MAT 2379B
Final Exam

December 18, 2012 **Professor: Rafal Kulik**
Time: 180 minutes

Student Number:_____

Family Name: _____ **First Name:** _____

- **This is a closed book examination. Only TI 30 and Casio calculators are permitted.**
- **Record your answer to each question in the table below. Each question is worth 1 mark.**
- **At the end of the examination, hand in only this page.**

Question	Answer	Question	Answer
1	-	14	B
2	E	15	B
3	B	16	D
4	D	17	B
5	E	18	B
6	A	19	B
7	A	20	A
8	E	21	E
9	B	22	B
10	A	23	B
11	E	24	C
12	C	25	B
13	C		

GOOD LUCK !!!
****

Professor's use only: Grade=_____/25

1. Consider the following R output:

```
> pbinom(35,125,0.35)
[1] 0.05909366
> pbinom(36,125,0.35)
[1] 0.08553514
> pbinom(37,125,0.35)
[1] 0.1197826
> pbinom(42,125,0.35)
[1] 0.4108588
> pbinom(43,125,0.35)
[1] 0.4850622
> pbinom(44,125,0.35)
[1] 0.559525
```

Let X be a binomial random variable with $n = 125$ and $p = 0.35$. Using the R output above, calculate $P(37 < X < 43)$.

A) 0.3659 B) 0.5152 C) 0.3463 D) 0.3253 E) 0.3015

2. Hydrocarbons emitted by the exhaust systems of automobiles are some of the major contributors to air pollution. Let X be the number of grams of hydrocarbons emitted by an automobile per mile. Assume that X is normally distributed with a mean of 1 g and a standard deviation of 0.25 g. We claim that 89.97% of automobiles emit more than x_0 grams of hydrocarbons per mile. What is the value of x_0?

A) 0.75 B) 1.25 C) 0.59 D) 1.95 E) 0.68

3. It is known that in Canada, the blood types have the following distribution: 46% O, 42% A, 9% B, 3% AB. What is the probability that in a randomly chosen Canadian couple, the man and woman do not have the same blood type? Assume that the man's blood type is independent of the woman's blood type.

A) 0.397 B) 0.603 C) 0.2116 D) 0.1764 E) 0.7512

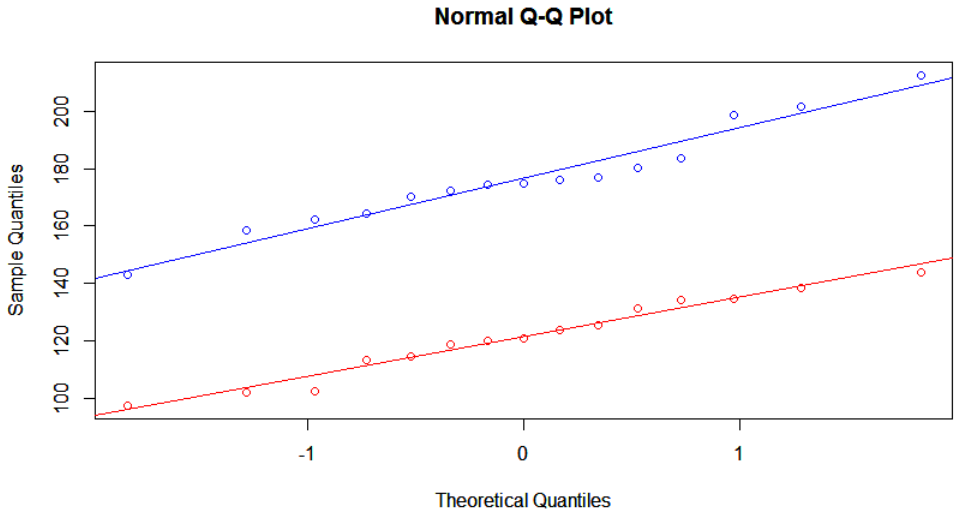
4. Recent data suggests that the number of heart attacks are higher in the days following a severe winter storm, due to snow shoveling. The maximum heart rate $\mathrm{HR}_{\max}$, the highest heart rate an individual can achieve without severe problems through exercise stress, depends on age and is calculated using the formula $\mathrm{HR}_{\max} = 220 - \mathrm{age}$. Using this formula, we may assume that the average $\mathrm{HR}_{\max}$ for the population of healthy adults of age 45 is $\mu = 175$. Assume that the standard deviation of the $\mathrm{HR}_{\max}$ in the same population is $\sigma = 20$. Let $\bar{X}$ be the average $\mathrm{HR}_{\max}$ for a sample of $n = 64$ individuals of age 45. Give an approximation for $P(\bar{X} > 180)$.

A) 0.6255 B) 0.3745 C) 0.0101 D) 0.0228 E) 0.9964

5. We continue with the situation in Question 4. We consider of group of 30 volunteers of age 45, which is split randomly into two sub-groups of size 15 each. The 15 persons in the first group are asked to snow shovel for 20 minutes. The values of $\mathrm{HR}_{\max}$ for these persons are recorded in R in the variable x. The other 15 volunteers are asked to use an electric snowblower for 20 minutes; their $\mathrm{HR}_{\max}$ is recorded in R in the variable y. Below is the summary of these data.

```
> mean(x)
[1] 189.5593
> sd(x)
[1] 11.08633
> mean(y)
[1] 173.0736
> sd(y)
[1] 9.895271
```

The picture below gives the overlayed QQ-plots for x and y.



Which one of the following statements is correct? (Only one statement is correct.)
A) Using these QQ-plots, we cannot infer that the two populations are normally distributed, but we can infer that two populations have the same variance.
B) Using these QQ-plots, we can infer that the two populations are normally distributed, but we cannot infer that the two populations have the same variance.
C) We cannot use these QQ-plots because our data are paired measurements. Hence, we should calculate instead the differences between x and y, and then produce a QQ-plot for these differences to verify that they are normally distributed.
D) We cannot use these QQ-plots because our data are paired measurements. For paired data, it is not necessary to verify that the differences between the x and y measurements are normally distributed.
E) Using these QQ-plots, we can infer that the two populations are normally distributed and have the same variance.

6. In a study of angina in rats, 18 animals with a history of angina were given an experimental drug, which might affect their oxygen intake, measured in milliliters per minute. A normal value for the oxygen intake in these rats is around 1600 ml/min. For the 18 rats, the sample mean was $\bar{x} = 1702$ ml/min, with a sample standard deviation $s = 181$ ml/min. Assuming that the data is normally distributed, is there enough evidence that the oxygen intake is affected significantly by this drug? Report the observed value of the test statistic (t_0) and the range of the p-value. Use the significance level $\alpha = 0.05$.

   A) $t_0 = 2.39$; p-value is between 0.02 and 0.05; the drug affects significantly the oxygen intake
   B) $t_0 = 2.39$; p-value is between 0.01 and 0.025; the drug affects significantly the oxygen intake
   C) $t_0 = 0.136$; p-value is between 0.01 and 0.05; the drug affects significantly the oxygen intake
   D) $t_0 = 0.136$; p-value is between 0.02 and 0.10; we cannot draw any conclusion
   E) $t_0 = 2.39$; p-value is between 0.05 and 0.10; the drug does not affect significantly the oxygen intake

7. The following table shows the distribution of blood types in the Australian population:

Blood Type	A	B	AB	0
Proportion	7%	41%	3%	49%

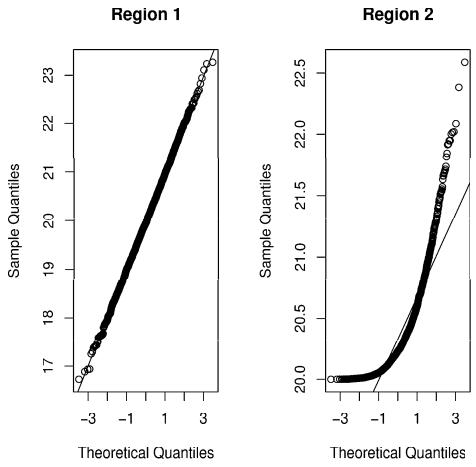
   A sample of 10 Australians is randomly selected. Find the probability that in this sample, at most 2 persons have blood type A or AB.

   A) 0.93 B) 0.07 C) 0.01 D) 0.54 E) 0.38

8. A diagnostic test for a certain disease has a sensitivity of 0.95 and a specificity of 0.85. Assume that the incidence rate of this disease is 0.2%. What is the positive predictive value of this test?

   A) 0.95 B) 0.0329 C) 0.9670 D) 0.9875 E) 0.0125

9. We would like to compare the average length of snakes in Forest Valley (Region 1) and Dry Valley (Region 2). Samples from both valleys have been studied and the following QQ-plots have been generated to assess the normality of the length of snakes in both valleys. Which one of the following



   statements is correct? (Only one statement is correct.)

   A) The lengths are normally distributed in both regions.
   B) The length is normally distributed in Region 1, but is not normally distributed in Region 2.
   C) The length is normally distributed in Region 2, but is not normally distributed in Region 1.
   D) The length is not normally distributed in either one of the two regions.

10. Consider a variable x which gives the yield of corn cob, in kg per hectare. We have created the variable x in R, for which we have used the command "summary". Below is the R output.

```
> x=c(2134,2170,2142,2799,2364,2199,2310,1620,1808,1476,1695)
> summary(x)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1476    1752    2142    2065    2254    2799
```

Which one of the following statements is correct? (Only one statement is correct.)

A) There are no outliers.
B) The value 1476 is an outlier.
C) The value 2799 is an outlier.
D) The values 1476 and 2799 are outliers.
E) The mean is larger than the median.

11. Before preparation for microscopic study, tissue cells are exposed to 20 minutes of anoxia (an abnormally low oxygen supply). These cells are then graded for the extent of damage as follows: 0=undamaged, 1=slightly damaged, 2=moderately damaged, 3=extensively damaged. Let X be the classification value for the extent of damage in a randomly chosen cell. Below is the table of the cumulative distribution function of X:

x	0	1	2	3
$F(x)$	0.15	0.40	0.90	1.00

Find the expected value of X and the variance of X.

A) $E(X) = 5.20, \mathrm{Var}(X) = 14.041$
B) $E(X) = 1.55, \mathrm{Var}(X) = 3.15$
C) $E(X) = 5.20, \mathrm{Var}(X) = 0.1573$
D) $E(X) = 2.45, \mathrm{Var}(X) = 1.52$
E) $E(X) = 1.55, \mathrm{Var}(X) = 0.7475$

12. Among the patients with lung cancer, at most 10% will survive for three years after the diagnostic. Roche Holding, one of the world's largest manufacturers of cancer drugs claims that its Tarceva medicine can extend the lifetime of patients with non-small-cell lung cancer. Assume that in a study of 150 patients with lung cancer treated with Tarceva, 22 survived for 3 years. Compute a 85% confidence interval for the proportion p of lung cancer patients treated with Tarceva, who will survive for 3 years after the diagnostic. Based on this interval, can we conclude that the proportion p is higher than 10%?

A) [9.0%; 20.3%]; we cannot conclude that p is higher than 10%
B) [9.9%; 19.4%]; we cannot conclude that p is higher than 10%
C) [10.5%; 18.8%]; the proportion p is higher than 10%
D) [11.3%; 17.5%]; the proportion p is higher than 10%
E) [8.7%; 23.1%]; we cannot conclude that p is higher than 10%

13. We continue with the situation in Question 12. Using a test of hypothesis of level $\alpha = 0.05$, is there enough evidence that the proportion p is higher than 10%? Report the observed value of the test statistic and the p-value.

    A) $z_0 = 1.41$, p-value=0.0793; there is not enough evidence that p is higher than 10%
    B) $z_0 = 2.53$, p-value=0.0057; there is enough evidence that p is higher than 10%
    C) $z_0 = 1.91$, p-value=0.0281; there is enough evidence that p is higher than 10%
    D) $z_0 = 1.62$, p-value=0.0526; there is not enough evidence that p is higher than 10%
    E) $z_0 = 1.04$, p-value=0.1492; there is not enough evidence that p is higher than 10%

14. A pharmaceutical company has produced a new drug and claims that it helps reducing the systolic blood pressure. To test this claim, a physician prescribes this drug to 10 of his patients. The patients' systolic blood pressure, before and after the drug treatment, are recorded in the variables "Before" and "After" given below:

```
Before=c(140,135,122,150,126,138,141,155,128,130)
After=c(135,136,120,148,122,136,140,153,120,128)
```

    The physician types into the R console: `t.test(Before,After,alternative="greater")` and obtains the following output:

```
data:  Before and After
t = 0.5499, p-value = 0.2946
alternative hypothesis: true difference in means is greater than 0
sample estimates: mean of x mean of y
    136.5     133.8
```

    The physician's assistant types `t.test(Before,After,paired=TRUE,alternative="greater")` and obtains the following output:

```
data: Before and After
t = 3.4825, df = 9, p-value = 0.003456
alternative hypothesis: true difference in means is greater than 0
sample estimates: mean of the differences
                    2.7
```

    Which one of the following statements is correct? (Only one statement is correct.)

    A) The assistant uses the correct command. There is not enough evidence that the new drug reduces the systolic blood pressure.

    B) The assistant uses the correct command. There is enough evidence that the new drug reduces the systolic blood pressure.

    C) The physician uses the correct command. There is not enough evidence that the new drug reduces the systolic blood pressure.

    D) The physician uses the correct command. There is enough evidence that the new drug reduces the systolic blood pressure.

    E) Neither the physician, nor the assistant are using the correct command. The t-test should not be used in this situation.

15. Assume that among the students in a certain college, 55% are overweight, 20% have high blood pressure, and 60% are overweight or have high blood pressure. A student is randomly selected from this college. Let O be the event that the student is overweight and H be the event that the student has high blood pressure. Compute $P(O \cap H)$. Is the fact that a student is overweight independent of the fact that this student has high blood pressure?

    A) The events are independent, and $P(O \cap H) = 0.15$
    B) The events are dependent, and $P(O \cap H) = 0.15$
    C) The events are independent, and $P(O \cap H) = 0.6$
    D) The events are dependent, and $P(O \cap H) = 0.6$
    E) The events are independent, and $P(O \cap H) = 0.11$

16. Let X be the amount of calcium (in miligrams) per deciliter of blood for a randomly chosen healthy adult. In a sample of 25 individuals, it was found that the average amount of calcium was 9.8 mg, with a standard deviation of 0.5 mg. Assuming that the data is normally distributed, find a 80% confidence interval for the average amount of calcium per deciliter of blood in the general population of healthy adults.

    A) [9.6289; 9.9711] B) [9.5936; 10.0064] C) [9.1432; 9.9876]
    D) [9.6682; 9.9318] E) [8.7654; 10.1632]

17. A class consists of 490 female and 510 male students. The students are classified according to their final grade in this course, as follows:

	Passed	Did not pass
Female	430	60
Male	410	100

    A student is randomly selected from this class. What is the probability that the student did not pass the course, given that this student is a female?

    A) 0.06 B) 0.12 C) 0.41 D) 0.81 E) 0.88

18. A sample of 20 fish was catched in lake A and their length (in cm) was measured. Another sample of 12 fish was catched in lake B and their length (in cm) was also measured. The data are summarized in the table below:

	sample size	sample mean	sample variance
Lake A	20	13.09	1.222
Lake B	12	10.29	1.186

    Let μ_1 and μ_2 be the average fish length in lake A, respectively lake B. We would like to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$. Assume that the two populations are normally distributed with equal variances. Find the range of the p-value for this test.

    A) the p value is between 0.025 and 0.05
    B) the p-value is between 0.1 and 0.2
    C) the p-value is between 0.95 and 0.975
    D) the p-value is between 0.05 and 0.1
    E) the p-value is between 0.25 and 0.5

19. A survey was conducted on 1000 adults, among which 48% were men. The results of the survey show that 15% of men and 30% of women are afraid of flying. Using a contingency table, test the hypothesis that sex is independent of the fact that a person is afraid of flying. Report the range of the p-value and your conclusion at level $\alpha = 0.05$.

    A) p-value < 0.005. The sex and the fact that a person is afraid of flying are independent.
    B) p-value < 0.005. There is an association between the sex and the fact that a person is afraid of flying.
    C) $0.005 < p$-value < 0.01. There is an association between the sex and the fact that a person is afraid of flying.
    D) $0.05 < p$-value < 0.10. The sex and the fact that a person is afraid of flying are independent.
    E) p-value > 0.10. The sex and the fact that a person is afraid of flying are independent.

20. A scientist's job is to detect a situation in which the mean bacteria count has risen above the maximum safe level of 70 in Ottawa River. If the bacteria count exceeds 70, water is considered unsafe. To test whether the water is unsafe, the following sample of size 9 is drawn:

    $$69 \quad 74 \quad 75 \quad 70 \quad 72 \quad 73 \quad 71 \quad 73 \quad 68.$$

    Note that the sample standard deviation is $s = 2.35$. Assume that the data is normally distributed. Is there enough evidence that the water is unsafe? Determine the p-value and use $\alpha = 0.05$.

    A) The water is unsafe and $0.025 < p -$ value < 0.05.
    B) The water is unsafe and $p -$ value < 0.025.
    C) The water is safe and $0.025 < p -$ value < 0.05.
    D) The water is safe and $p -$ value < 0.025.
    E) The water is unsafe and $p -$ value > 0.95.

21. Consider the following data sets:

    ```
    x=c(1,4,6,12,10,16,20)
    y=c(2,4,5,8,11,14)
    ```

    The sample standard deviation for the first data set is 6.74. Find the pooled variance.
    Hint: First compute the sample variance for the second data set.

    A) 22.6 B) 11.3 C) 6.7 D) 33.1 E) 16.8

22. Let $X_1, X_2, \ldots, X_{28}$ be a random sample from a normal population with mean $\mu = 125$. Let $\bar{X}$ be the mean of this sample and S^2 be its variance. Find a value c such that

    $$P\left(\frac{\bar{X} - 125}{S/\sqrt{n}} \leq c\right) = 0.95.$$

    A) $c = 1.701$ B) $c = 1.703$ C) $c = 1.645$ D) $c = -1.701$ E) $c = -1.703$

23. It is known that the height of the adults in a certain community follows a normal distribution with variance $\sigma^2 = 625$ centimeters^2. 25 adults were randomly selected from this community and their heights were measured. The data yielded a sample mean of $\bar{x} = 165$ centimeters. Construct a 90% confidence interval for the average height of the adults living in this community, measured in meters.

A) [157; 173] B) [1.57; 1.73] C) [155; 175] D) [121; 193] E) [1.55; 1.75]

24. A study is conducted to identify the major risk factors in the development of coronary-artery disease. The following data gives the cholesterol level (in miligrams per deciliter) for 7 women aged 20 to 29.

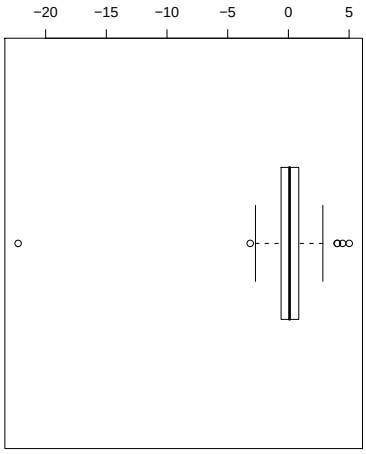
$$175.3 \quad 171.2 \quad 175.9 \quad 176.2 \quad 170.2 \quad 185.3 \quad 174.7$$

Compute: (a) a point estimate for the average cholesterol level of women aged 20 to 29; (b) the estimated standard error of the mean.

A) (a) 175.54; (b) 4.898
B) (a) 170.31; (b) 4.898
C) (a) 175.54; (b) 1.851
D) (a) 170.31; (b) 3.253
E) (a) 173.25; (b) 1.851

25. We include below the summary and the boxplot of data produced by R for a particular data set.

```
      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-22.27000  -0.58500   0.09504   0.02696   0.84570   5.01100
```



Which one of the following statements is correct? (Only one statement is correct.)

A) The IQR is 27.281.
B) There are three outliers greater than 0, two outliers smaller than 0 and the extreme values -22.27 and 5.011 are among the outliers.
C) The range of the data is 1.4307
D) The distance between the two fences is 1.4307
E) -22.27 is an outlier, but 5.011 is not an outlier.