

Question 1. Hotel Bookings (25%)

← Binominal distribution

Siz 6/5 169

During high season, hotels often take on more bookings than the number of rooms they have. The reason for this is that often a certain portion of bookings get cancelled last minute. However, it can happen that everyone shows up and there are not enough rooms available. This overbooking situation can be costly for the hotel as they need to send their possibly unhappy guests to another hotel.

Hotel Reykjavik Natura is a newly renovated hotel next to Reykjavik's domestic airport and Reykjavik University. It has 220 rooms and is relatively conveniently located for downtown, only 5 min driving away.

The hotel is very busy in June, especially on the 17 June, Iceland's Independence Day. However, in the past they have had some guests cancelling last minute opting to travel out of town instead.

Hotel Reykjavik Natura is interested in understanding the implication of taking on more bookings than there are rooms. They estimate the cancelling probability on 17 June to be 4%.

The so-called binomial distribution can be used to describe uncertain situations of this kind where there are two possible outcomes; a guest shows up or not. Fortunately, the binomial distribution can be approximated by the normal distribution such that the number of guests that show up is normally distributed with a mean of $n \cdot p$ and a standard deviation of $\sqrt{n \cdot p \cdot (1 - p)}$ where n is the number of rooms booked and p is the probability of showing up.

- a) *If the hotel books 225 rooms on 17 June what is the probability that there will be an overbooking situation?*
- b) *If the hotel books 233 rooms on 17 June what is the probability of having empty rooms?*
- c) *If the hotel wants only 2% probability of an overbooking situation on 17 June, how many rooms should it book?*
- d) *You have been hired as a consultant for Hotel Reykjavik Natura. You have understood the probability structure of the overbooking problem and the next step is to advise them on how many overbooked rooms to allow during peak season. Please make a list of the data you would need to get from the hotel in order to make your recommendation.*

Question 2. Presidential Election (25%)

The Borealis Rowing Club, the largest rowing club in the UK, elects a president every four years. The elections are coming up next month and there are two candidates competing for the prestigious role, Mr. Grump and Mrs. Flintstone. An opinion poll was done where 120 members were asked whom they would vote for. The results can be found in the file Grump and Flintstone.xlsx. The first column indicates, which candidate would get the vote of the individual asked with 1 representing Mr. Grump. The second column represents the age of the individual asked. The third column indicates how many medals the individual asked has won in rowing competitions.

- a) *Which candidate do you expect to win the election?*
- b) *Mrs. Flintstone seems to relate better to the younger members of the club. Do you agree?*
- c) *Mr. Grump's emphasis in his campaign has been on making rowing an attractive sport to the general public while Mrs. Flintstone's focus is on providing support for exceptional rowers. It is expected that the members of the rowing club that have themselves done well are more likely to support Mrs. Flintstone. Does the opinion poll support that idea?*
- d) *After the recent US presidential election the media claimed that the opinion polls failed. Provide two arguments supporting that claim and two arguments against.*

Question 3. Location Choice (25%)

Café Like is a coffee shop chain that focuses on reasonably priced coffee and comfortable atmosphere. Café Like is considering expanding further and is evaluating two different sites for a new coffee shop. It already owns 47 shops and has collected data on those to help with its decision making. The data includes information on last year's revenues (in million pounds), management and staff quality (on a 9 point scale), type of location (1 for prime location, 0 for non-prime location), local per capita annual income (in thousand pounds), and an index of competitor density (in number of competitors per km²). The data is available in the file CafeLike.xlsx.

- a) *Build and analyze a regression model that can be used to forecast the revenues of a new coffee shop.*
- b) *Two sites have been offered to Café Like. Comment on which, if either, you would recommend that the company should go for.*

Site 1: On a pleasant street off a high street in a rural but up-and-coming village. There aren't many competitors there yet with a competition density of 0.2 (2 pubs within 10 km²). The staff available in this area is usually loyal and hardworking. Many commuters to a relatively close-by big city have moved to the village with an average per capita income of £42,000. There was a coffee shop there before that recently closed down due to the age of the owner and the site is in good shape and does not need much work. The estimated running costs including rent are about £50,000.

Site 2: A corner location on a high street in town that used to house a popular hardware store. There are nine coffee shops within 10 km². Staffing is not thought to be a big problem but competent managers have been a hard find. The average per capita income is £55,000. The running costs (inclusive) are thought to be about £150,000.

Question 4. Cash Withdrawal (25%)

A bank wants to analyse what variables have an effect on the amount of cash withdrawn from automatic teller machines (ATMs) located in residential neighbourhoods. A sample of total daily withdrawals from ATMs has been collected, together with information suspected to affect withdrawals. This information includes the median value of homes in the neighbourhood, the median family income in the neighbourhood, the average checking balance of customers in the neighbourhood, the distance to the next nearest ATM, and whether or not the withdrawals occurred on a weekend. A part of the data set is given in Table 1.

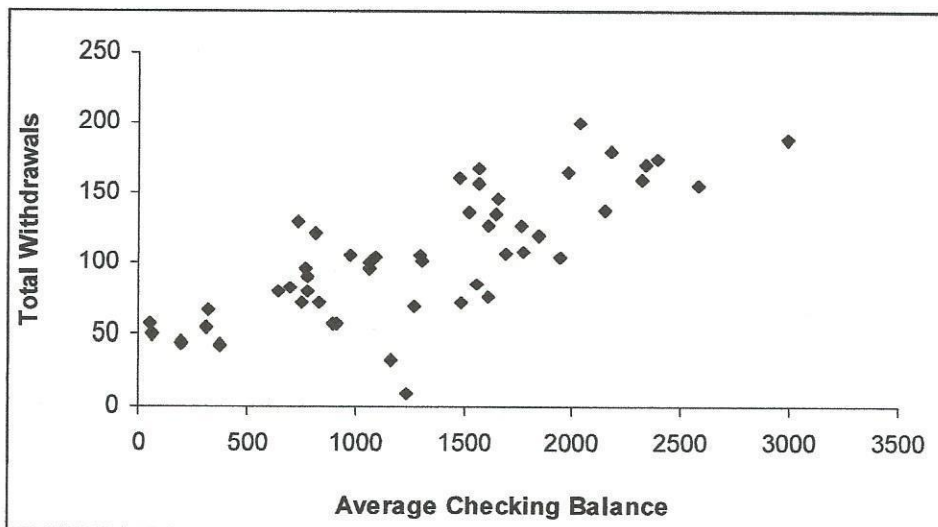
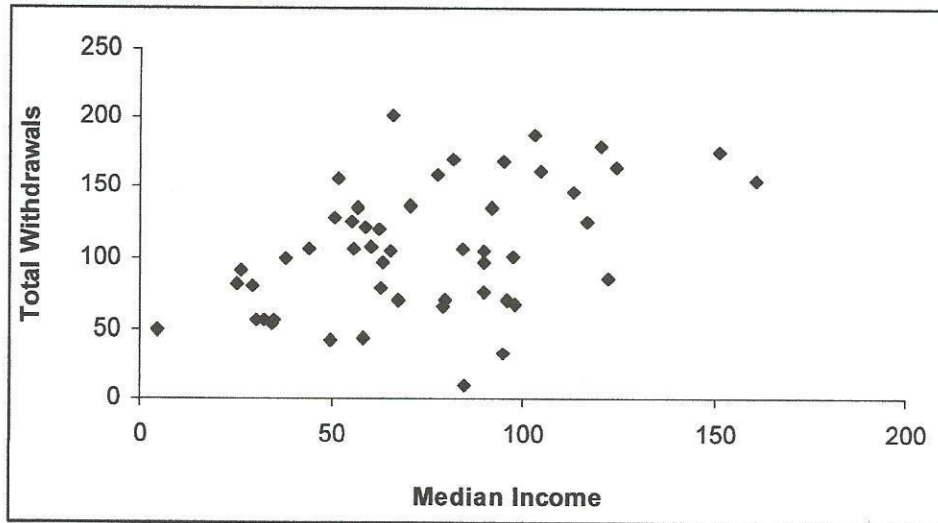
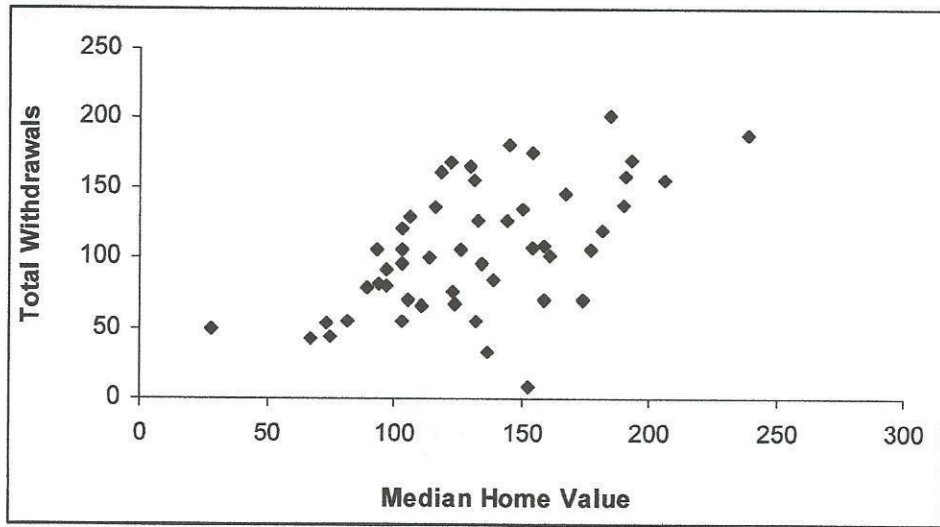
Table 1. Cash Withdrawal data

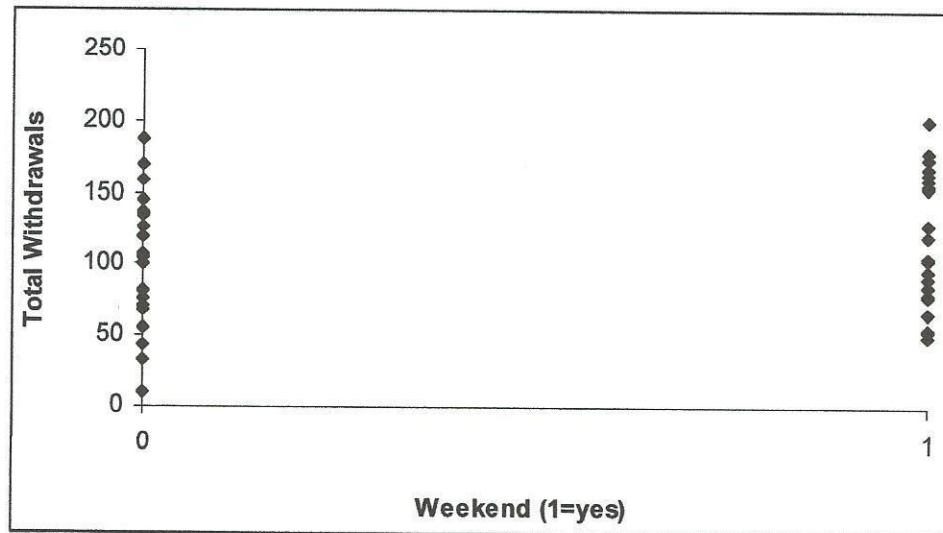
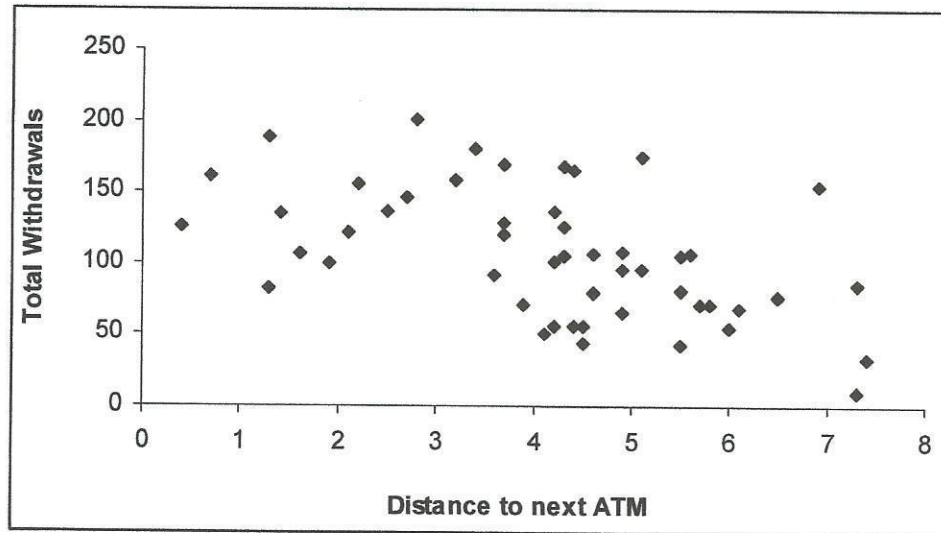
Total Daily Withdrawals (£1,000s)	Median Home Value (£1,000s)	Median Income (£1,000s)	Average Checking Balance (£1,000s)	Distance to Next ATM (miles)	Weekend
107	154	55.9	1690	4.6	No
56	82	30.6	891	4.4	No
50	28	4.5	60	4.1	Yes
135	150	56.7	1651	1.4	No
56	103	32.5	914	4.5	No
?	?	?	?	?	?
33	137	94.7	1164	7.4	No
105	103	90.1	1093	4.3	Yes
68	124	98	1269	6.1	No
126	133	116.6	1762	4.3	No
71	174	95.8	1483	5.8	No

Table 2 contains the correlation coefficients between the different variables, where a dummy variable is used to indicate whether the withdrawal occurred on a weekend (1) or not (0). Also scatter plots between all the independent variables and total withdrawals are given.

Table 2. Correlation matrix

	Withdrawals	Med. Home Value	Med. Income	Avg. Checking Balance	Distance Next ATM	Weekend
Withdrawals	1.00					
Med. Home Value	0.56	1.00				
Med. Income	0.45	0.52	1.00			
Avg. Checking Balance	0.76	0.84	0.63	1.00		
Distance Next ATM	-0.51	-0.06	0.26	-0.15	1.00	
Weekend	0.21	-0.30	0.09	-0.20	0.09	1.00





a) Explain in practical terms the meaning of the correlation coefficient between the 'Distance to Next ATM' variable and the total withdrawals.

A multiple regression analysis was performed, with the results given in Table 3.

Table 3. Multiple Regression Results

Regression Statistics						
Multiple R	0.95					
R Square	0.91					
Adjusted R Square	0.89					
Standard Error	14.72					
Observations	50					

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	69.527	10.118	6.87	0.00	49.135	89.919
Median Home Value	0.004	0.100	0.04	0.97	-0.198	0.206
Median Income	0.105	0.096	1.09	0.28	-0.089	0.298
Average Checking Balance	0.048	0.007	7.08	0.00	0.034	0.062
Distance to Next ATM	-12.076	1.400	-8.63	0.00	-14.898	-9.254
Dummy Weekend	35.163	4.631	7.59	0.00	25.829	44.497

b) Do the results suggest that there is a relationship between the median family income in the neighbourhood and total ATM withdrawals in that neighbourhood?

Based on the results in Table 3, a new regression model was run. The results are given in Table 4.

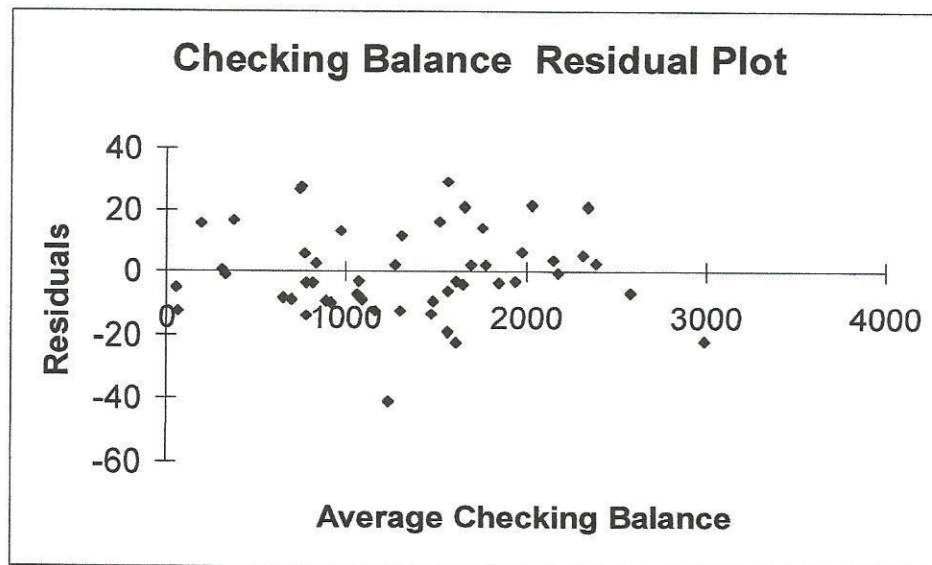
Table 4. Multiple Regression Results – Revised Model

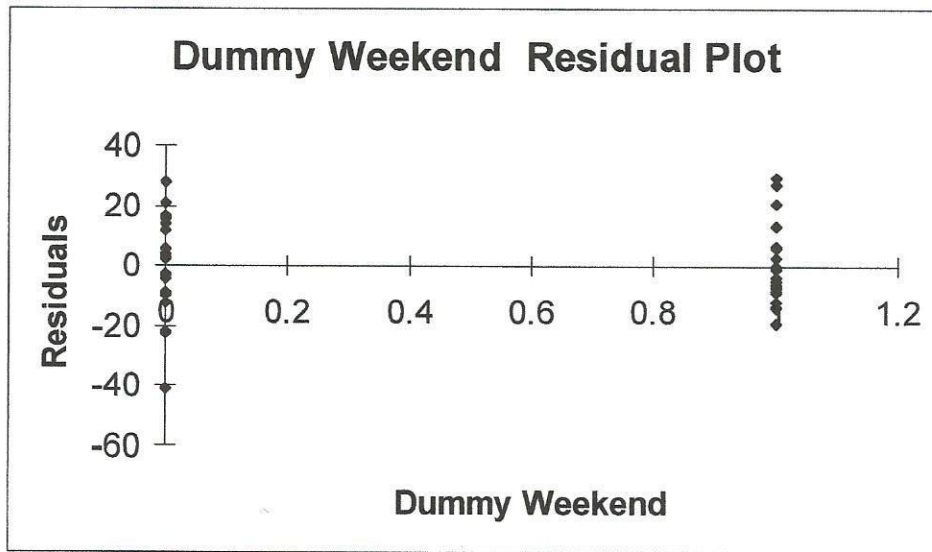
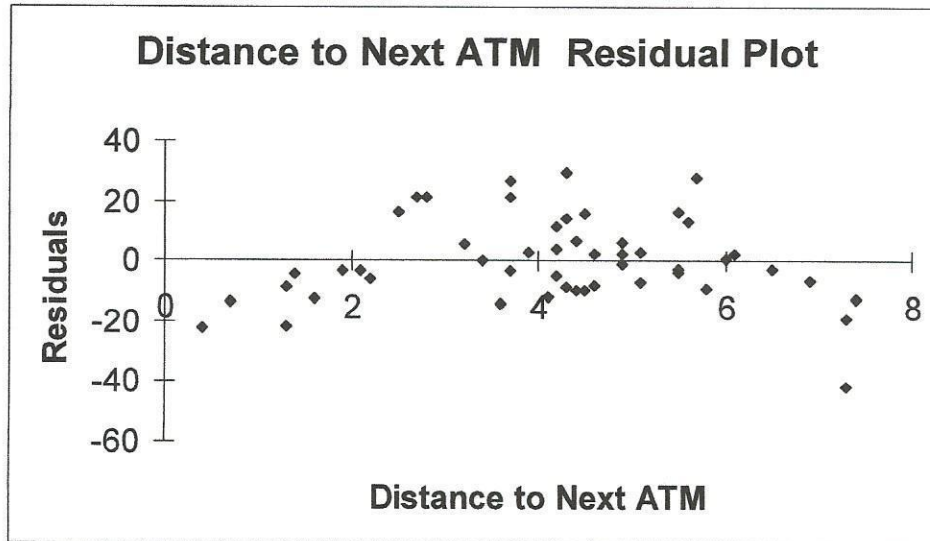
<i>Regression Statistics</i>						
Multiple R	0.95					
R Square	0.90					
Adjusted R Square	0.90					
Standard Error	14.59					
Observations	50					

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	69.171	7.646	9.05	0.00	53.780	84.561
Average Checking Balance	0.052	0.003	16.50	0.00	0.046	0.058
Distance to Next ATM	-11.380	1.223	-9.31	0.00	-13.841	-8.919
Dummy Weekend	36.547	4.252	8.59	0.00	27.988	45.106

c) For forecasting purposes, would you prefer the first model (with the results in Table 3), or the second model (with the results in Table 4)? Why?

Below, the three residual (error) plots for the revised model are given.





- d) Do any of the residual plots highlight any problems with the regression analysis? If yes, interpret them and discuss how these could be resolved.
- e) Provide a forecast for the total withdrawals on a weekday from an ATM located 5 miles from the nearest other ATM, in a neighbourhood where the average checking balance of the customers living in that neighbourhood is 1,000, the median value of the homes is 150 and the median family income is 100.