	1028.6294	1534.1206	343.00	1913.00
Houston	17	1879.6471	350.87657	85.10007	1699.2430	2060.0511	1125.00	2409.00
Dallas	14	1359.3571	683.48289	182.66849	964.7259	1753.9884	32.00	2276.00
San Lucas	13	1423.4615	709.22195	196.70278	994.8830	1852.0401	167.00	2557.00
Total	60	1499.8667	596.90491	77.06009	1345.6698	1654.0636	32.00	2557.00

SPSS descriptive statistics by city

Guide: Read city N, mean, SD, SE and mean CI. Explanation: Houston has the highest sample mean; the four group sizes and spreads are not identical. These descriptive intervals are separate from adjusted pairwise comparison intervals.

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Account Balance (\$)	Based on Mean	2.961	3	56	.040
	Based on Median	2.407	3	56	.077
	Based on Median and with adjusted df	2.407	3	44.293	.080
	Based on trimmed mean	2.896	3	56	.043

SPSS homogeneity-of-variance tests

Guide: Read the mean-based Levene row and compare Sig. with .05. Explanation: $p = .040$ rejects equal city variances. Use Welch for the omnibus mean comparison and Games-Howell for pairwise comparisons rather than pooled-variance Tukey.

Robust Tests of Equality of Means

Account Balance (\$)				
	Statistic ^a	df1	df2	Sig.
Welch	6.521	3	28.026	.002
Brown-Forsythe	3.549	3	40.262	.023

a. Asymptotically F distributed.

SPSS robust equality-of-means tests:
use Welch

Guide: Read the Welch row, Statistic, both degrees of freedom and Sig. Explanation: $F(3, 28.026) = 6.521$, $p = .002$ rejects equal city mean balances. An omnibus result says at least one mean differs, not that every pair differs.

Question 8 — City mean comparisons: Games–Howell

Answer and analysis

Yes. Welch ANOVA shows that at least one city mean differs: $F(3, 28.026) = 6.521$, $p = .002$. Reject H_0 at .05. Games–Howell identifies Houston as significantly higher than Jamestown by \$598.27 ($p = .002$; 95% CI: \$199.39 to \$997.16). No other city pair is significant at .05 after these comparisons. Houston has the highest observed mean, but the result does not establish that banking in Houston causes higher balances; customer composition and other factors could explain the difference.

Multiple Comparisons

Dependent Variable: Account Balance (\$)

Games-Howell

(I) Banking city	(J) Banking city	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Jamestown	Houston	-598.27206*	145.95563	.002	-997.1575	-199.3867
	Dallas	-77.98214	217.78154	.984	-681.1271	525.1628
	San Lucas	-142.08654	229.68023	.925	-784.4599	500.2868
Houston	Jamestown	598.27206*	145.95563	.002	199.3867	997.1575
	Dallas	520.28992	201.51873	.079	-47.6311	1088.2109
	San Lucas	456.18552	214.32220	.185	-155.0328	1067.4039
Dallas	Jamestown	77.98214	217.78154	.984	-525.1628	681.1271
	Houston	-520.28992	201.51873	.079	-1088.2109	47.6311
	San Lucas	-64.10440	268.43949	.995	-803.1514	674.9426
San Lucas	Jamestown	142.08654	229.68023	.925	-500.2868	784.4599
	Houston	-456.18552	214.32220	.185	-1067.4039	155.0328
	Dallas	64.10440	268.43949	.995	-674.9426	803.1514

*. The mean difference is significant at the .05 level.

SPSS Games–Howell multiple comparisons

Guide: Use the Games-Howell Sig. column and each pair's 95% CI. Explanation: Only Houston-Jamestown is significant ($p = .002$), with Houston about \$598.27 higher (CI \$199.39 to \$997.16). Other city pairs are not significant at .05.

Interpretation

The significant pair is Houston–Jamestown. Houston–Dallas has $p = .079$ and Houston–San Lucas has $p = .185$, so neither is significant at .05. An omnibus difference does not imply that every city differs from every other city. The repeated reverse-direction rows are the same comparisons with opposite signs; they are not additional independent findings.

Question 8 Ordinary ANOVA and city mean confidence intervals

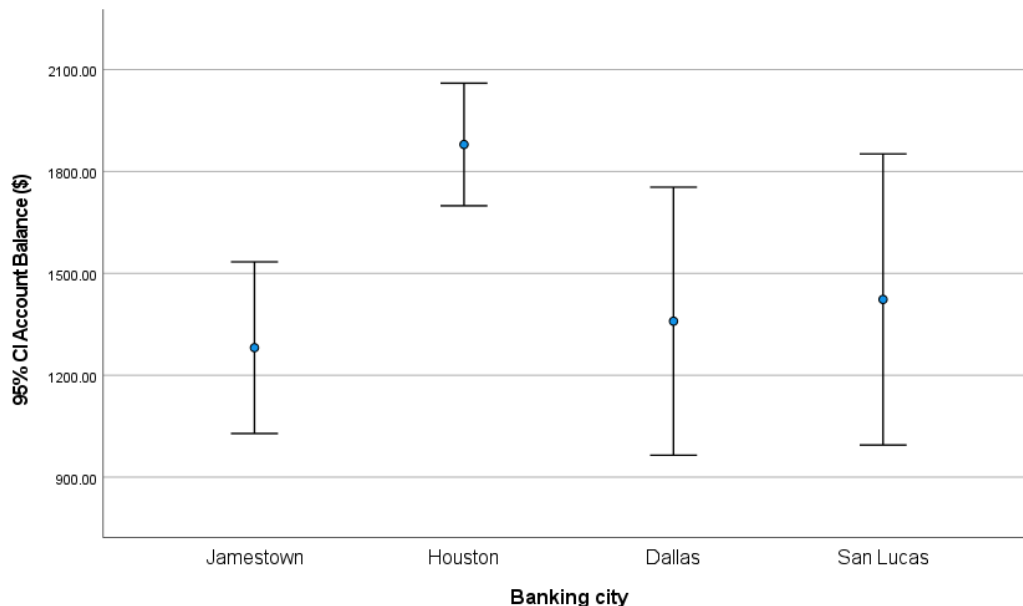
ANOVA

Account Balance (\$)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3568072.856	3	1189357.619	3.816	.015
Within Groups	17453360.08	56	311667.144		
Total	21021432.93	59			

SPSS ordinary ANOVA table: unequal-variance caveat

Guide: Read the ordinary ANOVA F, df and Sig., then check its variance assumption. Explanation: $F(3, 56) = 3.816$, $p = .015$ is shown for lecture coverage. Because Levene $p = .040$, Welch remains the primary result.



SPSS city mean balances with individual 95% confidence intervals

Guide: Points/bars show city means; error bars are individual 95% confidence intervals, not ± 1 SD or SE. Explanation: Houston has the largest estimated mean. CI overlap is not an adjusted pairwise test; use Games-Howell for pair decisions.

Interpretation

The lecture-style ordinary one-way ANOVA gives $F(3, 56) = 3.816$, $p = .015$. It is shown for completeness, but its equal-variance assumption is rejected (Levene $p = .040$), so Welch $F(3, 28.026) = 6.521$, $p = .002$ remains the primary result. The plotted intervals describe each city mean separately. Their overlap is not the Games–Howell pairwise test; use the adjusted comparison table for pairwise significance. Tukey is not used because it pools the unequal variances.

Question 8 Kruskal–Wallis sensitivity and pairwise comparisons

Ranks			
	Banking city	N	Mean Rank
Account Balance (\$)	Jamestown	16	22.56
	Houston	17	42.18
	Dallas	14	27.21
	San Lucas	13	28.54
	Total	60	

SPSS Kruskal–Wallis ranks

Guide: Read each city's Mean Rank and N rather than interpreting ranks as dollars.
 Explanation: Houston's mean rank is highest (42.18), while Jamestown's is lowest (22.56).
 The next table tests the omnibus rank difference.

Test Statistics ^{a,b}	
	Account Balance (\$)
Kruskal-Wallis H	11.564
df	3
Asymp. Sig.	.009

a. Kruskal Wallis Test

b. Grouping Variable:
Banking city

SPSS Kruskal–Wallis test statistics

Guide: Read Kruskal-Wallis H, df and significance.
 Explanation: $H(3) = 11.564$, $p = .009$ detects a city distribution/rank difference.
 This sensitivity analysis is not interchangeable with Welch's hypothesis about means.

Pairwise Comparisons of Banking city

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig. ^a
Jamestown-Dallas	-4.652	6.391	-.728	.467	1.000
Jamestown-San Lucas	-5.976	6.521	-.916	.359	1.000
Jamestown-Houston	-19.614	6.083	-3.224	.001	.008
Dallas-San Lucas	-1.324	6.727	-.197	.844	1.000
Dallas-Houston	14.962	6.303	2.374	.018	.106
San Lucas-Houston	13.638	6.434	2.120	.034	.204

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.

Asymptotic significances (2-sided tests) are displayed. The significance level is .050.

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

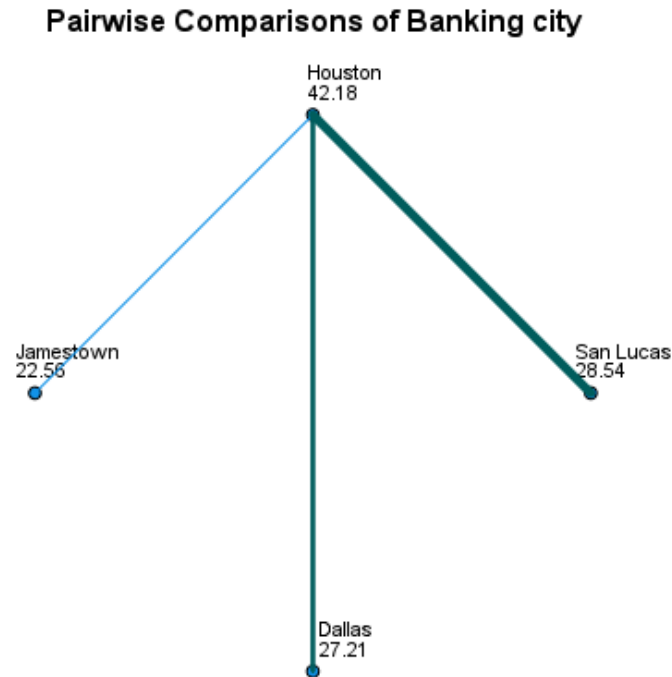
SPSS rank pairwise comparisons: Bonferroni-adjusted significance

Guide: Read Adj. Sig., not the unadjusted Sig., for the six city pairs. Explanation: After Bonferroni adjustment, only Jamestown-Houston is significant ($p = .008$). Adjustment controls the familywise error across these pairwise comparisons.

Interpretation

The rank-based sensitivity analysis gives $H(3) = 11.564$, $p = .009$. It also detects a city difference, but tests balance distributions/ranks rather than equal means. The Bonferroni-adjusted pairwise result is significant only for Jamestown–Houston ($p = .008$); Houston has the higher mean rank. Use the Adj. Sig. column, not the unadjusted Sig. column.

Question 8 Rank pairwise comparison diagram



Each node shows the sample average rank of Banking city.

SPSS rank pairwise comparison diagram

Guide: Node labels show cities' mean ranks; consult the adjusted table to interpret pairwise significance. Explanation: Houston-Jamestown is the only significant adjusted pair. Lines and node positions do not represent geography, dollar differences or causation.

Interpretation

Each city node shows its mean rank: Houston 42.18, San Lucas 28.54, Dallas 27.21 and Jamestown 22.56. These are ranks, not dollar balances. The adjusted pairwise table identifies only Houston-Jamestown as significant ($p = .008$); the other city comparisons are not significant at .05. The diagram complements the table and does not imply causation or geographical distance.

Question 9 — Excel correlation: balance and ATM transactions

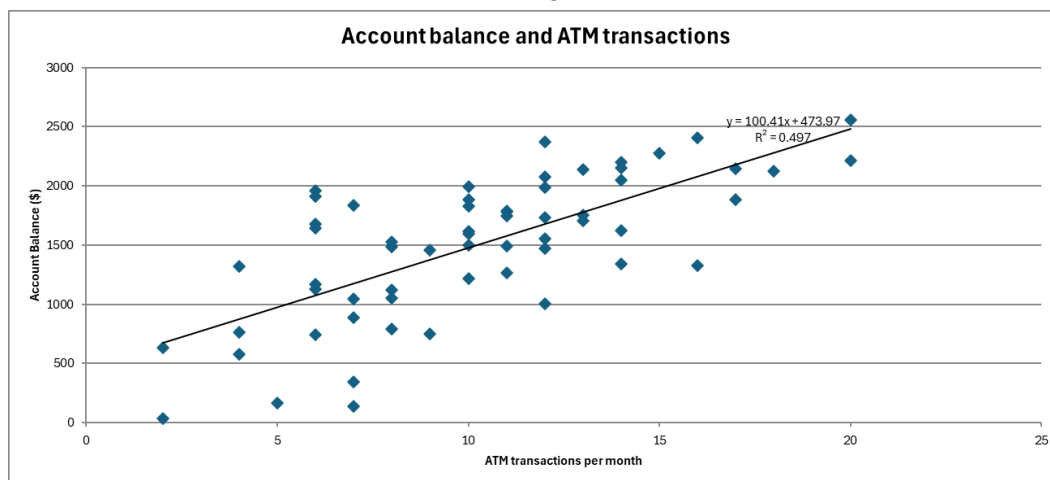
Using Excel, find the correlation coefficient between the Account Balance and Number of ATM transactions per month? Interpret your result.

Method and hypotheses

Pearson correlation calculated in Microsoft Excel: =CORREL(Data!A2:A61,Data!B2:B61). Both variables are quantitative, and the scatterplot is used to assess the direction and approximate linear form.

Checks and test selection

The calculation includes all 60 customer pairs. Correlation describes a linear relationship; independent observations and the usual inferential assumptions are needed for its p-value. The course classifies a correlation above .70 as strong.



Microsoft Excel scatterplot with a simple linear trendline

Guide: Each point is one customer's ATM use and dollar balance; the fitted line summarizes their linear relationship. Explanation: The positive trend accompanies $r = .70499$ and simple-model R-squared about .497. This line is not the two-predictor model used for Q10.

Correlations

		Account Balance (\$)	ATM transactions per month
Account Balance (\$)	Pearson Correlation	1	.705
	Sig. (2-tailed)		<.001
	N	60	60
ATM transactions per month	Pearson Correlation	.705	1
	Sig. (2-tailed)	<.001	
	N	60	60

SPSS correlation output

Guide: Read Pearson Correlation, Sig. (2-tailed) and N for balance versus ATM use. Explanation: $r = .70499$, $p < .001$, $N = 60$ indicates a strong positive linear association under the lecture's convention. SPSS checks the Excel result, but Q9 is calculated in Excel.

Answer and analysis

There is a strong positive linear relationship: $r = 0.70499$ (approximately .705), $n = 60$, $p < .001$. Customers with more ATM transactions tend to have higher account balances in this sample. The squared correlation is about .497, so the single-predictor linear relationship accounts for about 49.7% of the observed balance variation. The plotted trendline uses ATM alone; it is not the two-predictor equation in Question 10. Association does not establish that increasing ATM transactions will increase a customer's balance.

Question 10 — Excel multiple regression and prediction

Using Excel, estimate: Account Balance = a + b (Number of ATM transactions) + c (Number of other bank services used). (i) Write the linear equation of the model. (ii) Predict Account Balance when ATM transactions per month = 15 and other bank services used = 4.

Method and hypotheses

Ordinary least-squares multiple regression calculated in Microsoft Excel using LINEST with an intercept: =LINEST(Data!A2:A61,Data!B2:C61,TRUE,TRUE). LINEST returns predictor coefficients in reverse column order, so the services and ATM coefficients are assigned to the correct variables. Excel t- and F-distribution formulas calculate inference.

Checks and test selection

The predictors are ATM transactions and other services, not the numeric city or yes/no codes. SPSS model output and residual plots accompany the Excel results. The requested input values (15 ATM transactions, 4 services) are within the observed ranges.

Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.739 ^a	.545	.530	409.43302

a. Predictors: (Constant), Other bank services used, ATM transactions per month

b. Dependent Variable: Account Balance (\$)

SPSS model summary

Guide: Read R-squared, adjusted R-squared and Std. Error of the Estimate. Explanation: The two predictors explain about 54.5% of sample balance variation (adjusted 53.0%); residual SE is about \$409.43. Explained variation is not prediction certainty.

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11466215.38	2	5733107.690	34.200	<.001 ^b
	Residual	9555217.554	57	167635.396		
	Total	21021432.93	59			

a. Dependent Variable: Account Balance (\$)

b. Predictors: (Constant), Other bank services used, ATM transactions per month

SPSS regression ANOVA

Guide: Read regression F, model/residual df and Sig. Explanation: $F(2, 57) = 34.200$, $p < .001$ rejects the hypothesis that both predictor slopes are zero. This regression ANOVA is different from the city-group ANOVA in Q8.

Coefficients ^a										
		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Collinearity Statistics	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	247.031	167.805		1.472	.146	-88.993	583.055		
	ATM transactions per month	93.133	13.058	.654	7.132	<.001	66.985	119.281	.949	1.054
	Other bank services used	68.224	27.681	.226	2.465	.017	12.794	123.654	.949	1.054

a. Dependent Variable: Account Balance (\$)

SPSS coefficients for the same fitted model

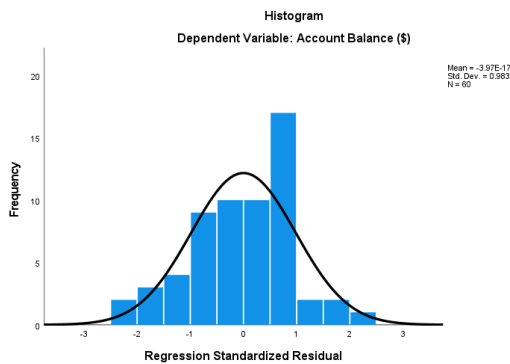
Guide: Use Unstandardized B for the equation; t, Sig. and CI describe each slope holding the other predictor fixed. Explanation: Balance = 247.0307 + 93.1334 x ATM + 68.2241 x services. At 15 and 4, Excel predicts \$1,916.93 using unrounded coefficients.

Answer and analysis

(i) Predicted Account Balance (\$) = $247.0307 + 93.1334 \times \text{ATM transactions} + 68.2241 \times \text{other services}$. (ii) At 15 ATM transactions and 4 services, the predicted balance is \$1,916.93, using the unrounded Excel coefficients. Holding the other predictor constant, one additional ATM transaction is associated with \$93.13 higher predicted balance, and one additional service with \$68.22. The model is significant, $F(2, 57) = 34.200$, $p < .001$; $R^2 = .545$ and adjusted $R^2 = .530$. ATM has $p < .001$ and services $p = .017$. The prediction is an estimated conditional mean, not a guaranteed individual balance; residual standard error is about \$409.43. These coefficients are associations, not causal effects.

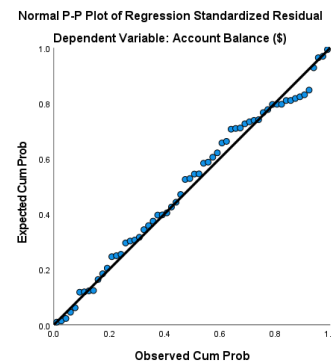
Excel prediction calculation: $247.0307233 + 93.1334126 \times 15 + 68.2240555 \times 4 = 1,916.9281336$. The Analysis sheet retains the live CORREL, LINEST and prediction formulas.

Question 10 — Regression diagnostics



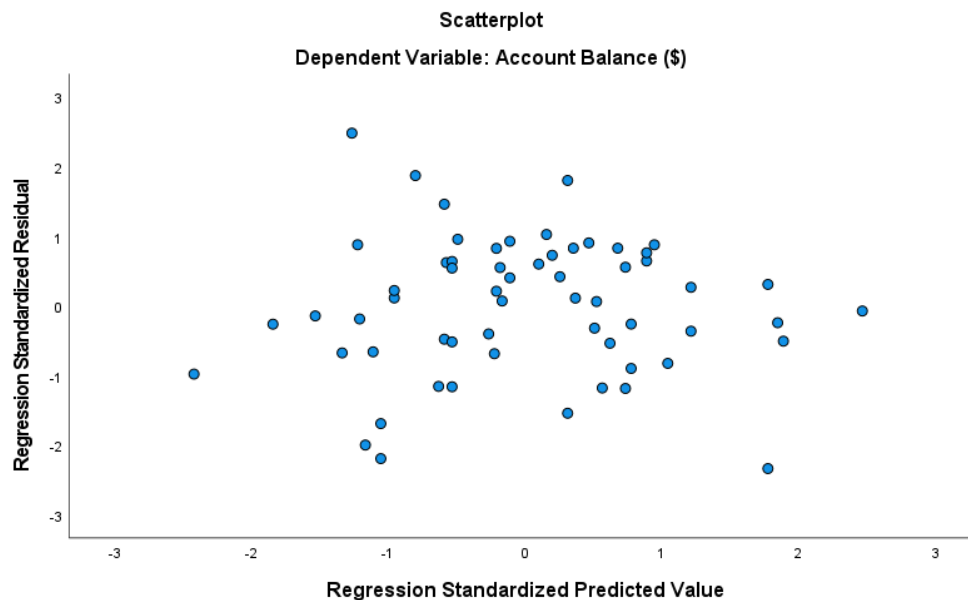
SPSS standardized residual histogram

Guide: Read the distribution of standardized residuals, not raw balances. Explanation: There is no pronounced non-normal pattern, but a histogram is only a visual diagnostic. Approximate residual normality supports conventional regression inference, not causal claims.



SPSS normal P-P plot of residuals

Guide: Points close to the diagonal indicate residual probabilities near a normal reference. Explanation: No pronounced systematic departure appears. The P-P plot supplements the histogram; it does not establish independence or constant variance.



SPSS residuals versus fitted values

Guide: Look for curvature, a widening funnel or isolated extreme points around zero residual. Explanation: No obvious strong curve or funnel appears, though this cannot prove linearity/homoscedasticity. Sampling independence still requires study-design information.

Interpretation

The residual histogram and P-P plot show no pronounced departure from approximate normality. The residual-versus-fitted plot shows no obvious curved pattern or strong funnel, although visual checks cannot prove linearity or constant variance. Both VIFs are 1.054, so there is little evidence of multicollinearity. Standardized residuals range from -2.337 to 2.481 and maximum Cook's distance is .196: no case exceeds the conventional Cook's-distance threshold of 1, but influential cases still deserve review. Independence cannot be established from these plots; sampling and collection details remain important.

Exploratory two-way ANOVA of city and debit-card status

This extension jointly examines the factors used in Q6 and Q8. Account balance is the outcome; city has four levels and debit-card ownership has two. The full factorial Type III model tests city, debit status and their interaction. It is supplementary to the ten specified questions, not a replacement for their unadjusted tests.

All eight cells contain 5–9 customers. Shapiro–Wilk does not reject cell normality ($p = .186$ to $.961$), but these small cell samples have limited power. Mean-based Levene gives $F(7, 52) = 2.799$, $p = .015$, so conventional pooled-variance F tests and confidence intervals require caution. The variance-appropriate Welch/Games–Howell analysis remains the main conclusion about cities.

Descriptive Statistics				
Dependent Variable: Account Balance (\$)				
Banking city	Has a debit card	Mean	Std. Deviation	N
Jamestown	No	1266.3333	582.02685	9
	Yes	1300.7143	331.64977	7
	Total	1281.3750	474.31675	16
Houston	No	2013.8889	298.38836	9
	Yes	1728.6250	361.12837	8
	Total	1879.6471	350.87657	17
Dallas	No	1216.8750	814.42468	8
	Yes	1549.3333	458.46861	6
	Total	1359.3571	683.48289	14
San Lucas	No	1195.1250	619.48307	8
	Yes	1788.8000	752.52455	5
	Total	1423.4615	709.22195	13
Total	No	1435.8235	671.15068	34
	Yes	1583.6154	483.11465	26
	Total	1499.8667	596.90491	60

SPSS city × debit-card cell descriptives

Guide: Read N, mean and SD for each city-by-debit cell. Explanation: All eight cells are populated, with 5-9 customers each. Small cells make distribution checks weak and estimated interactions less stable than a large factorial sample.

Levene's Test of Equality of Error Variances ^{a,b}					
		Levene Statistic	df1	df2	Sig.
Account Balance (\$)	Based on Mean	2.799	7	52	.015
	Based on Median	1.451	7	52	.205
	Based on Median and with adjusted df	1.451	7	31.813	.220
	Based on trimmed mean	2.689	7	52	.019

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: Account Balance (\$)

b. Design: Intercept + city + debit + city * debit

SPSS factorial Levene test

Guide: Read Levene F, df and Sig. for the eight-cell factorial model. Explanation: $F(7, 52) = 2.799$, $p = .015$ rejects equal cell variances. Conventional Type III F tests and pooled-model CIs are therefore exploratory, not the main city conclusion.

Tests of Normality

		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	city_debit	Statistic	df	Sig.	Statistic	df	Sig.
Account Balance (\$)	Jamestown No	.208	9	.200 [*]	.912	9	.329
	Jamestown Yes	.128	7	.200 [*]	.971	7	.908
	Houston No	.140	9	.200 [*]	.957	9	.763
	Houston Yes	.167	8	.200 [*]	.945	8	.665
	Dallas No	.238	8	.200 [*]	.879	8	.186
	Dallas Yes	.163	6	.200 [*]	.945	6	.701
	San Lucas No	.130	8	.200 [*]	.980	8	.961
	San Lucas Yes	.236	5	.200 [*]	.933	5	.617

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

SPSS normality tests for the eight city × debit cells

Guide: Read Shapiro-Wilk for each of the eight cells. Explanation: All p-values are .186-.961, so normality is not rejected. With only 5-9 cases per cell, this is limited evidence; the unequal-variance finding still warrants caution.

Factorial ANOVA and estimated marginal means

Tests of Between-Subjects Effects

Dependent Variable: Account Balance (\$)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	5380792.86 ^a	7	768684.694	2.556	.024	.256
Intercept	131408864.8	1	131408864.8	436.891	<.001	.894
city	3237669.347	3	1079223.116	3.588	.020	.172
debit	411985.302	1	411985.302	1.370	.247	.026
city * debit	1568394.420	3	522798.140	1.738	.171	.091
Error	15640640.08	52	300781.540			
Total	155997434.0	60				
Corrected Total	21021432.93	59				

a. R Squared = .256 (Adjusted R Squared = .156)

SPSS Type III two-way ANOVA and effect sizes

Guide: Read city, debit and city x debit rows, their F, Sig. and partial eta-squared. Explanation: City $p = .020$; debit $p = .247$; interaction $p = .171$. Only the city effect is detected, and heteroscedasticity limits the conventional model's inference.

1. Banking city

Dependent Variable: Account Balance (\$)

Banking city	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Jamestown	1283.524	138.193	1006.220	1560.828
Houston	1871.257	133.246	1603.880	2138.634
Dallas	1383.104	148.095	1085.931	1680.278
San Lucas	1491.962	156.328	1178.267	1805.658

SPSS city estimated marginal means

Guide: Estimated marginal city means average the two debit groups with equal weights. Explanation: They differ slightly from raw Q1 means because group sizes differ. Their model-based CIs assume the conventional factorial variance model and require caution.

2. Has a debit card

Dependent Variable: Account Balance (\$)

Has a debit card	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
No	1423.056	94.219	1233.991	1612.120
Yes	1591.868	109.217	1372.709	1811.028

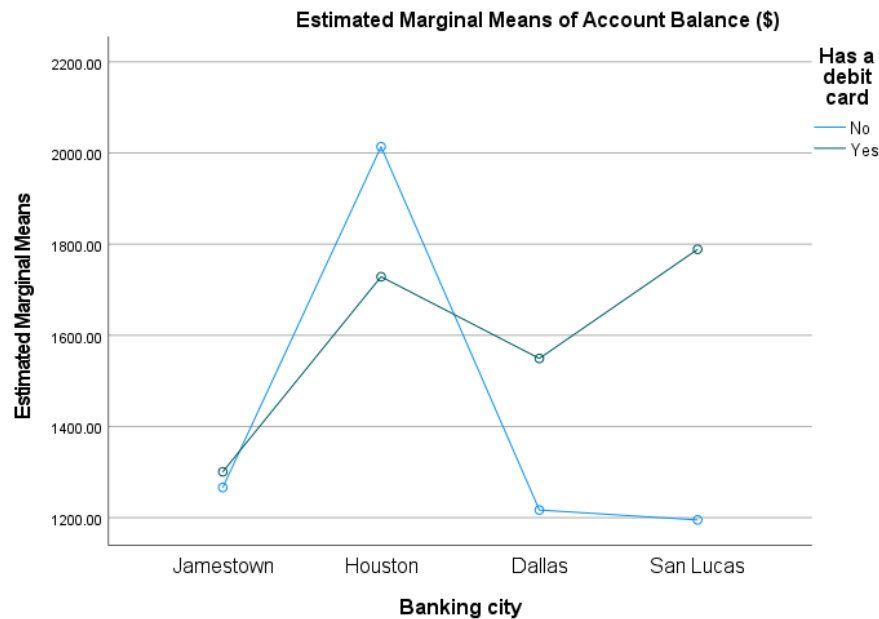
SPSS debit-card estimated marginal means

Guide: Estimated debit-group means average across the four cities with equal weights. Explanation: These are adjusted model summaries, not raw group averages. The debit main-effect test is non-significant ($p = .247$); do not infer significance from plotted or tabulated means alone.

Interpretation

The conventional two-way ANOVA detects a city main effect, $F(3, 52) = 3.588$, $p = .020$, partial eta squared = .172. It does not detect a debit-card main effect, $F(1, 52) = 1.370$, $p = .247$, or a city $\times$ debit-card interaction, $F(3, 52) = 1.738$, $p = .171$. The profile lines are not parallel, but that visual pattern is not statistically significant evidence of interaction. Estimated marginal city means give equal weight to the two debit groups; they therefore differ slightly from the raw city means in Q1/Q8. These observational and heteroscedastic results do not establish causal effects.

City × debit-card interaction profile



SPSS city × debit-card interaction profile

Guide: Follow debit-group profile lines across nominal city categories, not time. Explanation: Nonparallel lines suggest a possible interaction visually, but the interaction test is not significant ($p = .171$). This plot cannot establish city or debit causal effects.

3. Banking city * Has a debit card

Dependent Variable: Account Balance (\$)

Banking city	Has a debit card	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Jamestown	No	1266.333	182.812	899.495	1633.172
	Yes	1300.714	207.289	884.758	1716.670
Houston	No	2013.889	182.812	1647.050	2380.728
	Yes	1728.625	193.901	1339.534	2117.716
Dallas	No	1216.875	193.901	827.784	1605.966
	Yes	1549.333	223.898	1100.049	1998.617
San Lucas	No	1195.125	193.901	806.034	1584.216
	Yes	1788.800	245.268	1296.634	2280.966

SPSS estimated cell means and 95% confidence intervals

Guide: Read each city-debit estimated mean, SE and 95% CI. Explanation: These describe the eight model cells, not eight independent significance decisions. Small cell sizes and unequal variances make the pooled-model intervals exploratory.

The horizontal axis contains four nominal city categories, not time. The lines connect estimated means to compare debit groups across cities. Crossing or nonparallel lines alone do not establish interaction; the interaction F test has $p = .171$.

Lecture 3 mean precision checks

The lecture's descriptive rule requires $\text{mean} \geq 2.5 \times \text{SD}$ and $\text{mean} \geq 10 \times \text{SE}$. The equivalent ratios are checked below for the three quantitative variables across all 60 customers. SD measures variation among customers; $\text{SE} = \text{SD} / \text{square root of } n$ measures sample-mean precision.

Report								
	variable	n	Mean	SD	SE	Mean / SD	Mean / SE	Course criteria
1	Balance (\$)	60	1499.867	596.905	77.060	2.513	19.464	Both met
2	ATM transactions	60	10.217	4.191	.541	2.438	18.884	Not both met
3	Other services	60	4.417	1.977	.255	2.234	17.305	Not both met

SPSS descriptive mean/SD and mean/SE course checks

Guide: Compare Mean / SD with 2.5 and Mean / SE with 10; both must meet the lecture rule.

Explanation: Balance meets both (2.513, 19.464). ATM use (2.438, 18.884) and services (2.234, 17.305) miss the SD criterion, not the SE criterion. This heuristic does not invalidate their data or determine test validity.

Balance meets both criteria, narrowly for mean/SD. ATM use and other services meet mean/SE but miss mean/SD. Report their medians, spread and distributions alongside means rather than concealing this limitation. The ratios depend on the measurement origin and scale; this course heuristic is not a universal validity test, normality test or reason to discard observations. It does not replace the question-specific assumption checks and confidence intervals.

Lecture technique coverage

The following mapping distinguishes methods used in the project from lecture methods requiring a different data design. Supplementary analyses are identified separately from the ten required answers.

Lecture 1 Data types and descriptive statistics

Included

Variable definitions distinguish three scale variables from three numerically coded nominal variables. Q2 reports mean, median, all tied modes, SD, range, minimum and maximum. Explore output provides SE, variance and distribution summaries where relevant.

Lecture 2 Charts in Excel

Included; selected chart forms

Q1 uses a mean bar chart; Q5 a grouped frequency bar chart; Q9 an Excel scatterplot and fitted straight line. Bars compare nominal categories without treating city codes as quantities. Time-trend, pie and 3-D variants are discussed in the exclusions.

Lecture 3 Validity of the mean: SD and SE checks

Added after full audit

The appendix calculates mean/SD and mean/SE for balance, ATM use and other services, with native SPSS output. Course thresholds are 2.5 and 10. This is a relative-variation/precision heuristic, not a formal normality test or a condition that determines whether a t-test is valid.

Lecture 4 Pearson correlation, direction and strength

Included

Q9 calculates Pearson r in Excel and checks it in SPSS. The scatterplot supports a positive approximately linear relationship; $r = .70499$ is strong under the lecture's classification. Correlation does not establish causation.

Lecture 5 Simple and multiple linear regression

Included

Q9 shows the simple linear trendline and its R-squared as a descriptive companion to correlation. Q10 fits the required two-predictor model, reports coefficients and model fit, and predicts balance at 15 ATM transactions and four services. The trendline is not substituted for multiple regression.

Lecture 6 Excel correlation and regression examples

Included

Q9/Q10 retain live CORREL and LINEST formulas, coefficient SEs, t tests, CIs, model F, R-squared, adjusted R-squared and residual SE. The equation and requested prediction are explained alongside SPSS output for the same fitted model.

Lecture 7 Hypotheses, significance and inference

Included

Each inferential question states its null and alternative hypotheses and uses $\alpha = .05$. Decisions use p-values and, where available, confidence intervals. A non-significant result means failure to reject, not proof that populations are identical.

Lecture 8 SPSS setup, frequencies, Explore and normality plots

Included

Variable definitions and numeric value labels are shown. Descriptives, normality tests, histograms, boxplots and Q-Q plots accompany relevant analyses; Q10 adds residual diagnostics. Duplicated stem-and-leaf and menu screenshots are addressed in the exclusions.

Lecture 9 Parametric/nonparametric test selection

Included

Selections use measurement scale, independent-group count, the requested hypothesis and distribution/variance checks. Mean-based tests and rank-based sensitivity analyses are distinguished; rank tests do not automatically test population means.

Lecture 10 One-sample and independent-samples t-tests

Included; paired form excluded

Q4 tests services against seven and Q7 balance against \$2,000. Q3 compares interest groups using Welch's row; Q6 includes a Welch sensitivity test of means. No specified question involves the same outcome measured twice on a customer.

Lecture 11 One-way ANOVA, post-hoc tests and uncertainty charts

Included; repeated-measures form excluded

Q8 contains ordinary ANOVA, Levene, primary Welch ANOVA, Games-Howell and a mean chart with individual 95% CIs. Tukey is replaced because variances differ. Repeated-measures ANOVA and sphericity checks require repeated outcomes absent here.

Lecture 12 Factorial ANOVA and interaction

Supplementary two-way analysis

City-by-debit Type III two-way ANOVA adds cell descriptives, normality, Levene, partial eta-squared, estimated marginal/cell means and an interaction profile. It is exploratory because variances differ. Reasons for excluding a full three-way model and extra pooled-variance follow-ups are given below.

Lecture 13 Mixed-design ANOVA

Excluded: study design does not match

No within-customer repeated outcome or time factor is supplied. Balance, ATM counts and service counts are different attributes, not occasions of the same outcome. Mixed ANOVA cannot answer the specified questions.

Lecture 14 Wilcoxon, Mann-Whitney and Kruskal-Wallis

Included; related-sample forms excluded

Q4 includes one-sample Wilcoxon and signed-rank output. Q6 includes Mann-Whitney ranks, U test and boxplot. Q8 includes Kruskal-Wallis ranks, H test, Bonferroni pairwise table and rank diagram. Paired Wilcoxon/Friedman require comparable repeated outcomes absent here.

Lecture 15 Categorical association and related binary tests

Chi-square included; McNemar/Cochran Q excluded

Q5 contains observed/expected crosstabulation, chi-square, Cramer's V and grouped bars; expected counts exceed five. McNemar tests paired marginal rates rather than city-debit association. Cochran Q needs three or more comparable binary conditions. Detailed exclusions follow.

Excluded methods and why

Excluded means a technique is not used to answer this project, not that it was missed in the lecture review. Some need a different study design; others are replaced by a variance-appropriate method or would duplicate an existing result. No missing measurements or groups are invented.

Paired t-test and paired Wilcoxon signed-rank (Lecture 10, 14)

Excluded: each customer has one balance measurement, with no before/after or matched measurements of the same outcome. ATM transactions and services are different attributes, so the same row does not create comparable outcome pairs. Independent-group tests answer Q3/Q6 instead.

Repeated-measures ANOVA and Friedman (Lecture 11, 14)

Excluded: there are no three or more repeated quantitative measurements/conditions for each customer. Four cities are independent groups, not repeated occasions. One-way/Welch ANOVA and Kruskal-Wallis answer Q8 instead.

Mauchly, Greenhouse-Geisser/Huynh-Feldt corrections and within-subject contrasts (Lecture 11, 13)

Excluded because no multi-level within-subject factor exists. These checks concern repeated-measurement covariance, not unequal variances between cities. Levene and Welch address the latter.

Mixed ANOVA, Box's M and time-by-group plots (Lecture 13)

Excluded: mixed ANOVA needs both a grouping factor and a repeated outcome. Grouping factors exist but no repeated outcome is supplied. A time-by-group plot would invent time measurements. The included city-by-debit profile is a between-subject two-way interaction plot, not a time trend.

Complete city-by-debit-by-interest three-way ANOVA (Lecture 12)

Not fitted: the full $4 \times 2 \times 2$ design has 16 cells, including two empty cells and several with only 2-3 customers. Some full-factorial effects cannot be estimated independently, and small cells make checks/interaction estimates fragile. No required question asks for this joint model. A reduced model would need a separately justified specification; the eight-cell city-by-debit extension is reported instead.

Tukey HSD and additional pooled-variance factorial follow-ups (Lecture 11, 12)

Tukey is replaced, not overlooked: city variances differ (Levene $p = .040$), so Welch/Games-Howell is used. Rank comparisons use Bonferroni adjustment. Extra factorial pairwise/simple-effect tests are not added: interaction is not detected ($p = .171$), factorial variances differ ($p = .015$), and no planned simple-effect question was specified. The factorial result remains exploratory.

McNemar test (Lecture 15)

Excluded because Q5 asks about city-debit association, not equality of two paired binary marginal rates. Debit and interest can mathematically form a paired 2×2 table, but they describe different account attributes; equality of those rates is an additional, unspecified question. McNemar cannot replace the city-debit chi-square test.

Cochran Q test (Lecture 15)

Excluded: only two binary attributes (debit/interest) are supplied, not three or more comparable binary conditions per customer. City has four nominal categories and cannot become a third binary response without changing the question.

Time-trend charts, pie/3-D variants and duplicate distribution displays (Lecture 2, 8)

Time-trend charts are excluded because no date/time series exists. Pie charts are optional for counts, but Q5 already shows counts/percentages; city mean balances are not parts of one total. 3-D adds no analytical information. Histograms, boxplots and Q-Q displays avoid duplicate stem-and-leaf pages. CI bars are included and labeled; SE-bar versions would repeat the same means on another uncertainty scale.

Duplicate simple regression reports, generic rejection diagrams and menu screenshots (Lecture 5-8)

The simple trendline is included for Q9; Q10 requires two predictors, so another full one-predictor inferential report would not answer a new question. Generic textbook rejection diagrams and repeated menu dialogs are not project results. Every result image instead has a guide explaining its statistic and decision.