MAT 2379 A
Final Examination (with Solutions)

December 12, 2010 **Professor Raluca Balan**
Time: 3 hours

Student Number:____________________ **Seat Number:** ______________

Family Name: ______________________ **First Name:** ________________

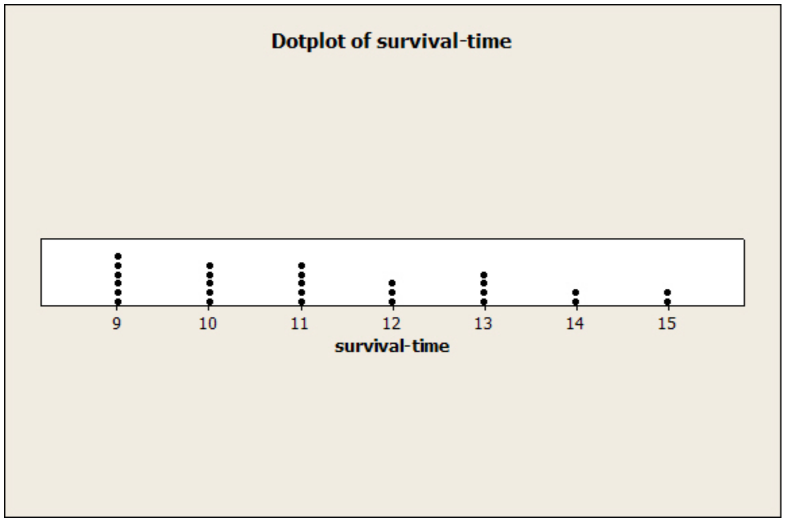
- **This is an open book examination. Only TI 30 calculators are permitted.**
- **Record your answer to each question in the table below. Each question is worth 4 marks.**
- **At the end of the examination, hand in only this page.**

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

****

Professor's use only: Grade=______/25

1. The following graph is the dotplot of the data which gives the survival time (in weeks) after chemotherapy of 27 terminally-ill cancer patients.



Find the mean ($\bar{x}$) and the median ($\tilde{x}$).

A) $\bar{x} = 11.296, \quad \tilde{x} = 11$

B) $\bar{x} = 12.563, \quad \tilde{x} = 12$

C) $\bar{x} = 10.541, \quad \tilde{x} = 11$

D) $\bar{x} = 11.575, \quad \tilde{x} = 13$

E) $\bar{x} = 10.332, \quad \tilde{x} = 12$

Solution: The mean is

$$\bar{x} = \frac{9 \times 6 + 10 \times 5 + 11 \times 5 + 12 \times 3 + 13 \times 4 + 14 \times 2 + 15 \times 2}{27} = 11.296$$

The median is $\tilde{x} = y_{14} = 11$. The answer is A.

2. A chemical spill occurring in a river affected the fish downstream. It is estimated that 30% of the fish are contaminated, i.e. contain more than 1.5 mg of the chemical per kilogram of body weight. A fisherman has caught 5 fish. What is the probability that at least 2 fish are contaminated?

A) 0.5282 B) 0.4718 C) 0.1631 D) 0.3601 E) 0.8369

Solution: Let X be the number of contaminated fish, in the sample of 5 fish. X has a binomial distribution with $n = 5$ trials and $p = 0.30$.

The desired probability is:

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - (0.7)^5 - 5(0.3)(0.7)^4$$

$$= 1 - 0.1680 - 0.3601 = 1 - 0.5282 = 0.4718$$

The answer is B.

3. Let X be the number of smokers in a family composed of a husband and wife. X is a random variable which takes the values 0, 1, 2, with respective probabilities p_0, p_1, p_2. The average number of smokers per family is 0.63. 40% of families contain at least one smoker. Find the probabilities p_0, p_1 and p_2.

   A) $p_0 = 0.6,\ p_1 = 0.35,\ p_2 = 0.05$
   B) $p_0 = 0.6,\ p_1 = 0.2,\ p_3 = 0.2$
   C) $p_0 = 0.6,\ p_1 = 0.17,\ p_2 = 0.23$
   D) $p_0 = 0.4,\ p_1 = 0.4,\ p_3 = 0.2$
   E) $p_0 = 0,\ \ p_1 = 0.31,\ p_2 = 0.16$

   Solution: We know that $P(X \geq 1) = p_1 + p_2 = 0.4$. Hence $p_0 = 0.6$. We have:

$$E(X) = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 = p_1 + 2p_2 = 0.63.$$

   We obtain that $p_2 = 0.63 - 0.4 = 0.23$ and $p_1 = 0.4 - 0.23 = 0.17$. The answer is C.

4. The data summarized by the Minitab output below represents the time (in minutes) to the onset of an allergic reaction to a bee sting for 42 patients.

```
Descriptive Statistics: reaction-time

Variable        N  N*    Mean  SE Mean  StDev  Minimum      Q1
reaction-time  42   0  15.248    0.659  4.272    4.100  12.500

Variable       Median      Q3   Maximum
reaction-time  15.050  18.000    27.000
```

   Give the interquartile range (IQR) and identify some outliers, if they exist.
   A) $IQR = 22.9$, there are no outliers

B) $IQR = 13.9$, there are no outliers
C) $IQR = 14.5$, the value 4.10 is an outlier
D) $IQR = 6.5$, the values 12.50 and 18.00 are outliers
E) $IQR = 5.5$, the values 4.10 and 27.00 are outliers

Solution: $IQR = Q_3 - Q_1 = 18.0 - 12.5 = 5.5$. To identify some outliers, we calculate the fences:

$$Q_1 - (1.5)(IQR) = 12.5 - 8.25 = 4.25, \quad Q_3 + (1.5)(IQR) = 18.0 + 8.25 = 26.25$$

The values 4.1 and 27.0 are outliers, since they lie beyond the fences. The answer is E.

5. The meteorological data collected for a Canadian city show that the amount of snow during the month of December has a normal distribution with standard deviation $\sigma = 25$ cm. During the last 10 years, this city received an average amount of 82.6 cm of snow, with a standard deviation of 24.1 cm. Find a 96% confidence interval for the average amount of snow received by this city in December.

   A) [63.6, 101,6] B) [66.4 , 98.8] C) [64.3, 100.9] D) [65.4, 99.8]
   E) [67.1, 98.1]

   Solution: Since σ is known, the interval is based on the Z distribution. From Table 3, we need to find a value z such that $P(Z < z) = 0.98$. We find $P(Z < 2.05) = 0.9798$ and $P(Z < 2.06) = 0.9803$. We take the midpoint between 2.05 and 2.06, i.e. $z = 2.055$. The interval is:

$$82.6 \pm 2.055 \left(\frac{25}{\sqrt{10}} \right) = 82.6 \pm 16.2 = [66.4, 98.8]$$

   The answer is B.

6. Let μ be the average systolic blood pressure of a large population of hypertensive patients who have been taking a certain medication for one month. Using the data supplied by a sample of 12 patients, it was found that a 95% confidence interval for μ is [124.5, 135.1]. Find s, the sample standard deviation.

   A) 5.30 B) 2.41 C) 18.36 D) 9.37 E) 8.34

Solution: Since σ is not known, the interval is calculated based on the t distribution. From Table 4, we find $t_{0.025}(11) = 2.201$. The interval is given by the formula:

$$\bar{x} \pm 2.201 \left(\frac{s}{\sqrt{12}} \right)$$

From this general formula, we infer that the length of the interval is equal to $2(2.201)(s/\sqrt{12})$. In the case of our data, the length of the interval is 135.1-124.5=10.6. Hence, we must have:

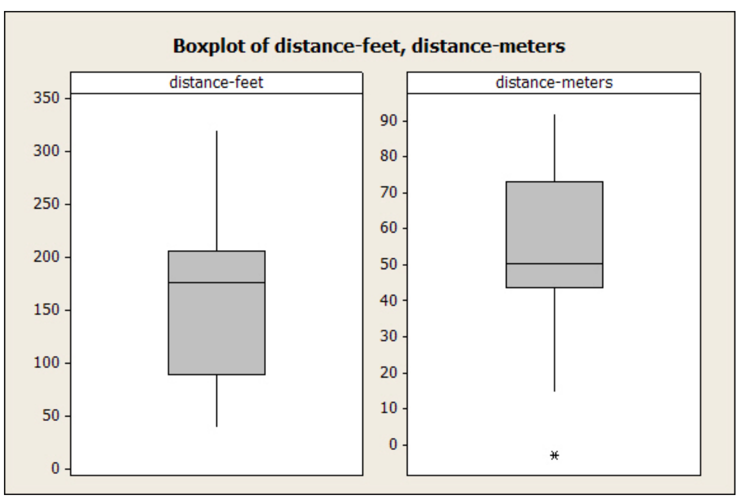
$$2(2.201) \left(\frac{s}{\sqrt{12}} \right) = 10.6$$

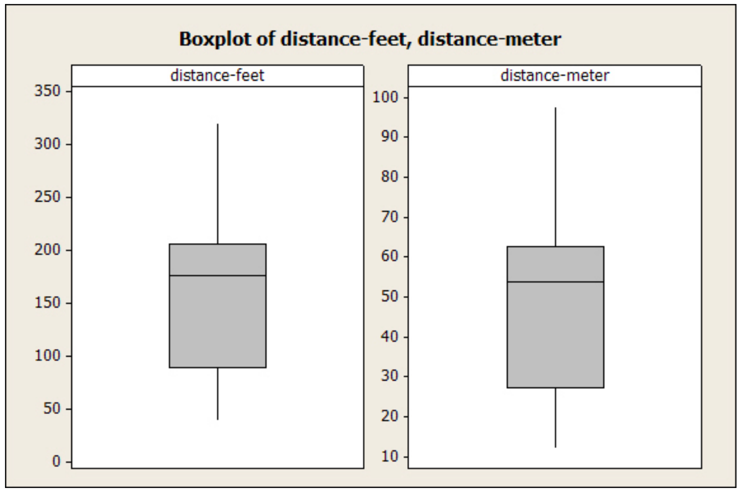
Solving for s, we obtain:

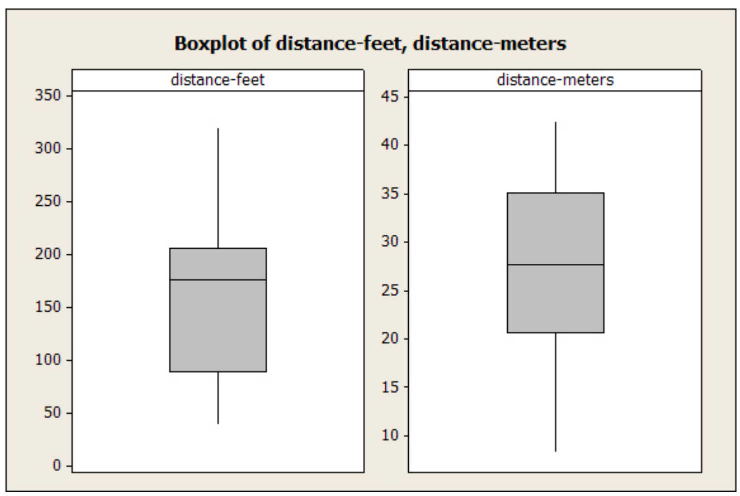
$$s = \frac{(10.6)\sqrt{12}}{2(2.201)} = 8.34$$

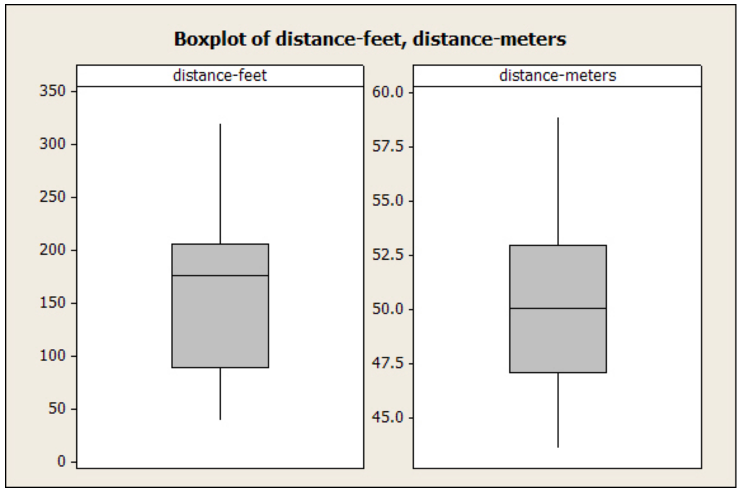
The answer is E.

7. To study the behavior of morning doves, researchers captured 25 birds, tagged them, released them, and then recorded the flight distance from the point of release to the first stop. The flight distances in feet and meters are recorded in Minitab in columns C1, respectively C2. Which one of the following four pictures gives the correct boxplots of the data in columns C1 and C2? (1 foot=0.3048 meters)

A) 

B) 
Boxplot of distance-feet, distance-meter
distance-feet
distance-meter
350
300
250
200
150
100
50
0
100
90
80
70
60
50
40
30
20
10

C) 
Boxplot of distance-feet, distance-meters
distance-feet
distance-meters
350
300
250
200
150
100
50
0
45
40
35
30
25
20
15
10

D) 

Solution: The data in column C2 is a linear transformation of the data in column C1. More precisely, $x'_i = (0.3048)x_i$ for each $i = 1, \ldots, 25$. The median and the quartiles are calculated using the same transformation. From the boxplot of C1, we see that $(Q_1, \tilde{x}, Q_3)$ are approximately $(90, 175, 200)$. Hence, for the transformed data in column C2, $(Q'_1, \tilde{x}', Q'_3)$ should be approximately (0.3048)90=27.43, (0.3048)175=53.34, (0.3048)200=60.96. The picture B) is the only one which corresponds to this calculation.

8. A large percentage of the trees in British Columbia have been affected by forest fires or western spruce budworms in the recent years. It is estimated that 75% of the forest remains in healthy condition, 12% has been devastated by a fire but not by budworms, and 5% has been damaged by budworms but not by fires. What is the probability that a randomly chosen tree has been affected by both fires and budworms?

   A) 0.60 B) 0.24 C) 0.08 D) 0.07 E) 0.13

   Solution: Let A be the event that the tree has been affected by a fire and B be the event that the tree has been affected by budworms. We know that $P(A^c \text{ and } B^c) = 0.75$. Hence $P(A \text{ or } B) = 1 - 0.75 = 0.25$. Using Venn diagrams, we see that:

$$0.25 = P(A \text{ or } B) = P(A \text{ and } B^c) + P(B \text{ and } A^c) + P(A \text{ and } B)$$

$$= 0.12 + 0.05 + P(A\text{and } B)$$

Hence,

$$P(A \text{ and } B) = 0.25 - 0.12 - 0.05 = 0.08$$

The answer is C.

9. According to the Ontario legislation, passengers aged 13 or older can travel in the front seat of a motor vehicle. The following table gives the extent of injuries and the passenger position for 1000 accidents.

Extent of injury	Front Seat	Back Seat
None	188	70
Minor	232	295
Major	102	75
Death	23	15
Total	545	455

Based on this data, what is the probability of a passenger dying in a motor vehicle accident, given that the passenger was traveling in the front seat? Is death independent of the passenger position?

A) 0.04; yes B) 0.15; yes C) 0.05; no D) 0.042; no E) 0.5; yes.

Solution: Let D be the event that the passenger dies, and F be the event that the passenger travels is the front seat. We have:

$$P(D|F) = \frac{P(D \text{ and } F)}{P(F)} = \frac{23/1000}{545/1000} = \frac{23}{545} = 0.042$$

Since $P(D) = 38/1000 = 0.038 \neq 0.042 = P(D|F)$, death is not independent of the passenger position. The answer is D.

10. Based on the data given in problem 9, what is the probability that a passenger traveled in the back seat, given that the passenger did not have any injuries?

A) 0.15 B) 0.36 C) 0.27 D) 0.56 E) 0.21

Solution: Let B be the event that the passenger traveled in the back seat and N be the event that the passenger did not have any injury. We have:

$$P(B|N) = \frac{P(B \text{ and } N)}{P(N)} = \frac{70/1000}{(188 + 70)/1000} = \frac{70}{258} = 0.27$$

The answer is C. (A wrong answer is $P(N|B) = 70/455 = 0.15$, which is given by A.)

11. The weight of a watermelon is normally distributed with mean 1.5 kg and standard deviation 0.7 kg. Due to overproduction and lack of transportation, a farmer has to discard a quarter of his watermelons. He decides to discard all watermelons with a weight smaller than x_0. What is x_0?

A) 0.93 B) 0.70 C) 1.25 D) 0.50 E) 1.03

Solution: Let X be the weight of a randomly chosen watermelon. We have to find x_0 such that $P(X < x_0) = 0.25$. By standardization,

$$0.25 = P(X < x_0) = P\left(\frac{X - 1.5}{0.7} < \frac{x_0 - 1.5}{0.7}\right) = P\left(Z < \frac{x_0 - 1.5}{0.7}\right)$$

In Table 3, we search a value z_0 such that $P(Z < z_0) = 0.25$. We find $P(Z < -0.67) = 0.2514$ and $P(Z < -0.68) = 0.2483$. We take the midpoint between -0.67 and -0.68, i.e. $z_0 = -0.675$. We conclude that:

$$\frac{x_0 - 1.5}{0.7} = -0.675$$

Hence

$$x_0 = 1.5 - (0.7)(0.675) = 1.5 - 0.4725 = 1.0275 \approx 1.03$$

The answer is E.

12. The mutation of the BRCA gene seems to be related to breast cancer. In a study which involved 393 women, the mutation was found in 14 women. Give a 99% confidence interval for the proportion p of women with a mutation of the BRCA gene.

A) [0.017, 0.053] B) [0.012, 0.060]
C) [0.014 , 0.057] D) [0.014, 0.051]
E) [0.02,0.051]

Solution: An estimate for p is $\hat{p} = 14/393 = 0.0356$. The interval is based on the Z table. We find to find z such that $P(Z < z) = 0.995$.

In Table 3, we find $P(Z < 2.57) = 0.9949$ and $P(Z < 2.58) = 0.9951$. We take the midpoint $z = 2.575$. The interval is:

$$0.0356 \pm (2.575)\sqrt{\frac{(0.0356)(1-0.0356)}{393}} = 0.0356 \pm 0.024 = (0.0116, 0.0596).$$

(Using the value $z = 2.57$ or $z = 2.58$ gives the same answer.) The answer is B.

13. Maple syrup is a non-negligible nutritional source of manganese. It is made from the sap of two species of maple: the sugar maple and the black maple. To see if there is a significant difference between the manganese content of these two species of maple, we measured the manganese content (in mg per 100g syrup) for a sample of sugar maples and a sample of black maples. The data is summarized by the Minitab output below:

```
Two-sample T for Sugar Maple - Black Maple

group          N    Mean   StDev  SE Mean

Sugar Maple   20    3.29   0.22    0.049

Black Maple   12    3.55   0.29    0.084
```

Find: (i) a point estimate for $\mu_1 - \mu_2$, where μ_1 and μ_2 are the average manganese contents of the sugar maple, respectively the black maple; (ii) the standard error of $\bar{X}_1 - \bar{X}_2$.

A) (i) -0.07; (ii) 0.101
B) (i) -0.26; (ii) 0.188
C) (i) -0.26; (ii) 0.097
D) (i) -0.26; (ii) 0.101
E) (i) -0.07; (ii) 0.035

Solution: (i) A point estimate for $\mu_1 - \mu_2$ is $\bar{x}_1 - \bar{x}_2 = 3.29 - 3.55 = -0.26$. (ii) The standard error of $\bar{X}_1 - \bar{X}_2$ is:

$$SE_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(0.22)^2}{20} + \frac{(0.29)^2}{12}} = 0.097.$$

The answer is C.

14. The Easter tiger salamander is considered extinct in Ontario, but it persists in the prairies and British Columbia. Adults are usually dark with spots of color grey, green or black. In a random sample of 210 tiger salamanders, 60 have grey spots, 80 have green spots, and 70 have black spots. We would like to test the null hypothesis that the three colors of salamander spots are equally likely: $P\{\text{grey}\} = P\{\text{green}\} = P\{\text{black}\} = 1/3$. Give the range of the P-value and the conclusion of the test at level $\alpha = 0.05$.

    A) P-value< 0.01; the 3 colors are not equally likely
    B) $0.01 <$P-value< 0.02; the 3 colors are not equally likely
    C) $0.02 <$P-value< 0.05; the 3 colors are not equally likely
    D) $0.05 <$P-value< 0.10; the 3 colors are equally likely
    E) P-value> 0.10; the 3 colors are equally likely.

    Solution: We will test H_0: "the three colors are equally likely", against the alternative H_A : the three colors are not equally likely". The expected numbers under H_0 are:

    $$\hat{E}_1 = \hat{E}_2 = \hat{E}_3 = (210)(1/3) = 70$$

    The observed value of the test statistic is:

    $$\chi^2 = \frac{(60-70)^2}{70} + \frac{(80-70)^2}{70} + \frac{(70-70)^2}{70} = \frac{200}{70} = 2.857$$

    P-value$= P(Y > 2.857)$ where Y has a chi-square distribution with $3 - 1 = 2$ d.f. From Table 9, we see that P-value is greater than 0.20. Since the P-value is larger than $\alpha = 0.05$, we fail to reject H_0. We conclude that the 3 colors are equally likely. The answer is E.

15. A botanist is testing a new tomato fertilizer. He was growing two different groups of 8 plants each, using the standard fertilizer for the first group, and the new fertilizer for the second group. After 70 days, he measured the tomato yield (in kg) for each plant. The data is given in the table below:

Plant	Standard Fertilizer	Plant	New Fertilizer
1	4.76	1	5.60
2	4.25	2	4.98
3	3.98	3	5.12
4	3.44	4	3.86
5	3.87	5	4.56
6	4.78	6	5.76
7	3.99	7	4.21
8	3.21	8	4.05
Mean	4.035	Mean	4.767
Standard deviation	0.56	Standard deviation	0.71

Find a 95% confidence interval for $\mu_1 - \mu_2$, where μ_1 is the average tomato yield per plant using the standard fertilizer, and μ_2 is the average tomato yield per plant using the new fertilizer.

A) [-1.42, -0.04]
B) [-0.96,-0.51]
C) [-1.87, 0.21]
D) [-1.35, -0.31]
E) [-2.44, 0.78]

Solution: The two samples are independent. The sample of differences will not be used. The confidence interval is based on the t distribution. The number of degrees of freedom is:

$$\nu = \frac{(0.56^2/8 + 0.71^2/8)^2}{(0.56^2/8)^2/7 + (0.71^2/8)^2/7} = 13.28$$

We take $\nu = 13$. From Table 4, we find $t_{0.025}(13) = 2.160$. The interval is

$$4.035 - 4.767 \pm 2.160\sqrt{\frac{(0.56)^2}{8} + \frac{(0.71)^2}{8}} = -0.732 \pm 0.69 = [-1.422, -0.042]$$

The answer is A. The incorrect answer B is obtained using the paired sample interval: $t_{0.025}(15) = 2.131$

$$-0.732 \pm 2.131\left(\frac{0.3}{\sqrt{8}}\right) = -0.732 \pm 0.226 = [-0.958, -0.506]$$

16. The weight of an adult landlocked Atlantic salmon has a normal distribution with mean 1.4 kg and standard deviation 0.4 kg. A fisherman has caught 4 adult salmons in one day. Calculate:
(i) the probability that a randomly chosen adult salmon has a weight between 1.5 kg and 2.0 kg;
(ii) the probability that the average weight of the 4 salmons is between 1.5 kg and 2.0 kg.

A) (i) 0.3345 (ii) 0.0836
B) (i) 0.0877 (ii) 0.0777
C) (i) 0.3345 (ii) 0.3072
D) (i) 0.3345 (ii) 0.3345
E) (i) 0.3345 (ii) 0.1587

Solution: (i) Let X be the weight of an adult Atlantic salmon. By standardization,

$$\begin{aligned} P(1.5 < X < 2.0) &= P\left(\frac{1.5-1.4}{0.4} < \frac{X-1.4}{0.4} < \frac{2.0-1.4}{0.4}\right) \\ &= P(0.25 < Z < 1.5) \\ &= P(Z < 1.5) - P(Z < 0.25) \\ &= 0.9332 - 0.5987 = 0.3345 \end{aligned}$$

(ii) Let $\bar{X}$ be the average weight of the 4 salmons. By the central limit theorem, $\bar{X}$ has a normal distribution with mean 1.4 and standard deviation $0.4/\sqrt{4} = 0.2$. By standardization,

$$\begin{aligned} P(1.5 < \bar{X} < 2.0) &= P\left(\frac{1.5-1.4}{0.2} < \frac{\bar{X}-1.4}{0.2} < \frac{2.0-1.4}{0.2}\right) \\ &= P(0.5 < Z < 3.0) \\ &= P(Z < 3.0) - P(Z < 0.5) \\ &= 0.9987 - 0.6915 = 0.3072 \end{aligned}$$

The answer is C.

17. The Canadian cow which produces about 18 liters of milk per day has almost disappeared today, being replaced by breeds like Holstein, which produce about 30 liters of milk per day, with 3.25% fat. However, the milk of the Canadian cow is richer in fat and proteins, which is ideal for cheese making. Suppose that a farmer has 6 Canadian cows whose

milk has an average level of fat of 4.5%, with a standard deviation of 0.418%. Find a 98% confidence interval for the average percentage of fat found in the milk of Canadian cows.

A) [4.06, 4.94] B) [4.27, 4.73] C) [3.42, 5.57] D) [4.03, 4.97]
E) [3.93, 5.07]

Solution: Since σ is unknown, the interval is based on the t distribution. From Table 4, we find the value $t_{0.01}(5) = 3.365$. The interval is:

$$4.5 \pm 3.365\left(\frac{0.418}{\sqrt{6}}\right) = 4.5 \pm 0.57 = [3.93; 5.07]$$

The answer is E.

18. A feeding test is conducted on a sample of 24 milking cows to compare two diets: the first diet consists of field wilted alfalfa, the other diet consists of de-watered alfalfa. A sample of 12 cows, randomly selected from the herd, were given the field-wilted alfalfa; another sample of 12 cows were given the de-watered alfalfa. The milk production (in pounds) for each cow is given below:

Field-wilted alfalfa	44	44	56	46	47	38	58	53	49	35	46	30
De-watered alfalfa	35	47	55	29	40	39	32	41	42	57	51	39

Below is the Minitab output which gives the summary of the data:

```
group          N   Mean   StDev   SE Mean
Field-wilted  12  45.50    8.25      2.38
De-watered    12  42.25    8.74      2.52
```

Let μ_1 be the average milk production of cows which are on the field-wilted alfalfa diet, and μ_2 be the average milk production of cows which are on the de-watered alfalfa diet.
(i) Find a 95% confidence interval for $\mu_1 - \mu_2$;
(ii) Test $H_0 : \mu_1 = \mu_2$ against $H_A : \mu_1 \neq \mu_2$ at level $\alpha = 0.05$.

A) (i) $[-6.75, 10.75]$; (ii) 0.10 <P-value<0.20, we do not reject H_0

B) (i) $[-3.97, 10.47]$; (ii) $0.10 <$P-value<0.20, we do not reject H_0
C) (i) $[-3.97, 10.47]$; (ii) P-value> 0.20, we do not reject H_0
D) (i) $[-4.79, 11.29]$; ii) P-value<0.05, we reject H_0
E) (i) $[-3.95,\ 9.87]$; (ii) $0.05 <$P-value< 0.10, we do not reject H_0.

Solution: The two samples are independent. The c.i. and the test will be based on the t distribution. We calculate the number ν of d.f.:

$$\nu = \frac{[(8.25)^2/12 + (8.74)^2/12]^2}{[(8.25)^2/12]^2/11 + [(8.74)^2/12]^2/11} = 21.93$$

We take $\nu = 21$.

(i) From Table 4, we see that $t_{0.025}(21) = 2.08$. The 95% confidence interval is:

$$45.5 - 42.25 \pm 2.08\sqrt{\frac{(8.25)^2}{12} + \frac{(8.74)^2}{12}} = 3.25 \pm 7.22 = [-3.97, 10.47]$$

(ii) The observed value of the test statistic is:

$$t = \frac{45.5 - 42.25}{\sqrt{\frac{(8.25)^2}{12} + \frac{(8.74)^2}{12}}} = 0.937$$

From Table 4, we see that $P(T > 0.937)$ is between 0.10 and 0.20. Hence the P-value of the 2-tailed test is between 0.20 and 0.40. Since the P-value is larger than 0.05, we do not reject H_0. The answer is C.

19. International medical graduates (IMGs) who want to obtain a licence to practice medicine in Ontario have to go through the steps of a highly competitive selection process, despite the fact that they have earned a medical degree in their country of origin. Given the current lack of physicians in Ontario, some people argue that a larger number of IMGs should be given the right to practice. The following table summarizes the attitude towards IMGs of two groups: the first group consist of 50 physicians, the other group consists of 50 non-physicians.

	Physicians	Non-physicians
Larger number of IMGs: yes	30	35
Larger number of IMGs: no	20	15
Total	50	50

Is there enough evidence that the proportions of people who are in favor of giving the right to practice to a larger number of IMGs are different among the physicians, compared to the general population? Justify your answer using a test of hypothesis at level $\alpha = 0.05$. Report the range of the P-value.

A) The proportions are different; $0.01 <$ P-value < 0.02
B) The proportions are different; $0.02 <$ P-value < 0.05
C) The proportions are the same; $0.05 <$ P-value < 0.10
D) The proportions are the same; $0.10 <$ P-value < 0.20
E) The proportions are the same; P-value > 0.20.

Solution: We let p_1 be the proportion of people in favor of more IMGs in the population of physicians and p_2 be the proportion of people in favor of more IMGs in the general population (non-physicians). We would like to test $H_0 : p_1 = p_2$ against $H_A : p_1 \neq p_2$. The expected numbers under H_0 are:

$$\hat{E}_{11} = \hat{E}_{12} = \frac{(50)(65)}{100} = 32.5, \quad \hat{E}_{21} = \hat{E}_{22} = \frac{(50)(35)}{100} = 17.5$$

The observed value of the test statistic is:

$$\chi^2 = \frac{(30-32.5)^2}{32.5} + \frac{(35-32.5)^2}{32.5} + \frac{(20-17.5)^2}{17.5} + \frac{(15-17.5)^2}{17.5} = 1.099$$

From Table 9, we see that the P-value is greater than 0.20. Since the P-value is greater than $\alpha = 0.05$, we do not reject H_0. We conclude that the proportions are the same. The answer is E.

20. 16 subjects participated in an experiment to study the effectiveness of a diet program. For each subject, let x be the weight at the beginning of the program, y be the weight at the end of the program, and $d = x - y$ be the difference between the weight before the program and the weight after the program. Suppose that the sample mean and sample variance of the difference data set are $\overline{d} = 0.75$, respectively $s_d^2 = 1.44$. Is there enough evidence that the program was effective in reducing the weight? Justify your conclusion using a test of hypothesis at level $\alpha = 0.05$. Report the range of the P-value.

A) the program was effective, $0.01 <$ P-value < 0.02

B) the program was effective, $0.02 < \text{P-value} < 0.05$
C) the program was not effective, $0.05 < \text{P-value} < 0.10$
D) the program was not effective, $0.10 < \text{P-value} < 0.20$
E) the program was not effective, $\text{P-value} > 0.20$.

Solution: This is a paired test. We would like to test $H_0 : \mu_X = \mu_Y$ against $H_A : \mu_X > \mu_Y$. The observed value of the test statistic is:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{0.75}{\sqrt{1.44}/\sqrt{16}} = 2.5$$

P-value$= P(T > 2.5)$, where T has a t distribution with 15 d.f. From Table 4, we see that P-value is between 0.01 and 0.02. Since the P-value is smaller that $\alpha = 0.05$, we reject H_0, in favor of H_A. We conclude that the program was effective. The answer is A.

21. Crystalline forms of certain chemical compounds are used in various electronic devices. It is often more desirable to have large crystals rather than small ones. In a laboratory study, 14 crystals of the same initial size were allowed to grow for certain periods of time. The following data gives the weight y of the crystal (in grams) and the period x of time (in hours) which was used for each crystal.

Time	Weight	Time	Weight
2	0.08	16	8.4
4	1.12	18	8.81
6	4.43	20	10.81
8	4.98	22	11.16
10	4.92	24	10.12
12	7.18	26	13.12
14	5.57	28	15.04

For this data, we have:

$$\bar{x} = 70, \quad \bar{y} = 7.55, \quad \sum_{i=1}^{14}(x_i - \bar{x})^2 = 910, \quad \sum_{i=1}^{14}(x_i - \bar{x})(y_i - \bar{y}) = 458.12$$

$$\sum_{i=1}^{14}(y_i - \bar{y})^2 = 244.16$$

The time and weight are stored in columns C1, respectively C2. Below is the result of the linear regression analysis produced by Minitab:

```
Regression Analysis: C2 versus C1

The regression equation is C2 = 0.001 + 0.503 C1

Predictor     Coef   SE Coef      T       P
Constant    0.0014    0.5994   0.00   0.998
C1         0.50343   0.03520  14.30   0.000

S = 1.06177   R-Sq = 94.5%   R-Sq(adj) = 94.0%


Analysis of Variance

Source          DF      SS      MS       F      P
Regression       1  230.63  230.63  204.58  0.000
Residual Error  12   13.53    1.13
Total           13  244.16
```

Assume the linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$. Find a 95% confidence interval for β_1.

A) [-0.426, 0.426]
B) [0.056, 0.950]
C) [0.426, 0.580]
D) [0.405, 0.695]
E) [-0.5,0.5].

Solution: We have $b_0 = 0.001$ and $b_1 = 0.503$. The residual standard deviation is: $s_{Y|X} = 1.06177$. From Table 4, we find $t_{0.025}(12) = 2.179$. The 95% confidence interval for β_1 is:

$$0.503 \pm 2.179 \left(\frac{1.06177}{\sqrt{910}} \right) = 0.503 \pm 0.077 = [0.426, 0.580]$$

The answer is C.

22. Continuing with the situation in question 21, which one of the following statements is not true? (There is only one statement which is not true.)

A) Based on the regression model, we expect to have heavier crystals with extra time.

B) The P-value of the test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ is smaller than 0.05.
C) The average weight of a crystal which was allowed to grow for 15 hours is 7.546g.
D) The correlation coefficient is -0.9721.

Solution: A) is true since the slope b_1 is positive.

b) The observed value of the test statistic is:

$$t = \frac{0.503}{1.06177/\sqrt{910}} = 14.31$$

P-value= $2P(T > 14.31)$ where T has a t distribution with 12 d.f. From Table 4, P-value is smaller than 2(0.0005)=0.001. B) is true.

The average weight of a crystal which was allowed to grow for 15 hours is: $0.001 + (0.503)(15) = 7.546$g. C) is true.

The correlation coefficient is:

$$r = \frac{458.12}{\sqrt{(910)(244.16)}} = 0.97$$

Alternatively, $r = \sqrt{0.945} = 0.9721$. D) is not true. The answer is D.

23. Continuing with the situation in question 21, what is the average change in weight of a crystal for every additional 2 hours?

A) 0.002 B) 1.006 C) 0.503 D) 1.007 E) 2.001

Solution: The average change for one hour is $b_1 = 0.503$. The average change for 2 hours is $2(0.503) = 1.006$. The answer is B.

24. In Minitab, 50 samples of size $n = 100$ are generated from a normal distribution with the mean $\mu = 5$ and the standard deviation $\sigma = 5$. The confidence intervals for the mean were constructed with the same confidence levels assuming that σ is known. The results are given below:

```
One-Sample T: C1, C2, C3, C4, C5, C6, C7, C8, ...,C50

Variable    N   Mean  StDev  SE Mean      C.I.
```

C1	100	4.977	5.459	0.546	(3.893, 6.060)
C2	100	5.234	4.826	0.483	(4.276, 6.191)
C3	100	4.596	5.253	0.525	(3.553, 5.638)
C4	100	4.954	5.084	0.508	(3.946, 5.963)
C5	100	4.552	5.356	0.536	(3.489, 5.615)
C6	100	5.394	4.454	0.445	(4.510, 6.278)
C7	100	5.548	5.123	0.512	(4.531, 6.565)
C8	100	5.590	5.418	0.542	(4.515, 6.665)
C9	100	5.414	5.314	0.531	(4.360, 6.469)
C10	100	4.687	6.153	0.615	(3.466, 5.908)
C11	100	6.192	4.826	0.483	(5.234, 7.150)
C12	100	3.938	5.039	0.504	(2.939, 4.938)
C13	100	4.561	4.921	0.492	(3.584, 5.537)
C14	100	4.898	4.976	0.498	(3.911, 5.886)
C15	100	5.858	4.333	0.433	(4.998, 6.718)
C16	100	5.124	4.477	0.448	(4.236, 6.012)
C17	100	5.569	5.364	0.536	(4.505, 6.634)
C18	100	5.507	5.162	0.516	(4.483, 6.531)
C19	100	5.118	5.036	0.504	(4.119, 6.118)
C20	100	4.301	5.163	0.516	(3.277, 5.326)
C21	100	4.799	4.699	0.470	(3.867, 5.732)
C22	100	5.162	4.515	0.452	(4.266, 6.058)
C23	100	4.868	4.577	0.458	(3.960, 5.776)
C24	100	5.701	5.086	0.509	(4.692, 6.710)
C25	100	5.386	5.079	0.508	(4.378, 6.394)
C26	100	4.938	4.725	0.472	(4.000, 5.875)
C27	100	5.215	4.575	0.457	(4.307, 6.122)
C28	100	4.517	4.087	0.409	(3.706, 5.328)
C29	100	5.555	5.169	0.517	(4.530, 6.581)
C30	100	5.182	5.634	0.563	(4.064, 6.300)
C31	100	4.596	4.366	0.437	(3.729, 5.462)
C32	100	5.290	4.635	0.463	(4.371, 6.210)
C33	100	4.534	4.716	0.472	(3.598, 5.469)
C34	100	4.815	5.397	0.540	(3.745, 5.886)
C35	100	4.934	4.873	0.487	(3.967, 5.901)
C36	100	4.599	5.005	0.500	(3.606, 5.592)
C37	100	4.306	4.657	0.466	(3.382, 5.230)
C38	100	4.581	5.056	0.506	(3.578, 5.585)

C39	100	5.190	4.597	0.460	(4.278, 6.102)
C40	100	4.797	4.969	0.497	(3.811, 5.783)
C41	100	6.329	5.054	0.505	(5.326, 7.332)
C42	100	4.883	4.820	0.482	(3.927, 5.840)
C43	100	5.635	4.951	0.495	(4.652, 6.617)
C44	100	5.534	4.863	0.486	(4.569, 6.498)
C45	100	5.109	5.096	0.510	(4.098, 6.121)
C46	100	4.628	4.611	0.461	(3.713, 5.543)
C47	100	4.993	4.869	0.487	(4.027, 5.959)
C48	100	4.406	4.712	0.471	(3.471, 5.341)
C49	100	4.467	4.873	0.487	(3.500, 5.434)
C50	100	5.422	5.133	0.513	(4.403, 6.440)

Estimate the confidence level of the intervals based on the results produced by Minitab.

A) 94% B) 99% C) 80% D) 90% E) 70%.

Solution: Only C11, C12 and C41 do not cover the actual mean. 50-3=47 intervals cover the actual mean. The percentage is 94%. The answer is A.