

**Question 4: The following data provide the survival among cancer patients. The data provides 35 common comorbidities for patients who have or don't have stomach cancer. Use both logistic and ordinary regression to analyze these data and report the difference of the findings, in particular:**

- Using logistic regression, calculate the propensity to have cancer.

```
Call:
glm(formula = StomachCancer$Cancer ~ ., family = binomial, data = Stoma
chCancer[,
      (3:26)])
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.3721	-0.0631	-0.0572	-0.0451	4.0626

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-6.33528	0.11854	-53.445	< 2e-16	***
I305.1	-0.54365	0.20087	-2.706	0.006801	**
I309.81	-1.09108	0.32968	-3.310	0.000934	***
I311.	-0.13418	0.22506	-0.596	0.551037	
IE849.7	0.51442	0.24582	2.093	0.036382	*
I150.9	2.18940	0.36313	6.029	1.65e-09	***
I276.1	-0.65321	0.35933	-1.818	0.069088	.
I276.8	0.21606	0.25336	0.853	0.393777	
I530.81	0.33163	0.16681	1.988	0.046801	*
I263.9	0.86061	0.28623	3.007	0.002641	**
I276.51	0.29481	0.25158	1.172	0.241274	
IV15.82	0.22798	0.26422	0.863	0.388233	
I511.9	0.79222	0.29313	2.703	0.006880	**
I401.9	0.11641	0.15024	0.775	0.438422	
I787.20	1.26492	0.24062	5.257	1.47e-07	***
I564.00	-0.24512	0.28118	-0.872	0.383331	
I272.4	-0.28226	0.16171	-1.746	0.080889	.
I280.9	1.01211	0.23635	4.282	1.85e-05	***
I285.9	-0.13561	0.22255	-0.609	0.542316	
I496.	-0.01069	0.20064	-0.053	0.957510	
I458.9	-0.19664	0.32585	-0.603	0.546211	
I486.	-0.81208	0.31403	-2.586	0.009711	**
IV58.61	-0.07988	0.29444	-0.271	0.786177	
I197.7	2.42379	0.22666	10.694	< 2e-16	***
I578.9	0.91813	0.32789	2.800	0.005108	**

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2972.2 on 99999 degrees of freedom  
Residual deviance: 2715.3 on 99975 degrees of freedom  
AIC: 2765.3

Number of Fisher Scoring iterations: 10

```
> pscore<-model2$fitted.values
```

```
> summary(pscore)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000041	0.001017	0.001634	0.002070	0.001988	0.609900

- Group the diagnoses using SQL. Within the naturally occurring groups of diagnoses, calculate probability of cancer. Calculate the logit of the probability. Regress the logit function on the diagnoses using ordinary regression. [SQL](#)

```
Call:
lm(formula = Logit ~ I305.1 + I309.81 + I311 + IE849.7 + I150.9 +
  I276.1 + I276.8 + I530.81 + I263.9 + I276.51 + IV15.82 +
  I511.9 + I401.9 + I787.20 + I564.00 + I272.4 + I280.9 + I285.9 +
  I496 + I458.9 + I486 + IV58.61 + I197.7 + I578.9, data = StomachCan
cerCleaned)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-3.8894 -0.4494  0.2658  0.6486  1.5234
```

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-3.78093	0.08250	-45.829	< 2e-16	***
I305.1	-0.04615	0.08543	-0.540	0.589229	
I309.81	0.02136	0.10292	0.208	0.835643	
I311	0.17804	0.10379	1.715	0.086730	.
IE849.7	0.64776	0.17340	3.736	0.000202	***
I150.9	NA	NA	NA	NA	
I276.1	0.68612	0.20339	3.373	0.000783	***
I276.8	0.61840	0.17213	3.593	0.000350	***
I530.81	0.13216	0.08685	1.522	0.128535	
I263.9	0.19741	0.98235	0.201	0.840789	
I276.51	0.63477	0.23767	2.671	0.007740	**
IV15.82	0.49201	0.16091	3.058	0.002314	**
I511.9	0.79323	0.49543	1.601	0.109795	
I401.9	0.00105	0.07569	0.014	0.988937	
I787.20	0.84143	0.33219	2.533	0.011523	*
I564.00	0.61642	0.17754	3.472	0.000548	***
I272.4	0.19032	0.07755	2.454	0.014365	*
I280.9	0.58381	0.31522	1.852	0.064434	.
I285.9	0.64023	0.14223	4.501	7.88e-06	***
I496	0.52331	0.10138	5.162	3.17e-07	***
I458.9	0.84465	0.37594	2.247	0.024958	*
I486	0.62174	0.20363	3.053	0.002347	**
IV58.61	0.63369	0.20085	3.155	0.001672	**
I197.7	1.73684	0.57038	3.045	0.002412	**
I578.9	0.66744	0.56946	1.172	0.241563	

---

signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9789 on 714 degrees of freedom  
 Multiple R-squared: 0.1229, Adjusted R-squared: 0.09462  
 F-statistic: 4.349 on 23 and 714 DF, p-value: 1.026e-10