1. Since (a) checks if Age is a risk factor for being a case and (b) checks whether or not tobacco usage is associated with age and both are not equal to 0 (as seen from the values of β_1 for both (a) and (b)) this must mean that Age is a confounder.

(a) $\beta_1 = 1.3886$ is the log odds ratio for being a case in subjects with x = 1 (older than 45) versus x = 0 (younger than 45). The Odds Ratio would therefore be 4.009

(b) $\beta_1 = 0.6874$ is the log odds ratio for subjects who consumed any amount of tobacco with x = 1 (older than 45) versus x = 1 (younger than 45). The Odds Ratio would therefore be 1.989

2.

- $\beta_0 = -4.6732$ is the log odds ratio of being a case for subjects with x_1= 0 (no tobacco consumption) and x_2= 0 (age < 55 years old) and x_3= 0 (low or no alcohol consumption).
- $\beta_1 = 2.1786$ is the log odds ratio (odds ratio of 8.834) of being a case in subjects with x_1 = 1 (any tobacco consumption) versus x_1= 0 (no tobacco consumption) adjusted for (holding at constant values) x_2 age group and $x_3$alcohol intake.
- $\beta_2 = 1.2813$ is the log odds ratio (odds ratio of 3.601) of being a case in subjects with x_2 = 1 (age > 55) versus x_2= 0 (age < 55) adjusted for (holding at constant values) x_1 tobacco intake and $x_3$alcohol intake.
- $\beta_3 = 0.8162$ is the log odds ratio (odds ratio of 2.262) of being a case for subjects with x_1= 1 (any tobacco consumption) and x_2= 1 (age > 55 years old) and x_3= 1 (high alcohol consumption). of being a case in subjects with x_3= 1 (high alcohol consumption) versus x_3= 0 (low alcohol consumption) adjusted for (holding at constant values) $x_1$tobacco intake and x_2 age group.
- Alcohol consumption has an odds ratio of 2.262 which means that there is a 2.26x higher likelihood for heavy drinkers to get esophageal cancer

```
DATA cancer;
INFILE '/home/u58684395/tutorial7/tuynsc.txt';
INPUT Case Age AgeGP TobaccoGP TobaccoAMT logT1 Beer Cider Wine Aperitif Digestive
TotalA LogTotalA1;
AlcoholGP = 0;
AgeGP2 = 0;
TobaccoGP2 = 0;
IF TotalA ge 10 THEN AlcoholGP=1;
IF AgeGP>3 then AgeGP2=1;
IF TobaccoGP>0 then TobaccoGP2=1;
RUN;
```

```
PROC FORMAT;
VALUE CCFmt 1='Case' 0='Control';
VALUE AGEGPFmt 1='25-34' 2='35-44' 3='45-54' 4='55-64' 5='65-74' 6='75+';
VALUE TOBGPFmt 0='None' 1='1-4' 2='5-9' 3='10-14' 4='15-19' 5='20-29' 6='30-39'
7='40-49' 8='50+' 9='Unknown';
VALUE TOBGP2Fmt 0='None' 1='Any';
VALUE AGEGP2Fmt 0='< 55' 1='55+';
RUN;
```

```
PROC LOGISTIC DATA=cancer DESCENDING;
    MODEL Case=AgeGP2;
RUN;
PROC LOGISTIC DATA=cancer DESCENDING;
    MODEL TobaccoGP2=AgeGP2;
    WHERE TobaccoGP<9; /* 9 is unknown */
RUN;
```

```
PROC LOGISTIC DATA=cancer DESCENDING;
    MODEL Case=TobaccoGP2 AgeGP2 AlcoholGP;
    WHERE TobaccoGP<9;
RUN;
```

The LOGISTIC Procedure

Model Information	
Data Set	WORK.CANCER
Response Variable	Case
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	978
Number of Observations Used	978

Response Profile		
Ordered Value	Case	Total Frequency
1	1	200
2	0	778

Probability modeled is Case=1.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	992.863	924.221
SC	997.749	933.992
-2 Log L	990.863	920.221

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	70.6421	1	<.0001
Score	69.4797	1	<.0001
Wald	63.8477	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.1358	0.1413	228.4712	<.0001
AgeGP2	1	1.3886	0.1738	63.8477	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
AgeGP2	4.009	2.852	5.636

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	43.9	Somers' D	0.329
Percent Discordant	10.9	Gamma	0.601
Percent Tied	45.2	Tau-a	0.107
Pairs	155600	c	0.665

The LOGISTIC Procedure

Model Information	
Data Set	WORK.CANCER
Response Variable	TobaccoGP2
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	976
Number of Observations Used	976

Response Profile

Ordered Value	Response Profile TobaccoGP2	Total Frequency

Ordered Value	TobaccoGP2	Total Frequency
1	1	712
2	0	264

Probability modeled is TobaccoGP2=1.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1141.475	1121.817
SC	1146.358	1131.584
-2 Log L	1139.475	1117.817

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	21.6579	1	<.0001
Score	21.2986	1	<.0001
Wald	20.9363	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.7045	0.0924	58.1248	<.0001
AgeGP2	1	0.6874	0.1502	20.9363	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
AgeGP2	1.989	1.481	2.669

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	33.3	Somers' D	0.166
Percent Discordant	16.8	Gamma	0.331
Percent Tied	49.9	Tau-a	0.065
Pairs	187968	c	0.583

The LOGISTIC Procedure

Model Information	
Data Set	WORK.CANCER
Response Variable	Case
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	976
Number of Observations Used	976

Response Profile		
Ordered Value	Case	Total Frequency
1	1	200
2	0	776

Probability modeled is Case=1.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	991.947	852.072

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
SC	996.831	871.606
-2 Log L	989.947	844.072

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	145.8749	3	<.0001
Score	122.8314	3	<.0001
Wald	96.7857	3	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.6732	0.4780	95.5848	<.0001
TobaccoGP2	1	2.1786	0.3532	38.0464	<.0001
AgeGP2	1	1.2813	0.1792	51.1300	<.0001
AlcoholGP	1	0.8162	0.3436	5.6431	0.0175

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
TobaccoGP2	8.834	4.421	17.652
AgeGP2	3.601	2.535	5.117
AlcoholGP	2.262	1.153	4.436

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	61.9	Somers' D	0.495
Percent Discordant	12.3	Gamma	0.668
Percent Tied	25.8	Tau-a	0.162
Pairs	155200	c	0.748