Midterm Stat2507 (D, E, F) Winter 2013 Duration: 90 minutes Closed book
Simple calculators allowed. Pages are <u>two-sided</u>. Total of marks=100.

Last Name———————————————— **First**————————————-
Student # ———————————— **Section:**———

Multiple choice questions part: every question is worth 5 marks. Please circle only one answer.

The following four questions are based on the following stem-and-leaf plot.

```
Stem-and-leaf of DATA  N  = 49
Leaf Unit = 0.10

 2     1  68
 4     2  12
 12    2  55578899
 15    3  114
(13)   3  5566677778999
 21    4  00012234
 13    4  5678999
 6     5  11
 4     5  667
 1     6  1
```

1. The upper (3rd) quartile and the median are respectively
   a) 9 and 11.5 **b)** 45.5 and 38 **c)** 4.55 and 3.8 **d)** 3.55 and 5.1.

2. The 13th largest observation is
   a) 3.1 **b)** 3.5 **c)** 35 **d)**4.5

3. The standard deviation can roughly be approximated by
   a) 4.5 **b)** 6.1 **c)**1.125 **d)**2.25

4. The lower (1st) quartile and the interquartile range (IQR) are respectively

   a) 2.5 and 4.5 **b)** 3 and 4.5 **c)** 3 and 1.55 **d)** 3.5 and 1.55.

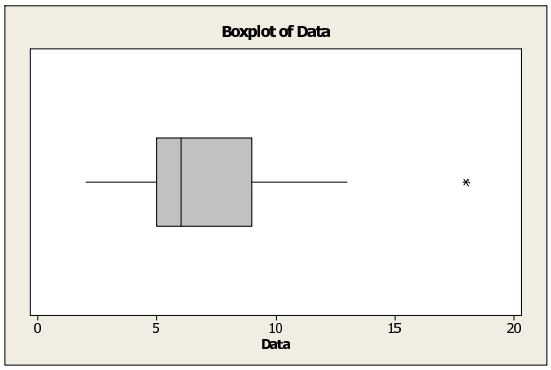
   The following two questions are based on the next boxplot (See the other side.)

5. Which of the following statements about the distribution below is correct?

   a) Symmetric and has no outlier **b)** skewed to the left and has outlier

   c) Skewed to the right and has outlier **d)** Symmetric and has outlier.

6. Which one of the following statements is correct?
   a) The mean is smaller than the median **b)** The 1st quartile is approximately 5
   c) The 3rd quartile is approximately 14 **d)** The maximum value is 13.5



7. According to the empirical rule, if a data set is fairly symmetric with mean 3 and standard deviation 1.5 then, about 95% of observations fall between

   a) 1.5 and 4.5 **b)** 0 and 6 **c)** -1.5 and 7.5 **d)** -3 and 9.

8. If A and B are independent and $P(A) = .6$, $P(B) = .3$ then $P(A \cup B)$ is

   a) .9 **b)** .8 **c)** .72 **d)** .18

9. A box contains three red toys and 6 white toys. A child is asked to draw two toys. The probability that both are red is

   a) 1/12 **b)** 1/24 **c)** 2/3 **d)** 1/3.

10. Two variables x and y were measured on a sample of 20 individuals and we observed the following quantities:

    $\sum x = 150, \quad \sum y = 220, \quad \sum x^2 = 1250, \quad \sum y^2 = 2500, \quad \sum xy = 1735.$

    Then the correlation coefficient between x and y is

    a) .99 **b)** .97 **c)** .95 **d)** .85

11. Two variables x and y were measured on a sample of 25 individuals and they yielded the following quantities:

    $\sum x = 180, \quad \sum y = 200, \quad \sum x^2 = 2200, \quad \sum y^2 = 2300, \quad r = .91.$

    Then the regression (best fitting) line is given by

    a) $y = -.8 + 201.23x$, **b)** $y = 201.23 + .8x$, **c)** $y = 2.24 + .8x$, **d)** $y = 5.21 + 1.2x$.

12. Steve takes either a bus or the subway to go to work with probabilities .3 and .7, respectively. When he takes the bus, he is late 30% of the days, and when he takes the subway, he is late 20% of the days. If Steve is late on a particular day, what is the probability that he took the bus?

    a) .23 **b)** .39 **c)** .61 **d)** .09

Long answer questions: show all your work

13. Consider the following set of observations: 9, 4, -3, 0, 3.

a) [12] Compute the mean $\bar{x}$ and the standard deviation s of this set of data.

Solution

$$\bar{x} = \frac{9+4-3+0+3}{5} = 2.6.$$

$$s^2 = \frac{91+16+9+0+9-(9+4-3+0+3)^2/5}{4} = 20.3 \qquad \text{so } s = \sqrt{20.3} = 4.5.$$

b) [4] Compute the z-score of the observation $x = 9$ and check if it is an outlier.

Solution

$$z\text{-score} = \frac{9-2.6}{4.5} = 1.42 < 3, \qquad \text{and therefore 9 is not an outlier.}$$

14. A discrete random variable X has the following pmf table

x	-2	0	3	5
$p(x)$	?	.1	.2	.4

a) [4] Compute $p(-2)$=**1-(.1+.2+.3)=.3**

b) [12] Compute the mean and the variance of X.

Solution $\mathbb{E}(X) = \mu = (-2)(.3) + 0(.1) + 3(.2) + 5(.4) = 2.$.

$$\mathbb{E}(X^2) = (4)(.3) + 0 + (9)(.2) + (25(.4) = 13, \qquad \text{therefore } \sigma^2 = 13 - (2^2) = 9.$$

15. A smoke-detector system uses two devices A and B. If smoke is present it will be detected by device A with probability .95, by device B .98, and by both devices .94.

a) [4] If a smoke is present, what is the probability that it will be detected (by at least one of the devices)?

Solution $P(A \cup B) = P(A) + P(B) - P(A \cap B) = .95 + .98 - .94 = .99$

b)[4] What is the probability that the smoke will not be detected.

Solution $1 - P(A \cup B) = 1 - .99 = .01.$

Formula sheet: for Ch.1-4

• A and B are independent means that $P(A \cap B) = P(A)P(B)$ or $P(A|B) = P(A)$.

• If A and B are mutually exclusive then $P(A \cup B) = P(A) + P(B)$,
otherwise, in general $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

• $P(A) = 1 - P(A')$ where A' is the complement event. If you are using A^c notation for complement event, you have $P(A) = 1 - P(A^c)$.

• Bayes rule:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

• Total probability law : if sample space can be decomposed into S_1, S_2, S_3 then

$$P(A) = P(A|S_1)P(S_1) + P(A|S_2)P(S_2) + P(A|S_3)P(S_3).$$

(if only two events S_1 and S_2, just ignore S_3).

• z-score for an observation x:

$$z\text{-score} = \frac{x - \bar{x}}{s}.$$

• Sample variance $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \dfrac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$ (short-cut formula)

• Bivariate data:

$$s_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}{n-1}, \quad r = \frac{s_{xy}}{s_x s_y}, \quad b = r\left(\frac{s_y}{s_x}\right), \quad a = \bar{y} - b\bar{x}, \quad y = a + bx.$$

• Expected value and variance of random variable X:

$$\mathbb{E}(X) = \mu = \sum_x xp(x), \qquad \sigma^2 = \sum_x (x-\mu)^2 p(x) = \left(\sum_x x^2 p(x)\right) - \mu^2$$