462 Marekting Models HW 1

Group 6: Brianna Chen, Sheryl Xu, Yang Zhou

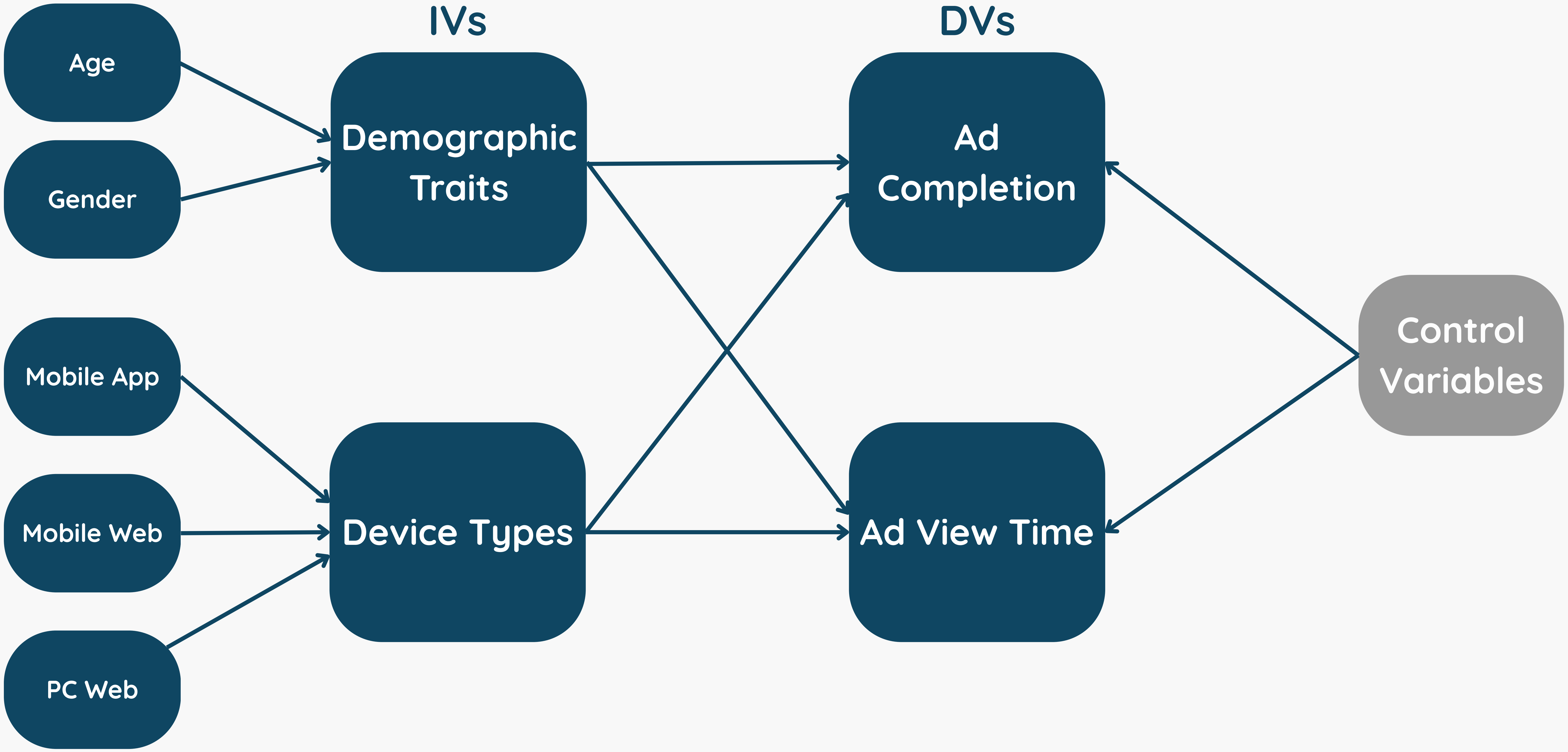
27 April, 2025

Research Question

How technical and personal factors influence ad engagement?

- How does the viewer's **device type (PC Web, Mobile Web, Mobile App)** influence both the probability of completing pre-roll skippable ads and the duration of ad viewing time?
- How do **demographic characteristics (age and gender)** impact both the probability of ad completion and the length of ad viewing time?

Conceptual Framework



Yang Zhou

Data

Dataset: pre-roll_ad.csv

Individual-level time-stamped behavioral secondary data of ad and content viewing records

- Sample population: 2,078,090 users
- Sample size: 10000 users
- Sample period: July 1 to 28, 2019.
- Company: one of the global top 10 portal companies in South Korea

Variables - Question 1

How does the viewer's <u>device type</u> (PC Web, Mobile Web, Mobile App) influence both the probability of completing pre-roll skippable ads and the duration of ad viewing time?

- Dependent Variables: [ad completion]ad_complete, [ad viewing time]l_ad_stop, u_ad_stop
- Independent Variable: [device type] media_platform
- Control Variables: genre_new + ad_brand_cat + age_new + gender_new + day + tt

Variables - Question 2

How do demographic characteristics (age and gender) impact both the probability of ad completion and the length of ad viewing time?

- Dependent Variables: [ad completion]ad_complete, [ad viewing time]l_ad_stop, u_ad_stop
- Independent Variables: [demographics] age_new + gender_new
- Control Variables: media_platform + genre_new + ad_brand_cat + day + tt

Descriptive Statistics

```
# A tibble: 1 × 13
  mean_age sd_age min_age max_age mean_clip_duration sd_clip_duration min_clip_duration max_clip_duration mean_ad_view_time
     <dbl>  <dbl>   <dbl>   <dbl>              <dbl>            <dbl>             <dbl>             <dbl>             <dbl>
1     32.4   12.3      10      60               176.             133.                33               866              7.63
  sd_ad_view_time min_ad_view_time max_ad_view_time ad_completion_rate
            <dbl>            <dbl>            <dbl>              <dbl>
1            3.66                0               15               21.1
> 
```

Viewer Age

- **Average Age:** 32.4 years
- **Age Range:** 10 to 60 years

Clip Duration

- **Average Clip Length:** 176 seconds (~2.9 minutes)
- **Range:** 33 to 866 seconds

Viewing Time

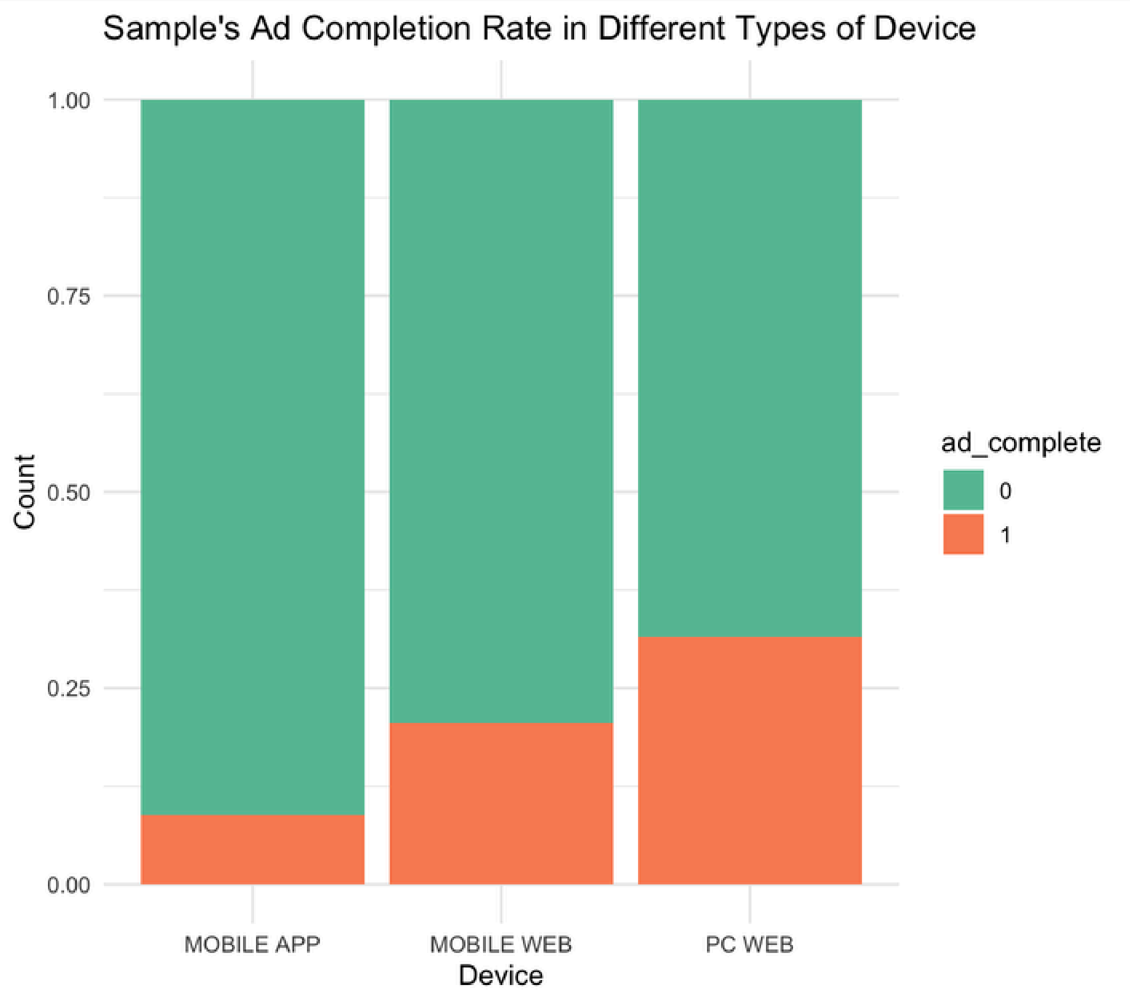
- **Average Viewing Time:** 7.63 seconds
- **Range:** 0 to 15 seconds

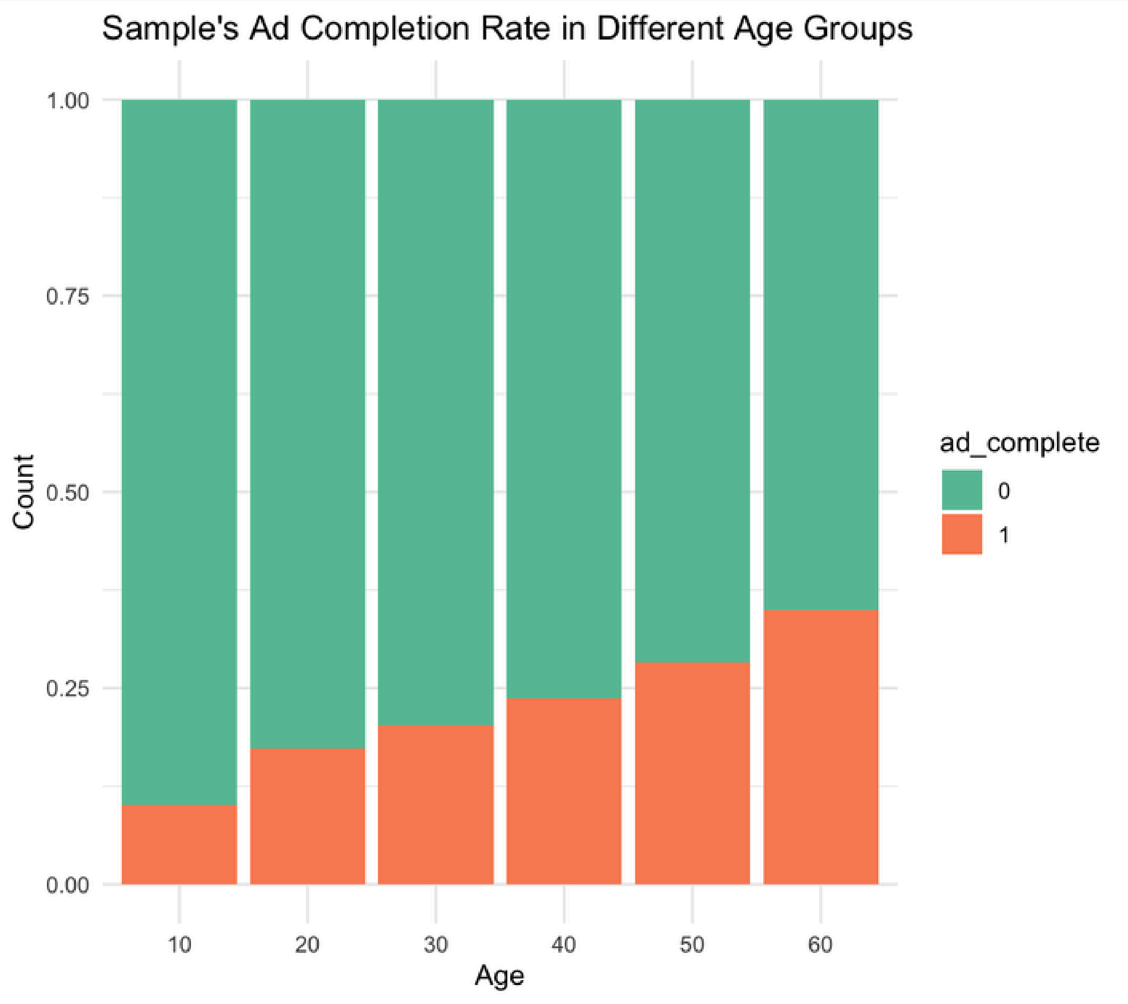
Ad Completion Rate

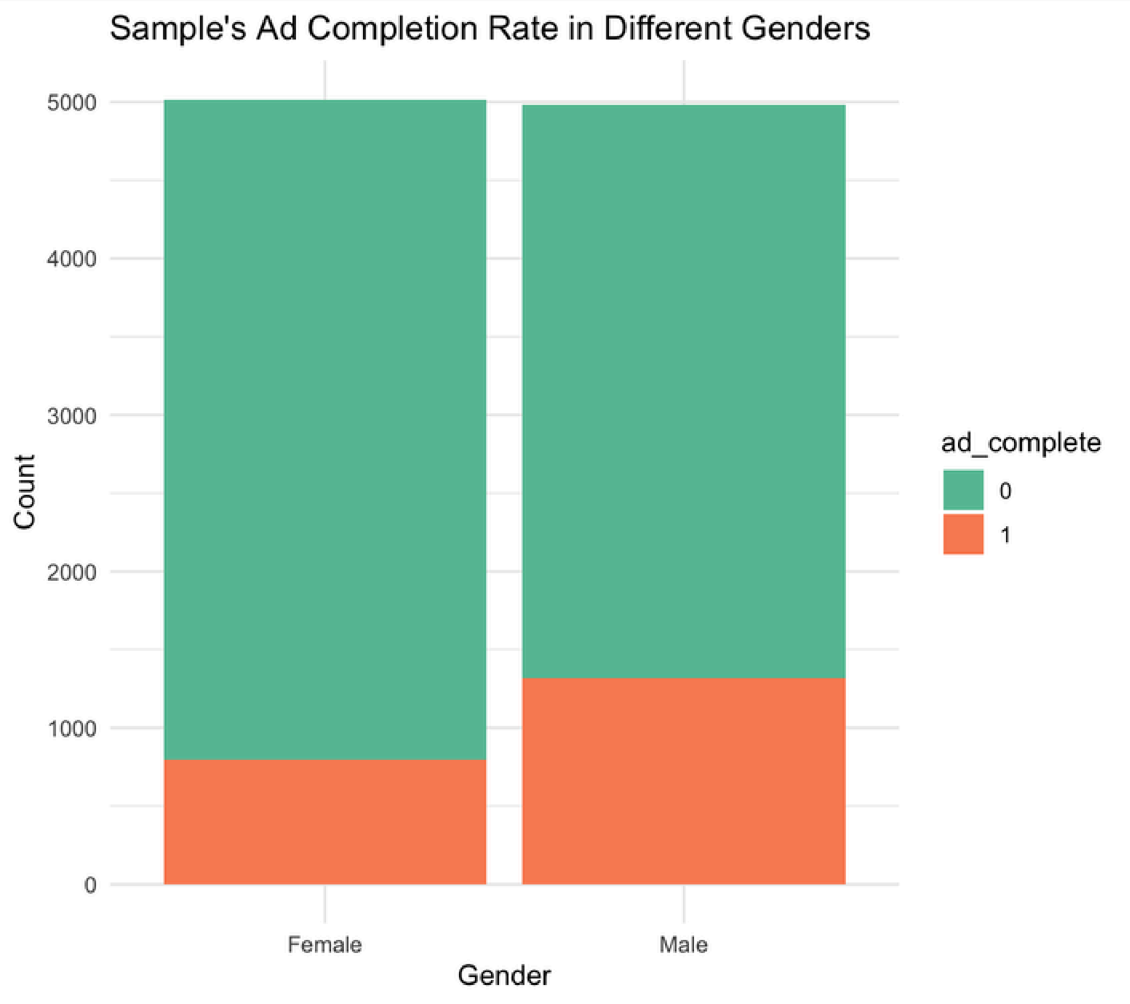
- **Completed Ads:** 21.1% of viewers finished the ad

Brianna Chen

Descriptive Statistics







Question 1: Linear Probability Model

device type, ad completion

```
Call:
lm(formula = ad_complete ~ media_platform + genre_new + ad_brand_cat +
    age_new + day + tt + clip_duration, data = ad)

Residuals:
     Min       1Q   Median       3Q      Max
-0.59738 -0.21826 -0.12713 -0.02789  1.14314

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -3.327e-01  5.917e-02  -5.623 1.93e-08 ***
media_platformMOBILE WEB  2.577e-02  1.919e-02   1.343 0.179444
media_platformPC WEB      1.079e-01  2.254e-02   4.788 1.71e-06 ***
```

Interpretation:

- **Mobile Web:** not statistically significant, meaning we do not have strong evidence that Mobile Web users are significantly more likely to complete the ad compared to Mobile App users, holding all other control variables constant.
- **PC Web:** PC Web users are **10.79% more likely** to complete a pre-roll skippable ad compared to Mobile App users, holding all other control variables constant.

Question 1: Logistic Regression Model

device type, ad completion

```
Call:
glm(formula = ad_complete ~ media_platform + genre_new + ad_brand_cat +
    age_new + day + tt + clip_duration, family = binomial(link = "logit"),
    data = ad)

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)           -5.4055476  0.5304418 -10.191  < 2e-16 ***
media_platformMOBILE WEB  0.3773240  0.1733768   2.176 0.029531 *
media_platformPC WEB   0.8670380  0.1868751   4.640 3.49e-06 ***
```

```
> 1 / (1 + exp(-0.3773240))
[1] 0.5932275
> 1 / (1 + exp(-0.8670380))
[1] 0.704129
```

Interpretation:

- **Mobile Web:** Being on Mobile Web increases the log-odds of completing the ad by **0.377 (59.3% of probability)** compared to Mobile App holding all other control variables constant.
- **PC Web:** Being on PC Web increases the log-odds of completing the ad by **0.867 (70.4% of probability)** compared to Mobile App, holding all other control variables constant.

Question 1: Probit RegressionModel

device type, ad completion

```
Call:
glm(formula = ad_complete ~ media_platform + genre_new + ad_brand_cat +
    age_new + day + tt + clip_duration, family = binomial(link = "probit"),
    data = ad)

Coefficients:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)               -3.0507512  0.2723895 -11.200  < 2e-16 ***
media_platformMOBILE WEB   0.1927932  0.0887740   2.172 0.029876 *
media_platformPC WEB       0.4850537  0.0979751   4.951 7.39e-07 ***
```

```
> pnorm(0.1927932)
[1] 0.5764395
> pnorm(0.4850537)
[1] 0.6861809
```

Interpretation:

- **Mobile Web:** Mobile Web users have a higher probit score of **0.193 (57.6% of probability)** to complete the ad than Mobile App users, holding other factors constant.
- **PC Web:** PC Web users have a higher probit score of **0.485 (68.6% of probability)** to complete the ad than Mobile App users, holding other factors constant.

Question 1: Interval Regression Model

device type, ad viewing time

```
Call:
survreg(formula = Surv(l_ad_stop, u_ad_stop, cens, type = "interval") ~
    media_platform + genre_new + ad_brand_cat + age_new + day +
        tt + clip_duration, data = ad, dist = "gaussian")
                          Value Std. Error     z       p
(Intercept)             3.271626   0.542224  6.03 1.6e-09
media_platformMOBILE WEB  0.336010   0.175855  1.91 0.05604
media_platformPC WEB     1.089014   0.206436  5.28 1.3e-07
```

exp(1.0890)≈2.97

Interpretation:

- **PC Web:** Highly statistically significantly, PC Web users' /sformed ad stop time is **1.0890 higher**(about **2.97 seconds later**) for users on Mobile Web, holding other factors constant.

Question 2: Linear Probability Model

age + gender, ad completion

```
Call:
lm(formula = ad_complete ~ gender_new + age_new + media_platform +
    day + tt + ad_brand_cat + genre_new + clip_duration, data = ad)

Residuals:
     Min       1Q   Median       3Q      Max
-0.62979 -0.21766 -0.12672 -0.02787  1.14664

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)         -3.108e-01  5.934e-02  -5.237 1.66e-07 ***
gender_newMale      -1.669e-02  1.090e-02  -1.531  0.12584
age_new20            4.752e-02  1.610e-02   2.952  0.00317 **
age_new30            4.811e-02  1.586e-02   3.033  0.00243 **
age_new40            7.330e-02  1.607e-02   4.563 5.11e-06 ***
age_new50            1.079e-01  1.830e-02   5.898 3.79e-09 ***
age_new60            1.722e-01  2.585e-02   6.663 2.83e-11 ***
```

Interpretation:

- **Age:** Compared with people under 20, all age groups have significantly higher completion rate holding the other variables fixed. Furthermore, the higher the age, the bigger the coefficient, suggesting **a positive relationship between age and completion.**
- **Gender:** The p value is higher than 0.05(0.13), which means that there is **no significant difference** between men and women in ad completion.

Question 2: Logistic Probability Model

age + gender, ad completion

```
Call:
glm(formula = ad_complete ~ gender_new + age_new + clip_duration +
    media_platform + age_new + day + tt + ad_brand_cat + genre_new +
    gender_new, family = binomial(link = "logit"), data = ad)

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)       -5.3502117  0.5348318 -10.004  < 2e-16 ***
gender_newMale    -0.1183038  0.0770764  -1.535 0.124811
age_new20          0.4454864  0.1336276   3.334 0.000857 ***
age_new30          0.4523138  0.1305307   3.465 0.000530 ***
age_new40          0.6133490  0.1306479   4.695 2.67e-06 ***
age_new50          0.8198713  0.1405458   5.833 5.43e-09 ***
age_new60          1.1558995  0.1755312   6.585 4.54e-11 ***
```

Interpretation:

- Compared with people under 20, **all age groups have significantly higher completion rate** holding the other variables fixed.
- A person **in their 20s** has 0.445 higher log-odds of completing the ad than someone under 20, which is **60.9% in probability**.

Question 2: Probit Probability Model

age + gender, ad completion

```
Call:
glm(formula = ad_complete ~ gender_new + age_new + clip_duration +
    media_platform + day + tt + ad_brand_cat + genre_new, family = binomial(link = "probit"),
    data = ad)

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)             -2.9976411  0.2737018 -10.952  < 2e-16 ***
gender_newMale          -0.0681162  0.0430915  -1.581 0.113939
age_new20                0.2398714  0.0705661   3.399 0.000676 ***
age_new30                0.2491810  0.0690734   3.607 0.000309 ***
age_new40                0.3398643  0.0693521   4.901 9.56e-07 ***
age_new50                0.4668141  0.0758661   6.153 7.60e-10 ***
age_new60                0.6662382  0.0985042   6.764 1.35e-11 ***
```

Interpretation:

- Compared with people under 20, all age groups have significantly higher completion rate holding the other variables fixed.
- A person **in their 20s** has a 0.239 higher p-score of completing the ad than someone under 20, which is **59.5% in probability**.

Question 2 - Interval Regression Model

```
Call:
survreg(formula = Surv(l_ad_stop, u_ad_stop, cens, type = "interval") ~
    gender_new + age_new + media_platform + clip_duration + day +
        tt + ad_brand_cat + genre_new, data = ad, dist = "gaussian")
                         Value Std. Error     z       p
(Intercept)           3.369260   0.543926  6.19 5.9e-10
gender_newMale       -0.026624   0.099840 -0.27 0.78973
age_new20             0.406966   0.147497  2.76 0.00580
age_new30             0.571446   0.145315  3.93 8.4e-05
age_new40             0.712603   0.147183  4.84 1.3e-06
age_new50             0.882421   0.167595  5.27 1.4e-07
age_new60             1.302480   0.236584  5.51 3.7e-08
```

Interpretation:

- **Age:** Every age group except for 20s has significantly higher viewing time than people under 20, holding other factors constant.
- A person in his 30s watches the ad for **1.77 seconds**(exp(0.57)) than someone under 20.

Question 2 - Likelihood Ratio Test

```
Model 1: ad_complete ~ gender_new + media_platform + day + tt + ad_brand_cat +
    genre_new + clip_duration
Model 2: ad_complete ~ gender_new + age_new + media_platform + day + tt +
    ad_brand_cat + genre_new + clip_duration
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  31 -4662.9
2  36 -4629.2  5 67.298  3.736e-13 ***
```

```
Model 1: ad_complete ~ gender_new + clip_duration + media_platform + day +
    tt + ad_brand_cat + genre_new
Model 2: ad_complete ~ gender_new + age_new + clip_duration + media_platform +
    day + tt + ad_brand_cat + genre_new
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  30 -4629.6
2  35 -4597.0  5 65.169  1.034e-12 ***
```

```
Model 1: ad_complete ~ gender_new + clip_duration + media_platform + day +
    tt + ad_brand_cat + genre_new
Model 2: ad_complete ~ gender_new + age_new + clip_duration + media_platform +
    day + tt + ad_brand_cat + genre_new
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  30 -4642.9
2  35 -4608.3  5 69.185  1.514e-13 ***
```

```
Model 1: Surv(l_ad_stop, u_ad_stop, cens, type = "interval") ~ media_platform +
    day + gender_new + tt + ad_brand_cat + genre_new
Model 2: Surv(l_ad_stop, u_ad_stop, cens, type = "interval") ~ gender_new +
    age_new + media_platform + clip_duration + day + tt + ad_brand_cat +
    genre_new
  #Df LogLik Df  Chisq Pr(>Chisq)
1  30 -13611
2  36 -13587  6 48.253  1.052e-08 ***
```

Interpretation:

- **H0**: The coefficients for age groups are 0.
- All four types of model shows significant difference in log-likelihood values between full models and restricted models without age_new, supporting the **significance and robustness** of the models.

Results - Question 1

Technical factor of device types (PC Web, Mobile Web, Mobile App) influences probability of completing the ads and ad viewing time

- **Mobile Web:** In both the logit and probit models, the probability to complete the ad is higher for Mobile Web, but not as much as for PC Web.
- **PC Web: PC Web users are consistently more likely to complete the ad than Mobile App users across all three models.** The impact is strongest in the logit and probit models, where PC Web users are **70.4%** more likely to complete the ad in the logit model and **68.6%** more likely in the probit model. Additionally, PC Web users' ad stop time is about **2.97 seconds later** for users on Mobile Web, holding other factors constant.

PC Web is the most effective platform for both completing the ad and extending ad viewing time compared to the other devices.

Results - Question 2

Age signficantly influences both ad completion probability and viewing duration

- Age Impact: Teens had the lowest completion rates. As we went up the age chart, the older the age group, the higher the completion rates. This pattern was consistent across multiple models. Oldest viewers had the highest completion rates.

However, gender impact is insignificant.

Marketing Implications

Theoretically:

- This study shows that **demographic factors create significant viewing patterns**. The strong correlation between age and ad completion suggests that ad tolerance and viewing tenure varies across generations.

Practically:

- Advertisers can **optimize ad sequencing** by:
    - **Prioritizing PC Web placements** over mobile platforms to drive higher completion rates.
    - **Developing age-segmented ads** to increase ad completion rates amongst younger viewers.

Conclusions

Demographic and Technical Factors Both Matter:

- Device Type and Age influence ad viewing time and completion
- PC Web users are more likely to complete ads
- Older viewers have higher ad completion rates and longer viewing times

Relative Importance:

- Device type was the most impactful on ad completion and viewing time, with PC Web being the most effective platform
- Gender doesn't have a statistically significant effect on engagement

Limitations and Future Research

Limitations:

- Limited variables for measuring the demographic factor.
- Limited explanation for causality

Future Study:

- Although gender has no direct effect on ad completion, we can study its interaction with other variables.
- Introduce more variables to better measure device use and demographics.
- Expand sample across different platforms and countries.

Appendix - Question 1

Linear Regression Model:

```
> m1 <- lm(ad_complete ~ media_platform
+      + genre_new
+      + ad_brand_cat
+      + age_new
+      + day
+      + tt
+      + clip_duration,
+      ad)
> summary(m1)
```

```
Call:
lm(formula = ad_complete ~ media_platform + genre_new + ad_brand_cat +
    age_new + day + tt + clip_duration, data = ad)

Residuals:
     Min       1Q   Median       3Q      Max
-0.59738 -0.21826 -0.12713 -0.02789  1.14314

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -3.327e-01  5.917e-02  -5.623 1.93e-08 ***
media_platformMOBILE WEB  2.577e-02  1.919e-02   1.343 0.179444
media_platformPC WEB      1.079e-01  2.254e-02   4.788 1.71e-06 ***
genre_newdrama            1.097e-01  5.255e-02   2.088 0.036788 *
genre_newentertainment    8.692e-02  5.207e-02   1.669 0.095084 .
genre_newinformation      9.838e-02  6.080e-02   1.618 0.105688
genre_newkids             1.878e-01  1.325e-01   1.417 0.156432
genre_newmusic            6.816e-02  5.589e-02   1.220 0.222679
genre_newsports           2.276e-01  5.284e-02   4.308 1.66e-05 ***
ad_brand_catdrink         2.889e-01  1.281e-02  22.547  < 2e-16 ***
ad_brand_cateducation     1.507e-01  2.326e-02   6.480 9.60e-11 ***
ad_brand_catelectronics   1.944e-01  2.749e-02   7.070 1.66e-12 ***
ad_brand_catfinance       3.149e-01  3.857e-02   8.165 3.61e-16 ***
ad_brand_catfood          1.961e-01  1.749e-02  11.212  < 2e-16 ***
ad_brand_catgovernment    2.108e-01  2.365e-02   8.911  < 2e-16 ***
ad_brand_cathealth_food   2.000e-01  1.853e-02  10.792  < 2e-16 ***
ad_brand_catman           2.513e-01  7.302e-02   3.441 0.000582 ***
ad_brand_catmedicine      1.908e-01  2.525e-02   7.556 4.52e-14 ***
ad_brand_catmovie         2.224e-01  2.266e-02   9.815  < 2e-16 ***
ad_brand_catpet           1.876e-01  2.764e-02   6.787 1.21e-11 ***
ad_brand_catwoman         1.824e-01  3.377e-02   5.399 6.85e-08 ***
age_new                   2.547e-03  3.368e-04   7.562 4.32e-14 ***
dayMon                   -1.260e-02  1.457e-02  -0.865 0.387163
daySat                    9.308e-02  1.360e-02   6.845 8.07e-12 ***
daySun                    6.375e-02  1.458e-02   4.372 1.24e-05 ***
dayThu                   -9.483e-03  1.549e-02  -0.612 0.540424
dayTue                    2.413e-02  1.474e-02   1.637 0.101620
dayWed                    9.633e-02  1.542e-02   6.249 4.30e-10 ***
tt                        1.805e-03  6.323e-04   2.854 0.004326 **
clip_duration             8.961e-05  2.958e-05   3.029 0.002461 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3852 on 9970 degrees of freedom
Multiple R-squared:  0.1122,	Adjusted R-squared:  0.1096
F-statistic: 43.44 on 29 and 9970 DF,  p-value: < 2.2e-16
```

Logistic Regression Model:

```
> m2 <- glm(ad_complete ~ media_platform
+                 + genre_new
+                 + ad_brand_cat
+                 + age_new
+                 + day
+                 + tt
+                 + clip_duration,
+                 family = binomial(link = "logit"),
+                 ad)
> summary(m2)
```

```
Call:
glm(formula = ad_complete ~ media_platform + genre_new + ad_brand_cat +
    age_new + day + tt + clip_duration, family = binomial(link = "logit"),
    data = ad)

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)             -5.4055476  0.5304418 -10.191  < 2e-16 ***
media_platformMOBILE WEB  0.3773240  0.1733768   2.176 0.029531 *
media_platformPC WEB      0.8670380  0.1868751   4.640 3.49e-06 ***
genre_newdrama            0.9477841  0.4781318   1.982 0.047450 *
genre_newentertainment    0.7625921  0.4755832   1.603 0.108827
genre_newinformation      0.8776254  0.5201587   1.687 0.091560 .
genre_newkids             1.4044375  0.8546049   1.643 0.100305
genre_newmusic            0.6416079  0.4990771   1.286 0.198587
genre_newsports           1.5078790  0.4778848   3.155 0.001603 **
ad_brand_catdrink         1.8414119  0.1029710  17.883  < 2e-16 ***
ad_brand_cateducation     0.8100157  0.2005177   4.040 5.35e-05 ***
ad_brand_catelectronics   1.2807858  0.2079073   6.160 7.26e-10 ***
ad_brand_catfinance       1.9428188  0.2221198   8.747  < 2e-16 ***
ad_brand_catfood          1.3017352  0.1349168   9.648  < 2e-16 ***
ad_brand_catgovernment    1.4152428  0.1767680   8.006 1.18e-15 ***
ad_brand_cathealth_food   1.3237383  0.1421364   9.313  < 2e-16 ***
ad_brand_catman           1.7402996  0.4804987   3.622 0.000292 ***
ad_brand_catmedicine      1.2301918  0.1993477   6.171 6.78e-10 ***
ad_brand_catmovie         1.5041881  0.1673094   8.990  < 2e-16 ***
ad_brand_catpet           1.2488235  0.2117498   5.898 3.69e-09 ***
ad_brand_catwoman         1.1668601  0.2635440   4.428 9.53e-06 ***
age_new                   0.0171213  0.0022881   7.483 7.28e-14 ***
dayMon                   -0.1224040  0.1094442  -1.118 0.263390
daySat                    0.6174426  0.0936382   6.594 4.28e-11 ***
daySun                    0.4660776  0.1014447   4.594 4.34e-06 ***
dayThu                   -0.0524049  0.1139335  -0.460 0.645545
dayTue                    0.1731010  0.1041158   1.663 0.096396 .
dayWed                    0.6022223  0.1037825   5.803 6.52e-09 ***
tt                        0.0141960  0.0046446   3.056 0.002240 **
clip_duration             0.0004715  0.0001764   2.673 0.007516 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10313  on 9999  degrees of freedom
Residual deviance:  9205  on 9970  degrees of freedom
AIC: 9265

Number of Fisher Scoring iterations: 5
```

Probit Regression Model:

```
> m3 <- glm(ad_complete ~ media_platform
+           + genre_new
+           + ad_brand_cat
+           + age_new
+           + day
+           + tt
+           + clip_duration,
+           family = binomial(link = "probit"),
+           ad)
> summary(m3)
```

```
Call:
glm(formula = ad_complete ~ media_platform + genre_new + ad_brand_cat +
    age_new + day + tt + clip_duration, family = binomial(link = "probit"),
    data = ad)

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)             -3.0507512  0.2723895 -11.200  < 2e-16 ***
media_platformMOBILE WEB  0.1927932  0.0887740   2.172 0.029876 *
media_platformPC WEB      0.4850537  0.0979751   4.951 7.39e-07 ***
genre_newdrama            0.5238937  0.2443478   2.144 0.032029 *
genre_newentertainment    0.4302006  0.2427762   1.772 0.076394 .
genre_newinformation      0.4806775  0.2704928   1.777 0.075561 .
genre_newkids             0.7878610  0.4890747   1.611 0.107197
genre_newmusic            0.3664257  0.2564367   1.429 0.153029
genre_newsports           0.8548381  0.2445632   3.495 0.000473 ***
ad_brand_catdrink         1.0305205  0.0542461  18.997  < 2e-16 ***
ad_brand_cateducation     0.4658554  0.1036540   4.494 6.98e-06 ***
ad_brand_catelectronics   0.7081944  0.1125501   6.292 3.13e-10 ***
ad_brand_catfinance       1.0872491  0.1326760   8.195 2.51e-16 ***
ad_brand_catfood          0.7245047  0.0725284   9.989  < 2e-16 ***
ad_brand_catgovernment    0.7857061  0.0960599   8.179 2.85e-16 ***
ad_brand_cathealth_food   0.7356736  0.0765823   9.606  < 2e-16 ***
ad_brand_catman           0.9632810  0.2725157   3.535 0.000408 ***
ad_brand_catmedicine      0.6762993  0.1064377   6.354 2.10e-10 ***
ad_brand_catmovie         0.8257126  0.0914008   9.034  < 2e-16 ***
ad_brand_catpet           0.6990833  0.1138295   6.141 8.17e-10 ***
ad_brand_catwoman         0.6490515  0.1413818   4.591 4.42e-06 ***
age_new                   0.0100100  0.0012985   7.709 1.27e-14 ***
dayMon                   -0.0446634  0.0597892  -0.747 0.455055
daySat                    0.3350643  0.0529007   6.334 2.39e-10 ***
daySun                    0.2457470  0.0571257   4.302 1.69e-05 ***
dayThu                   -0.0222253  0.0628297  -0.354 0.723535
dayTue                    0.0905830  0.0584487   1.550 0.121193
dayWed                    0.3432965  0.0590869   5.810 6.25e-09 ***
tt                        0.0067767  0.0025542   2.653 0.007975 **
clip_duration             0.0002667  0.0001051   2.537 0.011178 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10313.5  on 9999  degrees of freedom
Residual deviance:  9227.6  on 9970  degrees of freedom
AIC: 9287.6

Number of Fisher Scoring iterations: 5
```

Interval Regression Model:

```
> ad$type = relevel(as.factor(ad$media_platform), ref="MOBILE APP")
> ad$cens = 3
>
> m4 <- survreg(Surv(l_ad_stop, u_ad_stop, cens, type = "interval") ~
+             media_platform
+           + genre_new
+           + ad_brand_cat
+           + age_new
+           + day
+           + tt
+           + clip_duration,
+           ad, dist = "gaussian")
> summary(m4)
```

```
Call:
survreg(formula = Surv(l_ad_stop, u_ad_stop, cens, type = "interval") ~
    media_platform + genre_new + ad_brand_cat + age_new + day +
        tt + clip_duration, data = ad, dist = "gaussian")
                           Value Std. Error      z        p
(Intercept)             3.271626   0.542224   6.03 1.6e-09
media_platformMOBILE WEB 0.336010   0.175855   1.91 0.05604
media_platformPC WEB    1.089014   0.206436   5.28 1.3e-07
genre_newdrama          1.130376   0.481597   2.35 0.01892
genre_newentertainment  0.889212   0.477244   1.86 0.06243
genre_newinformation    0.732220   0.557173   1.31 0.18879
genre_newkids           1.842149   1.214039   1.52 0.12917
genre_newmusic          0.601068   0.512229   1.17 0.24062
genre_newsports         1.724776   0.484270   3.56 0.00037
ad_brand_catdrink       1.967625   0.117311  16.77 < 2e-16
ad_brand_cateducation   1.061901   0.213021   4.98 6.2e-07
ad_brand_catelectronics 1.504192   0.251750   5.97 2.3e-09
ad_brand_catfinance     2.140055   0.352833   6.07 1.3e-09
ad_brand_catfood        1.349554   0.160145   8.43 < 2e-16
ad_brand_catgovernment  1.629083   0.216516   7.52 5.3e-14
ad_brand_cathealth_food 1.474679   0.169681   8.69 < 2e-16
ad_brand_catman         2.238220   0.667615   3.35 0.00080
ad_brand_catmedicine    1.309842   0.231196   5.67 1.5e-08
ad_brand_catmovie       1.785223   0.207438   8.61 < 2e-16
ad_brand_catpet         1.261444   0.253117   4.98 6.2e-07
ad_brand_catwoman       1.218442   0.309273   3.94 8.2e-05
age_new                 0.020204   0.003083   6.55 5.6e-11
dayMon                 -0.086750   0.133440  -0.65 0.51562
daySat                  0.697380   0.124478   5.60 2.1e-08
daySun                  0.583389   0.133505   4.37 1.2e-05
dayThu                 -0.045300   0.141816  -0.32 0.74940
dayTue                  0.218650   0.134967   1.62 0.10523
dayWed                  0.778252   0.141106   5.52 3.5e-08
tt                      0.017346   0.005789   3.00 0.00273
clip_duration           0.000386   0.000271   1.43 0.15371
Log(scale)              1.210995   0.007820 154.86 < 2e-16

Scale= 3.36

Gaussian distribution
Loglik(model)= -13588.3   Loglik(intercept only)= -13930.9
        Chisq= 685.24 on 29 degrees of freedom, p= 1.2e-125
Number of Newton-Raphson Iterations: 3
n= 10000
```

Appendix - Question 2

Linear Regression Model:

```
fit_mlr <- lm(ad_complete ~ gender_new
              +age_new
              +media_platform
              +day+tt
              +ad_brand_cat
              +genre_new
              +clip_duration, data = ad)
summary(fit_mlr)

fit_mlr_2 <- lm(ad_complete ~ gender_new
                +media_platform
                +day
                +tt
                +ad_brand_cat
                +genre_new
                +clip_duration, data = ad)
lrtest(fit_mlr_2, fit_mlr)
```

Logistic Regression Model:

```
fit_logit <- glm(ad_complete ~ gender_new
                 +age_new
                 +clip_duration
                 +media_platform
                 +day
                 +tt
                 +ad_brand_cat
                 +genre_new,
               family = binomial(link = 'logit'), data = ad)
summary(fit_logit)

fit_logit_2 <- glm(ad_complete ~ gender_new
                   +clip_duration
                   +media_platform
                   +day
                   +tt
                   +ad_brand_cat
                   +genre_new,
                 family = binomial(link = 'logit'), data = ad)
lrtest(fit_logit_2, fit_logit)
```

Probit Regression Model:

```
fit_probit <- glm(ad_complete ~ gender_new
                  +age_new
                  +clip_duration
                  +media_platform
                  +day
                  +tt
                  +ad_brand_cat
                  +genre_new,
                family = binomial(link = 'probit'), data = ad)
summary(fit_logit)

fit_probit_2 <- glm(ad_complete ~ gender_new
                    +clip_duration
                    +media_platform
                    +day
                    +tt
                    +ad_brand_cat
                    +genre_new,
                  family = binomial(link = 'probit'), data = ad)
lrtest(fit_logit_2, fit_logit)
```

Interval Regression Model:

```
fit_int = survreg(formula = Surv(l_ad_stop, u_ad_stop, cens, type = 'interval')~ gender_new
                 +age_new
                 +media_platform
                 +clip_duration
                 +day
                 +tt
                 +ad_brand_cat
                 +genre_new,
                 data = ad, dist = 'gaussian')
summary(fit_int)

fit_int_2 = survreg(formula = Surv(l_ad_stop, u_ad_stop, cens, type = 'interval')~ gender_new
                 +media_platform
                 +clip_duration
                 +day
                 +tt
                 +ad_brand_cat
                 +genre_new,
                 data = ad, dist = 'gaussian')
lrtest(fit_int_2, fit_int)
```

Thank you