```
from sklearn.linear_model import LogisticRegression
import pandas as pd
import numpy as np
from scipy.stats import norm
```

A.

$$\pi_i = \pi(\beta X_i) = \pi(\beta_0 + \beta_{alcohol_i} * X_{alcohol_i} + \beta_{BMI_i} X_{BMI_i} + \beta_{age_i} X_{age_i}$$

B.

```
#load the data
df = pd.read_csv("/data/workspace_files/ex4.csv")
X = df[["alcohol", "BMI", "age"]]
X.insert(0, 'one', 1, True)
y = df["chd"]
```

```
reg = LogisticRegression(penalty="none", fit_intercept=False).fit(X, y)
beta_star_hat = reg.coef_[0]
a= [print(f"the beta_{x[0]} is {x[1]}") for x in zip(["zero", "alcohol", "BMI",
```

```
the beta_zero is -9.906057116302838
the beta_alcohol is 0.019612208107966704
```

```
the beta_BMI is -0.021686645217299128
the beta_age is 0.20531261330705375
```

C.

```
all_log_probs = reg.predict_log_proba(X)
sum_log_probs = sum([all_log_probs[i][yi] for i, yi in enumerate(y)])
print(f"the log likelyhood is {sum_log_probs}")
```

```
the log likelyhood is -6.827115831144858
```

D.

```
all_probs = reg.predict_proba(X)
thing = []
for p_1, p_2 in all_probs:
    thing.append(p_1*p_2)
V = np.diag(thing)
var = np.linalg.inv(X.T@V@X)

print(f"The variance is {var}")
```

```
The variance is [[ 5.87559147e+01 -2.59985608e-02 -1.17187406e+00 -5.2853908
 [-2.59985608e-02  7.44952162e-04  7.03496142e-04 -1.78907336e-04]
 [-1.17187406e+00  7.03496142e-04  5.35554305e-02 -5.50322653e-03]
 [-5.28539086e-01 -1.78907336e-04 -5.50322653e-03  1.35966443e-02]]
```

E.

```
X_sample = np.array([1, 8, 27,50])

logit = X_sample@np.array(beta_star_hat).T
p = np.exp(logit)/(1+np.exp(logit))
```

```
0.48273980924207194
```

```
se = (X_sample @ var @ X_sample)**0.5
z_alpha = norm.ppf(0.975)
almost_ci = [p-se*z_alpha, p+se*z_alpha]
ci = [1/(1+np.exp(-x)) for x in almost_ci]
print(se)
print(f"the confindence interval is {ci}")
```

```
0.7668761040764567
the confindence interval is [0.26496707223277977, 0.8792969588274144]
```