MAT 2379 A, Final Examination (with Solutions)

December 10, 2013 **Professor Raluca Balan**
Time: 3 hours

Student Number:_____ **Seat Number:** _____

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

- **This is a closed book examination. Only faculty-approved calculators are permitted: TI 30, TI 34, Casio fx-260 and Casio fx-300.**
- **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	A	14	D
2	B	15	D
3	C	16	D
4	E	17	D
5	D	18	A
6	C	19	A
7	E	20	C
8	B	21	B
9	E	22	D
10	D	23	A
11	C	24	D
12	A	25	B
13	B		

1. The American lobster (Homarus Americanus) has a length that can be modeled by a normal distribution with a mean of 59 cm and a standard deviation of 3.5 cm. Let X be the length of a randomly chosen lobster. Find a length x_0 such that $P(X < x_0) = 0.90$.

   A) 63.5 B) 54.5 C) 66.3 D) 51.4 E) 67.2

   Solution: By standardization,

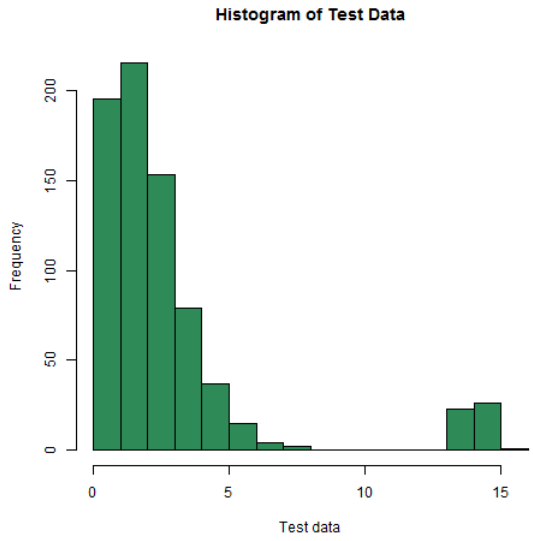
   $$0.90 = P(X < x_0) = P\left(\frac{X - 59}{3.5} < \frac{x_0 - 59}{3.5}\right) = P\left(Z < \frac{x_0 - 59}{3.5}\right)$$

   From Table 17.3 we see that

   $$\frac{x_0 - 59}{3.5} = 1.28$$

   Hence $x_0 = 59 + (3.5)(1.28) = 59 + 4.48 = 63.48$. The answer is A.

2. Consider the following histogram of a data set of 800 observations.



   Below are the descriptive statistics for this data set given by R:

```
Min.      1st Qu.   Median    Mean      3rd Qu.   Max.
0.04699   0.98040   1.74700   2.75900   2.99600   15.25000
```

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

A) The data has a symmetric distribution.
B) The values greater than 6.02 are outliers.
C) The range of the data is 2.0156.
D) The IQR for this data is 15.203.
E) If we draw repeatedly samples of size 5 from the previous data set, we would expect the averages of these samples to follow a normal distribution.

Solution: IQR=2.996-0.9804=2.0156 and (1.5)IQR=3.0234. Fence1 = $0.9804 - 3.0234 = -2.043$ and Fence2 = $2.996 + 3.0234 = 6.0194$. The values greater than 6.0194 are outliers. The answer E is not correct since the central limit theorem does not apply for samples of size 5 (too small). The range of the data is $15.25 - 0.04699 = 15.203$. The answer is B.

3. Consider two samples from independent normal populations. The populations have equal variances and equal means. Let $\overline{X}_1$ and $\overline{X}_2$ be the respective sample means. Let S_p be the pooled sample standard deviation, that is

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}},$$

where S_1 and S_2 are the respective sample standard deviations. Assuming that the sample sizes are $n_1 = n_2 = 13$, find the value b such that

$$P\left(\frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{2/13}} > b\right) = 0.05.$$

A) $b = 1.706$ B) $b = 1.645$ C) $b = 1.711$ D) $b = 2.064$
E) $b = 2.056$

Solution: Under the given conditions,

$$\frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{2/13}} \text{ has a } T(n_1+n_2-2) = T(24) \text{ distribution.}$$

Using Table 17.4, we see that $b = t_{0.05,24} = 1.711$. The answer is C.

4. pH is the negative logarithm (base 10) of the hydrogen ion activity. If x is the hydrogen ion activity, then

$$\text{pH} = -\log_{10}(x) = -\frac{\ln(x)}{\ln(10)}$$

We have a sample of 15 pH measurements in a water based solution at room temperature. The mean pH is 7.5 and the standard deviation for the pH is 1.3. We consider the following linear transformation:

$$y = -\ln(10)\text{pH}$$

Based on this data, what is the sample mean ($\bar{y}$) and sample variance (s_y^2) of the y measurements?

A) $\bar{y} = -16.58,\ s_y^2 = 4.25$
B) $\bar{y} = -15.45,\ s_y^2 = 3.23$
C) $\bar{y} = -12.56,\ s_y^2 = 9.33$
D) $\bar{y} = -14.26,\ s_y^2 = 5.29$
E) $\bar{y} = -17.27,\ s_y^2 = 8.96$

Solution: The sample mean of y is $\bar{y} = (-\ln(10))(7.5) = -17.27$. The sample variance of y is $s_y^2 = [-\ln(10)]^2(1.3)^2 = 8.96$. The answer is E.

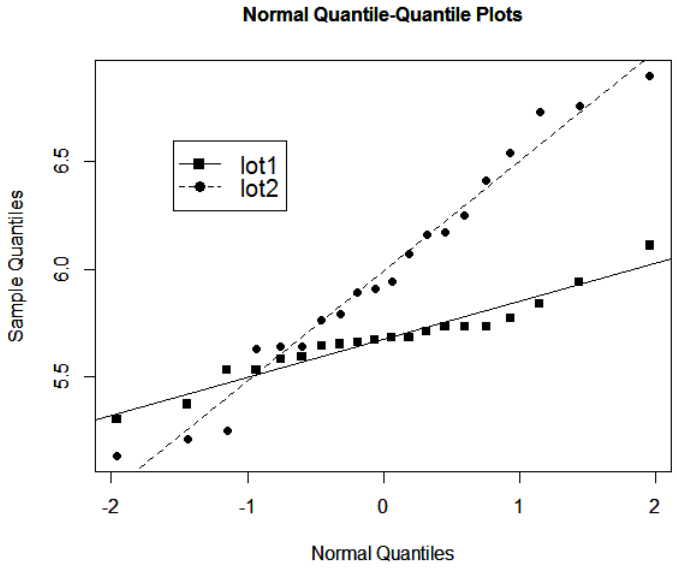
5. A pH level of the soil between 5.3 and 6.5 is optimal for strawberries. To measure the pH level, a field is divided into two lots. In each lot, we select randomly 20 samples of soil. The data are recorded in R as follows:

```
Lot1=c(5.66,5.73,5.68,5.77,5.73,5.71,5.68,5.58,6.11,5.37,
5.67,5.53,5.59,5.94,5.84,5.53,5.64,5.73,5.30,5.65);
Lot2=c(5.25,6.73,6.25,5.21,5.63,6.41,5.89,6.76,5.13,5.64,
5.94,6.16,5.64,6.54,5.79,5.91,6.17,6.90,5.76,6.07)
```

We computed sample means and sample variances for both lots:

	sample size	sample mean	sample variance
Lot1	20	5.672	0.03
Lot2	20	5.989	0.26

We also computed the sample mean and the sample variance for the differences between the values in Lot1 and the values in Lot2: $\bar{d} = -0.317$; $s_D^2 = 0.3044$. Furthermore, we produced the overlayed QQ-plots for the two data sets:



Let μ_1 and μ_2 be the average pH level in Lot1, respectively Lot2. We would like to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$. 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 0.05 and 0.1
C) the p-value is between 0.95 and 0.975
D) the p-value is between 0.01 and 0.025
E) the p-value is between 0.25 and 0.5

Solution: From the QQ-plot we may assume that the two populations are normal with unequal variances. The observed value of the test statistic is:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = \frac{5.672 - 5.989}{\sqrt{0.03/20 + 0.26/20}} = -2.63.$$

The p-value is $2P(T < -2.63) = 2P(T > 2.63)$, where T has an approximate $T(\nu)$ distribution where

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_1^2/n_1)^2/(n_2 - 1)}$$

$$= \frac{(0.03/20 + 0.26/20)^2}{(0.03/20)^2/19 + (0.26/20)^2/19} = 23.33.$$

We will use $\nu = 23$ degrees of freedom. Since $0.005 < P(T > 2.63) < 0.01$, then $0.01 < p\text{-value} < 0.02$. The answer is D.

6. Consider the following data sets:

```
x1=c(1,4,6,12,10,16,20);
x2=c(2,4,5,8,11,14)
```

Note that the sample variance for the first data set is 45.47619. Find the pooled standard deviation for the two data sets.

A) 22.6 B) 11.3 C) 5.85 D) 34.2 E) 34.2

Solution: The sample mean for the second data set is:

$$\bar{x}_2 = \frac{1}{6}(2 + 4 + 5 + 8 + 11 + 14) = 7.33.$$

The sample variance for the second data set is:

$$s_2^2 = \frac{1}{5}[2^2 + 4^2 + 5^2 + 8^2 + 11^2 + 14^2 - 6(7.33)^2] = 20.66667$$

The sample sizes are $n_1 = 7$ and $n_2 = 6$. The pooled variance is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{6(45.47619) + 5(20.66667)}{11} = 34.199$$

The pooled standard deviation is $s_p = \sqrt{34.199} = 5.85$. The answer is C.

7. Buffered aspirin contains on average 325.05 mg of aspirin per tablet, with a standard deviation of 0.5 mg. We select 55 buffered aspirin tablets at random. Approximate the probability that the mean quantity of aspirin per tablet (among the 55 tablets) is between 324.85 mg and 325.25 mg.

A) 0.3108 B) 0.6892 C) 0.9750 D) 0.8777 E) 0.9970

Solution: Let $\bar{X}$ be the mean quantity of aspiring among the 55 tablets. Since $n = 55$ is large, by the central limit theorem with $\mu = 325.05$ and $\sigma = 0.5$, we know that

$$\frac{\bar{X} - 325.05}{0.5/\sqrt{55}} \text{ has a approximately a standard normal distribution.}$$

Hence

$$\begin{aligned} P(324.85 < \bar{X} < 325.25) &\approx P\left(\frac{324.85 - 325.05}{0.5/\sqrt{55}} < Z < \frac{325.25 - 325.05}{0.5/\sqrt{55}}\right) \\ &= P(-2.97 < Z < 2.97) \\ &= P(Z < 2.97) - P(Z - 2.97) \\ &= 0.9985 - 0.0015 = 0.9970 \end{aligned}$$

using Tables 17.2 and 17.3. The answer is E.

8. The probability that a drug A is efficient for the treatment of ear infections is 0.7. For drug B, this probability is 0.8. A large clinical trial is conducted on patients with chronical ear infections: 40% of the patients are treated with drug A, the others are treated with drug B. If the drug has been efficient for a certain patient, what is the probability that this patient was treated with drug A?

A) 0.500 B) 0.368 C) 0.719 D) 0.639 E) 0.576

Solution: Let E be the event that the drug is efficient. We need to calculate $P(A|E)$. Using Bayes' rule, we have

$$P(A|E) = \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|B)P(B)} = \frac{(0.7)(0.4)}{(0.7)(0.4) + (0.8)(0.6)} = 0.368.$$

The answer is B.

9. With a growing aging population, the demand for hip replacement surgery in Canada has been increased in the last decade. The following table gives the distribution of the waiting time (in months) for hip replacement surgery in Ontario in 2013:

Percentage of patients	Waiting time
10%	4 months
18%	5 months
53%	6 months
12%	7 months
7%	8 months

What is the probability that a randomly chosen patient who needs a hip replacement has to wait more than the average waiting time for this operation?
Hint: First compute the average waiting time for this operation.

A) 0.90 B) 0.65 C) 0.19 D) 0.53 E) 0.72

Solution: Let X be the waiting time of the randomly chosen patient. X has a discrete distribution with the following probabilities:

x	4	5	6	7	8
$P(X = x)$	0.1	0.18	0.53	0.12	0.07

The average waiting time is:

$$E(X) = 4(0.1) + 5(0.18) + 6(0.53) + 7(0.12) + 8(0.07) = 5.88.$$

The desired probability is:

$$P(X > 5.88) = P(X = 6)+P(X = 7)+P(X = 8) = 0.53+0.12+0.07 = 0.72.$$

The answer is E.

10. Large mouth bass is one type of fish raised by the aquaculture industry. Suppose that the length of this fish has a normal distribution with a mean of 19.6 cm and a standard deviation of 2.75 cm. Let ℓ be a value (in cm) such that 75% of these fish have a length smaller than ℓ. Which one of the following R commands gives the correct value for ℓ?

A) `qnorm(0.25,19.6,7.5625)`
B) `qnorm(0.75,19.6,7.5625)`
C) `2.75*qnorm(0.75,0,1)`
D) `2.75*qnorm(0.75,0,1)`+19.6

E) 1-qnorm(0.25,19.6,2.75)

Solution: We want to find ℓ such that $0.75 = P(X \leq \ell)$, where $X \sim N(19.6, 2.75)$. By the standardization theorem,

$$0.75 = P\left(Z < \frac{\ell - 19.6}{2.75}\right).$$

Thus, $(l-19.6)/2.75 = z$ where $P(Z < z) = 0.75$, i.e. $z = \text{qnorm}(0.75, 0, 1)$. It follows that $\ell = 2.75\, z + 19.6$. The answer is D.

11. Suppose that 23% of adults are smokers. It is known that 57% of smokers and 13% of non-smokers will develop lung cancer. If a person has lung cancer, what is the probability that he/she was not a smoker?

A) 0.670 B) 0.567 C) 0.433 D) 0.231 E) 0.357

Solution: Let S be the event that a person is a smoker and C be the event that a person develops lung cancer. We are given

$$P(S) = 0.23; \quad P(C|S) = 0.57; \quad P(C|S') = 0.13.$$

We want $P(S'|C)$. By the total probability rule,

$$P(C) = P(C|S)P(S) + P(C|S')P(S') = (0.57)(0.23) + (0.13)(0.77)$$

$$= 0.1311 + 0.1001 = 0.2312.$$

By the multiplication rule,

$$P(S' \cap C) = P(C|S')P(S') = (0.13)(0.77) = 0.1001$$

Hence,

$$P(S'|C) = \frac{P(S' \cap C)}{P(C)} = \frac{0.1001}{0.2312} = 0.433.$$

The answer is C.

12. Dravet syndrome is a rare and severe form of epilepsy. Children with Dravet syndrome experience multiple life-threatening seizures that are resistent to most anti-epileptic medication. The treatment with cannabis (medical marijuana) seems to give positive results in some patients.

This treatment has not yet been approved by Health Canada, but was approved in some states in the United States. We denote by p the proportion of patients treated with cannabis who experience a great reduction of their seizures. It is thought that this proportion is approximately 50%. If the proportion p is higher than 50%, we say that cannabis is effective in reducing the number of seizures. We want to test the hypotheses $H_0 : p = 0.5$ against $H_1 : p > 0.5$. Explain when type I error or type II error occur by choosing the correct statement from the list below.(Only one statement is correct.)

A) Type I error occurs when we decide that cannabis is effective in reducing the number of seizures, when in fact it is not.
B) Type II error occurs when we decide that cannabis is effective in reducing the number of seizures, when in fact it is not.
C) Type I error occurs when we are not able to gain enough evidence that cannabis is effective in reducing the number of seizures, when in fact it is.
D) Type II error occurs when decide that cannabis is not effective in reducing the number of seizures, when indeed it is not effective.

Solution: The answer is A.

13. Milk is an important nutrient for the development of healthy bones. We would like to estimate the average amount μ of milk consumed daily by the children in a certain elementary school. A sample of 44 children yielded a mean consumption of 463 ml of milk per day, with a sample standard deviation of 132 ml. Based on this data, give a 90% confidence interval for μ. Assume that the milk consumption is normally distributed with known standard deviation $\sigma = 140$ ml.

A) [400.26; 455.74] B) [428.28; 497.72] C) [408.76; 517.24]
D) [443.48; 482.52] E) [447.85; 478.15]

Solution: The interval is:

$$463 \pm 1.645 \frac{140}{\sqrt{44}} = [428.28; 497.72]$$

The answer is B.

14. It is reported that the average daily nutrient intake in healthy young

women is 2300 kcal. A study is run on $n = 27$ women to test the validity of this claim. In this study, it was found that the average daily nutrient intake for the 27 women was $\bar{x} = 2363$ kcal with a sample standard deviation $s = 237$ kcal. Is there enough evidence that the the average daily nutrient is significantly different from the value 2300 kcal? Use an appropriate test of hypotheses at level $\alpha = 0.10$. Report the observed value of the test statistic (t_0) and the range of the p-value. (Assume that the daily nutrient intake is normally distributed.)

A) $t_0 = 1.10$ and $0.05 < p\text{-value} < 0.1$. The average daily nutrient intake is significantly different than 2300.
B) $t_0 = 1.10$ and $0.1 < p\text{-value} < 0.2$. The average daily nutrient intake is not significantly different than 2300.
C) $t_0 = 1.38$ and $0.05 < p\text{-value} < 0.1$. The average daily nutrient intake is significantly different than 2300.
D) $t_0 = 1.38$ and $0.1 < p\text{-value} < 0.2$. The average daily nutrient intake is not significantly different than 2300.
E) $t_0 = 2.19$ and $0.05 < p\text{-value} < 0.1$. The average daily nutrient intake is significantly different than 2300.

Solution: Let μ be the the average daily nutrient intake. We want to test $H_0 : \mu = 2300$ versus $H_1 : \mu \neq 2300$. The observed value of the test statistic is:

$$t_0 = \frac{2363 - 2300}{237/\sqrt{27}} = 1.38$$

The p-value is $2P(T_{26} > 1.38)$. From Table 17.4 we see that $P(T_{26} > 1.38)$ is between 0.05 and 0.1. Hence the p-value is between 0.1 and 0.2. Since the p-value is larger than 0.1, we fail to reject H_0. μ is not significantly different than 2300. The answer is D.

15. 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

We select one of these 1000 passengers at random. What is the probability that the passenger died in a motor vehicle accident, given that the passenger was traveling in the front seat? Is death independent of the passenger position?

A) 0.013; yes B) 0.154; 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.

16. A sample of $n = 60$ cigarettes of a certain brand was collected for measuring the level of carbon monoxide, in milligrams per cigarette. The data was imported into R. The following descriptive statistics were produced with R.

```
> names(table)
[1] "Carbone.Monoxide"
> x=table$Carbone.Monoxide
> quantile(x,type=6)
    0%    25%    50%    75%   100%
 7.600 13.025 14.850 16.600 22.100
```

We would like to see if this data contains any outliers. Choose the correct statement from the list below. (Only one statement is correct.)

A) There are no outliers.
B) 7.6 is an outlier, but 22.1 is not an outlier.

C) 22.1 is an outlier, but 7.6 is not an outlier.
D) There are at least two outliers.
E) 13.025 and 16.6 are both outliers.

Solution: $\text{IQR} = Q_3 - Q_1 = 16.6 - 13.025 = 3.575$. The fences are

$$\text{lower fence} = Q_1 - 1.5\text{IQR} = 13.025 - 1.5(3.575)$$

$$= 13.025 - 5.3625 = 7.6625$$

$$\text{upper fence} = Q_3 + 1.5\text{IQR} = 16.6 + 1.5(3.575)$$

$$= 16.6 + 5.3625 = 21.9625.$$

Since 7.6 and 22.1 are outside the fences, then both of these values are outliers. So there are at least two outliers. The answer is D.

17. Selective serotonin reuptake inhibitors (SSRIs) are a group of medications frequently used in the treatment of depression, anxiety and other mental health disorders. A group of researchers would like to estimate the proportion of patients who use SSRIs in the adult population in the Ottawa area. They review the charts of 839 patients chosen at random from the adult patients attending some health clinics in Ottawa. Among the 839 patients, 179 are currently taking an SSRI. Find a 95% confidence interval for the proportion p of patients in the Ottawa area who are currently taking an SSRI.

A) [0.190; 0.237] B) [0.199; 0.228] C) [0.175; 0.201]
D) [0.186; 0.241] E) [0.213; 0.233]

Solution: A point estimate of p is $\hat{p} = 179/839 = 0.213$. The 95% confidence interval for p is:

$$0.213 \pm 1.96\sqrt{\frac{(0.213)(1-0.213)}{839}} = [0.186;\ 0.241]$$

The answer is D.

18. A researcher wants to determine if there is an association between regular cell phone use while driving and involvement in car accidents. The table below gives a summary of the data for a sample of 180 drivers.

regular phone use while driving	accident in the last two years yes	no	total
yes	20	10	30
no	58	92	150
total	78	102	180

Based on these data, is there enough evidence that there is an association between regular cell phone use and involvement in car accidents? Give the observed value for the χ^2 test statistic and state your conclusion at level $\alpha = 0.05$.

A) $u_0 = 7.98$. There is an association between regular cell phone use and involvement in car accidents.
B) $u_0 = 7.98$. There is not enough evidence that there is an association between regular cell phone use and involvement in car accidents.
C) $u_0 = 2.74$. There is an association between regular cell phone use and involvement in car accidents.
D) $u_0 = 2.74$. There is not enough evidence that there is an association between regular cell phone use and involvement in car accidents.
E) $u_0 = 8.73$. There is an association between regular cell phone use and involvement in car accidents.

Solution: We would like to test H_0 : "the cell phone use and the involvement in car accidents are independent" against H_1 : "there is an association between the cell phone use and the involvement in car accidents". The expected frequencies under the assumption of independence are:

$$\hat{E}_{11} = \frac{78 \cdot 30}{180} = 13 \quad \hat{E}_{12} = \frac{102 \cdot 30}{180} = 17$$
$$\hat{E}_{21} = \frac{78 \cdot 150}{180} = 65 \quad \hat{E}_{22} = \frac{102 \cdot 150}{180} = 85$$

regular phone use while driving	accident in the last two years yes	no	total
yes	20 (13)	10 (17)	30
no	58 (65)	92 (85)	150
total	78	102	180

The observed value of the χ^2 test statistic is

$$u_0 = \frac{(20-13)^2}{13} + \frac{(10-17)^2}{17} + \frac{(58-65)^2}{65} + \frac{(92-85)^2}{85} = 7.98.$$

The p-value is $P(U > 7.98)$, where U has a $\chi^2((2-1)(2-1)) = \chi^2(1)$ distribution. By Table 17.5, p value is less than 0.005. We reject H_0 in favor of H_1. There is an association between regular cell phone use and involvement in car accidents. The answer is A.

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

Blood Type	A	B	AB	O
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

Solution: Let X be the number of persons in this sample who have blood type A or AB. Then X has a binomial distribution with $n = 10$ trials and probability of success $p = 0.07 + 0.03 = 0.1$. The probability that at most 2 persons have blood of type A or AB is:

$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) =$$

$$\left(\begin{array}{c} 10 \\ 0 \end{array} \right) (0.1)^0 (0.9)^{10} + \left(\begin{array}{c} 10 \\ 1 \end{array} \right) (0.1)^1 (0.9)^9 + \left(\begin{array}{c} 10 \\ 2 \end{array} \right) (0.1)^2 (0.9)^8 =$$

$$0.34868 + 0.38742 + 0.19371 = 0.92981.$$

The answer is A.

20. In Canada, chicken eggs are sold in cartons of a dozen, rated by size. The large size is a popular size for cooking. To be deemed "large", an egg must weigh between 56g and 62g. The average weight for large eggs should be 59g. A grocery store manager believes that one of the suppliers is putting smaller eggs in the "large" cartons. A sample of 30 eggs selected randomly from "large" cartons yields a sample average

$\bar{x} = 58.7$g with a sample standard deviation $s = 0.3$g. These measurements appear to be normally distributed. Which one of the following situations is appropriate to gain evidence for the hypothesis that the average egg weight in a "large" carton is less than 59g? (Only one answer is correct.)

A) Perform a two-sided test of the null hypothesis that the average egg weight is 59g.
B) Perform a one-sided test of the null hypothesis that the average egg weight is 59g against the alternative that it is less than 59g. Assume that the variance of the egg weight is $\sigma^2 = 1\text{g}^2$. Use the standard normal distribution for the calculation of the p-value.
C) Perform a one sided test of the null hypothesis that the mean is 59g against the alternative that it is less than 59g. Estimate the variance from the data. Use the T distribution with 29 degrees of freedom for the calculation of the p-value.
D) Perform a one-sided test of the null hypothesis that the mean is 59g against the alternative that it is less than 56 g. Estimate the variance from the data. Use the T distribution with 29 degrees of freedom for the calculation of the p-value.
E) Perform a one-sided test of the null hypothesis that the mean is 59g against the alternative that it is less than 59g. Estimate the variance from the data. Use the standard normal distribution for the calculation of the p-value.

Solution: We test $H_0 : \mu = 59$ against $H_1 : \mu < 59$. The variance σ^2 of the population is unknown. We estimate σ^2 by s^2. The observed value of the test statistic is

$$t_0 = \frac{58.7 - 59}{0.3/\sqrt{30}} = -5.47$$

p-value=$P(T_{29} < -5.47)$. The answer is C.

21. Multiple sclerosis (MS) is an inflammatory disease in which the insulating covers of nerves cells in the brain and spinal cord are damaged. For most patients, MS begins with one or two isolated symptoms over a number of days. If 45% of patients have motor problems, 20% have optic neuritis (inflammation of the optic nerve) and 12% have both symptoms, what is the probability that an MS patient had neither one

of these symptoms at the onset of the disease?

A) 0.88 B) 0.47 C) 0.53 D) 0.65 E) 0.35

Solution: Let A be the event that a randomly chosen MS patient had motor problems, and B the event that the patient had optic neuritis. We know that $P(A) = 0.45$, $P(B) = 0.20$ and $P(A \cap B) = 0.12$. By the addition rule,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.45 + 0.20 - 0.12 = 0.53$$

The probability that the patient has neither one of the symptoms is $P(A' \cap B') = 1 - P(A \cup B) = 1 - 0.53 = 0.47$. The answer is B.

22. A researcher wants to see if there is a significant difference between the resting pulse rates for men and women. Here are summary statistics for the resting pulse rates for two samples selected within the two groups.

```
> summary(men)
  Min.   1st Qu.  Median  Mean   3rd Qu.  Max.
  64.00  70.00    73.00   73.03  76.00    81.00
> summary(women)
  Min.   1st Qu.  Median  Mean   3rd Qu.  Max.
  60.00  68.25    73.00   73.10  79.50    87.00
```

The researcher also used some graphical methods for summarizing the data, which are included below.

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

A) It is reasonable to assume that these are samples from two populations with equal variances.
B) The men pulse rates are much more dispersed than the pulse rates for women.
C) The medians in the two boxplots can be used for estimating the two population variances.
D) It is reasonable to assume that the pulse rates are normally distributed from populations with unequal variances.

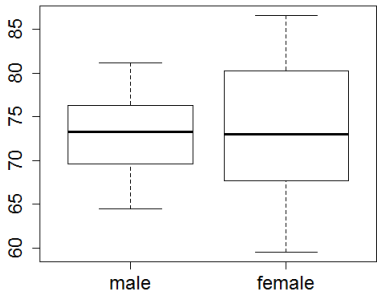


Figure 1: Side-by-side Boxplots

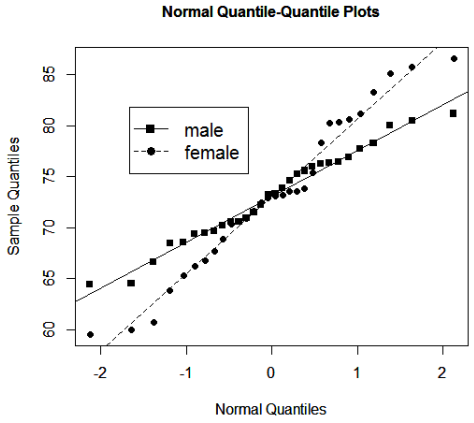


Figure 2: Overlayed QQ-Plots

E) The slopes of the two QQ-plots can be used for a comparison between the two population means.

Solution: Since both QQ plots seem to have a linear tendencies, it is reasonable to assume that both populations are normally distributed. Furthermore, it is not reasonable to assume that the populations have equal variances, since the slopes in the two plots are different. The boxplots further support the hypothesis that the variances are different within the two groups. The women pulse rates are much more dispersed than the pulse rates for men. The answer is D.

23. A study is run to determine the effects of hot summer weather on a large lake. Warm temperatures can lead to a decreased level of dissolved oxygen in the water, which in turn can cause suffocation in certain fish. After a particularly hot period, water samples were taken from 25 randomly selected locations in the lake, and the dissolved oxygen content was measured. This sample of 25 measurements was found to have a mean level of dissolved oxygen of 6.3 parts per million and a standard deviation of 1.7 parts per million. Suppose that the level of dissolved oxygen is normally distributed. Use the following R output to give a 97% confidence interval for the mean content of dissolved oxygen in the lake:

```
> qt(0.985,24)
```

```
[1] 2.306913
> qt(0.97,24)
[1] 1.973994
> qt(0.015,24)
[1] -2.306913
> pt(0.97,24)
[1] 0.8291393
> pt(0.985,24)
[1] 0.8327746
```

A) [5.52; 7.08] B) [5.63; 6.97] C) [5.06; 7.54]
D) [5.18; 7.42] E) [5.35; 7.25]

Hint: The command pt(2.5,3) gives the probability $P(T \leq 2.5)$ where T has a T-distribution with 3 degrees of freedom. The command qt(0.5,4) gives the value t such that $P(T \leq t) = 0.5$ where T has a T-distribution with 4 degrees of freedom.

Solution: The 97% confidence interval is:

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = 6.3 \pm t\frac{1.7}{\sqrt{25}}$$

where t is such that $P(-t \leq T \leq t) = 0.97$ and T has 24 degrees of freedom. To find t we argue as follows: $P(T < -t) = P(T > t) = (1 - 0.97)/2 = 0.03/2 = 0.015$ and $P(T < t) = 0.97 + 0.015 = 0.985$. Hence $t = \text{qt}(0.985, 24) = 2.306913$. The interval is:

$$6.3 \pm (2.306913)\frac{1.7}{\sqrt{25}} = 6.3 \pm 0.78435 = [5.51565; 7.08435]$$

The answer is A. The incorrect answer B is obtained using $t = \text{qt}(0.97, 24) = 1.973994$.

24. An important medical question is whether jogging leads to a reduction in the pulse rate. To test this hypothesis, 10 non-regular jogging volunteers agreed to start a jogging program for a month. At the end of the month, their pulse rates were measured and compared to their earlier values. The data is as follows:

```
Before=c(74,86,98,102,78,84,86,90,94,98)
After=c(70,85,90,100,71,90,84,88,90,94)
```

A researcher types into the R console:

```
t.test(Before,After,alternative="greater")
```

and obtains the following output:

```
data:  Before and After
t = 0.6761, df = 17.98, p-value = 0.2538
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -4.381868       Inf
sample estimates:
mean of x mean of y
     89.0      86.2
```

A student types into the R console:

```
t.test(Before,After,paired=TRUE,alternative="greater")
```

and obtains the following output:

```
data:  Before and After
t = 2.3155, df = 9, p-value = 0.02291
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.5833562       Inf
sample estimates:
mean of the differences
                   2.8
```

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

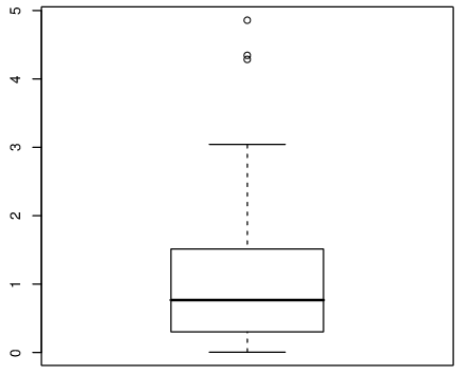
A) The researcher uses the correct command. When $\alpha = 0.05$, there is not enough evidence that jogging leads to a reduction in the pulse rate.

B) The student uses the correct command. When $\alpha = 0.05$, there is not enough evidence that jogging leads to a reduction in the pulse rate.
C) The researcher uses the correct command. When $\alpha = 0.05$, there is enough evidence that jogging leads to a reduction in the pulse rate.
D) The student uses the correct command. When $\alpha = 0.05$, there is enough evidence that jogging leads to a reduction in the pulse rate.
E) Neither the researcher, nor the student are using the correct command. The t-test should not be used in this situation.

Solution: These are paired data. The student uses the correct command. Since in the student's output, the p-value is smaller than 0.05, we reject $H_0 : \mu_X = \mu_Y$ in favor of $H_1 : \mu_X > \mu_Y$. (μ_X is the average pulse rate before, μ_Y is the average pulse rate after one month.) We conclude that the pulse rate has been reduced. The answer is D.

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

```
Min.      1st Qu.   Median    Mean      3rd Qu.   Max.
0.006432  0.303200  0.767000  1.033000  1.513000  4.860000
```



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

A) The interquartile range is 4.8535.
B) There are three outliers.
C) There are no outliers.

D) The distance between the two fences is 0.7670.
E) The data has a symmetric distribution.

Solution: IQR=1.513-0.3032=1.2098. The distance between the fences is:

$$\begin{aligned} \text{Fence2} - \text{Fence1} &= (Q_3 + 1.5IQR) - (Q_1 - 1.5IQR) \\ &= (Q_3 - Q_1) + 3IQR = 4IQR = 4(1.2098) = 4.8392. \end{aligned}$$

The answer is B.