ANSWERS CHAPTER 11

THINK IT OVER

think it over

TIO 11.1:

$E(y|football) = b_0 + b_1x_1$

$E(y|cricket) = b_0 + b_2 + b_1x_1$

$E(y|rugby) = b_0 + b_3 + b_1x_1$

b_0

$b_0 + b_2$

$b_0 + b_3$
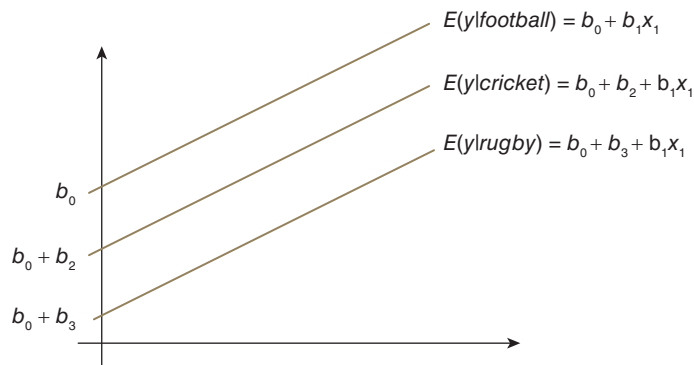
Since the gradients are the same, b_1x_1, the lines are parallel and only differ where they cross the vertical axis, i.e. the intercept.

TIO 11.2: The football weeks are the 'baseline' and all the other results are compared to them.

TIO 11.3: This would imply that $e^{-(b_0\cdots)} = 0$ which is not possible. In other words check your values or your arithmetic!

TIO 11.4: 0.5, since $e^0 = 1$

TIO 11.5: In probability theory the minimum value is 0, i.e. cannot occur and the maximum value is 1, definitely occurs. The logistic equation, because of the exponential term, can never reach 0 or 1 (only at infinity). If you use Excel to plot a graph of –10 to +10 in steps of 0.3 and adjust the output to 15 decimal places, you will see something similar to the 'S' shaped curve below.
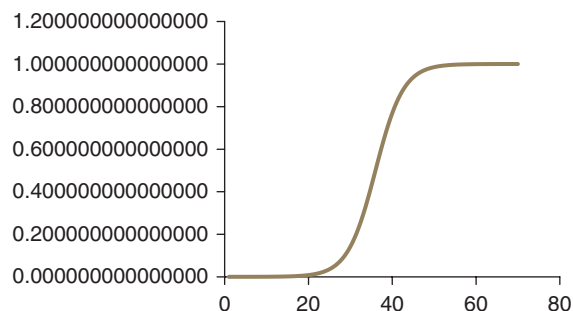
TIO 11.6: A linear relationship typically produces a straight-line graph and follows the rule $y = mx + c$. A non-linear relationship produces a non-straight-line graph, i.e. a nice curvy one!

EXERCISES

1.   (a)

Classification Table[a,b]

			Predicted		
			Sale		Percentage
	Observed		0	1	Correct
Step 0	Sale	0	27	0	100.0
		1	23	0	.0
	Overall Percentage				54.0

a. Constant is included in the model.

b. The cut value is .500

Initial model, i.e. no predictors included. This model will be based on not making a sale, since this was the most frequent outcome (from the classification table: 27 no sale, 23 sale). Accordingly, the model correctly classifies 54% of cases.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.160	.284	.319	1	.572	.852

b_0 = -0.16

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Email(1)	.144	1	.704
		Telephone(1)	2.013	1	.156
		Telandemail(1)	.593	1	.441
	Overall Statistics		3.357	3	.340

Overall Statistics (equivalent to the chi-squared statistic) is reported as 3.357 with a significance of 0.34. This level of significance says that the excluded variables would not make a significant difference to the predictive power of the model.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3.455	3	.327
	Block	3.455	3	.327
	Model	3.455	3	.327

The chi-squared value confirms that the new model, which includes the predictors, is not much better than the baseline model at predicting sales.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	65.539[a]	.067	.089

a. Estimation terminated at iteration number 3 because

parameter estimates changed by less than .001.

Classification Table[a]

		Predicted		
		Sale		Percentage
	Observed	0	1	Correct
Step 1	Sale 0	21	6	77.8
	1	12	11	47.8
	Overall Percentage			64.0

a. The cut value is .500

The overall percentage says the new model is a slight improvement on the baseline (64% correct classifications compared to 54%).

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Email(1)	-.268	.608	.194	1	.660	.765	.232	2.520
	Telephone(1)	-1.006	.617	2.659	1	.103	.366	.109	1.225
	Telandemail(1)	.646	.621	1.080	1	.299	1.908	.564	6.447
	Constant	.171	.547	.098	1	.755	1.186		

a. Variable(s) entered on step 1: Email, Telephone, Telandemail.

The Wald statistic tells us that follow up by email is not a significant predictor of a sale. A similar interpretation applies to the other predictors as well. It looks as if follow up by telephone had the most impact, but not significantly so.

Exp(B) tells us that the odds of getting a sale decreases if email alone is used more often (Exp(B) < 1). If email and telephone follow up is used then the odds of getting a sale improve (Exp(B) > 1).

The 95% confidence interval tells us that we can be fairly confident that the population value of exp *b* will lie somewhere between 0.232 and 2.52.

```
Observed Groups and Predicted Probabilities
    16  +
+
        I
I
        I
I
F       I
I
R   12     +
+
E       I
I
Q       I                            1
I
U       I                            1  1
I
E   8  +                             1  1     1
+
N   I  0                             1  1
I
C       I           1    1           0  1     1
I
Y       I           1    1           0  0     1
I
    4  +           0    0           0  0   0  1   1
+
        I           0    0    1     0  0   0  1   1
I
        I           0    0    1     0  0   0  0   1
I
        I           0    0    0     0  0   0  0   1
I
Predicted -----+-----+----+-----+-----+-----+-----+-----+---
--+-----+-----
      Prob:    0    .1   .2    .3    .4    .5    .6    .7
 .8    .9    1
Group:
0000000000000000000000000000000000000000000000000001111111111111111
111111111111111111111111111111111
          Predicted Probability is of Membership for 1
          The Cut Value is .50
          Symbols: 0 - 0
                   1 - 1
          Each Symbol Represents 1 Case.
```

The histogram confirms that the new model is not much better than the baseline since the cases are spread out across the graph and in particular there is clustering around the 0.5 probability level. In other words, the probability of making a sale due to follow up communication, is about 50%. Also, a lot of cases are misclassified, which again confirms the new model is not significantly better than the baseline at predicting sales by implementing follow up communications. A good model would have, in this case all the 1s (indicating a sale) clustered towards the right hand side.

(b) Overall, the statistics tell us that the new model is not significantly better at predicting sales than the baseline. However, it does indicate a slight improvement. So if you were selling high-value items, maybe Range Rover Sports, any extra sales are worth the effort!

2. No answer required.

3. No answer required.