KEY BIOSTAT EXAM

HS 167 Exam Key October 29, 2009

1. Data table [5 pts]

[INDICATES ONE POINT]

Table 1.2 Per capita cigarette consumption in 1930 (CIG1930) and lung cancer cases per 100,000 in 1950 (LUNGCA) in 11 countries.

COUNTRY	CIG1930	LUNGCA
USA	1300	20
Great Britain	1100	46
Finland	1100	35
Switzerland	510	25
Canada	500	15
Holland	490	24
Australia	480	18
Denmark	380	17
Sweden	300	11
Norway	250	9
Iceland	230	6

(a) What are the ~~individuals~~ (UNITS OF OBSERVATION) ~~in this data table~~?

[Countries]

(b) What variables are in this data table?

[CIG1930]
[LUNGCA]

(c) Identify the measurement scale of each variable as either *categorical* or *quantitative*.

[CIG1930 is quantitative]
[LUNGCA is quantitative]

NOTE: COUNTRY is not a measurement *per se*; it is a label. If COUNTRY is identified as a variable, it must be identified as categorical.

POINTS ON THIS PAGE 5

2. Health care expenditures [10 pts]

1280	1340	1770	1910	1940	2070	2220	2350
2410	2590	2840	2960	3000	3190	3860	5220

(a) Stemplot with split-stem values. Use an axis multiplier of 1000.

1|23
1|799
2|0234
2|589
3|01
3|8
4|
4|
5|2
x1000

[4 point A/B/C/D/F scale: points for stem and leaf accuracy]

(b) [Median = 2380]
(Depth is (16 + 1) / 2 = 8.5, placing it between 2350 and 2410)

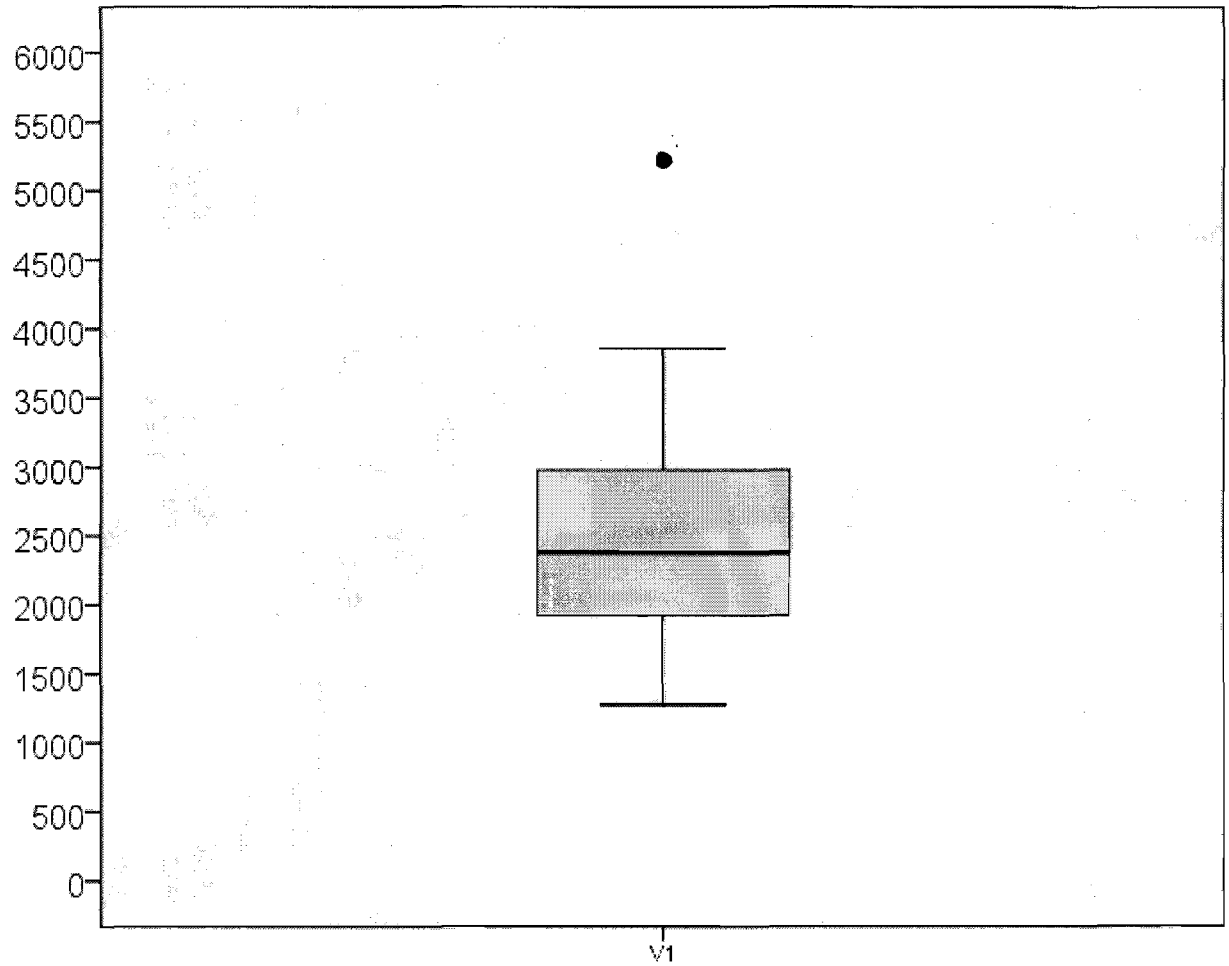
(c) Quartiles

[Q1 = 1925]
(Depth is (8 + 1) / 2 = 4.5, placing it between 1910 and 1940)

[Q3 = 2980]
(Depth is 4.5 from the top, placing it between 2960 and 3000)

/7

(d) Boxplot



Points for accurate depiction of:
[Upper outside value = 5220]
[Upper inside value = 3860]
[Lower inside value = 1280]

Calculations:
IQR = 2980 – 1925 = 1055
F_U = 2980 + (1.5)(1055) = 4562.5; 5220 is outside; 3860 is upper inside value
F_L = 1925 – (1.5)(1055) = 342.5; no lower outside values 1280 is the lower inside value.

/3

3. One-sentence response [1 pt]

[This indicates that the distribution has a positive-skew or high outlier(s).]

4. Tiny data set

I	X	Deviations	Squared deviations
1	10	$10-20=-10$	100
8	30	$30-20=10$	100
Σ	40	0	200

$$s^2 = \frac{1}{2-1} \cdot 200 = 200$$

$$s = \sqrt{200} = 14.142$$

[One point for deviations: −10 and 10]
[One point for sum of squares = 200]
[One point for the variance =200)]
[One point for standard deviation =14.14]

/5

5. Standard Normal, tails [5 pts].

(a) [Pr(Z < -1.19) = .1170]

(b) [Pr(Z > 1.19) = .1170]

(c) [Pr(Z < -1.19 or Z > 1.19) = .2340]

(d) Sketch not shown.
[One point for accurate curve centered on 0 with σ inflection points.]
[One point for tails shaded to right and left of ± 1.19]

6. Exercise 5.16 re-revisted [5 pts].

```
     |
.50  +             +-----------+
     |             |XXXXXXXXXXX|
     |             |XXXXXXXXXXX|
     |             |XXXXXXXXXXX|
.25  +-----------+XXXXXXXXXXX+-----------+
     |              XXXXXXXXXXXXXXXXXXXXXX|
     |              XXXXXXXXXXXXXXXXXXXXXX|
     |              XXXXXXXXXXXXXXXXXXXXXX|
     +-----------+-----------+-----------|
     0                                   3
```

[One point for curve]
[One point for shading]
[One point for:
Pr(X > 1) = area of shaded region
= (1 × ½ rectangle) + (1 × ¼ rectangle)
= ¾

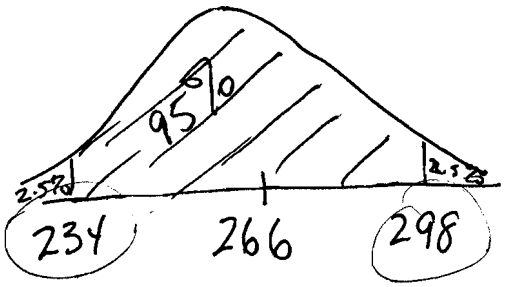
[TWO FREE POINTS] Yea!

/10

7. [10 pts] The length of uncomplicated human pregnancies from the last menstrual period to birth varies according to a distribution that is approximately Normal with mean 266 days and standard deviation 16 days.

(a) Draw a Normal curve on which the mean and standard deviation are correctly labeled. Shade the region corresponding to 95% of uncomplicated gestational lengths.

Sketch
Lower extent of shaded range = 266 – 2·16 = 234
Upper extent of shaded range = 266 + 2·16 = 298



(b) What proportion last less than 230 days?

State: Let X represent gestation from LMP to birth in days
X~N(266,16)
Pr(X < 230) = ?

Standardize: z = (230 – 266) / 16 = –2.25

Table B: Pr(Z < –2.25) = .0122

(c) WHAT IS 95th PERCENTILE ON THIS DISTRIBUTION ?
[95th percentile on X~N(266, 16); i.e., Pr(X < x) = .95, find x]
[Table B: $z_{.95}$ = 1.64 or 1.65 or 1.645]
[Unstandardize: x = 266 + (1.645)(16) = 266 + 26.32 = 292.32 days]

/10