**WEEK 5 Assignment**

**(a) Show that ∂ Pr[respi = 1] ∂agei + ∂ Pr[respi = 0] ∂agei = 0.**

$$\Pr(resp_i = 1) = [1 - \Pr(resp_i = 0)]$$

Therefore:

$$\frac{\partial \Pr(resp_i = 1)}{\partial age_i} = \frac{\partial\, [1 - \Pr(resp_i = 0)]}{\partial age_i} = -\frac{\partial \Pr(resp_i = 0)}{\partial age_i}$$

Hence:

$$\frac{\partial \Pr(resp_i = 1)}{\partial age_i} + \frac{\partial \Pr(resp_i = 0)}{\partial age_i} = 0$$

**(b) Assume that you recode the dependent variable as follows: respnew i =−respi+1. Hence, positive response is now defined to be equal to zero and negative response to be equal to 1. Use the odds ratio to show that this transformation implies that the sign of all parameters change.**

$$Odds\ Ratio = \frac{\Pr(resp_i = 1)}{\Pr(resp_i = 0)} = e^{(\beta_0 + \beta_1 * male_i + \cdots..)}$$

Since $resp_i^* = -resp_i + 1$

It follows: $\Pr(resp_i^* = 1) = \Pr(resp_i = 0)\ and\ \Pr(resp_i^* = 0) = \Pr(resp_i = 1)$

Therefore:

$$Odds\ Ratio^* = (Odds\ Ratio)^{-1} = e^{(-\beta_0 - \beta_1 * male_i - \cdots..)}$$

Since $\frac{1}{\exp(a*x)} = \exp(-a * x)$

**(c) Consider again the odds ratio positive response versus negative response: Pr[respi = 1] Pr[respi = 0] = exp(β0 + β1malei + β2activei + β3agei + β4(agei/10)2 ). During lecture 5.5 you have seen that this odds ratio obtains its maximum value for age equal to 50 years for males as well as females. Suppose now that you want to extend the logit model and allow that this age value is possibly different for males than for females. Discuss how you can extend the logit specification.**

As an idea I would separate the observations in two different sets, Male and Female.
Therefore, I would have two different models:
Response Male (1,0) with explanatory variables Activity|gender=Male and Age|gender=Male.

Response Female (1,0) with explanatory variables Activity|gender=Female and Age|gender=Female.
(in both I would also consider the squared (age/10).
In the case of the Male observations, I get:
Estimated Coefficients:

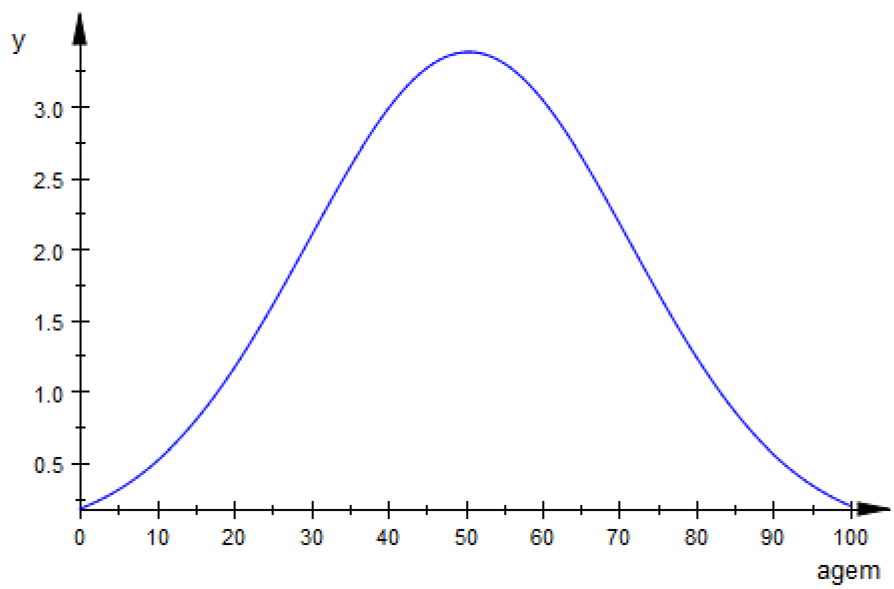
| | Estimate | SE | tStat | pValue |
|---|---|---|---|---|
| (Intercept) | -2.6058 | 1.0952 | -2.3793 | 0.017347 |
| activitym | 0.89559 | 0.21138 | 4.237 | 2.2654e-05 |
| agem | 0.11606 | 0.043919 | 2.6425 | 0.0082298 |
| agemsq | -0.11491 | 0.042437 | -2.7078 | 0.0067729 |

671 observations, 667 error degrees of freedom
Dispersion: 1
Chi^2-statistic vs. constant model: 27.5, p-value = 4.56e-06

From this I get the Odd Ratio associated to the Male dataset as:

$$e^{-0.0011491\,\mathrm{agem}^2+0.11606\,\mathrm{agem}+0.89559\,\mathrm{activitym}-2.6058}$$

For active male customers (activity=1) the relationship odd ratio / age shows again the max ratio at 50 years:




With the odd ratio below 1 only for very young or old customer, which is reasonable since the product here is some kind of investment vehicle.

As far as it concerns the Female data:

Estimated Coefficients:

| | Estimate | SE | tStat | pValue |
|---|---|---|---|---|
| (Intercept) | -0.39407 | 1.471 | -0.26789 | 0.78878 |
| activityf | 1.0846 | 0.38694 | 2.803 | 0.005063 |
| agef | -0.020842 | 0.058784 | -0.35455 | 0.72292 |
| agefsq | 0.020448 | 0.055231 | 0.37022 | 0.71122 |

254 observations, 250 error degrees of freedom

Dispersion: 1

Chi^2-statistic vs. constant model: 7.89, p-value = 0.0483

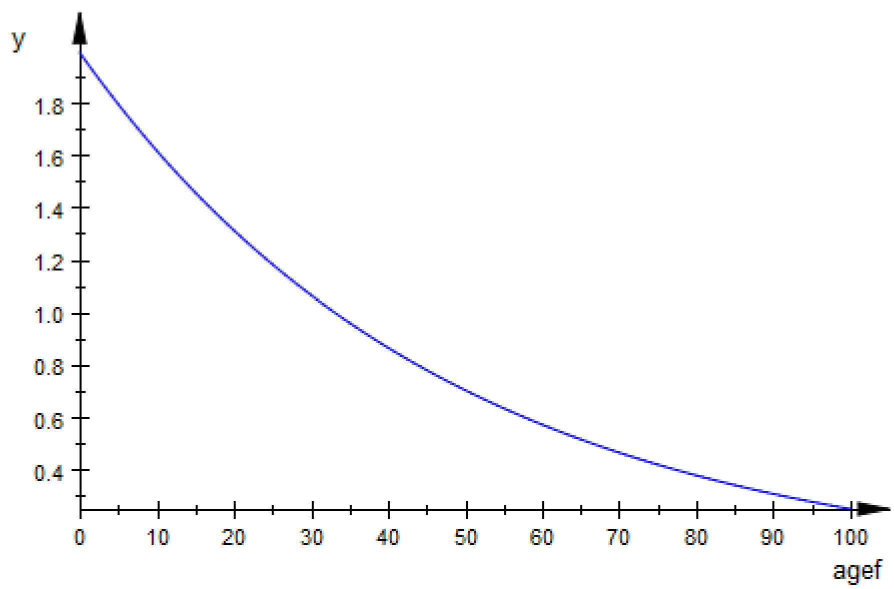
Which is indicating that the age related explanatory variables are not statistically significant with both a p-value well above 5%

In any case, the odd ratio associated to the female data would be:

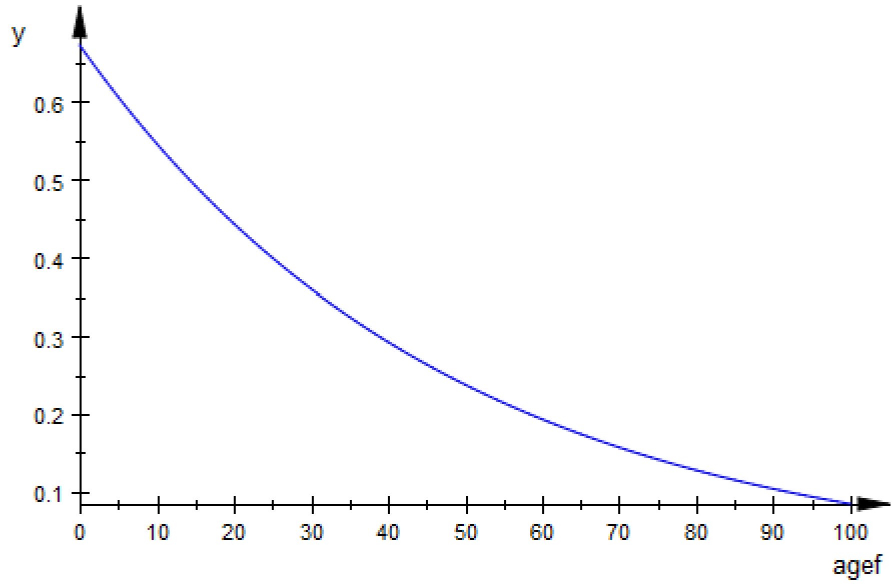
$$e^{0.0002045\ agef^{2} - 0.020842\ agef + 1.0846\ activityf - 0.39407}$$

So, for an active customer female the relationship between the odd ratio and age would indicate that the younger the active female customer, the more likely is to respond positively to the solicitation.

In particular, we see the odd ratio above 1 for ages below around 30 years:



For inactive female customers, we see a similar relationship but the odd ratio is always below 1:




As a conclusion, I would say that dividing the problem into 2 different sets, the male and the female responses, we see that younger female customers tend to respond more positively than older customers.
In any case, as it concerns female customers it appears that age is not a significant variable as suggested by the high p-values of age and age-squared.
Also, their inclusion doesn't add much at all to the explanatory ability of the activity variable on its own: the Chi-sq test is basically the same with or without the age variable.