MAT 2379 A
Solutions to the Final Examination

December 20, 2011 **Professor Raluca Balan**
Time: 3 hours

Student Number:____________________ **Seat 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		13	
2		14	
3		15	
4		16	
5		17	
6		18	
7		19	
8		20	
9		21	
10		22	
11		23	
12		24	

****

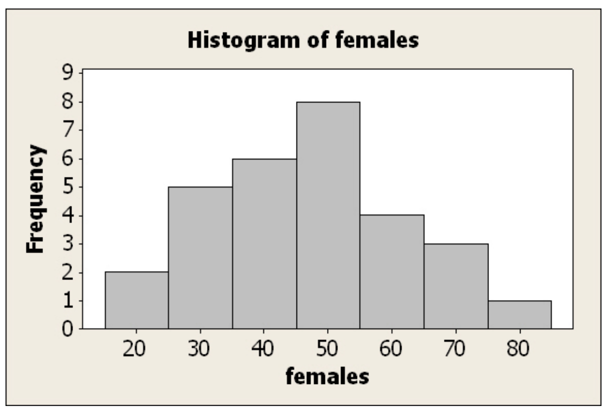
Professor's use only: Grade=______/24

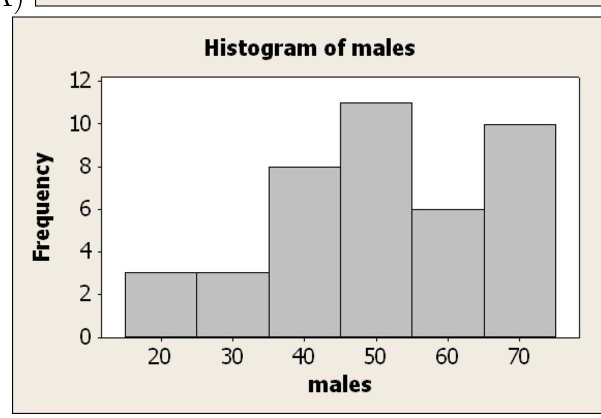
1. In a large city, a record is kept of the drug addicts arrested for the first time for minor offences, during a period of 6 months. These people are classified according to their age and gender. The summary of the data is given below:

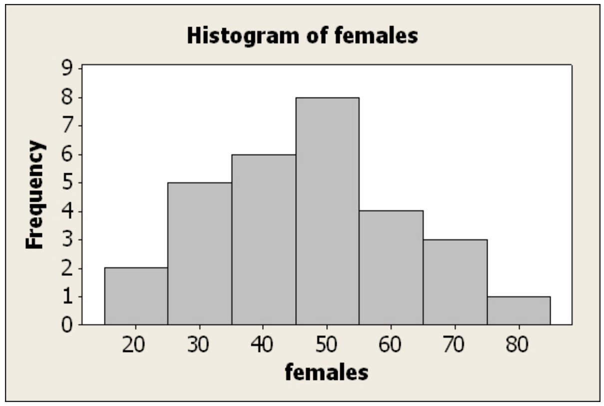
Age	Frequency (females)	Frequency (males)
15-24	2	3
25-34	5	4
35-44	6	9
45-54	8	12
55-64	4	6
65-74	3	7
75-84	1	0
Total	29	41

Which one of the pairs of histograms below represent the histograms of frequencies for this data? (Only one pair is correct.)

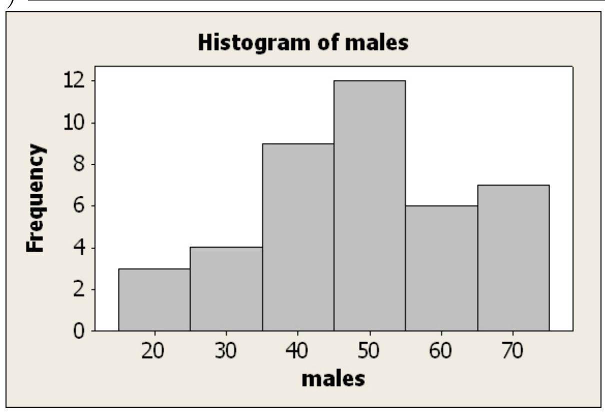
A)

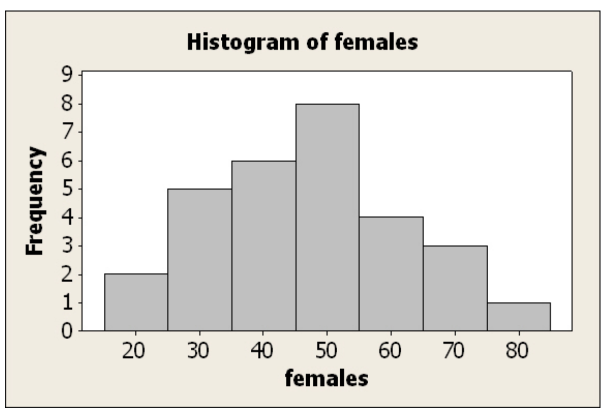




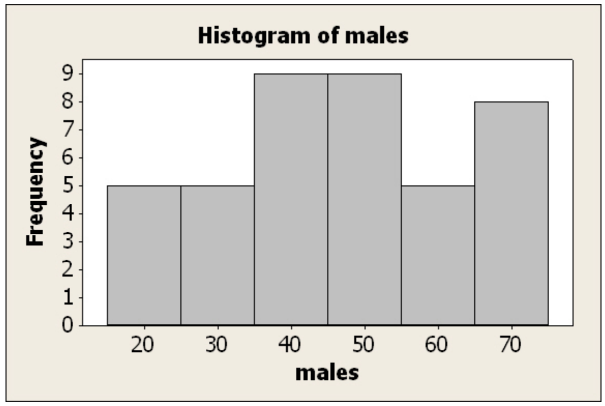

Histogram of females
Frequency
0
1
2
3
4
5
6
7
8
9
20
30
40
50
60
70
80
females

B)

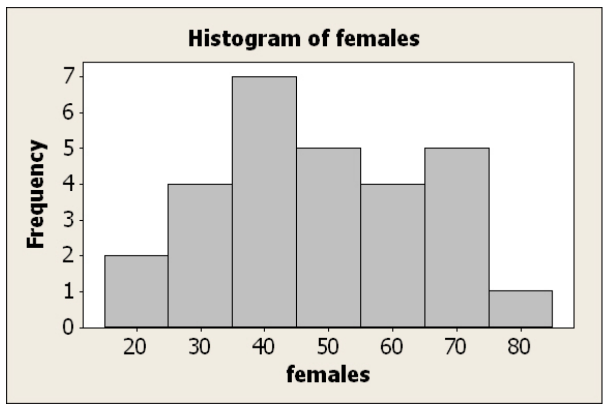

Histogram of males
Frequency
0
2
4
6
8
10
12
20
30
40
50
60
70
males

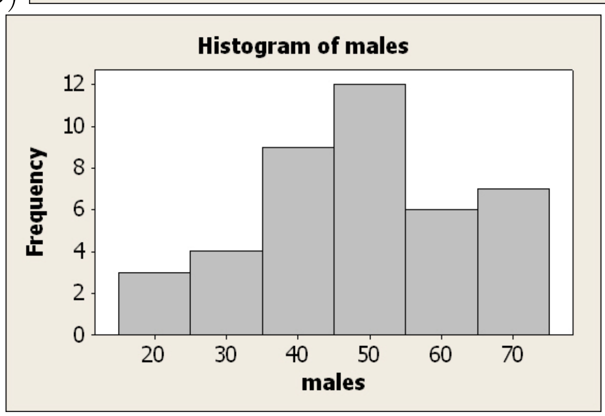

Histogram of females
Frequency
0
1
2
3
4
5
6
7
8
9
20
30
40
50
60
70
80
females

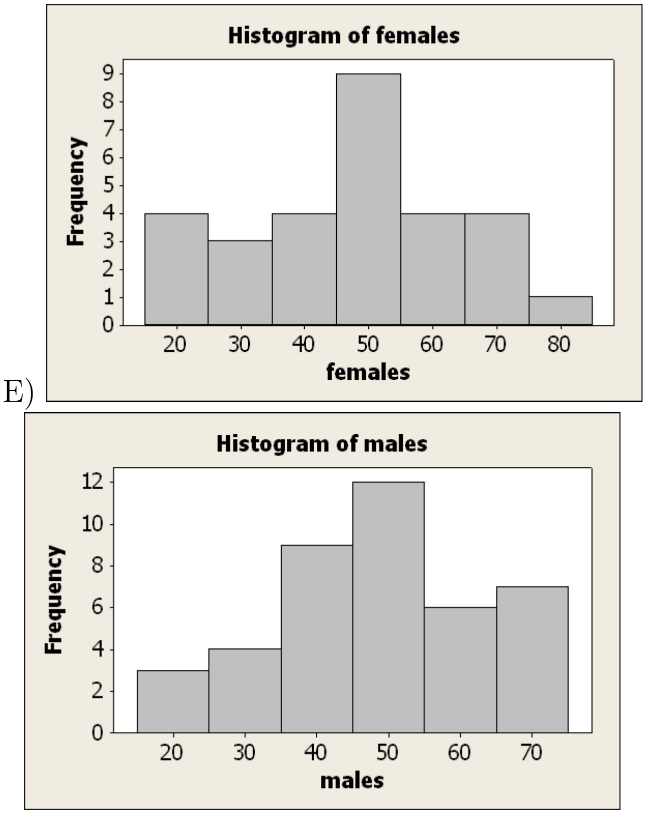
C)


Histogram of males
Frequency
males

D)


Histogram of females
Frequency
females


Histogram of males
Frequency
males

E) 

Solution: The only histograms which correspond to the given frequencies are given in B). The answer is B.

2. A recent study published by the Canadian Cancer Society estimates that, based on current incidence rates, 40% of women and 45% of men in Canada will develop cancer during their lifetimes. Assuming that in Canada, the ratio between the male and female populations is 0.82, what is the probability that a randomly selected person will develop cancer?

   A) 0.42 B) 0.55 C) 0.43 D) 0.45 E) 0.41

   Solution: Let A be the event that a randomly selected person will develop cancer and B be the event that the person is a female. We know that $P(A|B) = 0.40$ and $P(A|B') = 0.45$. We have to find $P(A)$. Note that $P(B) = p_1$, where p_1 is the percentage of females in the Canadian population. Let p_2 be the percentage of males. We know that $p_1 + p_2 = 1$ and $p_2/p_1 = 0.82$. We solve for p_1 and p_2. We have: $p_2 = (0.82)p_1$ and hence $p_1 + (0.82)p_1 = 1$, i.e. $(1.82)p_1 = 1$. We find $p_1 = 1/1.82 = 0.55$. Hence $P(B) = 0.55$ and $P(B') = 0.45$. By the total probability rule,

$$\begin{aligned} P(A) &= P(A|B)P(B) + P(A|B')P(B') \\ &= (0.40)(0.55) + (0.45)(0.45) = 0.42 \end{aligned}$$

The answer is A.

3. Professors from the Department of Mathematics are reporting that the grades of students in the course MAT 2379 are normally distributed with a mean of 80 and standard deviation $\sigma = 5$. However, a group of students believe that the average grade might be smaller. A sample of 25 randomly selected students from this course have yielded a sample mean of 77.71. Using a significance level $\alpha = 0.01$, is there enough evidence that the average grade in this course is less than 80? Determine the p-value for the related test of hypotheses.

   A) there is enough evidence for students' claim; $p-\text{value} = 0.011$
   B) there is enough evidence for students' claim; $p-\text{value} = 0.989$
   C) there is enough evidence for students' claim; $p-\text{value} = 0.022$
   D) there is not enough evidence for students' claim; $p-\text{value} = 0.011$
   E) there is not enough evidence for students' claim; $p-\text{value} = 0.022$

   Solution: Students wish to test $H_0 : \mu = 80$ versus $H_1 : \mu < 80$. The observed value of the test statistic is

$$z_0 = \frac{77.71 - 80}{5/\sqrt{25}} = -2.29.$$

   Dealing with a left-tailed test $p-\text{value} = P(Z < -2.29) = 0.011$. This number is larger than the level of significance 0.01. We fail to reject H_0. There is not enough evidence to support the students' claim. The answer is D.

4. According to a new rule under the Ontario's Tobacco Control Act, the retailers selling tobacco are no longer allowed to exhibit any tobacco products. It is estimated that prior to this rule, the percentage of

adolescent smokers was around 15%. One year after this rule was implemented, it was found that in a sample of 525 adolescent, 70 smoke. Is there enough evidence that the new rule has been effective in reducing the percentage of adolescent smokers? Report your conclusion at the significance level $\alpha = 0.10$. What is the range of the p-value?

A) yes, the rule was effective; p-value < 0.01
B) yes, the rule was effective; $0.01 < p$-value < 0.05
C) yes, the rule was effective; $0.05 < p$-value < 0.10
D) no, the rule was not effective; $0.10 < p$-value < 0.15
E) no, the rule was not effective; p-value > 0.15

Solution: We would like to test $H_0 : p = 0.15$ versus $H_1 : p < 0.15$. An estimate for p is $\hat{p} = 70/525 = 0.133$. The observed value of the test statistic is:

$$z_0 = \frac{0.133 - 0.15}{\sqrt{(0.15)(0.85)/525}} = -1.07.$$

From Table 17.3, we find that p-value$= P(Z < -1.07) = 0.1423$. Since the p-value is greater than 0.10, we fail to reject H_0. The rule was not effective. The p-value lies between 0.10 and 0.15.

5. Under controlled conditions, the probability that a white rat is infected by a virus A is 0.6, the probability that it is infected by a virus B is 0.7, and the probability that it is infected by both viruses is 0.5. Under these conditions, what is the probability that a randomly chosen rat becomes infected by at least one of the two viruses?

   A) 0.2 B) 0.5 C) 0.8 D) 0.7 E) 0.6

   Solution: Let A be the event that the rat is infected by virus A, and B be the event that the rat is infected by virus B. We know that $P(A) = 0.6$, $P(B) = 0.7$ and $P(A \cap B) = 0.5$. By the addition rule, $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.7 - 0.5 = 0.8$. The answer is C.

6. Suppose that the height of a tree is normally distributed with mean 225 cm and standard deviation of 20 cm. We would like to say that 75% of the trees have a height smaller than h. What is the value of h?

   A) 7.5 B) 230.6 C) 67.4 D) 268.8 E) 238.5

Solution: We have to find h such that $P(X < h) = 0.75$. By standardization,

$$0.75 = P(X < h) = P\left(\frac{X - 225}{20} < \frac{h - 225}{20}\right) = P\left(Z < \frac{h - 225}{20}\right)$$

From Table 17.3, we find $\frac{h-225}{20} = 0.675$ and $h = 225 + 20(0.675) = 238.5$. The answer is E.

7. A biologist wants to compare the number of eggs that two species of insects lay in one year. It is believed that the numbers of eggs for the two species of insects are normally distributed and have the same variance. The data is summarized in the following table:

	sample size	sample mean	sample variance
Species 1	$n_1 = 8$	$\bar{x}_1 = 43.14$	$s_1^2 = 71.65$
Species 2	$n_2 = 8$	$\bar{x}_2 = 47.79$	$s_2^2 = 52.66$

Construct a 95% confidence interval for the difference $\mu_1 - \mu_2$, where μ_1 is the average number of eggs laid by the first species, and μ_2 is the average number of eggs laid by the second species.

A) [3.81,13.11] B) [-13.11, 3.81] C) [43.14, 47.79]
D) [-21.55,12.25] E) [3.81, 15.78]

Solution: The two samples are independent with equal variances. The pooled sample variance and standard deviation are

$$s_p^2 = \frac{7(71.65) + 7(52.66)}{14} = 62.155 \quad \text{and} \quad s_p = \sqrt{62.155} = 7.8838.$$

The number of degrees of freedom is 14 and $t_{0.025,14} = 2.145$. Therefore, a 95% confidence interval for $\mu_1 - \mu_2$ is

$$43.14 - 47.79 \pm (2.145)(7.8838)\sqrt{\frac{1}{8} + \frac{1}{8}} = -4.65 \pm 8.46 = [-13.11, 3.81].$$

The answer is B.

8. We would like to study the effect of acid rain on the growth of plants. In a particular area, under normal conditions, a dogwood sapling will

grow 8 inches on average, during the first year. It is thought that the acid rain will slow down this growth. Set up the hypotheses for testing this claim. When does a type I error occur?

A) $H_0 : \mu = 8, H_1 : \mu < 8$. A type I error occurs when we conclude that the acid rain slows down the growth, when in fact, it does not.

B) $H_0 : \mu < 8, H_1 : \mu = 8$. A type I error occurs when we conclude that the acid rain does not slow down the growth, when in fact, it does.

C) $H_0 : \mu = 8, H_1 : \mu > 8$. A type I error occurs when we conclude that the acid rain does not slow down the growth, when in fact, it does.

D) $H_0 : \mu = 8, H_1 : \mu \neq 8$. A type I error occurs when we conclude that the acid rain slows down the growth, when in fact, it does not.

E) $H_0 : \mu < 8, H_1 : \mu > 8$. A type I error occurs when we conclude that the acid rain does not slow down the growth, when in fact, it does.

Solution: The answer is A.

9. Ecologists established that the eggs of a sea turtle have an average length of 7.1 cm, with a standard deviation of 0.37 cm. They measured the length of the 196 eggs found in a recently discovered nest. For these 196 measurements, the sum and the sum of squares have been calculated using Minitab. Below is the Minitab output:

```
Descriptive Statistics: length

Variable      N   N*        Sum     Sum of squares
length      196    0  1624.1000         13556.3500
```

Calculate the standard deviation for this sample.

A) 0.37 B) 1.255 C) 0.944 D) 0.506 E) 0.711

Solution:

$$s^2 = \frac{1}{n-1}[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2]$$

$$= \frac{1}{196-1}[13556.35 - \frac{1}{196}(1624.1)^2] = 0.50611$$

Hence, $s = \sqrt{0.50611} = 0.711$. The answer is E.

10. We continue with the situation in Question 9. Identify the sample mean ($\bar{x}$) and the population mean (μ).

A) $\bar{x} = 7.1$ cm, $\mu = 8.29$ cm
B) $\bar{x} = 8.29$ cm, $\mu = 7.1$ cm
C) $\bar{x} = 8.29$ cm, $\mu = 8.29$ cm
D) $\bar{x} = 7.1$ cm, $\mu = 7.1$ cm
E) $\bar{x} = 1624.1$ cm, $\mu = 13556.35$ cm

Solution: The answer is B.

11. For many years, a farmer has not kiln-dry his barley seeds before sowing. (To kiln-dry means to dry in an insulated chamber where airflow, temperature and humidity are controlled.) The non-kiln-dried seeds yield an average 672 kg of barley per 4000 m^2. This year, the farmer decides to kiln-dry his barley seeds before sowing. 10 varieties of kiln-dried barley seeds are sowed. For each of these varieties, the yield is measured in kg per 4000 m^2. These 10 measurements are used to produce a 95% confidence interval for the mean yield of kiln-dried barley. Below is the Minitab output which gives this interval:

```
Descriptive Statistics: yield

Variable   N  N*    Mean         95% CI
yield     10   0  674.27   (653.43, 695.11)
```

Find the sample standard deviation for the 10 measurements.

A) 9.21 B) 848.56 C) 58.26 D) 674.27 E) 29.13

Solution: The length of the interval is: $695.11 - 653.43 = 41.68$. On the other hand, this length is $2(2.262)s/\sqrt{10}$ (from Table 17.4, row 9, we see that $t = 2.262$). Hence,

$$2(2.262)\frac{s}{\sqrt{10}} = 41.68$$

We find $s = \frac{\sqrt{10}(41.68)}{2(2.262)} = 29.13$.

12. Nine patients are evaluated for pain on a scale of 0 to 10, after using a control medication for pain relief (0=no pain, 10=severe pain). One week, the same patients are evaluated again after being given a new medication for pain relief. The following results are obtained.

	Control (x_1)	New (x_2)	Difference ($d = x_1 - x_2$)
Mean	$\bar{x}_1 = 4.224$	$\bar{x}_2 = 2.98$	$\bar{d} = 1.244$
Standard deviation	$s_1 = 0.05$	$s_2 = 0.01$	$s_d = 0.03$

Construct a 95% confidence interval for the difference between the average pain level using the control medication (μ_1) and the average pain level using the new medication (μ_2). Using this confidence interval, can we say that the new treatment is effective in pain reduction?

A) the interval is [-1.267, -1.221]; the treatment is not effective
B) the interval is [1.221, 1.267]; the treatment is not effective
C) the interval is [1.221, 1.267]; the treatment is effective
D) the interval is [-1.051, -1.253]; the treatment is not effective
E) the interval is [-1.051, -1.253]; the treatment is effective.

Solution: The data sets are paired. We use t-distribution with 8 degrees of freedom. $t_{0.025,8} = 2.306$. The confidence interval is:

$$1.244 \pm 2.306 \left(\frac{0.03}{\sqrt{9}} \right) = 1.244 \pm 0.023 = [1.221; 1.267]$$

Since the interval contains only positive values, we can say that $\mu_1 > \mu_2$. Hence the treatment is effective: there is less pain with the new medication, on average. The answer is C.

13. It is estimated that 18% of university students suffer from depression, 2% consider suicide and 1% suffer from depression and consider suicide. What is the probability that a randomly selected student considers suicide, given that the student does not suffer from depression?

A) 0.056 B) 0.033 C) 0.025 D) 0.012 E) 0.005

Solution: Let A be the event that the student suffers from depression and B be the event that the student considers suicide. We know that $P(A) = 0.18$, $P(B) = 0.02$ and $P(A \cap B) = 0.01$. The desired probability is:

$$P(B|A') = \frac{P(B \cap A')}{P(A')} = \frac{P(B) - P(B \cap A)}{1 - P(A)}$$

$$= \frac{0.02 - 0.01}{1 - 0.18} = \frac{0.01}{0.82} = 0.012$$

The answer is D. Note that A) is the wrong answer:

$$P(A'|B) = \frac{P(A' \cap B)}{P(B)} = \frac{0.01}{0.18} = 0.056$$

14. The probability of colorblindness depends on a person's sex. A person can be male (event M) or female (event F), and we suppose $P(M) = P(F)$. Colorblind persons form 3% of the population. In addition, we know that the probability of colorblindness among males is 5%. Determine the probability that a person is male given that this person is colorblind.

    A) 0.833 B) 0.167 C) 0.05 D) 0.95 E) 0.50

    Solution: Note that $P(M) = P(F) = 0.5$. Let C be the event of colorblindness, $P(C) = 0.03$. In addition, $P(C|M) = 0.05$. Hence,

    $$P(M|C) = \frac{P(M \cap C)}{P(C)} = \frac{P(C|M)P(M)}{P(C)} = \frac{(0.05)(0.5)}{0.03} = 0.833.$$

    The answer is A.

15. A simple urine test was developed for a particular disease. A study involved 100 patients with the disease and 50 patients without the disease. Among the patients with the disease 97 had a positive result, while there were only 5 positive results among the subjects without the disease. For this test, determine: a) the false-positive rate; and b) the false-negative rate.

    A) a) 0.03; b) 0.1 B) a) 0.1; b) 0.03 C) a) 0.97; b) 0.9
    D) a) 0.9; b) 0.97 E) a) 0.5; b) 0.5

    Solution: We have the following table:

	Diseased	Non-diseased
Test+	97	5
Test-	3	45
Total	100	50

Therefore, false-positive-rate= $P(\text{Test+}|\text{True-}) = \frac{5}{50} = 0.1$ and false-negative-rate= $P(\text{Test-}|\text{True+}) = \frac{3}{100} = 0.03$. The answer is B.

16. We consider a random experiment which consists in crossing two yellow mice. The number X of yellow offsprings is a random variable with the following probability distribution:

x	0	1	2	3
$f(x)$	1/27	6/27	12/27	8/27

What is the expected number of yellow offsprings for a randomly chosen pair of yellow mice?

A) 0 B) 1 C) 2 D) 3 E) 1.5

Solution:

$$E(X) = 0(1/27) + 1(6/27) + 2(12/27) + 3(8/27) = 2.$$

The answer is C.

17. The blood type is determined by the alleles I^A, I^B and i of the gene I. A type A person can have genotype I^AI^A or I^Ai, a type B person can have genotype I^BI^B or I^Bi, a type O person has genotype i, a type AB person has genotype I^AI^B. In a couple, the woman has genotype I^Ai and the man has genotype I^Bi. This couple has 4 children. What is the probability that they have at least two children with blood type O?

A) 0.3164 B) 0.4219 C) 0.2109 D) 0.9492 E) 0.2617

Solution: Using a tree diagram, we see that the probability that a child has type O blood is 0.25. Let X be the number of children with type O blood in this family. X has a binomial distribution with $n = 4$ trials and probability of success $p = 0.25$. The desired probability is:

$$\begin{aligned} P(X \geq 2) &= 1 - P(X = 0) - P(X = 1) = 1 - (0.75)^4 - 4(0.25)(0.75)^3 \\ &= 1 - 0.3164 - 0.4219 = 1 - 0.7383 = 0.2617 \end{aligned}$$

The answer is E. A), B), C), D) are the wrong answer obtained by calculating:
$P(X = 0) = 0.3164$, $P(X = 1) = 0.4219$, $P(X = 2) = 0.2109$ and $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.9492$, respectively.

18. According to the Ontario Ministry of Agriculture, 2011 was the year with the lowest annual wheat yield for the province, since 2005. Suppose that the wheat yield of a randomly selected field of 10000 m^2 has a normal distribution with the mean $\mu = 462$ kg and standard deviation $\sigma = 50$ kg. We select randomly 10 wheat fields of this size and we record the wheat yield (in kg) for each of them. What is the probability that the average wheat yield for the 10 fields is between 420 kg and 500 kg?

    A) 0.2234 B) 0.9879 C) 0.9910 D) 0.7511 E) 0.6517

    Solution: The average yield $\overline{X}$ of the 10 fields has a normal distribution with mean $\mu = 462$ and variance $\sigma^2/n = 50^2/10$. Hence,

$$\begin{aligned} P(420 < \overline{X} < 500) &= P\left(Z \leq \frac{500-462}{50/\sqrt{10}}\right) - \left(Z \leq \frac{420-462}{50/\sqrt{10}}\right) \\ &= P(Z \leq 2.40) - P(Z \leq -2.66) \\ &= 0.9918 - 0.0039 = 0.9879 \end{aligned}$$

    The answer is B.

19. The following Minitab output summarizes the data representing the time (in months) until the onset of dementia for 35 patients in psychiatric care.

```
Descriptive Statistics: time

Variable   N  N*  Mean  SE Mean  StDev  Variance  Minimum
time      35   0  7.57     1.23   7.27     52.83     0.07

Variable     Q1  Median     Q3  Maximum
time       2.14    5.33  10.62    34.49
```

    A logarithmic transformation is applied to this data set. The following Minitab output gives the summary for the transformed measurements:

```
Descriptive Statistics: log-time
```

Variable	N	N*	Mean	SE Mean	StDev	Variance	Minimum
log-time	35	0	1.471	0.220	1.301	1.693	-2.664

Variable	Q1	Median	Q3	Maximum
log-time	0.761	1.673	2.363	3.541

Calculate: (a) the geometric mean; (b) the geometric standard deviation.

A) (a) 1.47; (b) 1.30
B) (a) 7.57; (b) 7.27
C) (a) 4.35; (b) 3.67
D) (a) 1.23; (b) 7.27
E) (a) 2.02; (b) 0.26

Solution: The geometric mean and geometric standard deviation are: $e^{\bar{y}} = e^{1.471} = 4.35$ and $e^{s_y} = e^{1.301} = 3.67$. The answer is C. The incorrect answer E is obtained by calculating ln(1.471) and ln(1.301).

20. We continue with the situation in Question 19. Determine the sample range (R) and the interquartile range (IQR) of the time (in months) until the onset of dementia.

A) $R = 34.42$; $IQR = 8.48$
B) $R = 8.48$; $IQR = 34.42$
C) $R = 0.98$; $IQR = 0.24$
D) $R = 0.24$; $IQR = 0.98$
E) $R = 6.20$; $IQR = 1.60$

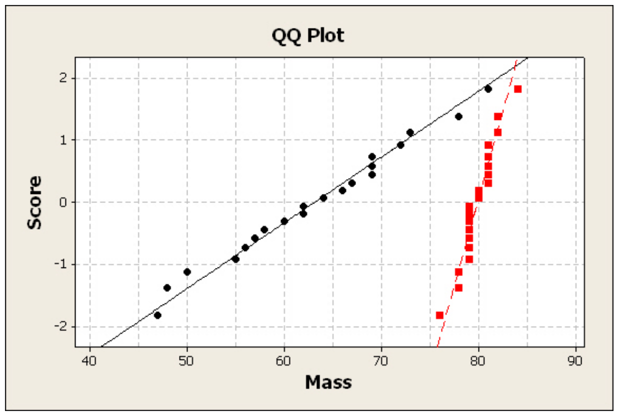
Solution: $R = 34.49 - 0.07 = 34.42$ and $IQR = Q_3 - Q_1 = 10.62 - 2.14 = 8.48$. The answer is A. The incorrect answer E is obtained as $R = 3.541 - (-2.664) = 6.205$ and $IQR = 2.363 - 0.761 = 1.602$

21. We would like to compare the egg masses of tortoises found on the islands of Malabar and Grande Terre in the Indian Ocean. The following Minitab output summarizes the egg-mass measurements for two samples collected on these islands.

Descriptive Statistics: Malabr Grade-Terre

Variable	N	N*	Mean	Variance
Malabar	20	0	63.15	89.40
Grande Terre	20	0	79.90	3.147

To verify the assumptions of normality and equality of variances, we reproduce below the overlayed qq-plots for the two samples.



Assume that the egg mass is normally distributed on both islands. Let μ_1 be the average egg mass for tortoises in the Malabar island and μ_2 be the average egg mass for tortoises in the Grande Terre island. We would like to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$. Calculate the observed value of the test statistic (t_0) and its number (ν) of degrees of freedom.

A) $t_0 = -7.79, \nu = 20$
B) $t_0 = -7.79, \nu = 38$
C) $t_0 = 7.79, \nu = 20$
D) $t_0 = 7.79, \nu = 38$
E) $t_0 = -7.95, \nu = 19$

Solution: The variances are not equal. We use Welch's test. The observed value of the test statistic is:

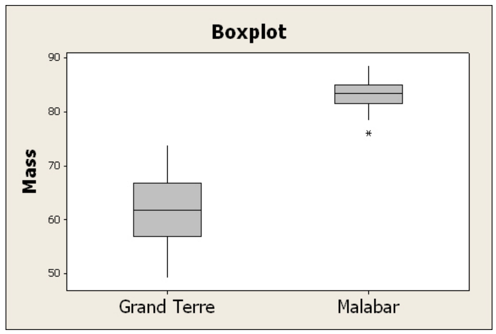
$$t_0 = \frac{63.15 - 79.9}{\sqrt{89.4/20 + 3.147/20}} = -7.79$$

The number of degrees of freedom is

$$\nu = \frac{[89.4/20 + 3.147/20]^2}{[89.4/20]^2/19 + [3.147/20]^2/19} = 20.33$$

We round it down to the closest integer, which is 20. The answer is A.

22. We continue with the situation in Question 21. The following Minitab output shows the box-plots of the egg-mass measurements for the two samples collected on these islands.



Choose the true statement. Only one statement is true.
(A) There is an outlier in the sample of Grand Terre.
(B) The location has no effect on the egg mass.
(C) There are no outliers in the sample of Malabar.
(D) These graphs are meaningless and should not be used to compare the egg masses on different islands.
(E) The median of egg masses is larger for Malabar. However, the egg mass on Malabar has less variability.

Solution: The correct answer is E.

23. The following observations give the total number of metric tons of salt per week, used on roadways in 5 randomly selected counties in Ontario, during the month of January:

$$4516 \quad 6313 \quad 5625 \quad 4462 \quad 3460$$

Assuming that these observations are normally distributed, find a 90% confidence interval for the average number of metric tons of salt per week used in a county.

A) [4058.3; 5692.1] B) [3562.9; 6456.8] C) [4514.5; 5745.3]
D) [3816.5; 5933.9] E) [3439.1; 6321.2]

Solution: For this data set, we have: $\bar{x} = 4875.2$, $s^2 = 1233005$ and $s = 1110.4$. From Table 17.4, we find $t = 2.132$. The interval is:

$$4875.2 \pm 2.132\left(\frac{1110.4}{\sqrt{5}}\right) = 4875.2 \pm 1058.7 = [3816.5; 5933.9]$$

The answer is D. The wrong answer A is obtained using $z = 1.645$.

24. The following table summarizes the data obtained for two random samples consisting of 100 men, respectively 200 women. Each individual in this study was classified according to the hair color and the sex.

	hair color				
Sex	black	brown	blond	red	Total
men	32	43	16	9	100
women	55	65	64	16	200
Total	87	108	80	25	300

The observed value of the test statistic is:

$$\begin{aligned} U_0 &= \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-\hat{E}_{ij})^2}{\hat{E}_{ij}} \\ &= \frac{(32-29)^2}{29}+\frac{(43-36)^2}{36}+\frac{(16-26.6667)^2}{26.6667}+\frac{(9-8.3333)^2}{8.3333}+ \\ &\quad \frac{(55-58)^2}{58}+\frac{(65-72)^2}{72}+\frac{(64-53.3333)^2}{53.3333}+\frac{(16-16.6667)^2}{16.6667} \\ &= 8.987. \end{aligned}$$

Is this test statistic used for a test of independence or a test of homogeneity? What is the range of the p-value?

A) test of independence; $0.025 < p$-value < 0.05
B) test of homogeneity; $0.025 < p$-value < 0.05
C) test of homogeneity; $0.01 < p$-value < 0.025
D) test of independence; $0.01 < p$-value < 0.025
E) test of homogeneity; p-value < 0.005.

Solution: This is a test of homogeneity. The p-value is $P(U > 8.987)$, where U has a χ^2 distribution with $(2-1)(4-1) = 3$ degrees of freedom. From Table 17.5, we see that the p-value is between 0.025 and 0.05. The answer is B.