Given the proportions of the three-letter words found in 8 Twain essays vs. 10 Snodgrass essays, let us investigate whether or not the Snodgrass essays were written by Mark Twain.

Part a

Let us first perform the Wald test for equality of the means. In other words, so solve for the p-value, we will solve $W(X) = z_{1-\frac{\alpha}{2}}$ such that $p(X) = 2\Phi(-W(X))$.

```
twain = [0.225, 0.262, 0.217, 0.240, 0.230, 0.229, 0.235, 0.217];
snodgrass = [0.209, 0.205, 0.196, 0.210, 0.202, 0.207, 0.224, 0.223, 0.220, 0.201];
n = length(twain);
m = length(snodgrass);

W = abs((mean(twain) - mean(snodgrass)) / sqrt((var(twain)/n) + (var(snodgrass)/m)));
p_value = 2*normcdf(-1*W);
disp("p-value: "); disp(p_value);
```

```
p-value:
   2.1260e-04
```

Since the p-value is very small and close to 0 (less than 0.01), we have very strong evidence against the null hypothesis. Thus, we conclude that it is very likely that Twain did not write the Snodgrass essays.

Part b

Let us now perform the permutation test in order to avoid large sample normality assumptions.

```
K = 10^5;
s_obs = abs(mean(twain) - mean(snodgrass));
Z = [twain snodgrass];

count = 0;
for i = 1:K
    Z_pi = Z(:,randperm(n + m));
    s_pi = abs(mean(Z_pi(1:n)) - mean(Z_pi(n+1:n+m)));
    if s_pi > s_obs
        count = count + 1;
    end
end
p = count / K;
disp("Estimated p-value: "); disp(p);
```

```
Estimated p-value:
   7.5000e-04
```

Furthermore, since this estimated p-value is also very small and close to zero, we make the same conclusions as in part a such that we have very strong evidence against the null hypothesis. Thus, it is very strongly suggested that Twain did not write the Snodgrass essays.