Irania Matos
Homework 8
Lab 8 +Lab Addendums
Group: Nancy Reyes Soto, Keyla Pereira, Kerra Sinanan, Heidy Collado.

Results:

LPM

	TRUE	
PRED	0	1
FALSE	203,909	9,306
TRUE	182,021	17,039

Logit

	TRUE	
PRED	0	1
FALSE	235,290	11,352
TRUE	150,640	14,993

Standardized LPM

	TRUE	
PRED	0	1
FALSE	180,978	8,244
TRUE	166,340	15,442

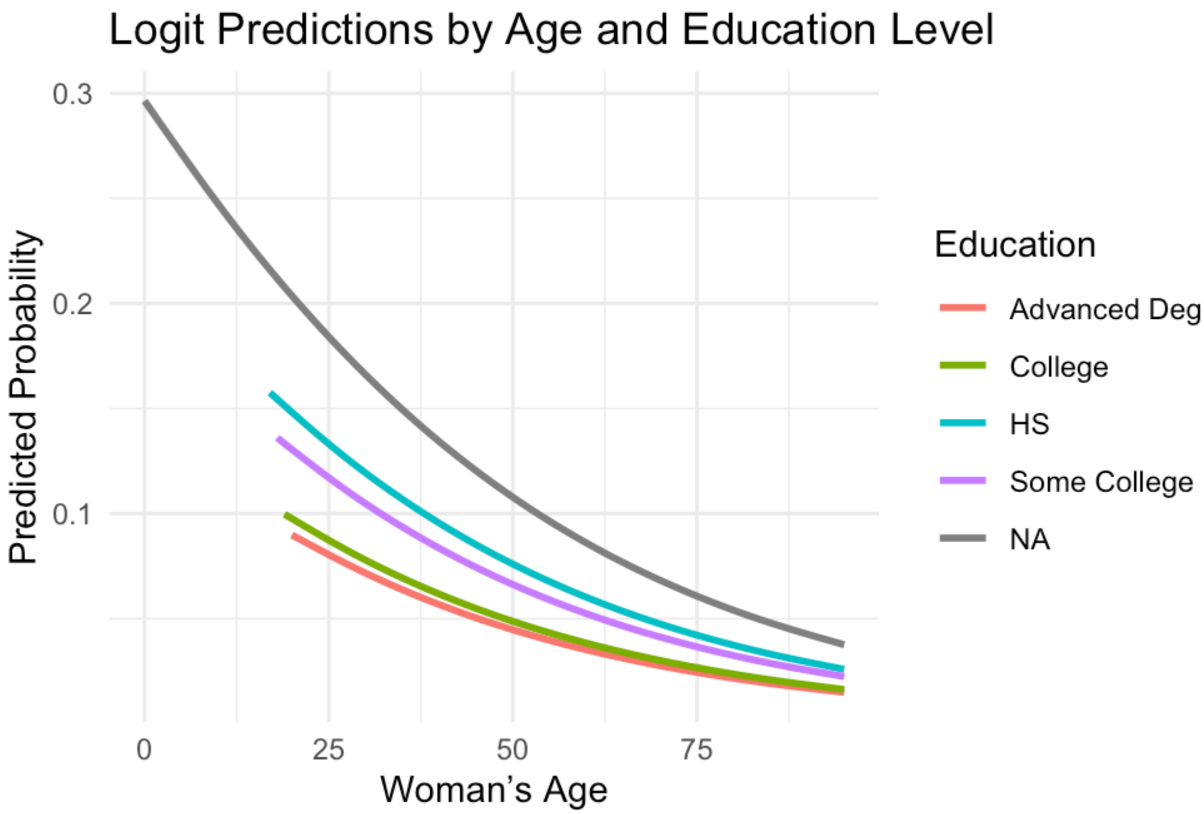
Standardized Logit

	TRUE	
PRED	0	1
FALSE	213,252	10,308
TRUE	134,066	13,378

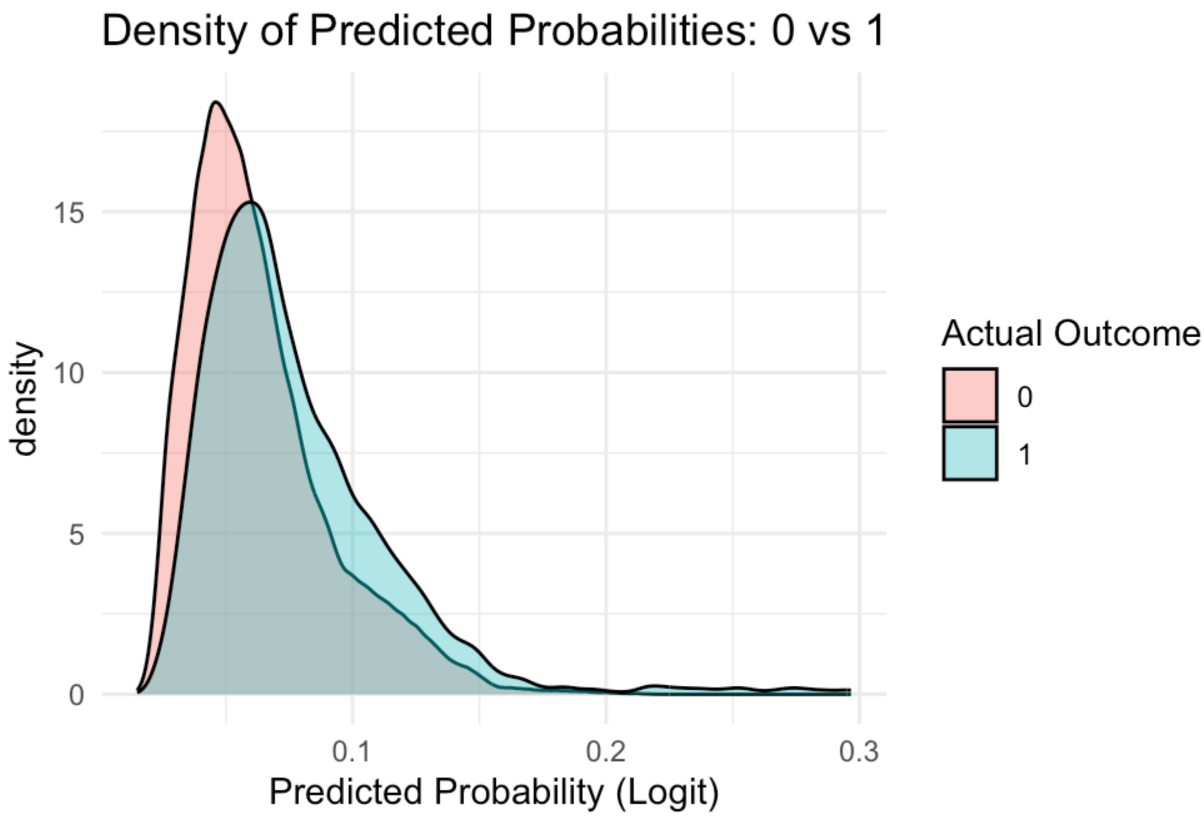
In the in-sample comparison, both the linear probability model and the logit model show the same broad prediction pattern: predicting a “0”is much easier than predicting a “1” man ≥10 years older.

The LPM predicts 17,039 true positives (cases where the man is ≥10 years older) and 9,306 false negatives. The logit model predicts slightly fewer true positives, 14,993, and more false negatives 11,352, but it compensates by producing fewer false positives.

This shows the trade-off between false positives and false negatives. The logit model is more conservative it predicts “1” less often. As a result, it lowers the false positive rate but increases false negatives. Both models agree that the outcome is rare, and both rely heavily on age and education as strong predictor



There is a clear monotonic pattern: as education increases, the predicted probability of a large age gap decline. Age and education jointly shape the estimated probability, with younger and less educated women having the highest predicted values.



The densities overlap heavily, showing why prediction is difficult
the logit model does not sharply separate couples with large vs small age gaps.

Articles for Project
Question: Still working on it but it will focus on females in sport

"Equal Pay on the Hardwood? Gender Differences in NCAA Division I Basketball Coaches' Salaries"
Humphreys analyzes gender differences in NCAA Division I coaching salaries using the federally reported Equity in Athletics Data System (EADA), which provides publicly accessible salary and program-funding information for both men's and women's teams. The study merges these salary data with team performance indicators such as winning percentage, postseason participation, and school revenues. The econometric strategy centers on log-salary regressions that control for experience, school characteristics, conference affiliation, and institutional fixed effects. The paper's main question is whether female coaches receive lower compensation than comparable male coaches after accounting for performance and program resources. This provides an important institutional perspective on gender pay inequality within sports labor markets.

Coates & Webber – “Pay and Performance in Men’s and Women’s Football: Comparing the MLS and NWSL”

Coates and Webber use a panel of player-level data from Major League Soccer (MLS) and the National Women’s Soccer League (NWSL) to compare how performance translates into pay for male versus female professional soccer players. Their dataset includes publicly available salary information released by the MLS Players Association and reported NWSL salary tiers, combined with match statistics such as goals, assists, minutes played, position, and team quality. Using these micro-data, the authors estimate log-linear earnings equations in which salary is regressed on performance metrics, demographic characteristics, and club-level controls, with league indicators and interaction terms to test whether the return to performance differs across genders. Their main research question focuses on whether women’s compensation in the NWSL is lower than men’s in the MLS once productivity and team factors are accounted for. The findings show significantly weaker returns to performance in the women’s league, reinforcing concerns about gender-based inequality in professional soccer.