

Part I. (60 points) Do all calculations in SAS. Use a word processor of your choice to write a report. Insert computer text output and graphics to support what you are saying, but you need to write something that looks like an academic paper — not a pile of computer output. Turn in a hard copy of your HW in class (i.e., don't email me your HW). Also:

1. Clearly specify parameters and hypotheses when appropriate.
2. Write a coherent conclusion.

(60^{pts}) **1. Basal body temperature:** These data give the average menstrual cycle length and average preovulatory basal body temperature for a sample of 21 women¹.

60 pts

Cycle length (days)	BBT (degrees C)
22.9	36.44
26.3	36.21
26.6	36.71
26.6	36.11
26.9	36.52
26.9	36.53
27.1	36.52
27.1	36.53
27.1	36.45
27.1	36.53
28.0	36.31
28.4	36.63
28.4	36.55
28.8	36.44
28.8	36.55
28.8	36.52
28.8	36.52
28.8	36.52
28.8	36.52
29.1	36.48
29.1	36.48
29.1	36.39
29.1	36.39
30.0	36.37
30.3	36.77
31.2	36.76
31.8	36.50

Create a SAS program, and provide the program as part of your writeup, to read in the data, and carry out the subsequent parts of the HW. It would be preferable for you create a text file with the data in it and use an INFILE statement to read the data in from that file. Include relevant comments as part of the SAS program; see the class notes for how to do this.

(a) (10 pts) In the DATA step, create two new variables from the cycle length and BBT. The variables to be created are BBT in degrees Fahrenheit, and a CHARACTER variable called CL, defined by

```
CL = long if cycle length greater than 28
    = short if cycle length less than or equal to 28.
```

Note: If you precede the IF ... THEN ... ELSE statements that you should use to create CL by the SAS statement:

```
length cl $ 5;
```

then 5 characters will be used when printing out the levels of CL. Otherwise, the number of characters used in the print out will equal the length of the first level of CL defined in the If statement (4 instead of 5).

- (b) (10 pts) Print out the entire data set, with an appropriate title given, and labels for the variables.
- (c) (10 pts) Print out univariate numerical and graphical summaries for the numerical cycle length variable (not CL!). Describe relevant features of the data set.
- (d) (10 pts) Print out univariate numerical and graphical summaries for the BBT variable in degrees Fahrenheit BY levels of CL. Discuss the side-by-side box plots.
- (e) (10 pts) Plot the BBT in degrees Fahrenheit against cycle length (not CL). Describe what the plot tells you.
- (f) (10 pts) Let μ be the population mean menstrual cycle length for women. It is commonly accepted that $\mu = 28$ days. Devise a way to use the UNIVARIATE procedure to compute a p-value for a two-sided test of $H_0 : \mu = 28$. No hand calculations allowed. Would you reject H_0 at the 5% level? Comment on the results of the test.

¹Source: Royston, J.P. and Abrams, R.M. (1980). An objective method for detecting the shift in basal body temperature in women. *Biometrics*, 36, 217-224.

60 pts