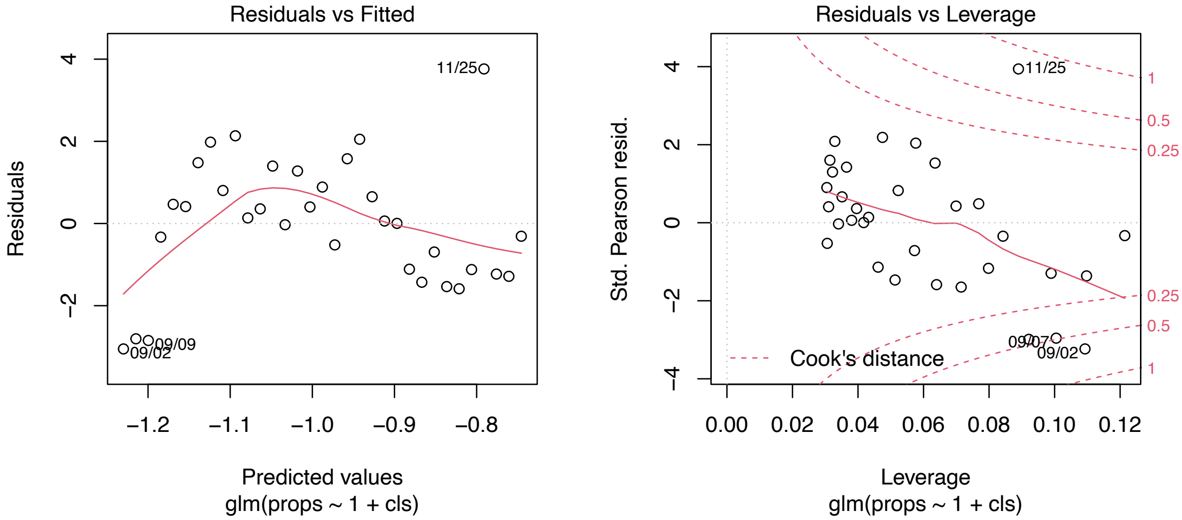


1 Class Attendance, 28 points

Across the $n = 33$ lectures of STAT 378, Zoom tracked the number of student attendees. In this question, we consider modeling Zoom attendence using Logistic Regression. We first consider

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1(\text{class}) \qquad \text{(Model 1)}$$

where π_i is the percent of the class on Zoom on day i, and (class) is the class number, i.e. $1, 2, 3, \ldots, 33$. The fitted coefficients are

	Est	Std Err	Test Stat	p-value
β_0	-1.25	0.10	-12.5	<2e-16
β_1	0.015	0.005	3.1	0.0023

1. **(4 points)**

   Based on the estimated parameters for Model 1, how has Zoom attendance changed as the class number increases, i.e. as the term progresses?

   increase by 0.015 per Class

2. **(2 points)**

   Which distribution is used to compute the p-values in the final column?

   Normal Dist.

3. **(4 points)**

   Describe how this model fits the data based on the above figures.

We next consider

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1(\text{class}) + \beta_2(\text{Day} < \text{Sept 10}) + \beta_3(\text{Day} > \text{Nov 13}) \qquad \text{(Model 2)}$$

where (Day $<$ Sept 10) is a boolean for the first 2 week of class and (Day $>$ Nov 13) is a boolean for classes after reading week. The fitted coefficients are

	Est	Std Err	Test Stat	p-value
β_0	-0.920	0.14	-6.60	3.25e-11
β_1	0.003	0.01	0.34	0.732
β_2	-1.503	0.28	-5.34	9.13e-08
β_3	-0.102	0.17	-0.60	0.548

4. **(4 points)**

   Based on the estimated parameters for Model 2, how has Zoom attendance changed across the term?

β_2 is significant, attendance before sept 2 was lower

The Analysis of Deviance table comparing Model 1 to Model 2 is

	Res DF	Res Deviance	DF	Deviance	p-value
Model 1	$\star$	80.81			
Model 2	$\star$	38.55	$\star$	42.26	6.65e-10

5. **(2 points)**

   What distribution is used to compute the above p-value and what are the degrees of freedom?

   4

   χ^2 distribution

6. **(4 points)**

   What do you conclude from the above Analysis of Deviance table.

   Reduction means added predictors have explanatory power

7. Considering the class on 11/25/2022[1]

   (a) **(4 points)**
   This data point has a very large Cook's distance; explain what that means.

   (b) **(4 points)**
   This data point has a very large DFFIT; explain what that means.

[1] i.e. Black Friday

2 The Great Least Squares Estimator, 48 points

In this question, we consider the standard linear regression model $Y = X\beta + \varepsilon$ for $Y, \varepsilon \in \mathbb{R}^n$, $\beta \in \mathbb{R}^{p+1}$, and $X \in \mathbb{R}^{n\times(p+1)}$ where the first column of X is all 1's.

1. **(2 points)**

   Write down the least squares estimator for $\hat{\beta}$.

   $\hat{\beta} = (X^T X)^{-1} X^T Y$

2. **(4 points)**

   What assumptions are required for $\hat{\beta}$ to be the Minimal Variance Unbiased Estimator?

   $E[\varepsilon_i] = 0$

   $\mathrm{var}[\varepsilon_i] = \sigma^2$

   $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$

3. **(6 points)**

   Write down the formula for the model residuals, r, in terms of X and Y. Compute the mean and variance for r using the assumptions from question (2) above.

$r = Y - \hat{Y} = Y - X\hat{\beta} = Y - X(X^TX)^{-1}X^TY = (I - X(X^TX)^{-1}X^T)Y$

$= (I - P)Y$

$E[r] = (I-P)X\beta = X\beta - PX\beta = X\beta - X\beta = 0$

$\mathrm{Var}[r] = \mathrm{Var}[(I-P)Y] = (I-P)\mathrm{Var}(Y)(I-P)^T$

$= (I-P)\sigma^2 I (I-P)$

$= (I-P)^2 \sigma^2 I$

$= \sigma^2 (I-P)$

4. **(4 points)**

   Derive the covariance between r and $\hat{\beta}$.

$$\mathrm{cov}(r, \hat{\beta}) = \mathrm{cov}\left[(I-P)Y, (X^TX)^{-1}X^TY\right]$$
$$= (I-P)\,\mathrm{cov}(Y,Y)\,(X^TX)^{-1}X^T$$
$$= (I-P)\,\sigma^2 I\,[X(X^TX)^{-1}]$$
$$= (X-X)\,\sigma^2 (I)\,(X^TX)^{-1}$$
$$= \underline{\underline{0}}$$

5. **(4 points)**

   Compute the average of the residuals, i.e. $\frac{1}{n}\sum_{i=1}^{n} r_i$.

$$\frac{1}{n}\sum Y - \hat{Y} = \frac{1}{n}\cdot(I-P)Y$$
$$= \frac{1}{n}(Y - PY) \quad \downarrow \text{idempotent}$$
$$= \frac{1}{n}\cdot 0 = \underline{\underline{0}}$$

For the next part, consider the shrinkage estimator from the written assignment, i.e. $\tilde{\beta} = c\hat{\beta}$ for some $c \in (0, 1)$.

6. **(8 points)**

   Derive a formula for the residuals $\tilde{r}$ of $\tilde{\beta}$ in terms of X, Y, and c. Compute the mean and variance for $\tilde{r}$ using the assumptions from question (2) above.

$$\tilde{r} = Y - \tilde{Y} = Y - X\tilde{\beta} = Y - Xc\hat{\beta}$$

$$= Y - Xc(X^TX)^{-1}X^TY = Y - cX(X^TX)^{-1}X^TY$$

$$\tilde{r} = Y - cHY = (I - cH)Y$$

$$E[\tilde{r}] = E[(I - cH)Y] = (I - cH)E[Y]$$

$$= (I - cH)X\beta$$

$$Var[\tilde{r}] = (I - cH)Var\,Y(I - cH)^T$$

7. **(6 points)**

   Derive the covariance between $\tilde{r}$ and $\tilde{\beta}$.

$$Cov[(I - cH)Y, c(X^TX)^{-1}X^TY)$$

$$(I - cH)\,Cov(Y, Y)\,[c(X^TX)^{-1}X^T]^T$$

$$(I - cH)\,Var(Y)\,cX(X^TX)^{-1}$$

$$c\sigma^2(I - cH)(X^TX)^{-1}X$$

8. **(6 points)**

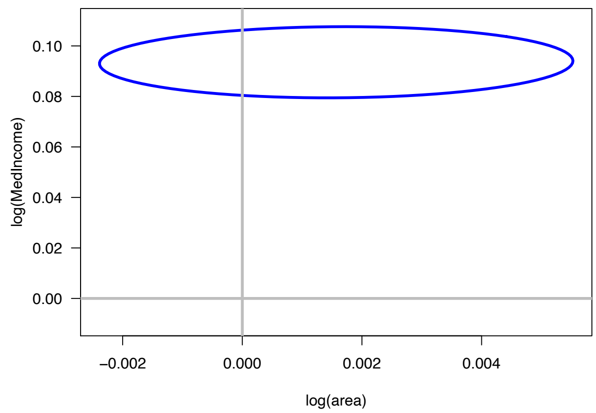
   Compute the average of the residuals, i.e. $\frac{1}{n}\sum_{i=1}^{n}\tilde{r}_i$.

9. **(8 points)**

   Assuming that $\bar{Y} = 0$, show that

   $$SS_{\text{res}}(\tilde{\beta}) := \sum_{i=1}^{n}(Y_i - \tilde{Y}_i)^2 = SS_{\text{res}} + (1-c)^2 SS_{\text{exp}}$$

   where SS_{res} and SS_{exp} are the sums of squares for the least squares estimator $\hat{\beta}$.



3 Choosing a Regression, 32 points

In this question, we attempt to predict which factors contribute to voter turnout in the USA across $n = 3104$ counties. This data comes from the 2016 election and self reporting on the 2010 census. The acronym "NHW" refers to non-hispanic white.

We first fit the model

$$\frac{\#\text{votes}}{\text{total pop}} = \beta_0 + \beta_1 \log(\text{area}) + \beta_2 \log(\text{median income})$$

and get the above 95% confidence ellipsoid for β_1 and β_2.

1. **(2 points)**

   Based on the above picture, how does increasing the area of a county affect voter turnout?

2. **(2 points)**

   Based on the above picture, how does increasing the median income of a county affect voter turnout?

We next fit the model

$$\frac{\#\text{votes}}{\text{total pop}} =$$

$$\beta_0 + \beta_1 \log(\text{area}) + \beta_2 \log(\text{median income}) + \beta_3 \frac{\#\text{males}}{\#\text{females}} + \beta_4 \log(\text{pop density}) +$$

$$\beta_5(\%\text{NHW Male}) + \beta_6(\%\text{NHW Female}) + \beta_7(\%\text{Trump Voters}) + \beta_8 \frac{\#\text{NHW males}}{\#\text{NHW females}}$$

Performing backwards variable selection with respect to AIC results in the removal of β_3 from the model. Here are the VIF values:

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
VIF Before Selection	1.55	1.38	16.27	2.39	275.0	279.1	1.96	8.72
VIF After Selection	1.55	1.37	–	2.37	55.2	56.9	1.95	1.25

3. **(6 points)**

   Use the above table to explain how the removal of β_3 affected the other variables in the model.

4. A $1-\alpha$ confidence interval for β_5 is

$$|\beta_5 - \hat{\beta}_5| \leq t_{1-\alpha/2}\sqrt{(X^{\mathrm{T}}X)^{-1}_{5,5} SS_{\mathrm{res}}/(n-p-1)}$$

When β_3 is removed from the model...

(a) **(2 points)**
What will likely happen to $t_{1-\alpha/2}$?

(b) **(2 points)**
What will likely happen to $(X^{\mathrm{T}}X)^{-1}_{5,5}$?

(c) **(2 points)**
What will likely happen to $SS_{\mathrm{res}}/(n-p-1)$?

Recall that if $Z \sim \mathcal{N}(\mu, \sigma^2 I_n)$, then the pdf is

$$f(z) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Z_i - \mu_i)^2 \right)$$

5. **(6 points)**

   For the standard linear regression model, $Y = X\beta + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$, show that the maximum log-likelihood is

   $$C - \frac{n}{2} \log\left(\frac{SS_{\text{res}}}{n} \right)$$

   for some constant C that does not depend on the data.

   Note: You don't have to rederive the MLEs $\hat{\beta}$ and $\hat{\sigma}^2$.

6. **(6 points)**

   AIC is $-2\log(\text{likelihood}) + 2(\#\text{parameters})$. If we have a regression model with p predictors and a model with $p-1$ predictors, but both have the same AIC, then find the relationship between their residual sums of squares, i.e. compute $SS_{\text{res}}(p-1)/SS_{\text{res}}(p)$.

7. **(4 points)**

   BIC is $-2\log(\text{likelihood}) + \log(n)(\#\text{parameters})$. Compute $SS_{\text{res}}(p-1)/SS_{\text{res}}(p)$ as in the previous question but for two models that have equal BIC values.

   Note: You can reuse the calculations from the previous question without rewriting them.