EMET8005 Tutorial exercises #1

Exercise 1 Download the file `CAschools.dta` from the Wattle course page. The units of observation are all the K-6 and K-8 school districts in California in 1999. The sample size is 420.

(a) Set the working directory to somewhere sensible.

(b) Load the data into Stata.

(c) Explore what variables are in the dataset. Use `browse`, `list` and `describe`. You may want to `set more on` so that things don't fly past too fast in the `Results` window.

(d) List the observations 56–60 for variables `dist_cod`, `county`, `stratio`, `testscr`.

(e) List the observations for variables `dist_cod`, `county`, `stratio`, `testscr` for Kern county.

(f) Create a histogram for average income. Create histograms with half as many bins and double the number of bins.

(g) Create a new variable for the logarithm of average income.

(h) Create a histogram for the logarithm of average income. Add a suitable title, and change the axis titles if needed. Make the bars blue.

(i) Find out how many districts have more than 1000 computers (or had in 1999).

(j) Check if `stratio` is equal to `enrl_tot` divided by `teachers` for all observations. *Hint:* One way to do this is to create a new variable as `stratio-enrl_tot/teachers` and check that the minimum and the maximum values are 0, apart from negligible rounding. Another way is to use the `assert` command, again with some allowance for rounding.

(k) Get summary statistics for `avginc` if `avginc` is lower than the median value. *Hint:* You can get the median value with the `summarize` command and the `details` option.

(l) Get summary statistics for `avginc` in Kern county.

(m) Use the following code to create a categorical variable for the student-teacher ratio:

```
gen cat=.
replace cat=1 if stratio<=17
replace cat=2 if 17<stratio&stratio<=20
replace cat=3 if 20<stratio
```

Tabulate `cat` and verify that there are no missing values left.

(n) Use the `display` command to find the binomial probability of getting 8 heads out of 19 coin tosses and the probability of getting 8 or more heads, when the coin comes up heads 53% of the time. *Hint:* Use the `help density functions` command to find the syntax for the functions `binomialp` and `binomial`.

(o) Get summary statistics for the variables in the dataset whose name begin with 'c'.

(p) Get summary statistics for the variables in the dataset whose name ends with 'pct'.

(q) Make a copy of the variable `stratio`, ie create a new variable that is equal to `stratio`. Then replace the values of the variable with missings for observations where average income is less than 20. Check your effort using `browse` or `list`. Summarise `stratio` and the new variable.

(r) Drop all the new variable you created from the dataset.

(s) Summarize the variable `testscr`.

(t) Use the `mean` command to compute the mean of `testscr`, the standard error of the mean, and a 95% confidence interval for the mean.

Exercise 2 Suppose V_1 and V_2 are random variables and c_1, c_2 and c_0 are constants. Show that

$$\mathrm{Cov}(c_1 V_1, c_2 V_2 + c_0) = c_1 c_2 \,\mathrm{Cov}(V_1, V_2).$$

Hint: Use the definitions of the variance and covariance and rule $\mathrm{E}(c_1 V_1 + c_2 V_2 + c_0) = c_1 \,\mathrm{E}(V_1) + c_2 \,\mathrm{E}(V_2) + c_0$.

Exercise 3 Let V be a random variable with density $p(v) = 1/(\omega - \alpha)$ for $\alpha \leq v \leq \omega$ and $p(v) = 0$ otherwise, where α and ω are positive parameters with $\alpha < \omega$.

(a) Derive the cumulative distribution function (cdf) of V.

(b) Derive the expected value $\mathrm{E}(V)$.

(c) Derive the variance $\mathrm{Var}(V)$.

Exercise 4 Let A and Z be discrete random variables with the following (joint) distribution:

Value of A	Value of Z	
	1	3
2	0.10	0.40
4	0.30	0.20

That is, $\mathrm{P}(A = 2, Z = 1) = 0.10$ and so forth.

(a) Calculate the (marginal) probability distribution, mean and variance of A.

(b) Calculate the (marginal) probability distribution, mean and variance of Z.

(c) Calculate the covariance of A and Z and the correlation coefficient.

(d) Define the random column vector $\boldsymbol{V} = (A, Z)'$. What are the mean vector $\mathrm{E}(\boldsymbol{V})$ and the variance matrix $\mathrm{Var}(\boldsymbol{V})$?