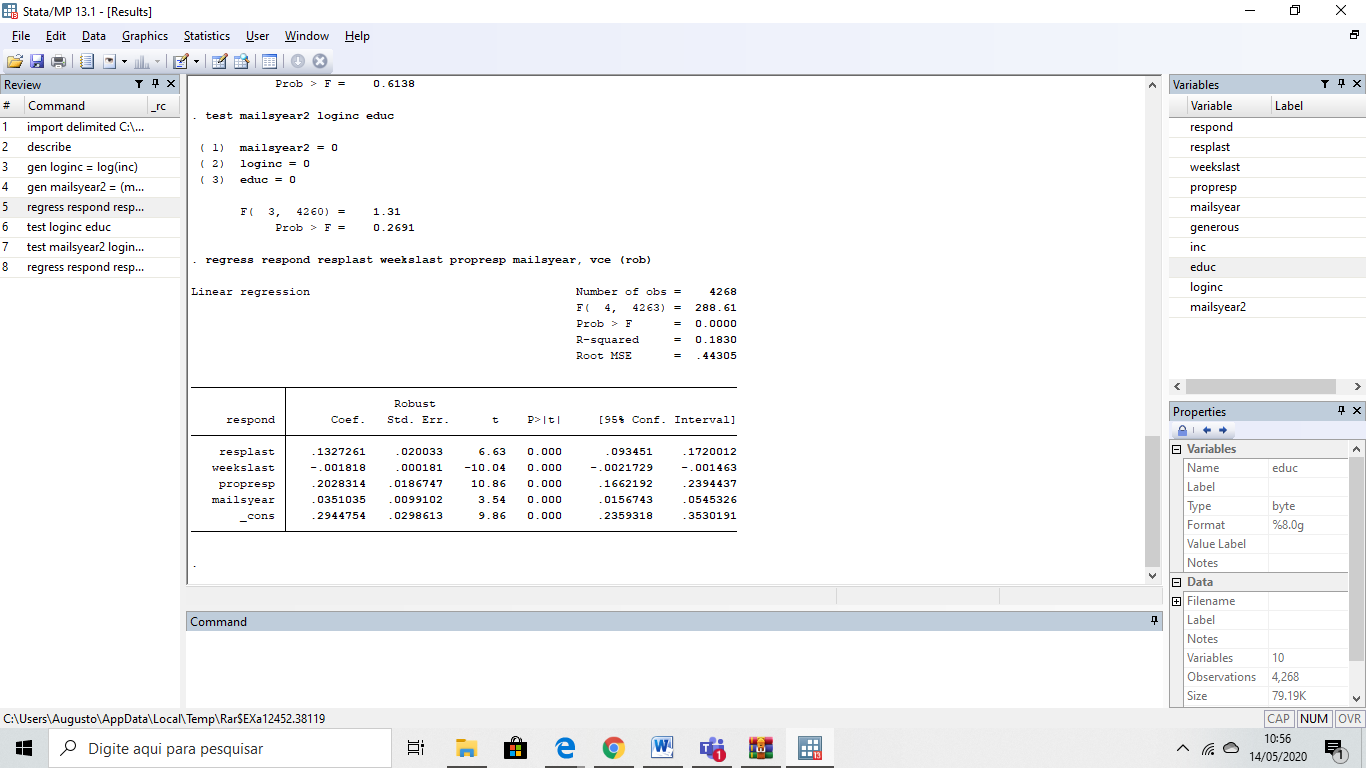
Esse trabalho tem como objetivo explicar a probabilidade de um indivíduo fazer uma doação para uma instituição de solidariedade social. O modelo para explicar tal relação pode ser descrito por:

**P (respond =1| resplast, weekslast, propresp, mailsyear, inc, educ) =**

**Φ (β0 + β1 resplast + β2 weekslast + β3 propresp + β4 mailsyear + β5 mailsyear2 + β6 ln(inc) + β7 educ)**

À luz do modelo de probabilidade linear (LPM) podemos estimar o impacto probabilístico de cada regressor sobre a variável dicotômica dependente **respond**. Obtemos as seguintes estimativas para os parâmetros e suas significâncias (ver anexo 2).

Pode-se perceber que ao nível de 5%, as variáveis **mailsyear2**, **loginc** e **educ** não são significativas individualmente, pelo que as testamos conjuntamente para averiguar essa significância. Elas se mostram estatisticamente desprezíveis (anexo 3), o que torna possível excluí-las da regressão. A interpretação da regressão abaixo (Modelo 1) é a seguinte:



* Se o indivíduo contribuiu na campanha anterior, em média impacta em um aumento de probabilidade de contribuir nessa campanha de 13 pontos percentuais.
* Para cada semana adicional que se passa desde a última contribuição, reduz em média a probabilidade da variável dependente em 0,1 ponto percentual.
* Para taxa de resposta com contribuição aos pedidos de donativos anteriores aumenta a probabilidade de contribuir nessa campanha em 20 pontos percentuais.
* Cada pedido adicional recebido para doar aumenta a probabilidade de contribuir nessa campanha em 3,5 pontos percentuais.

É importante contrastar as vantagens desse modelo e desvantagens. No aspecto positivo o método OLS aplicado é centrado e consistente, além de ter simples cálculo e simples interpretação. É visto como melhor aproximação linear para a resposta de probabilidade e usualmente estima bem os efeitos parciais. Por outro lado, o LPM pode ter valores de probabilidade ajustados que sejam superiores a 1 ou inferiores a 0. Possui uma relação rígida (linear) que muitas vezes não é razoável, o que faz com que não seja eficiente. Os erros são heterocedásticos, e por isso, se usa a matriz de variância-covariância robusta de White.

Tenta-se ainda correr um modelo agregando a variável **generous** (anexo 4), mas que não se demonstra estatisticamente significativo, pelo que ficamos temporariamente com o modelo 1.

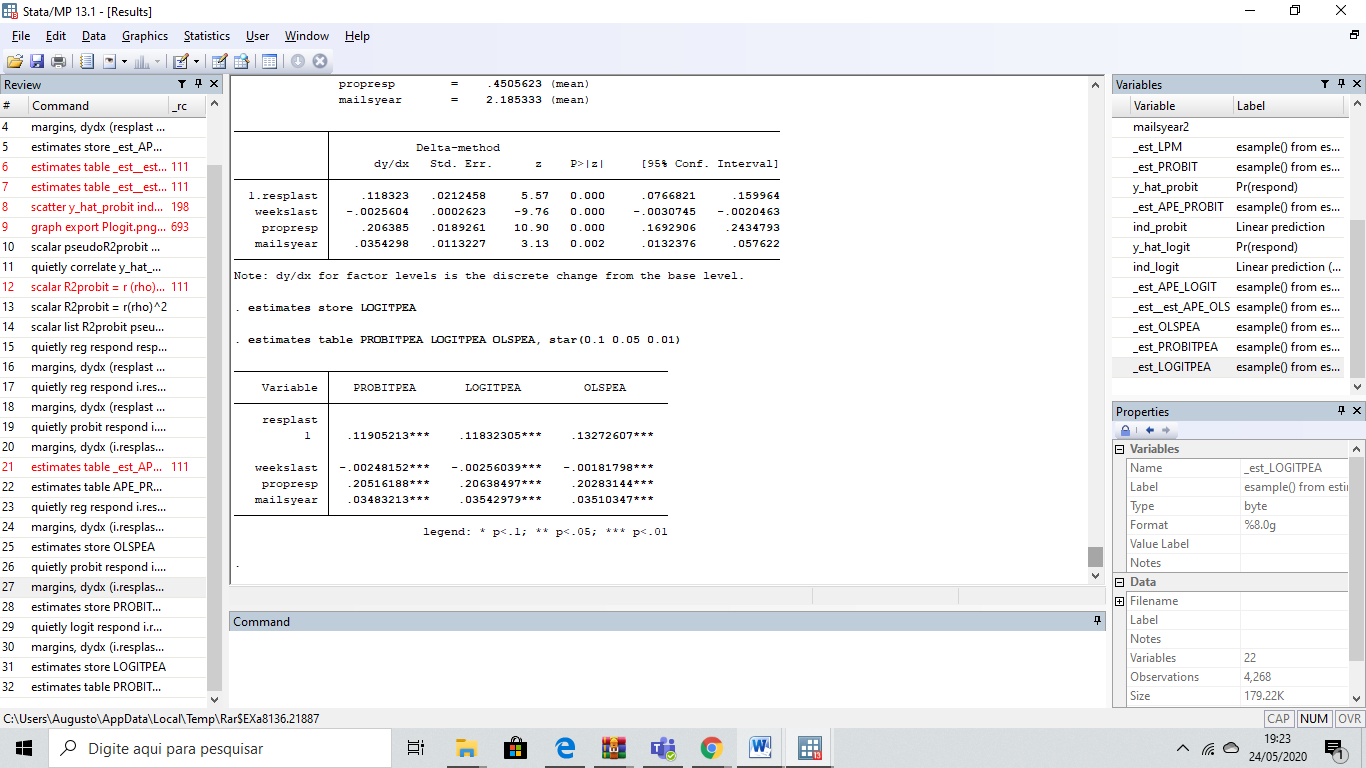
Há alternativas de modelos não lineares que se ajustam melhor aos dados e permitem relações mais flexíveis entre regressores e regressandos. Esses fatores nos inclinam a escolher um desses métodos como adequado. Exemplos deles são os modelos PROBIT e LOGIT. Esses não têm soluções analíticas e são resolvidos de forma iterativa.

Realizou-se alguns testes para verificar a correta especificação do modelo. Primeiramente realizou-se o Teste de Wald para verificar a significância conjunta de algumas variáveis que não foram significativas(loginc, educ), e de acordo com o anexo 14 não há evidências para a rejeição da hipótese nula e, portanto, essas variáveis de fato não são significativas e as mesmas podem ser retiradas do modelo. Posteriormente, realizou-se o teste de rácio de verossimilhanças, e de acordo com o anexo 15, verificou que não há evidencias para a rejeição da hipótese nula e, portanto, as variáveis não são significativas, o que converge com o resultado do teste de Wald.

O PROBIT e LOGIT com as variáveis significativas verificada nos testes de especificação do modelo nos oferece as estimativas que podem ser vistas nos anexos 5 e 7. Em seguida, verificou-se a qualidade de ajustamento do novo modelo proposto. Primeiramente, comparou-se o R2 com o pseudo - R2 (anexo 16). Verificou-se que não há tanta diferença entre os respectivos coeficientes de determinação. Logo depois, verificou-se a capacidade preditiva dos modelos PROBIT e LOGIT, que pode ser visto nos anexos 17 e 18. É importante ressaltar que o cutoff usado foi 0.5, também se testou o cutoff 0.40(média), mas o modelo perdeu capacidade preditiva. Comparando a capacidade de previsão dos dois modelos (anexo 19), nota-se que o modelo LOGIT tem uma ligeira vantagem na previsão, mas nada muito significativo.

Após a especificação do novo modelo e avaliação da capacidade de ajustamento, podemos obter os efeitos parciais médios (APE) pelo método PROBIT (anexo 6). Assim como estimamos em PROBIT os coeficientes e os efeitos parciais médios (APE), podemos fazer o mesmo para o LOGIT. O que muda de um modelo para o outro é o comportamento paramétrico dos erros ou a função de distribuição cumulativa que acompanha o modelo. No PROBIT temos que e ~ iid N(0,1), que segue distribuição normal; no LOGIT temos que e ~ iid logística padrão, tendo distribuição acumulada logística. O APE do LOGIT pode ser observado no anexo 8. Reparamos que os coeficientes de LPM (logo seus efeitos parciais) e os efeitos parciais dos métodos não lineares são relativamente semelhantes, com mesmo sinal. Sendo o LOGIT e PROBIT mais similares entre eles.

Já os efeitos parciais na média (PEA) do LPM se mantêm os mesmos (anexo 9) que o APE e os coeficientes. O mesmo não é válido para os estimadores não lineares (PROBIT e LOGIT). O APE e o PEA são distintos tanto para o PROBIT como para o LOGIT (anexos 10 e 11), mas possuem valores próximos como podem ver abaixo:

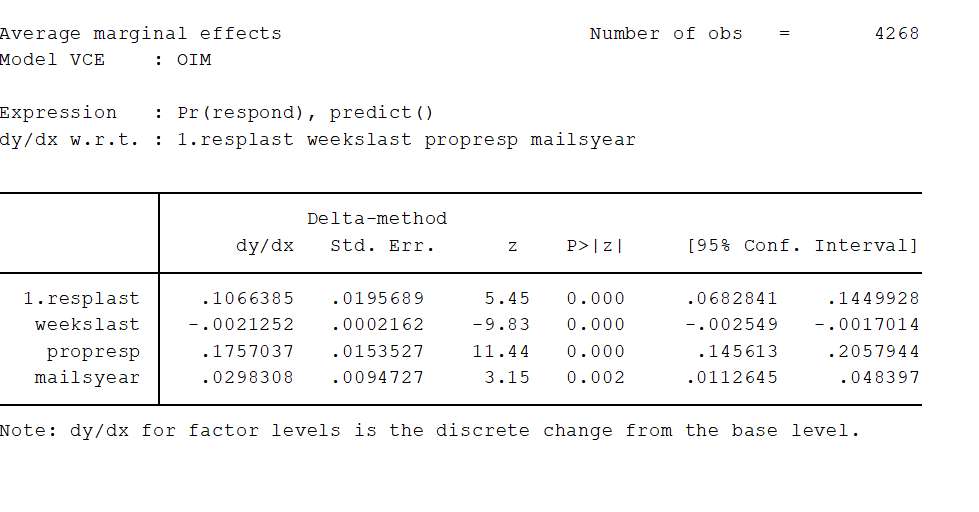


Quando se assume que a variável **resplast** é endógena, e portanto é correlacionada com o termo de erro, não podemos mais utilizar o método OLS de estimação, pois esse deixa de ser consistente. Nesse caso, devemos estimar a regressão em dois passos, utilizando **generous** como variável instrumental. O mesmo método de estimação com variáveis instrumentais é aplicado ao modelo PROBIT. Uma condição necessária para instrumentar uma variável e utilizar o método em dois passos é que o instrumento seja correlacionado com a variável a ser instrumentada, mas não seja correlacionado com o termo de erro. Essa condição é que garante a não existência de instrumento fraco. Os erros desse modelo devem seguir normalidade (u|z ~ Normal (0,1)) e ser homocedásticos.

Além disso, cabe ressaltar que se houver uma heterogeneidade não observada, como é o caso da generosidade, independente dos regressores, faz com que o PROBIT comum estime de maneira consistente o APE. O procedimento em dois passos, no entanto, por utilizar estimativas no seu processo, faz com que os erros padrão no segundo passo não sejam válidos. Devem ser corrigidos pelo método *Bootstrap*.

Pode-se ver as estimações dos coeficientes para o procedimentos em dois passos no OLS (anexo 12). Para o PROBIT com variável endógena, pode-se analisar as estimações no anexo 13.

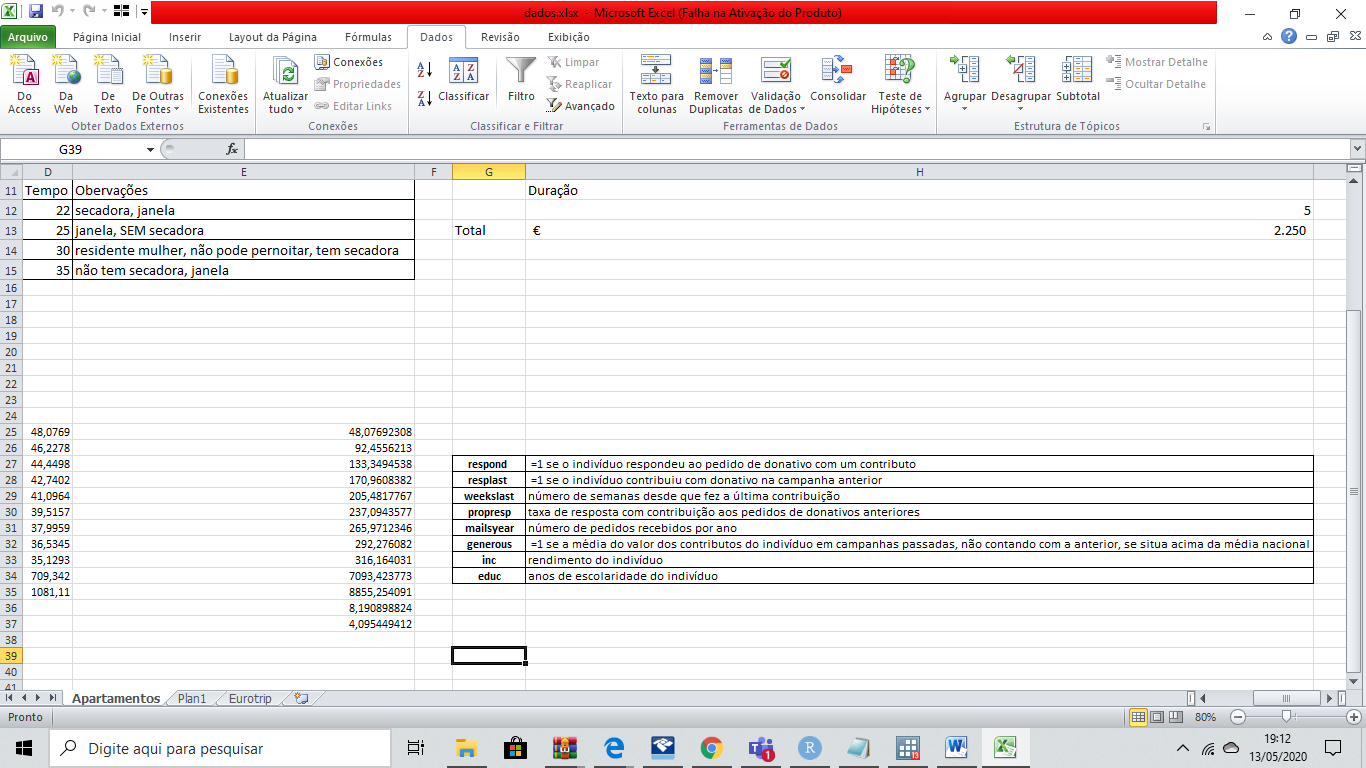
Optou-se pela escolha do modelo PROBIT comum, pois não temos evidências a priori que a variável resplat seja endógena, e que portanto, seja necessária uma estimação em dois passos. Como já assinalado, o PROBIT tem a vantagem sobre o LPM por não ter efeitos fixos lineares na probabilidade da variável dependente. No entanto, como forma de comparação ao modelo LPM, podemos utilizar o APE do PROBIT para observar os efeitos parciais médios que nos dizem que:



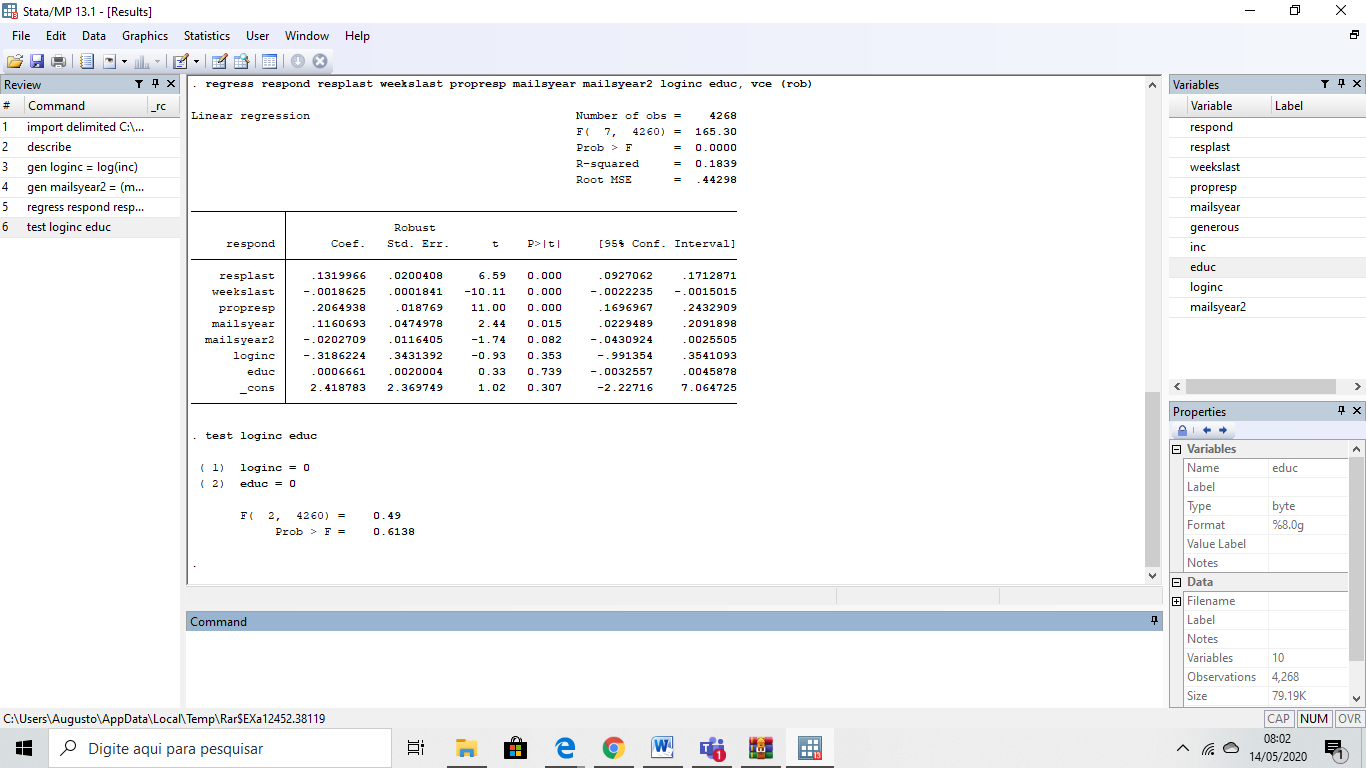
* Se o indivíduo contribuiu na campanha anterior, em média impacta em um aumento de probabilidade de contribuir nessa campanha de 10,5 pontos percentuais.
* Para cada semana adicional que se passa desde a última contribuição, reduz em média a probabilidade da variável dependente em 0,2 ponto percentual.
* Para taxa de resposta com contribuição aos pedidos de donativos anteriores aumenta em média a probabilidade de contribuir nessa campanha em 17,5 pontos percentuais.
* Cada pedido adicional recebido para doar aumenta em média a probabilidade de contribuir nessa campanha em 3 pontos percentuais.

Anexos

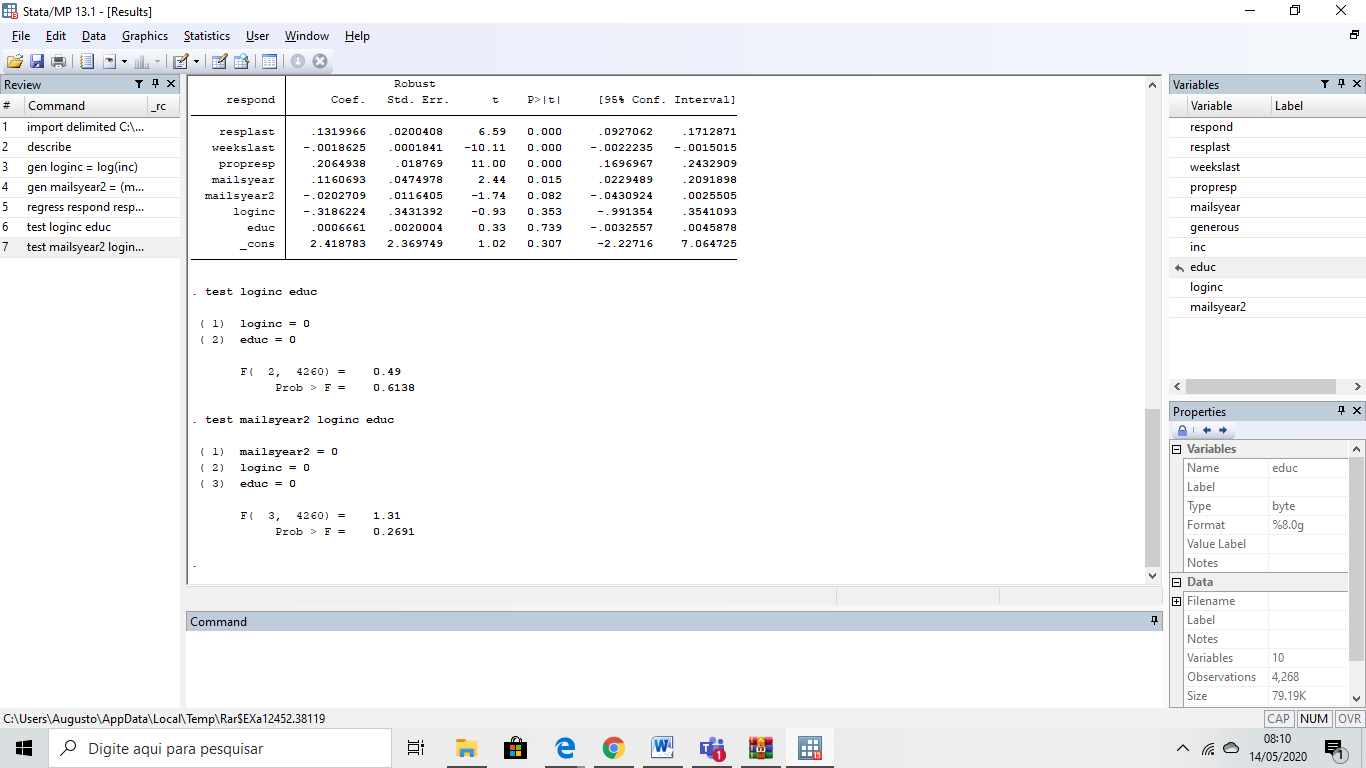
Anexo 1 – Referência das variáveis da base de dados



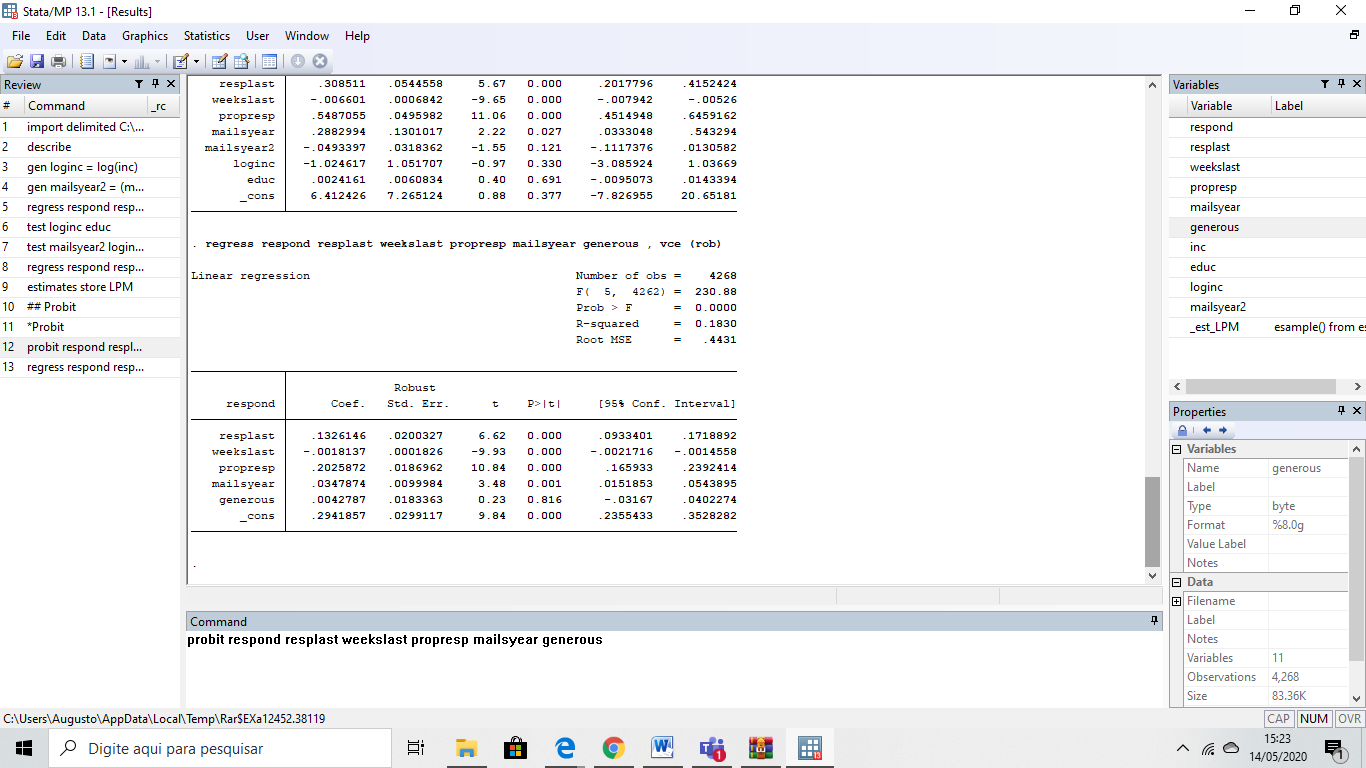
Anexo 2 – Estimativas para o LPM completo



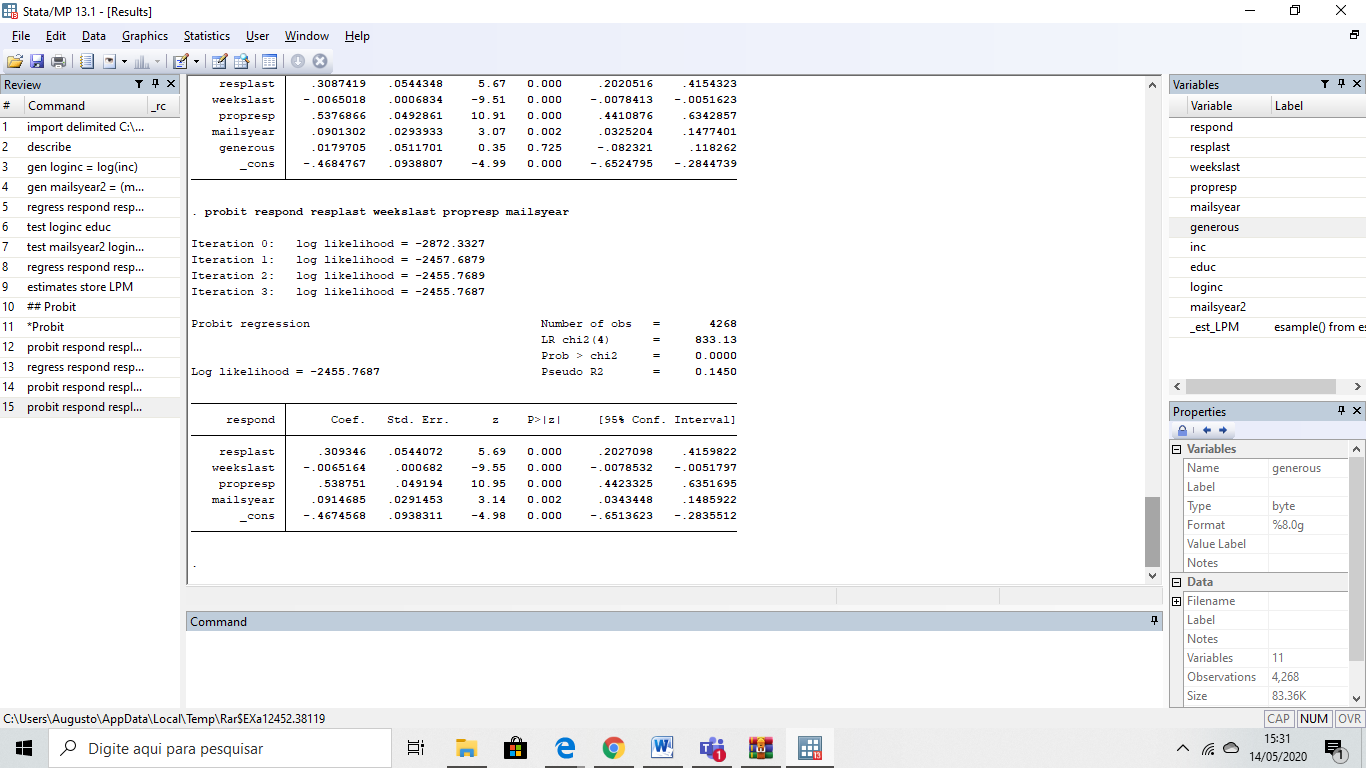
Anexo 3 – Teste de significância conjunta



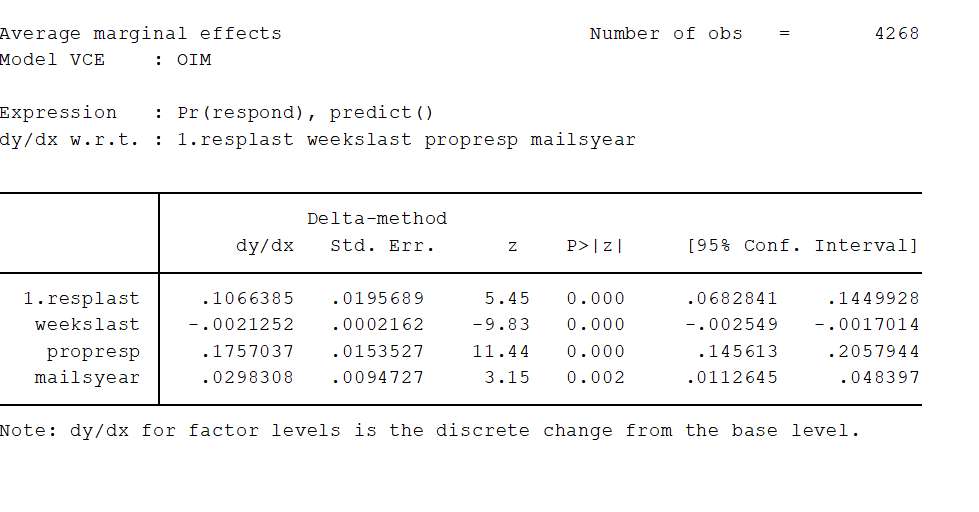
Anexo 4 – LPM com variável generous



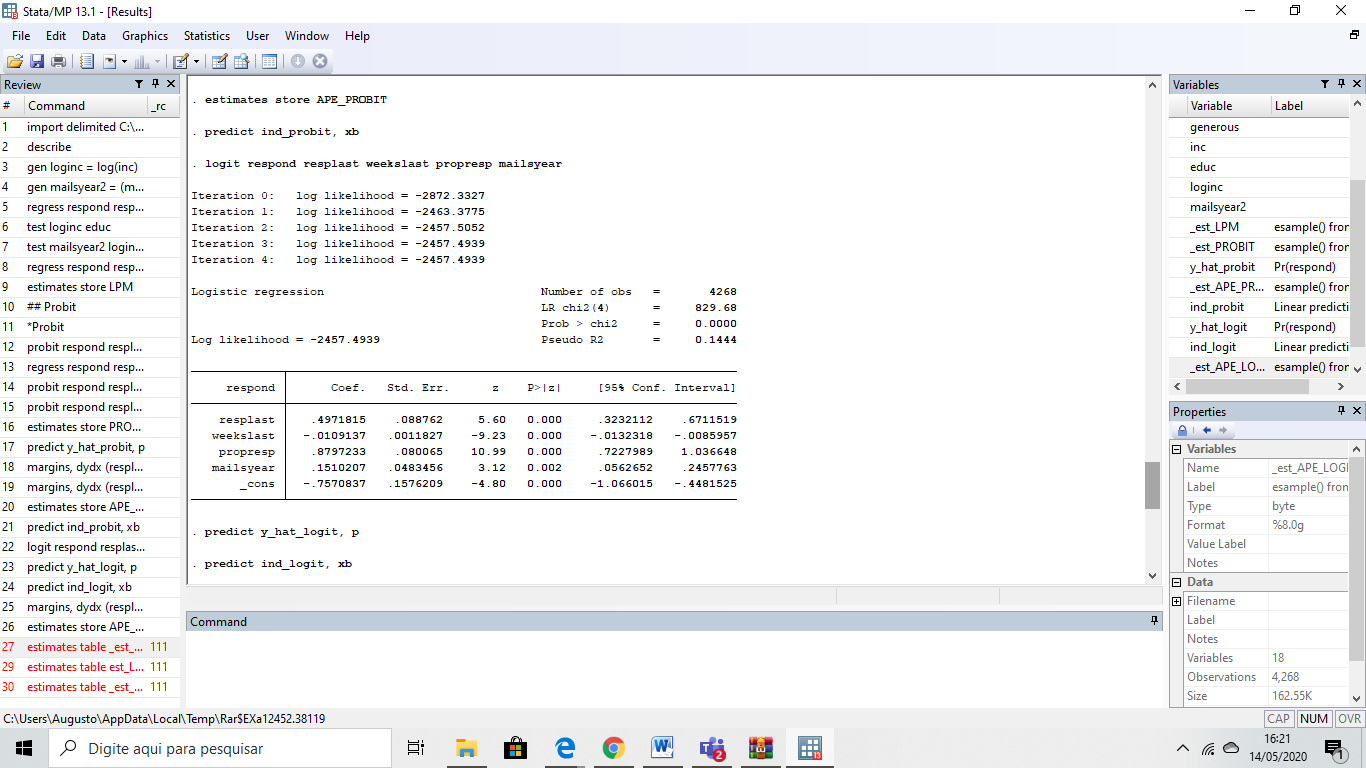
Anexo 5 – PROBIT com variáveis significativas de LPM



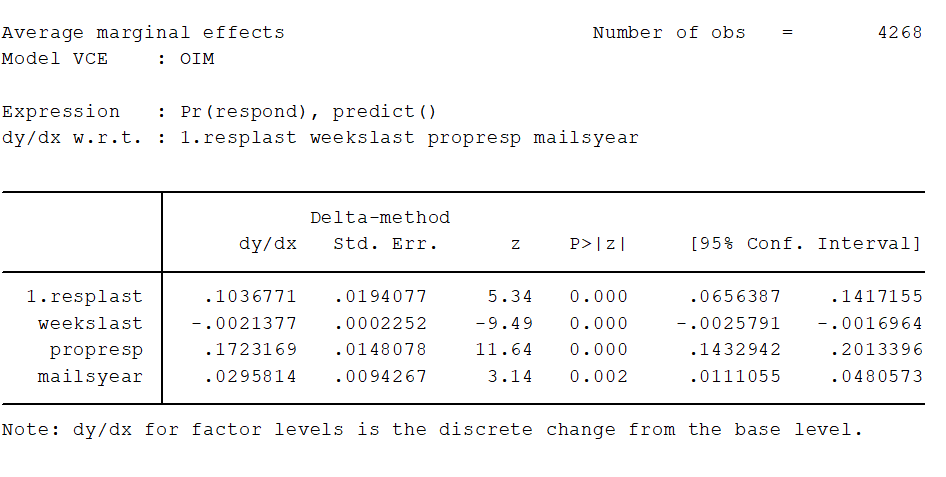
Anexo 6 – APE PROBIT



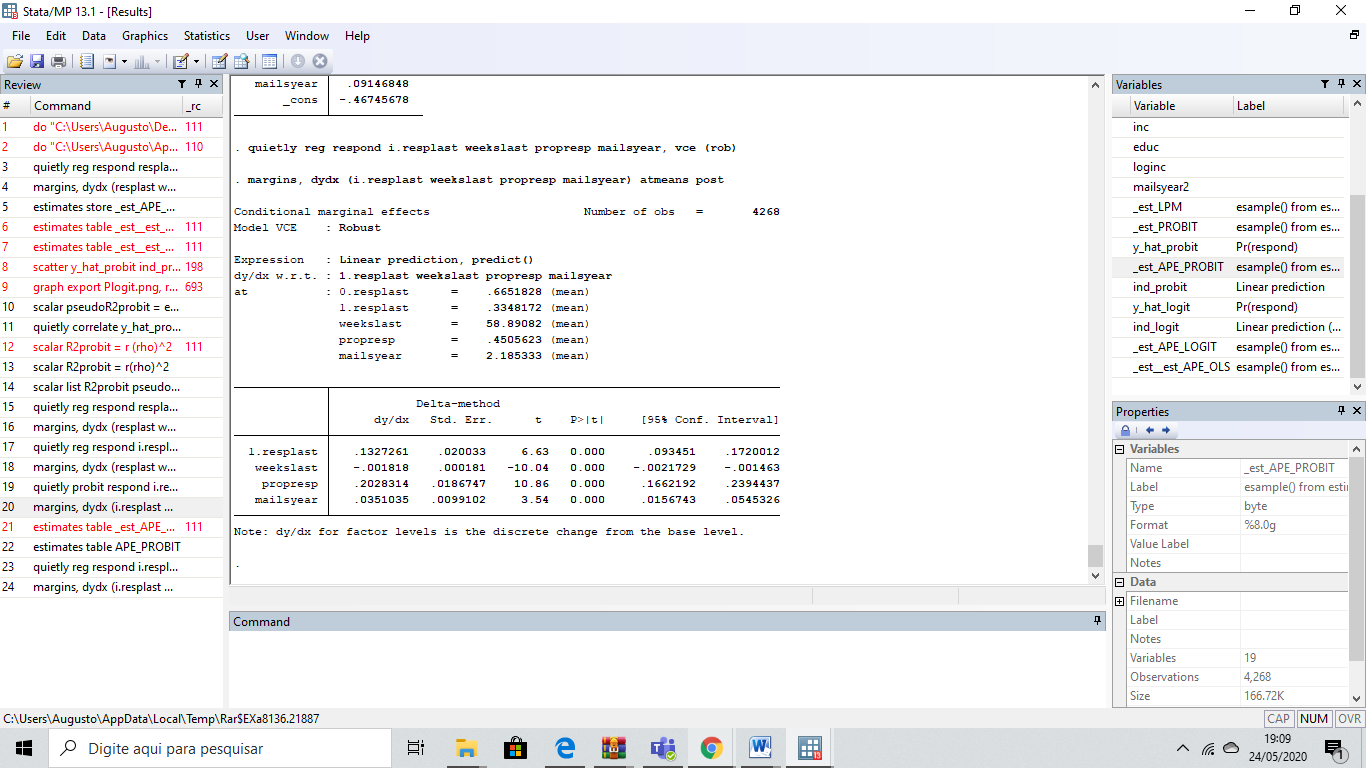
Anexo 7 – LOGIT com variáveis significativas de LPM



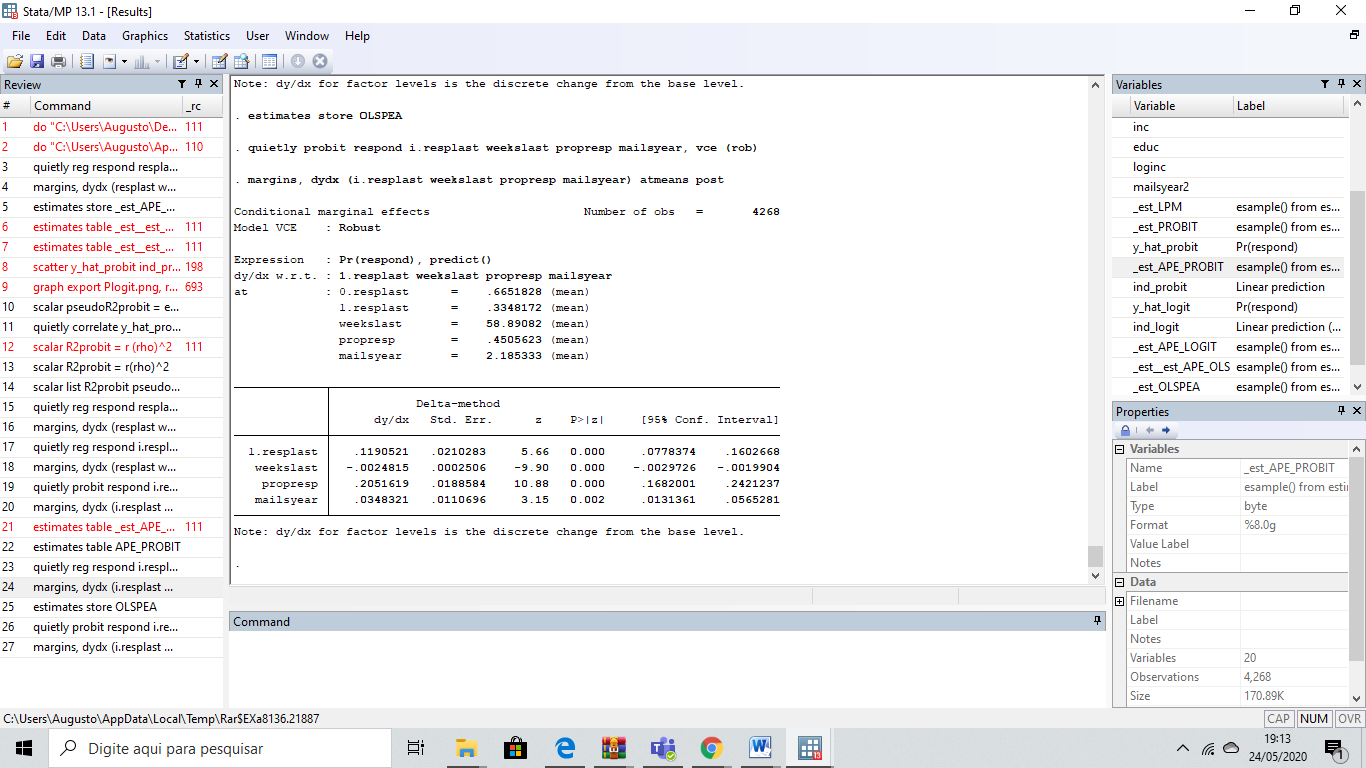
Anexo 8 – APE LOGIT



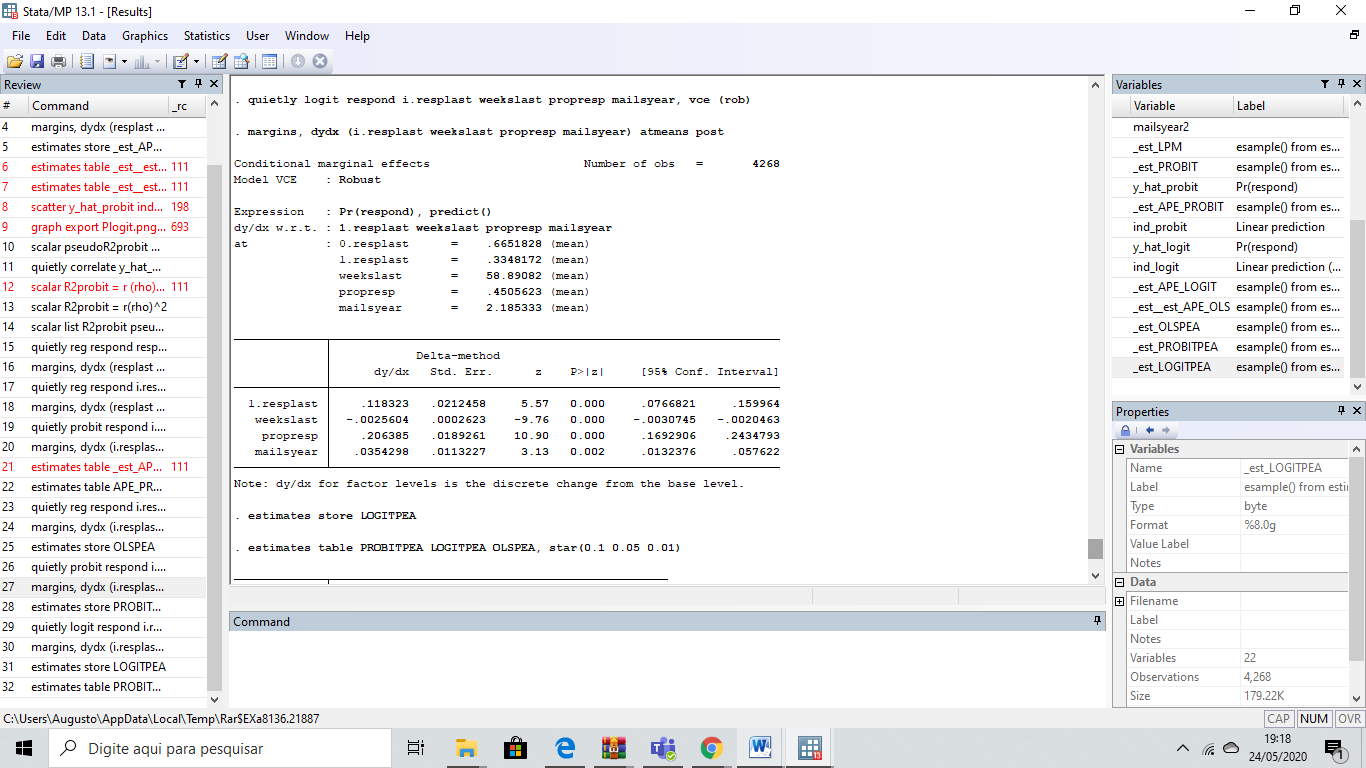
Anexo 9 – PEA LPM



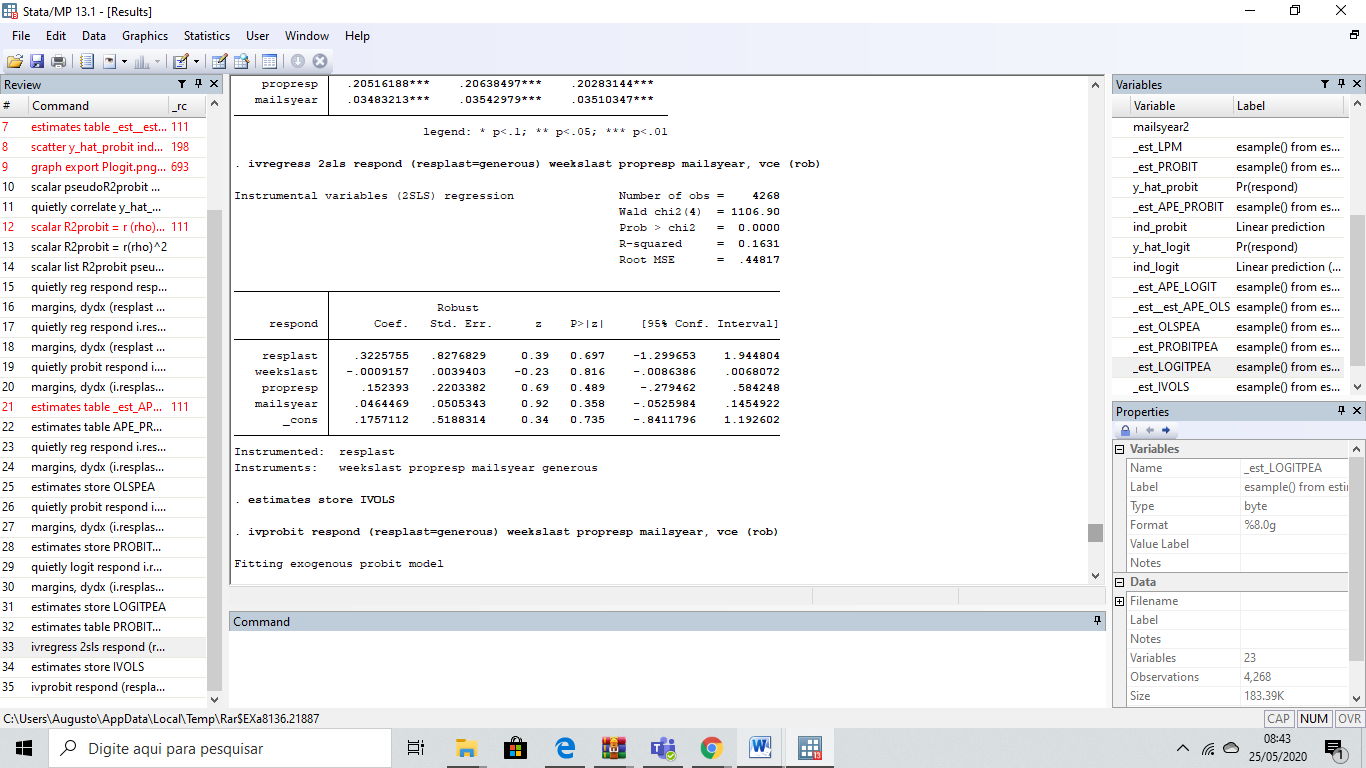
Anexo 10 – PEA PROBIT



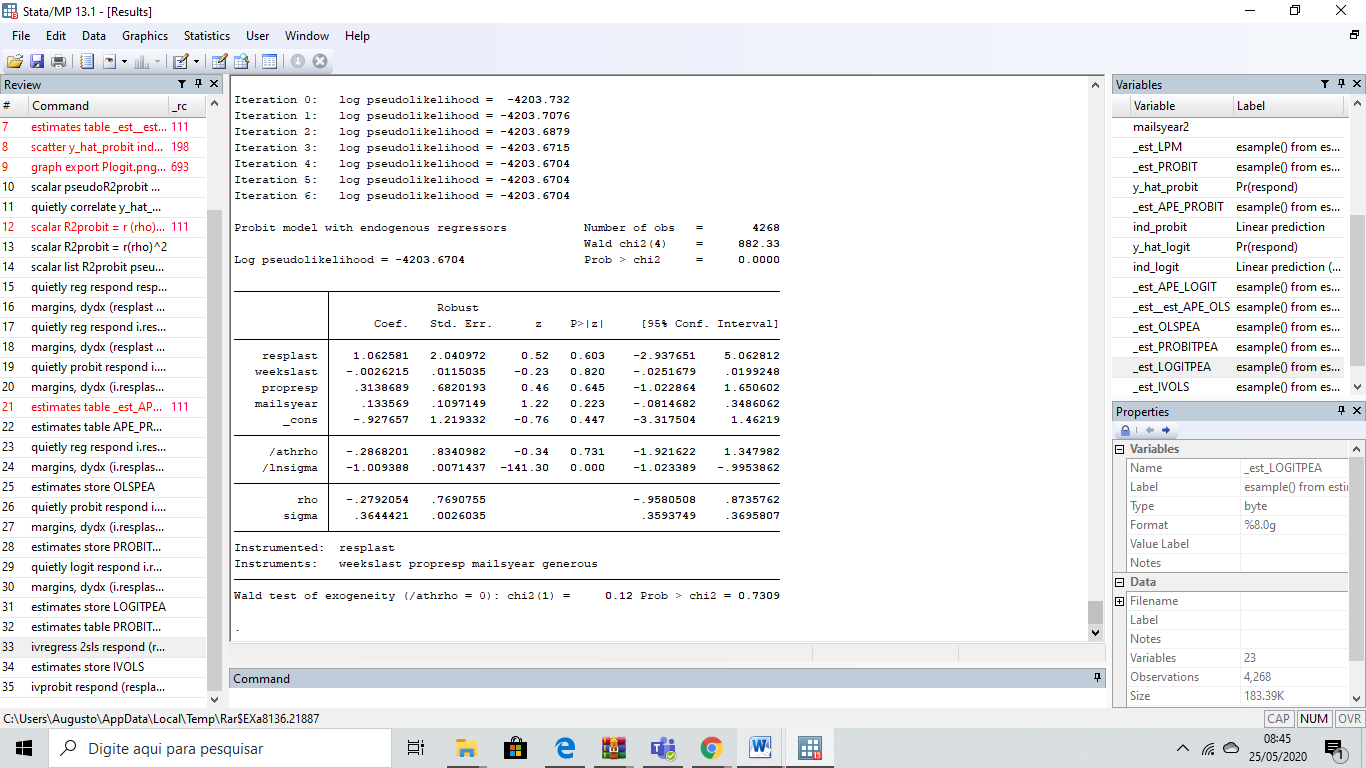
Anexo 11 – PEA LOGIT



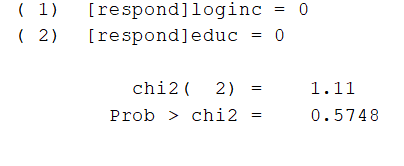
Anexo 12 – Estimação 2SLS



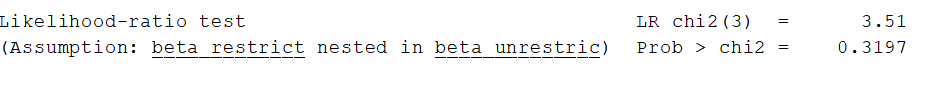
Anexo 13 – PROBIT com Variável Endógena



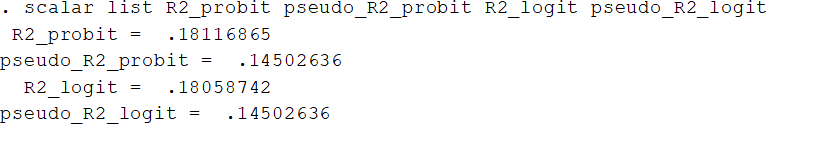
Anexo 14 – Teste de Wald



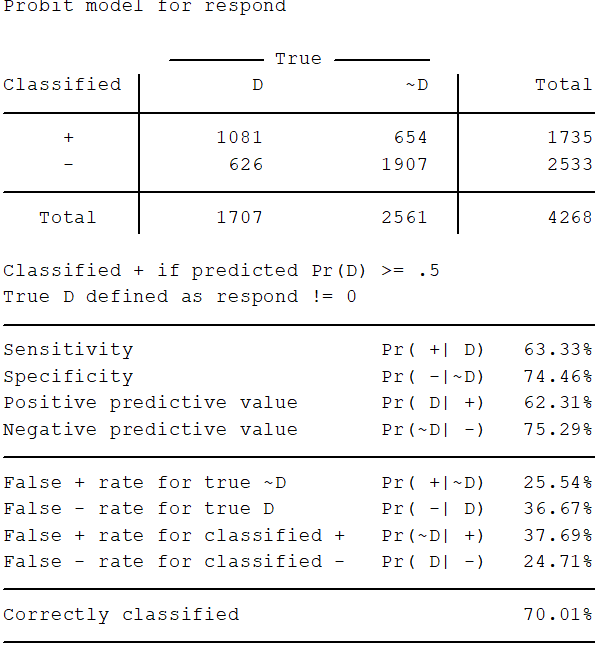
Anexo 15 – Teste do rácio da verossimilhança



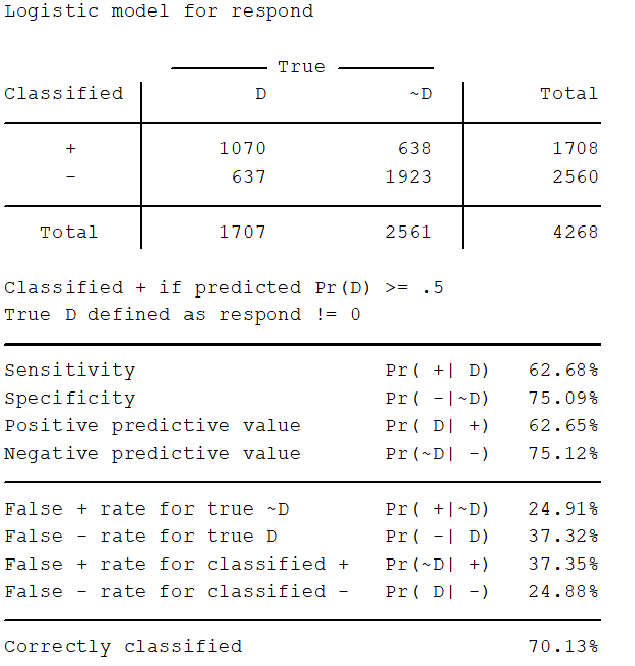
Anexo 16 – Comparação R2  com o pseudo- R2



Anexo 17 – Capacidade Preditiva PROBIT



Anexo 18 – Capacidade Preditiva LOGIT



Anexo 19 – Comparação da previsão dos modelos LOGIT e PROBIT

