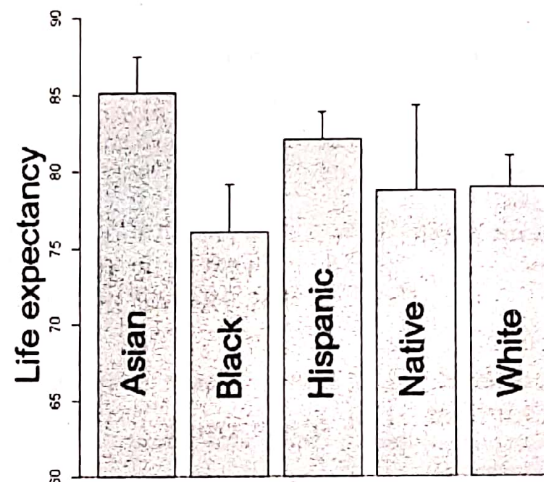**Question 1.** In order to describe how life expectancy depends on ethnic origins in US residents, life expectancy was measured in the 51 states for each ethnic group, and an ANOVA was conducted. The results are shown below.

Descriptive statistics:

Mean life expectancy ± standard deviation
Asian: 85.12 ± 2.35
Black: 76.06 ± 3.10
Hispanic: 82.06 ± 1.82
Native: 78.69 ± 5.57
White: 78.91 ± 2.02

ANOVA results:

```
> summary(AOVres)
            Df Sum Sq Mean Sq F value Pr(>F)
dat$Origin   4   2484   620.9   57.98 <2e-16 ***
Residuals  250   2677    10.7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


> TukeyHSD(AOVres)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = dat$LifeExp ~ dat$Origin)

$`dat$Origin`
                      diff        lwr       upr     p adj
Black-Asian     -9.0568627 -10.8375479 -7.276178 0.0000000
Hispanic-Asian  -3.0607843  -4.8414694 -1.280099 0.0000380
Native-Asian    -6.4274510  -8.2081361 -4.646766 0.0000000
White-Asian     -6.2058824  -7.9865675 -4.425197 0.0000000
Hispanic-Black   5.9960784   4.2153933  7.776764 0.0000000
Native-Black     2.6294118   0.8487267  4.410097 0.0006301
White-Black      2.8509804   1.0702953  4.631666 0.0001557
Native-Hispanic -3.3666667  -5.1473518 -1.585982 0.0000042
White-Hispanic  -3.1450980  -4.9257831 -1.364413 0.0000210
White-Native     0.2215686  -1.5591165  2.002254 0.9970467
```

Write the results like in a scientific report.

Question 2. We want to know whether the causes of death are the same between ethnic groups. On a given area and during a given time we count the number of death due to heart disease ("HeartDis"), cancers, stroke, violent death (crimes, injuries and suicides) and other causes. The results are expressed in count numbers and in percentage per group.

Raw data:

	White	Black	Native	Asian	Hispanic
HeartDis	254	327	165	145	165
Cancers	198	250	127	125	121
Stroke	52	82	40	53	39
ViolentDeath	60	65	80	27	45
Other	13	34	32	9	23

Percentage by ethnic group

	White	Black	Native	Asian	Hispanic
HeartDis	44.02	43.14	37.16	40.39	41.98
Cancers	34.32	32.98	28.60	34.82	30.79
Stroke	9.01	10.82	9.01	14.76	9.92
ViolentDeath	10.40	8.58	18.02	7.52	11.45
Other	2.25	4.49	7.21	2.51	5.85
	100.00	100.00	100.00	100.00	100.00

(handwritten annotations: 78 next to White, 75 next to Black, 65 next to Native, 74 next to Asian, 71 next to Hispanic)

The results of a Fisher exact test on raw data are shown below:

```
    Fisher's Exact Test for Count Data with simulated p-value (based on 50000
replicates)

data:  Causes
p-value = 0.00002
alternative hypothesis: two.sided
```

i. Write the results like in a scientific report.

ii. Which statistical tests would you perform to better support or verify your conclusions?

Question 3. The results of a correspondence analysis (CA) are shown below. As compared to principal component analysis,

- Inertia is used like eigen values
- only the relative locations of levels should be interpreted in CA (i.e. long distances between two labels indicate strong differences between the two levels respectively).

```
Inertia information:
Call: inertia.dudi(x = CA)

Decomposition of total inertia:
         inertia      cum    cum(%)
Ax1 1.964e-02 0.01964    79.10
Ax2 3.285e-03 0.02293    92.33
Ax3 1.897e-03 0.02482    99.97
Ax4 8.520e-06 0.02483   100.00
```
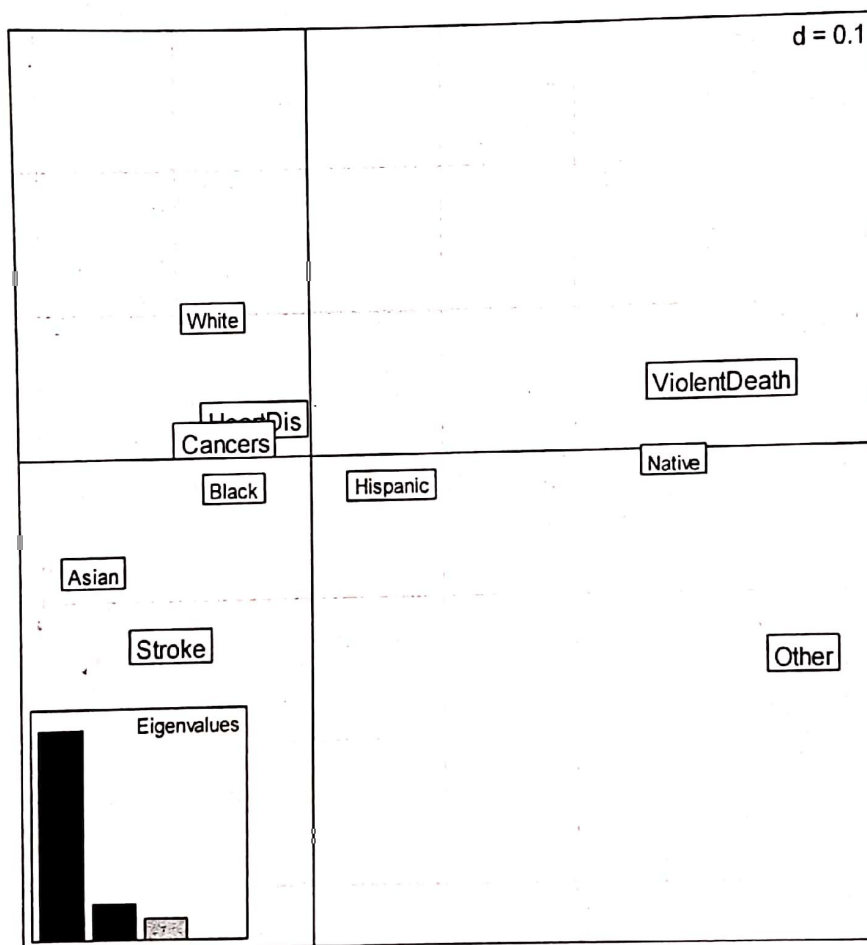

Describe the results of the correspondence analysis and compare them to your conclusions in question 2.