

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
26.6	36.21
26.8	36.71
26.8	36.11
26.9	36.25
26.9	36.53
27.5	36.41
27.5	36.51
27.6	36.45
27.7	36.53
28.0	36.51
28.4	36.53
28.4	36.44
28.5	36.22
28.8	36.62
28.8	36.40
29.4	36.43
29.9	36.39
30.0	36.37
30.6	36.77
31.3	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).

*Solution:* Program editor contents:

```
* options for sas session;
options ls=79 nodate nocenter;

* part (a) *****;
* read data, create variables, assign labels;
data menst;
 * read data file;
 infile 'F:\Dropbox\UNM\teach\ADA2_stat528\assess\ADA2_HW_01_bbt.dat';

 * assign variables to columns in infile;
 input cycle bbt;

 * create categorical cycle length variable;
 length cl $ 5; * a string $ of length 5;
```

<sup>1</sup>Source: 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 |

```

if cycle > 28
 then cl = 'long';
 else cl = 'short';

bbtF = bbt*9/5+32; * bbt in degrees Fahrenheit;

label cycle = 'menstrual cycle length (days)'
 bbt = 'basal body temp (degrees C)'
 bbtF = 'basal body temp (degrees F)'
 cl = 'menstrual cycle length (>28 days is long)';

run;

```

(b) (10 pts) Print out the entire data set, with an appropriate title given, and labels for the variables.

*Solution:* Program editor contents:

```

* part (b) *****;
* print dataset with title and variable labels;
title 'average menstrual cycle length and average preovulatory basal body temperature';
proc print data=menst label;
run;

```

Output window contents:

average menstrual cycle length and average preovulatory basal body temperature

| Obs | menstrual<br>cycle length<br>(days) | basal body<br>temp<br>(degrees<br>C) | menstrual<br>cycle length<br>(>28 days<br>is long) | basal body<br>temp<br>(degrees<br>F) |
|-----|-------------------------------------|--------------------------------------|----------------------------------------------------|--------------------------------------|
| 1   | 22.9                                | 36.44                                | short                                              | 97.592                               |
| 2   | 26.3                                | 36.21                                | short                                              | 97.178                               |
| 3   | 26.6                                | 36.71                                | short                                              | 98.078                               |
| 4   | 26.8                                | 36.13                                | short                                              | 97.034                               |
| 5   | 26.9                                | 36.25                                | short                                              | 97.250                               |
| 6   | 26.9                                | 36.53                                | short                                              | 97.754                               |
| 7   | 27.5                                | 36.41                                | short                                              | 97.538                               |
| 8   | 27.6                                | 36.53                                | short                                              | 97.754                               |
| 9   | 27.6                                | 36.45                                | short                                              | 97.610                               |
| 10  | 28.0                                | 36.31                                | short                                              | 97.358                               |
| 11  | 28.4                                | 36.63                                | long                                               | 97.934                               |
| 12  | 28.4                                | 36.54                                | long                                               | 97.772                               |
| 13  | 28.5                                | 36.52                                | long                                               | 97.736                               |
| 14  | 28.8                                | 36.62                                | long                                               | 97.916                               |
| 15  | 28.8                                | 36.40                                | long                                               | 97.520                               |
| 16  | 29.4                                | 36.48                                | long                                               | 97.664                               |
| 17  | 29.9                                | 36.39                                | long                                               | 97.502                               |
| 18  | 30.0                                | 36.37                                | long                                               | 97.466                               |
| 19  | 30.3                                | 36.77                                | long                                               | 98.186                               |
| 20  | 31.2                                | 36.76                                | long                                               | 98.168                               |
| 21  | 31.8                                | 36.50                                | long                                               | 97.700                               |

(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.

*Solution:* Program editor contents:

```

* part (c) *****;
title 'statistics for variable cycle';
proc univariate data=menst plot;
 var cycle;
run;

```

Mean cycle length is 28.2 with a standard error of the mean of 0.42 days. By the box plot, the distribution is right-skewed with an extreme left outlier at 22.9 days.

Output window contents:

statistics for variable cycle

The UNIVARIATE Procedure  
Variable: cycle (menstrual cycle length (days))

|                 |            | Moments          |            |
|-----------------|------------|------------------|------------|
| N               | 21         | Sum              | Weights    |
| Mean            | 28.2190476 | Sum Observations | 592.6      |
| Std Deviation   | 1.94515271 | Variance         | 3.78361905 |
| Skewness        | -0.5722681 | Kurtosis         | 1.7469272  |
| Uncorrected SS  | 16798.28   | Corrected SS     | 75.672381  |
| Coeff Variation | 6.89304874 | Std Error Mean   | 0.42446712 |

| Location | Basic Statistical Measures |                     |         |
|----------|----------------------------|---------------------|---------|
|          |                            | Variability         |         |
| Mean     | 28.21905                   | Std Deviation       | 1.94515 |
| Median   | 28.40000                   | Variance            | 3.78362 |
| Mode     | 26.90000                   | Range               | 8.90000 |
|          |                            | Interquartile Range | 2.50000 |

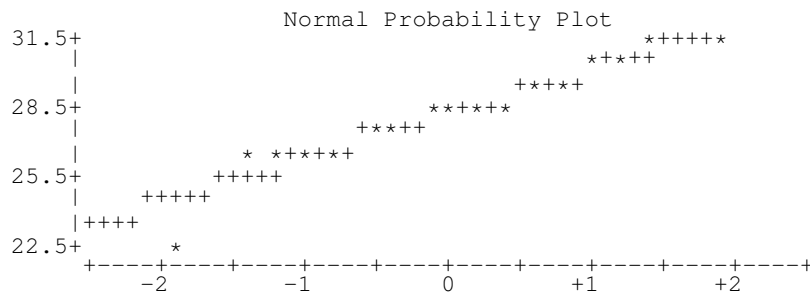
NOTE: The mode displayed is the smallest of 4 modes with a count of 2.

| Tests for Location: Mu0=0 |             |                   |        |
|---------------------------|-------------|-------------------|--------|
| Test                      | -Statistic- | -----p Value----- |        |
| Student's t               | t 66.48112  | Pr >  t           | <.0001 |
| Sign                      | M 10.5      | Pr >=  M          | <.0001 |
| Signed Rank               | S 115.5     | Pr >=  S          | <.0001 |

| Quantiles (Definition 5) |          |
|--------------------------|----------|
| Quantile                 | Estimate |
| 100% Max                 | 31.8     |
| 99%                      | 31.8     |
| 95%                      | 31.2     |
| 90%                      | 30.3     |
| 75% Q3                   | 29.4     |
| 50% Median               | 28.4     |
| 25% Q1                   | 26.9     |
| 10%                      | 26.6     |
| 5%                       | 26.3     |
| 1%                       | 22.9     |
| 0% Min                   | 22.9     |

| Extreme Observations |     |                |     |
|----------------------|-----|----------------|-----|
| ----Lowest----       |     | ----Highest--- |     |
| Value                | Obs | Value          | Obs |
| 22.9                 | 1   | 29.9           | 17  |
| 26.3                 | 2   | 30.0           | 18  |
| 26.6                 | 3   | 30.3           | 19  |
| 26.8                 | 4   | 31.2           | 20  |
| 26.9                 | 6   | 31.8           | 21  |

| Stem | Leaf   | # | Boxplot   |
|------|--------|---|-----------|
| 31   | 28     | 2 |           |
| 30   | 03     | 2 |           |
| 29   | 49     | 2 | +-----+   |
| 28   | 044588 | 6 | *---+---* |
| 27   | 566    | 3 |           |
| 26   | 36899  | 5 | +-----+   |
| 25   |        |   |           |
| 24   |        |   |           |
| 23   |        |   |           |
| 22   | 9      | 1 | 0         |



(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.

Solution: Program editor contents:

```

* part (d) *****;
title 'sort by cl';
proc sort data=menst;
 by cl;
run;

* calculate summaries of bbt by cl cycle length;
title 'statistics for bbt by cl';

```

```
proc univariate data=menst plot;
 var bbtF;
 by cl;
run;
```

The mean basal body temperature in degrees Fahrenheit (bbtF) is slightly higher (0.26 degrees higher) for those women with longer menstrual cycle lengths. Spread is slightly wider for shorter cycle lengths.

Output window contents:

```
statistics for bbtF by cl
menstrual cycle length (>28 days is long)=long
```

```
The UNIVARIATE Procedure
Variable: bbtF (basal body temp (degrees F))
```

|                 |            | Moments |                  |
|-----------------|------------|---------|------------------|
| N               |            | 11      | Sum Weights      |
| Mean            | 97.7785455 |         | Sum Observations |
| Std Deviation   | 0.24958059 |         | Variance         |
| Skewness        | 0.48775453 |         | Kurtosis         |
| Uncorrected SS  | 105167.706 |         | Corrected SS     |
| Coeff Variation | 0.25525088 |         | Std Error Mean   |

|          |          | Basic Statistical Measures |         |
|----------|----------|----------------------------|---------|
| Location |          | Variability                |         |
| Mean     | 97.77855 | Std Deviation              | 0.24958 |
| Median   | 97.73600 | Variance                   | 0.06229 |
| Mode     | .        | Range                      | 0.72000 |
|          |          | Interquartile Range        | 0.41400 |

```
menstrual cycle length (>28 days is long)=short
```

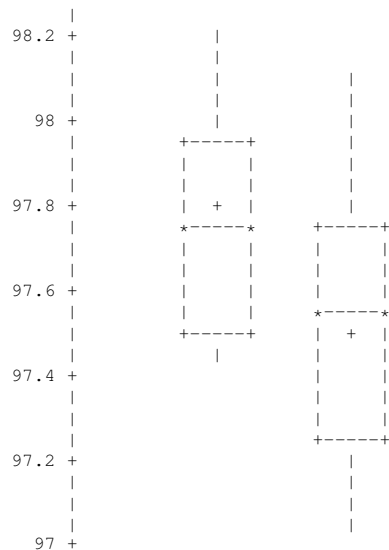
```
The UNIVARIATE Procedure
Variable: bbtF (basal body temp (degrees F))
```

|                 |            | Moments |                  |
|-----------------|------------|---------|------------------|
| N               |            | 10      | Sum Weights      |
| Mean            | 97.5146    |         | Sum Observations |
| Std Deviation   | 0.31384646 |         | Variance         |
| Skewness        | 0.16223608 |         | Kurtosis         |
| Uncorrected SS  | 95091.8586 |         | Corrected SS     |
| Coeff Variation | 0.32184561 |         | Std Error Mean   |

|          |          | Basic Statistical Measures |         |
|----------|----------|----------------------------|---------|
| Location |          | Variability                |         |
| Mean     | 97.51460 | Std Deviation              | 0.31385 |
| Median   | 97.56500 | Variance                   | 0.09850 |
| Mode     | 97.75400 | Range                      | 1.04400 |
|          |          | Interquartile Range        | 0.50400 |

```
The UNIVARIATE Procedure
Variable: bbtF (basal body temp (degrees F))
```

Schematic Plots





```
Tests for Normality
Test --Statistic-- -----p Value-----
Shapiro-Wilk W 0.953874 Pr < W 0.4022
Kolmogorov-Smirnov D 0.114305 Pr > D >0.1500
Cramer-von Mises W-Sq 0.038964 Pr > W-Sq >0.2500
Anderson-Darling A-Sq 0.320171 Pr > A-Sq >0.2500
```