Solutions to Homework 1

Solutions to Multiple Choice Questions

1. **D** – The definition of chance and assignable causes of variation are given in Section 1.4, Page 19. We can consider both seasonal change and holidays as predictable factors that each individually has an effect on the variable of interest (in this case, sales in the club). On the other hand, week to week changes of deviations from seasonal patterns are unpredictable in value and individually not very influential on the overall value.
2. **A** – The answer is obvious in view of a process view graph. See Section 1.1, Pages 3-6.
3. **B** – The goal is to identify any dependence on beer sales on the amount of rain. From Section 1.5, Pages 26-30 we know that scatterplots are very convenient graphical tools for detecting dependencies between two variables.
4. **E** – The definition of assignable causes of variation is given in Section 1.4, Page 19.
5. **A** – The normal model is discussed in Section 1.4, Pages 20-22. This model is characterized by the center expressed as the mean, and the spread expressed in terms of the standard deviation.
6. **B** – Regression models have two types of variable: explanatory and response. In the regression model, it is natural to assume that there is a cause-and-effect relationship between explanatory and response variables.
7. **D** – The mathematical formulation of the simple regression model is:

   $$Y = \alpha + \beta X + \varepsilon$$

X	=	Explanatory Variable
Y	=	Response Variable
$\alpha + \beta X$	=	Linear relationship in *X*
ε	=	Chance variation

8. **D** – The mathematical formulation of the prediction formula is conventionally:

   $$\hat{Y} = \alpha + \beta X \pm 2\sigma$$

X	=	Explanatory Variable
Y	=	Response Variable
$\alpha + \beta X$	=	Linear relationship in *X*
σ	=	Standard Deviation of Chance variation

9. **A** – Read the subsection *Interpreting Regression Coefficients*, Pages 43-44.
10. **C** – In the regression model, the value being predicted is the response variable. The response variable is dependent on explanatory variables in the cause-and-effect relationship.

Solutions to Problems

Answer to Question 1

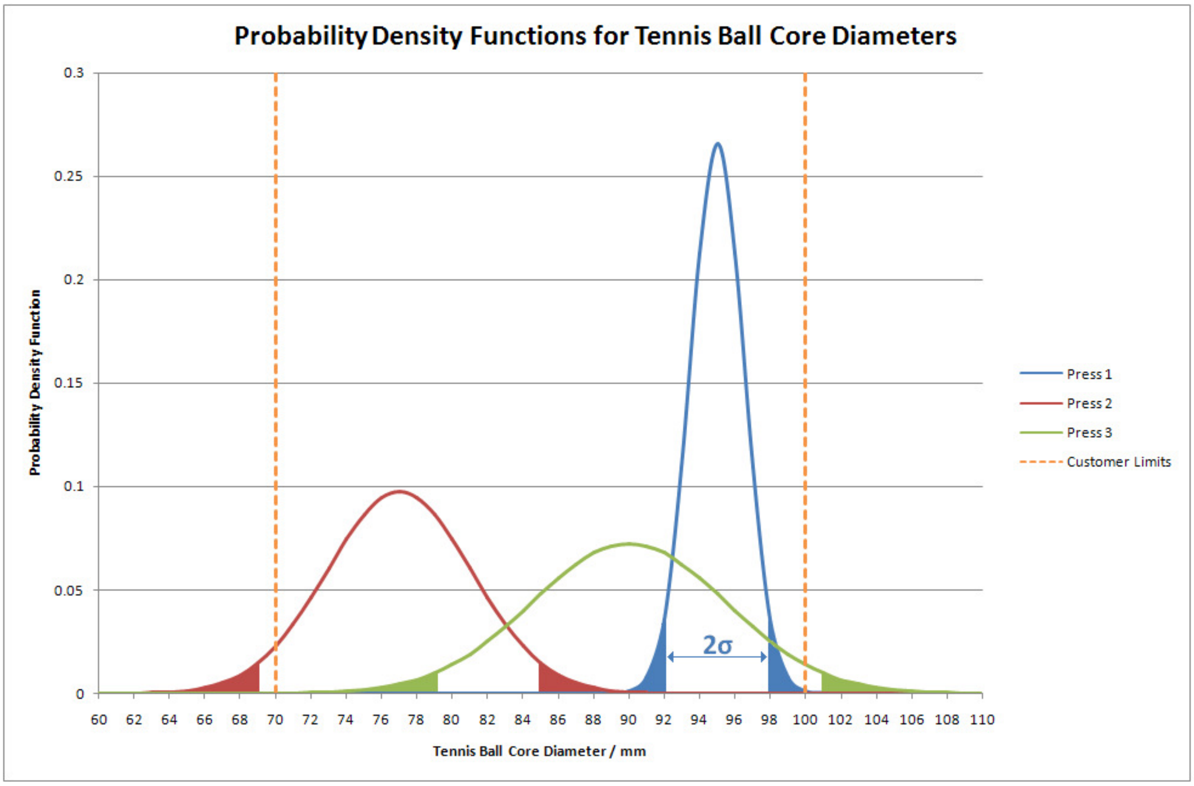
For this problem the 2σ rule set by the manufacturer means that each press ought to have a band of $\mu \pm 2\sigma$[1] for tennis ball core diameters, that lies within the customer band.

Press No.	Mean (μ)	Standard Deviation (σ)	Manufacturer 2σ Band	Customer Limits	Pass
1	95	1.5	(92, 98)	(70, 100)	yes
2	77	4.1	(68.8, 85.2)	(70, 100)	no
3	90	5.5	(79, 101)	(70, 100)	no

- The lower limit for the first press using the two sigma rule is 95 – 2(1.5) = 92
- The upper limit for the first press using the two sigma rule is 95 + 2(1.5) = 98
- The lower limit for the second press using the two sigma rule is 77 – 2(4.1) = 68.8
- The upper limit for the second press using the two sigma rule is 77 + 2(4.1) = 85.2
- The lower limit for the third press using the two sigma rule is 90 – 2(5.5) = 79
- The upper limit for the third press using the two sigma rule is 90 + 2(5.5) = 101

Only the first press passes the quality inspection. This is because the 2σ band for the tennis balls core diameters lies within the specified customer limits. If the manufacturer were to put a more stringent 3σ rule on the first press, a wider band of $\mu \pm 3\sigma$ for tennis ball core diameters ought to still lie within the customer band. The limits for the first press in this case would expand to (90.5, 99.5), which is still within the customer specified quality band of (70, 100).

[1] The 2σ band set by the manufacturer implies that 95.45% of tennis balls core diameters produced will fall within this range. This means that 4.55% of tennis balls will be outside the manufacturer's limits, and therefore possibly outside the customer's limits.



In the above graph we can clearly see that if we set a 2σ quality level on the presses, then only the first press is acceptable, since the band for this lies within the customer band, shown in red. It is not shown in the graph, but when we expand the limits to 3σ for the first press the band will still lie within the customer limits.

NOTE: The following section is not part of the answer to the question; it is merely an interesting observation on the flawed methodology of rating the presses.

From the above graph, if we were to integrate the area under any of the three curves between its 2σ limits, we would obtain 95.45% of the total area under the curve between the limits of plus and minus infinity. This is apparent from inspection since the 'tails' of the curve, outside the 2σ limits only contribute a small part to the total area under the curve, 4.55% in fact.

It is worth noting however, that if one of the presses produces tennis balls which are outside the 2σ limits, it may still be within the customer limits. This means that the tennis ball wouldn't be rejected by the manufacturer. This is clearly evident in the graph. So a better way of judging the presses would be to determine the actual quality level of tennis balls produced by each press (the inverse of the quality level for each press is known as the non-conformance rate). This means determining the area under each curve between the customer limits of 70 and 100. The results are shown in table format on the following page.

Press No.	Quality Level[2]	Defects per 10,000 Tennis Balls	Pass
1	99.9571%	4	yes
2	95.6118%	439	no
3	96.5344%	347	no
4 (μ = 85, σ = 7.4)	95.4500%	455	yes
5 (μ = 90, σ = 5.1)	97.5004%	250	no

If we had a fourth press with mean 85 and a standard deviation of 7.5, and if we set a 2σ limit we would have manufacturer limits of (70, 100) which just correspond exactly with the customer limits. This press would pass the quality inspection, but the quality level would be 95.4500% which translates to 455 defects per 100,000 tennis balls produced. This means the fourth press would be worse in terms of defects produced than presses 2 and 3 but will still pass the quality inspection! Now if we were to compare this with a fifth press with mean 90 and a standard deviation of 5.1, and if we set a 2σ limit we would have manufacturer limits of (79.8, 100.2) which are outside the customer limits. This press would fail the quality inspection, but the quality level would be 97.5004% which translates to 250 defects per 100,000 tennis balls produced. The fifth press has a 2.0504% quality improvement over the fourth press, but the fifth press will fail the quality inspection while the fourth press will pass! Another and more striking way of comparing the presses is the fact that the fifth press produces 45% less defective products than the fourth press.

So it should be remembered that when a manufacturer sets their limits they should be done with respect to the customer limits. In this question, tennis balls that have core diameters outside the manufacturer's limits, may still be within customer limits. Another reason for the incongruous results mentioned in the previous paragraph is due to the fact the manufacturing mean does not correspond to the customer mean. For the first press any tennis balls with core diameters inside the range of -16σ and +3σ are all within customer limits.

[2] In later tutorials we will see how to calculate the quality level and non-conformance rates for a normal curve.

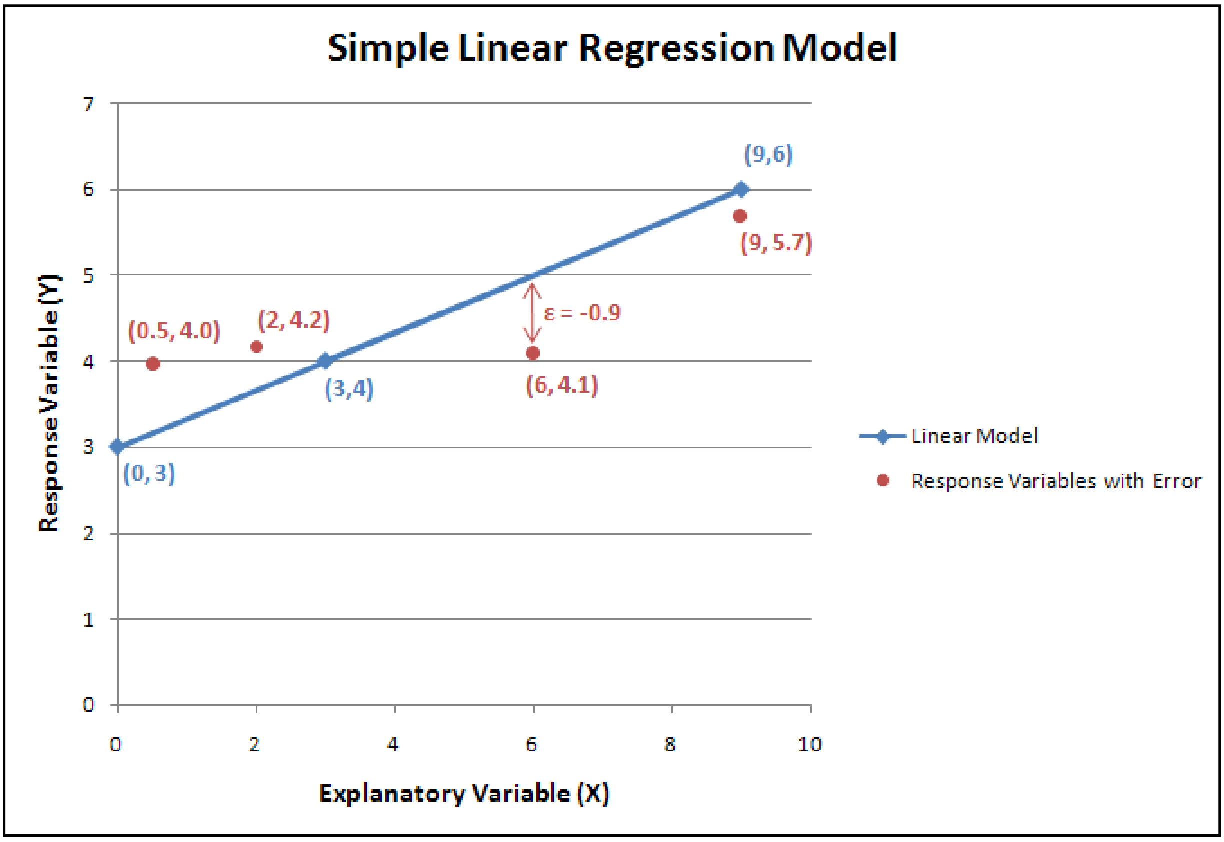
Answer to Question 2

The line generated from the equation $y = 3 + \frac{x}{3}$ passes through the three points (3,4), (0,3), (9,6). Therefore, we could actually determine the simple linear model from either the equation, or just two specified points. This is something to be aware of for an exam equation. We determine the response variables for four different explanatory variables, each of which has as an associated error term using the equation:

$$Y = 3 + \frac{X}{3} + \epsilon$$

The explanatory variables are X = 0.5, 2, 9, 6 and the error terms associated with each explanatory variable are ϵ = 0.8, 0.5, -0.3, -0.9. We determine the response variable for each X value using the equation and the associated error term for that X value. Looking at the below graph for each red point, the vertical distance from the red point to the blue line corresponds to the error value at that point.

X Value	ε Value	Corresponding Y Value
0.5	-0.8	4.0
2	0.5	4.2
9	-0.3	5.7
6	-0.9	4.1



Answer to Question 3

We firstly need to determine the predicted man-hours needed for period 6, 1963. Consulting the table in the book we see that the volume of mail at this time was 180 (millions of pieces of mail). According to the prediction formula the amount of man-hours (in thousands) for this period is given by:

$$\hat{Y}_{6,1963} = 50 + (3.3 * 180) \pm 20$$

$$\hat{Y}_{6,1963} = 644 \pm 20$$

The actual value observed during the period was 765. This value is well outside the 2σ band. The loss in terms of thousand man-hours is estimated as (765-644) ± 20 = 121 ± 20, i.e. between 101 and 141.

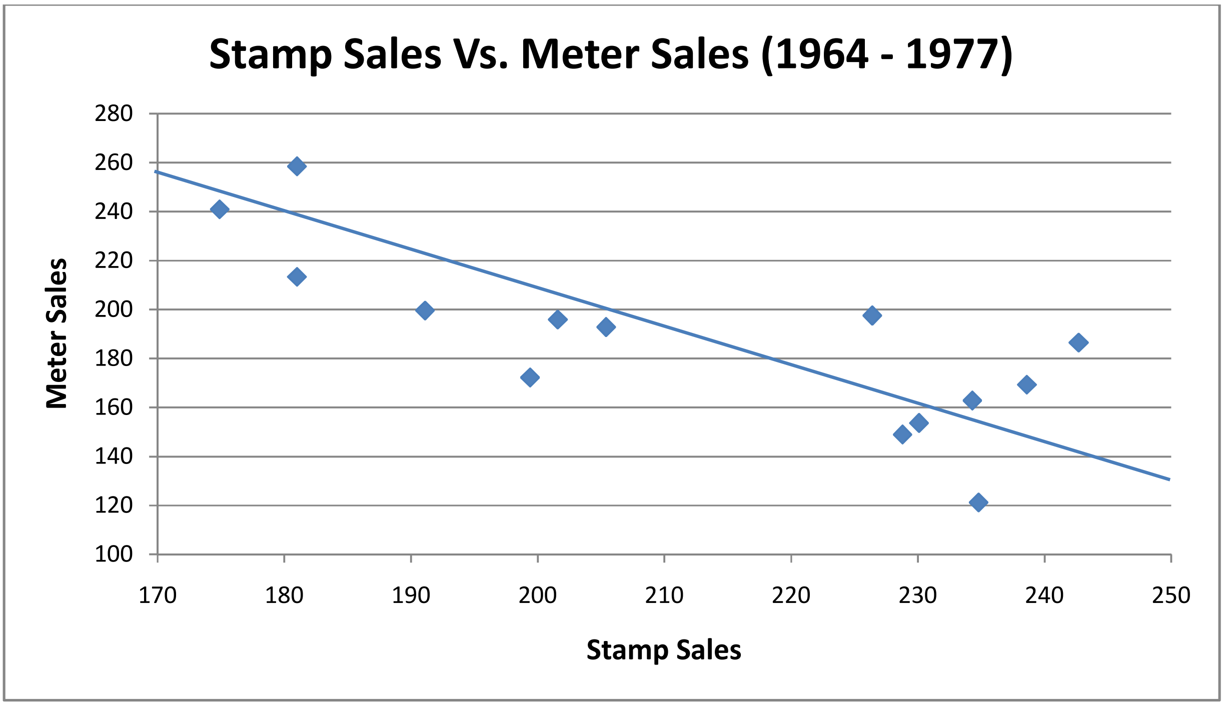
During the Christmas periods we observe the volumes of mail as 268 in 1962, and 270 in 1963. The volume prognosis for the Christmas periods are therefore, using the above prediction formula:

$$\hat{Y}_{7,1962} = 50 + (3.3 * 268) \pm 20 \qquad \hat{Y}_{7,1963} = 50 + (3.3 * 270) \pm 20$$

$$\hat{Y}_{7,1962} = 934.4 \pm 20 \qquad \hat{Y}_{7,1963} = 941 \pm 20$$

However the observed values are 1053 and 1070. Thus the extra requirement in terms of man-hours for the first Christmas period is estimated as (1053-934.4) ± 20 = 118.6 ± 20. The extra requirement in terms of man-hours for the second Christmas period is estimated as (1070-941) ± 20 = 129 ± 20. Thus it would be safe to say that the extra man power needed for a Christmas period will be roughly between the lowest lower limit and the highest higher limit of the extra man hour requirements for the Christmas periods of 1962 and 1963. The lowest lower limit will be from the first Christmas period, since 118.6 is less than 129. So the lowest lower limit is 118.6 – 20 = 96.6. The highest higher limit will be from the second Christmas period, since 129 is greater than 118.6. So the highest higher limit is 129 + 20 = 149. Therefore the extra man power needed is estimated to be between 96.6 an 149 thousand man-hours.

Answer to Question 4



There appears to be a relationship between stamp sales and meter sales. The equation for the straight line in the above graph will be given by:

$Y = a + bX$

Two points that lie roughly on the line are (250, 130) and (170, 255) therefore:

$130 = a + 250b$....................................... Using the first point

$a = 130 - 250b$ ....We now have an expression for the constant *a* in terms of *b*

$255 = a + 170b$....................................... Using the second point

$255 = (130 - 250b) + 170b$................Substituting in for *a* which we now have in terms of *b*

$255 = -80b + 130$

$b = -1.5625$.. We now know what *b* is

$a = 130 - 250(-1.5625)$.......................Since we know *b* we can determine *a*

$a = 520$

Therefore:

$Y = 520 - 1.5625X$

We evaluate prediction errors based on assuming that the stamp sales is the explanatory variable and meter sales is the response variable. We use the linear function we obtained to predict the meter sales and then we get the prediction error which is the difference between the predicted and actual values for meter sales.

Year	Stamp Sales[3]	Actual Meter Sales[2]	Predicted Meter Sales[2]	Prediction Error
1964	234.8	121.3	153.125	31.825
1965	228.8	149	162.5	13.5
1966	230.1	153.7	160.4688	6.76875
1967	234.3	162.8	153.9063	-8.89375
1968	238.6	169.3	147.1875	-22.1125
1969	242.7	186.5	140.7813	-45.7188
1970	226.4	197.5	166.25	-31.25
1971	199.4	172.2	208.4375	36.2375
1972	205.4	192.8	199.0625	6.2625
1973	201.6	195.9	205	9.1
1974	191.1	199.6	221.4063	21.80625
1975	181	213.3	237.1875	23.8875
1976	174.9	240.9	246.7188	5.81875
1977	181	258.4	237.1875	-21.2125

[3] Sales are recorded as millions of standard stamp equivalents, that is, total revenue in a year divided by the price of a stamp for a standard sealed letter for internal delivery and divided by 1,000,000.

Answer to Question 5

We want to predict stamp sales for the year 1985 using the equation:

$$Y_{1985} = 340 - 0.0316X_1 - 70.2X_2 \pm 8$$

X_1 = GNP (Gross National Product) for 1985
X_2 = RLP (Real Letter Price) for 1985

In order to do this we need to work out the GNP and RLP values for 1985. We are told that in 1985 the GNP is expected to increase by 1.5%. Since we know the GNP in 1984 is 1484.5 the predicted GNP for 1985 will be simply:

$$GNP_{1985} = 1484.5 + (1484.5 * 0.015) = 1506.767$$

Next we need to obtain the RLP value for 1985. We are told that the RLP is given by the SSV (Standard Stamp Value) divided by the CPI (Consumer Price Index):

$$RLP = \frac{SSV}{CPI}$$

In 1984 the RLP is 1.835

$$1.835 = \frac{SSV}{CPI_{1984}}$$

We assume that the SSV won't change in 1985:

$$SSV = 1.835 * CPI_{1984}$$

The RLP for 1985 is given by:

$$RLP_{1985} = \frac{SSV}{CPI_{1985}}$$

$$RLP_{1985} = \frac{1.835 * CPI_{1984}}{CPI_{1985}}$$

We also know that the CPI in 1985 will be 5.5% higher than the CPI in 1984

$$CPI_{1985} = CPI_{1984} + (0.055 * CPI_{1984})$$

$$CPI_{1985} = (1.055 * CPI_{1984}$$

$$RLP_{1985} = \frac{1.835 * CPI_{1984}}{1.055 * CPI_{1984}}$$

$$RLP_{1985} = \frac{1.835}{1.055}$$

$RLP_{1985} = 1.739336$

Now that we have the predicted GNP and RLP for 1985 we plug these values into the prediction formula for stamp sales, where X_1 denotes GNP and X_2 denotes RLP for 1985.

$Y_{1985} = 340 - 0.0316X_1 - 70.2X_2 \pm 8$

$Y_{1985} = 340 - (0.0316 * 1506.767) - (70.2 * 1.739336) \pm 8$

$Y_{1985} = 170.2848 \pm 8$ Million Standard Stamp Equivalents

Answer to Question 6

Firstly we will look at a 5% increase in stamp prices. Since we know that the RLP is proportional to the SSV, a 5% increase in stamp prices will result in a 5% increase in RLP in 1984.

$$RLP = \frac{SSV}{CPI}$$

In 1984 the RLP is 1.835 this will increase by 5%:

$$RLP_{1984,+5\%} = 1.835 * 1.05 = 1.92675$$

The GNP will remain unaffected by the increase in stamp price; therefore the predicted stamp sales for 1984 will be given by:

$$\hat{Y}_{1984,+5\%} = 340 - (0.0316 * 1484.5) - (70.2 * 1.92675) \pm 8$$

$$\hat{Y}_{1984,+5\%} = 157.8319 \pm 8$$

In 1985, to calculate the RLP, we must use the new higher RLP from 1984 that has increased due to the 5% increase in stamp prices. We also assume an inflation rate of 5.5% as given in Question 5. The GNP will increase by 1.5% as outlined in question 5. Again though, the increase in stamp price will not affect the GNP in 1985.

$$RLP_{1985,+5\%} = \frac{1.92675}{1.055} \ldots\ldots\ldots\ldots \textit{See the equation at the very end of page } 9$$

$$RLP_{1985,+5\%} = 1.826303$$

$$\hat{Y}_{1985,+5\%} = 340 - (0.0316 * 1506.767) - (70.2 * 1.826303) \pm 8$$

$\hat{Y}_{1985,+5\%} = 164.1797 \pm 8$ Million Standard Stamp Equivalents

Secondly we will look at a 10% increase in stamp prices. Since we know that the RLP is proportional to the SSV, a 10% increase in stamp prices will result in a 10% increase in RLP in 1984.

$$RLP = \frac{SSV}{CPI}$$

In 1984 the RLP is 1.835 this will increase by 10%:

$$RLP_{1984,+10\%} = 1.835 * 1.1 = 2.0185$$

The GNP will remain unaffected by the increase in stamp price; therefore the predicted sales for 1984 will be given by:

$$\hat{Y}_{1984,+10\%} = 340 - (0.0316 * 1484.5) - (70.2 * 2.0185) \pm 8$$

$$\hat{Y}_{1984,+10\%} = 151.3911 \pm 8$$

In 1985, to calculate the RLP, we must use the new higher RLP from 1984 that has increased due to the 10% increase in stamp prices. We also assume an inflation rate of 5.5% as given in question 5. The GNP will increase by 1.5% as outlined in question 5. Again though, the increase in stamp price will not affect the GNP in 1985.

$$RLP_{1985,+10\%} = \frac{2.0185}{1.055} \ldots\ldots\ldots\ldots \mathit{See\ the\ equation\ at\ the\ very\ end\ of\ page}\ 9$$

$$RLP_{1985,+10\%} = 1.91327$$

$$\hat{Y}_{1985,+10\%} = 340 - (0.0316 * 1506.767) - (70.2 * 1.91327) \pm 8$$

$\hat{Y}_{1985,+10\%} = 158.0746 \pm 8$ Million Standard Stamp Equivalents

Finally we will look at a 5% decrease in stamp prices. Since we know that the RLP is proportional to the SSV, a 5% decrease in stamp prices will result in a 4% decrease in RLP in 1984.

$$RLP = \frac{SSV}{CPI}$$

In 1984 the RLP is 1.835 this will decrease by 5%:

$$RLP_{1984,-5\%} = 1.835 * 0.95 = 1.74325$$

The GNP will remain unaffected by the decrease in stamp price; therefore the predicted sales for 1984 will be given by:

$$\hat{Y}_{1984,-5\%} = 340 - (0.0316 * 1484.5) - (70.2 * 1.74325) \pm 8$$

$$\hat{Y}_{1984,-5\%} = 170.7136 \pm 8$$

In 1985, to calculate the RLP, we must use the new lower RLP from 1984 that has decreased due to the 5% decrease in stamp prices. We also assume an inflation rate of 5.5% as given in question 5. The GNP will increase by 1.5% as outlined in question 5. Again though, the decrease in stamp price will not affect the GNP in 1985.

$$RLP_{1985,-5\%} = \frac{1.74325}{1.055} \dots \dots \dots \dots See\ the\ quation\ at\ the\ very\ end\ of\ page\ 9$$

$$RLP_{1985,-5\%} = 1.652370$$

$$\hat{Y}_{1985,-5\%} = 340 - (0.0316 * 1506.767) - (70.2 * 1.652370) \pm 8$$

$\hat{Y}_{1985,-5\%} = 176.3898 \pm 8$ Million Standard Stamp Equivalents

Answer to Question 7

We use the general equation to predict the number of housing completions per quarter for the years 1993 – 2000:

$$Y = 3248Q1 + 3901Q2 + 4174Q3 + 5031Q4 + 250t \pm 500$$

When we are making predictions for the first quarter of any year, Q1 will take on a value of one, and Q2, Q3, and Q4 will be zero. Similar logic applies when we are making predictions for the other quarters. Therefore the predicted housing completions for each of the four quarters are given by:

$$Y_{Q1} = 3248 + 250t \pm 500$$

$$Y_{Q2} = 3901 + 250t \pm 500$$

$$Y_{Q3} = 4174 + 250t \pm 500$$

$$Y_{Q4} = 5031 + 250t \pm 500$$

To predict housing completions for 2001 and 2002 we will start with *t = 33* (see page 49, Table 1.7) since *t = 32* corresponds to the last quarter in the year 2000. We then use the appropriate formula depending on what quarter we are in.

$Y_{2001,Q1} = 3248 + (250 * 33) \pm 500 = 11{,}498 \pm 500$ Housing Completions

$Y_{2001,Q2} = 3901 + (250 * 34) \pm 500 = 12{,}401 \pm 500$ Housing Completions

$Y_{2001,Q3} = 4174 + (250 * 35) \pm 500 = 12{,}924 \pm 500$ Housing Completions

$Y_{2001,Q4} = 5031 + (250 * 36) \pm 500 = 14{,}031 \pm 500$ Housing Completions

$Y_{2002,Q1} = 3248 + (250 * 37) \pm 500 = 12{,}498 \pm 500$ Housing Completions

$Y_{2002,Q2} = 3901 + (250 * 38) \pm 500 = 13{,}401 \pm 500$ Housing Completions

$Y_{2002,Q3} = 4174 + (250 * 39) \pm 500 = 13{,}924 \pm 500$ Housing Completions

$Y_{2002,Q1} = 5031 + (250 * 40) \pm 500 = 15{,}031 \pm 500$ Housing Completions

Due to the collapse of the construction bubble in Ireland, the construction business is in deep recession. Thus the linear growth model is no longer valid. Also the linear model is only justified for certain periods and for different periods different coefficients should be considered. Applying such a linear model today in 2009, which only applied to construction trends from 9 years ago, is simply not prudent.