Department of Mathematics and Statistics
University of Ottawa
MAT 2377C
FINAL EXAM SOLUTIONS

10 December 2015
Time: 180 minutes

Professor: Rafal Kulik

Student Number: ____________

Family Name: ____________

First Name: ____________

This is a closed book examination. Only non-programmable and non-graphic calculators are permitted. **Record your answer to each question in the table below.** Your package includes the title page, six pages with questions, the formula sheet, normal and t-tables. Number of questions: **24**. **NOTE: At the end of the examination, hand in only this page.**

Cellular phones, unauthorized electronic devices or course notes (unless an open-book exam) are not allowed during this exam. Phones and devices must be turned off and put away in your bag. Do not keep them in your possession, such as in your pockets. If caught with such a device or document, the following may occur: you will be asked to leave immediately the exam, academic fraud allegations will be filed which may result in you obtaining a 0 (zero) for the exam.By signing below, you acknowledge that you have ensured that you are complying with the above statement.

Your signature: ____________________________

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	

GOOD LUCK !!!

Q1. The events A and B are independent and $P(A) = 0.4$, $P(B^c) = 0.3$. Calculate $P(A \cup B)$

(a) 0.58 (b) 0.82 (c) 0.7

(d) 0.4 (e) None of the preceding

Solution to Q1:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B) = 0.4 + 0.7 - 0.4 * 0.7 = 0.82 \,.$$

Answer b). There is no rounding involved here. No other answer accepted.

Q2. A class consists of 490 female and 510 male students. The students are divided according to their marks

	Passed	Did not pass
Female	430	60
Male	410	100

If one person is selected randomly, the probability that it passed the exam given that it is male is:

(a) 0.067 (b) 0.804 (c) 0.410

(d) 0.196 (e) None of the preceding

Solution to Q2:

A - 'pass', A^c - did not pass, F - female, F^c - 'not female'=male. To compute:

$$P(A|F^c) = \frac{P(A \cap F^c)}{P(F^c)} = \frac{410/1000}{510/1000} = 0.8039216 \,.$$

Answer b). Rounding is obvious here. No other answer accepted.

Q3. A medical research team wished to evaluate a proposed screening test for Alzheimer's disease. The test was given to a random sample of 450 patients with Alzheimer's disease, in 436 cases the test result was positive. Also, the test was given to a random sample of 500 patients without the disease, only in 5 cases the result was positive. It is known that in the Canada 7.7% of the population aged 65 and over have Alzheimer's disease. Find the probability that a person **has the disease** given that the test was positive.

(a) 0.97 (b) 0.88 (c) 0.99

(d) 0.12 (e) None of the preceding

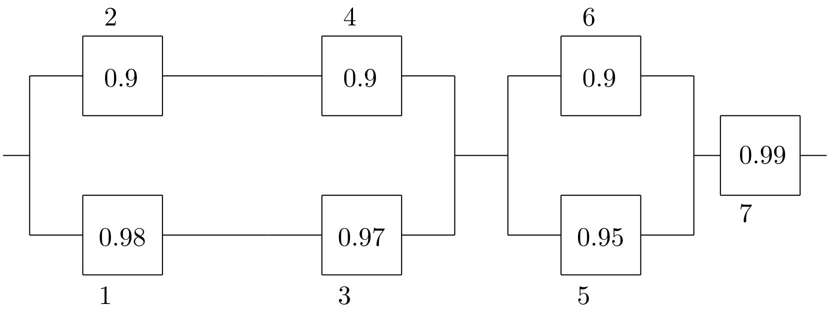
Solution to Q3:

A - test positive, D - a person has disease. Given: $P(A|D) = \frac{436}{450}$, $P(A|D^c) = \frac{5}{500}$, $P(D) = 0.077$. To find: $P(D|A)$ (Bayes' formula):

$$P(D|A) = \frac{P(A|D)P(D)}{P(A|D)P(D) + P(A|D^c)P(D^c)} = 0.8791474.$$

Answer b). Rounding is obvious here. No other answer accepted.

Q4. The following system operates only if there is a path of functional device from left to the right. The probability that each device functions is as shown. What is the probability that the circuit operates? Assume independence. Choose the closest answer.



(a) 0.96 (b) 0.84 (c) 0.78

(d) 0.99 (e) none of the preceding

Solution to Q4:

Let Box A: components 1,2,3,4; Box B: components 5,6; Box C: component 7.

$$P(\text{system works}) = P(\text{A works})P(\text{B works})P(\text{C works}).$$

Now, B is just parallel system, so that

$$P(\text{B works}) = 0.9 + 0.95 - 0.9 * 0.95 = 1.85 - 0.855 = 0.995.$$

Furthermore, $P(2 \text{ and } 4 \text{ work}) = 0.9 * 0.9 = 0.81$, $P(1 \text{ and } 3 \text{ work}) = 0.9506$. Now, A is the parallel system of $2, 4$ and $1, 3$, thus

$$P(\text{A works}) = P(2 \text{ and } 4 \text{ work}) + P(1 \text{ and } 3 \text{ work}) - P(2 \text{ and } 4 \text{ work})P(1 \text{ and } 3 \text{ work}) = 0.9906.$$

Final answer: 0.9757905.

Rounding up gives 0.98, rounding down gives 0.97. There is no such answer. Answer e) However, if you answered a) or d), I will also accept it

Q5. In a NiCd battery, a fully charged cell is composed of Nickelic Hydroxyde. Nickel is an element that has a multiple oxidation states. Let X be the nickel charge, which has the following probability mass function:

x	$f_X(x)$
0	.18
1	k
2	.33
3	.15

Determine the value of k and the variance of the nickel charge.

(a) 0.34, 2.000 (b) 0.15, 1.5235 (c) 0.34, 0.9075
(d) 1.45, 0.6045 (e) None of the preceding

Solution to Q5:

$k = 0.34$ (the sum of probabilities must be one). We have

$$\mu = \sum_x x f_X(x) = 1.45;$$

$$E(X^2) = \sum_x x^2 f_X(x) = 3.01;$$

$$Var(X) = E(X^2) - \mu^2 = 3.01 - 1.45^2 = 0.9075.$$

No issues with rounding here. Answer c)

Q6. ~~Among 100 items, 10 are defective.~~ In a large shipment, there are 10% defective items. We select items until we get one that is defective. What is the probability that we need to select 5 items before we get the one that is defective.

(a) 0.1 (b) 0.59 (c) 0.1^5
(d) 0.53 (e) None of the preceding

Solution to Q6:

The probability is

$$0.9^5 \times 0.1 = 0.059049 .$$

(5 good items and then the sixth one is defective). Answer e)

Q7. In the inspection of tin plate produced by a continuous electrolytic process, 1 imperfection is spotted per minute, on average. Find the probability of spotting at most two imperfections in 5 minutes. Assume that we can model the occurrences of imperfections as a Poisson process.

(a) $\frac{3}{2}\exp(-1)$ (b) $6\exp(-5)$ (c) $\frac{37}{2}\exp(-5)$
(d) $\exp(-1)$ (e) None of the preceding

Solution to Q7:

Rate per 5 minutes: $\lambda = 5$. If $X \sim Poisson(\lambda)$, then to compute $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$:

$$P(X = 0) + P(X = 1) + P(X = 2) = \exp(-\lambda) + \exp(-\lambda)\lambda + \lambda^2/2\exp(-\lambda) = \frac{37}{2}\exp(-5)$$

Answer c)

Q8. Assume that X is a continuous random variable with the density f_X:

$$f_X(x) = \begin{cases} 4x^3 & \text{if } x \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

The value of mean is:

(a)	0.5	(b)	0.8	(c)	0
(d)	3	(e)	None of the preceding.		

Solution to Q8:

$$\int_0^1 x f_X(x)\, dx = \int_0^1 4x^4\, dx = 4/5 = 0.8$$

Q9. A random sample of size $n_1 = 16$ is selected from a normal population with a mean of 75 and variance of 288. A second random sample of size $n_2 = 9$ is taken independently from another normal population with mean 80 and variance of 162. Let $\bar{X}_1$ and $\bar{X}_2$ be the two sample means. Find the probability that $\bar{X}_1 + \bar{X}_2$ **exceeds** 158.

(a)	0.691462	(b)	0.308538	(c)	0.738548
(d)	0.423501	(e)	None of the preceding		

Solution to Q9:

$\overline{X}_1 + \overline{X}_2 \sim N(75 + 80, 288/16 + 162/9) = N(155, 36)$.

$$P(\overline{X}_1 + \overline{X}_2 > 158) = P\left(\frac{\overline{X}_1 + \overline{X}_2 - 155}{\sqrt{36}} > \frac{158 - 155}{\sqrt{36}}\right) =$$
$$= P\left(Z > \frac{158 - 155}{\sqrt{36}}\right) = 1 - P(Z < 0.5) = 1 - 0.6914625 = 0.3085375.$$

This is the answer if you use R. If you read the table you get $1 - 0.6915 = 0.3085$. This should give you answer b).

Q10. Assume that we have a sample of size 10 from a population $N(4, 9)$. Denote by $\overline{X}$ and S^2, the sample mean and sample variance, respectively. Find c such that

$$P\left(\frac{\overline{X} - 4}{S/\sqrt{10}} \leq c\right) = .99.$$

(a)	1.833	(b)	2.326	(c)	1.645
(d)	2.821	(e)	None of the preceding		

Solution to Q10:

Equivalent statement: find c such that

$$P\left(\frac{\overline{X}-4}{S/\sqrt{10}} \geq c\right) = 0.01.$$

We have that $\dfrac{\overline{X}-4}{S/\sqrt{10}}$ has Student distribution with $n-1=9$ degrees of freedom. From the table we read that $P(t_9 > 2.821) = 0.01$, thus $c = 2.821$. Answer d).

Q11. For the following data set evaluate the sample median and the interquartile range.

```
2.6  3.7  0.8  9.6  5.8 -0.8  0.7  4.8  1.2  3.3  5.0  3.7  0.1 -3.1  0.6 0.3
```

(a) median=2.4; IQR=3.3; (b) median=1.9; IQR=3.8; (c) median=1.9; IQR=1.8;
(d) median=2.9; IQR=2.2; (e) None of the preceding

Q12. A textile fiber manufacturer is investigating a new drapery yarn, which the company claims has a mean thread elongation of $\mu = 12$ kilograms with standard deviation of $\sigma = 0.5$ kilograms. What should be the sample size, so that with probability 0.95 we will estimate the mean thread elongation with error at most 0.15 kg?

(a) 43 (b) 42 (c) 28
(d) 31 (e) None of the preceding.

Solution to Q12:

$$n = \left[\left(\frac{z_{.025}\,\sigma}{E}\right)^2\right] + 1 = \left[\left(\frac{(1.96)(0.5)}{0.15}\right)^2\right] + 1 = [42.68] + 1.$$

Thus $n = 43$. ($[\cdot]$ stands for *floor* function, that is rounding down.)

Q13. In a random sample of 1000 houses in a certain city, it is determined that 228 are heated by oil. Find a 99% confidence interval for the number of houses in this city that are heated by oil.

(a) [0.202, 0.254] (b) [0.197, 0.259] (c) [0.194, 0.262]
(d) [0.185, 0.247] (e) None of the preceding

Solution to Q13:

The confidence interval is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}\ ,$$

where $z_{\alpha/2} = z_{0.005} = 2.58$ and $\hat{p} = 228/1000$. The confidence interval is

$$[0.1938262, 0.2621738]$$

There are no issues with rounding.

Q14. An article in *Computers and Electrical Engineering* considered the speed-up of cellular neural networks (CNN) for a parallel general-purpose computing architecture. The data are as follows:

3.77 , 3.35 , 4.21 , 4.03 , 4.03 , 4.63
4.63 , 4.13 , 4.39 , 4.84 , 4.26 , 4.60

Assume that the population is normally distributed. The 99% confidence interval for the mean speed-up is:

(a) [4.155,4.323] (b) [3.863,4.615] (c) [4.040,4.438]
(d) [3.77,4.60] (e) None of the preceding.

Solution to Q14:

```
x=c(3.77 , 3.35 , 4.21 , 4.03 , 4.03 , 4.63,4.63 , 4.13 , 4.39 ,
4.84 , 4.26 , 4.60);

alpha=0.01; n=length(x);

mean(x)-qt(1-alpha/2,n-1)*sd(x)/sqrt(n) [1] 3.863531
mean(x)+qt(1-alpha/2,n-1)*sd(x)/sqrt(n) [1] 4.614802
```

Answer b). There is no issue with rounding.

Q15. The engineer measures weight of $n = 25$ pieces of steel and obtains $\bar{x} = 6$. The weight follows normal distribution with variance 16. The two-sided 95% confidence interval for the mean is:

(a) (-0.272,12.272) (b) (4.432,7.568)
(c) (3.250,8.750) (d) (4.120,7.522)
(e) None of the preceding.

Solution to Q15:

$$(\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}) = (6 - 1.96 * 4/5, 6 + 1.96 * 4/5) = (4.432, 7.568)$$

Answer b). There is no issue with rounding.

Q16. We want to test the hypothesis that the average content of containers of a particular lubricant equals 10 liters against the two-sided alternative. The contents of a random sample of 10 containers are

10.2	9.7	10.1	10.3	10.1
9.8	9.9	10.4	10.3	9.5

Find the p-value of this two-sided test. Assume that the distribution of contents is normal.

(a) $0.05 < p < 0.10$ (b) $0.10 < p < 0.20$ (c) $0.25 < p < 0.40$
(d) $0.50 < p < 0.80$ (e) None of the preceding

Solution to Q16:

We test $H_0 : \mu = 10$ vs. $H_1 : \mu \neq 10$. We have $\overline{x} = 10.03$ and $S^2 = 0.08678$. The p-value is

$$2P(\bar{X} > 10.03) = 2P\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} > \frac{10.03 - 10}{\sqrt{0.08678}/\sqrt{10}}\right)$$

$$2P(t_9 > 0.322) \in 2 \times (0.3, 0.4) = (0.6, 0.8).$$

The p-value is between 0.6 and 0.8, hence certainly between 0.5 and 0.8. Answer d).

Q17. The engineer measures weight of $n = 25$ pieces of steel and obtains $\bar{x} = 6$. The weight follows normal distribution with variance 16. He wants to test $H_0 : \mu = 5$ against $H_1 : \mu > 5$. The p-value for the test is:

(a)	0.2113	(b)	0.10565	(c)	0.89435
(d)	1.0000	(e)	None of the preceding.		

Solution to Q17:

$$P(\bar{X} > 6) = P\left(Z > \frac{6-5}{4/5}\right) = P(Z > 1.25) = 1 - 0.89435 = 0.10565.$$

This is the answer if you use R. If you read the table you get $1 - 0.8944 = 0.1056$. This should give you answer b).

Q18. Assume that random variables $X_1, \ldots, X_8$ follow normal distribution with mean 2 and variance 24. Independently, assume that random variables $Y_1, \ldots, Y_{16}$ follow normal distribution with mean 1 and variance 16. Let $\bar{X}$ and $\bar{Y}$ be the corresponding sample means. Then $P(\bar{X} + \bar{Y} > 4)$ is:

(a)	0.7721	(b)	0.30855	(c)	0.69165
(d)	0.9883	(e)	None of the preceding.		

Solution to Q18:

The question is incidentally very similar to Q9.
$\overline{X} + \overline{Y} \sim N(2 + 1, 24/8 + 16/16) = N(3, 4)$.

$$P(\overline{X} + \overline{Y} > 4) = P\left(\frac{\overline{X} + \overline{Y} - 3}{\sqrt{4}} > \frac{4-3}{\sqrt{4}}\right) = P\left(Z > \frac{4-3}{\sqrt{4}}\right) = 1 - P(Z < 0.5) = 1 - 0.6914625 = 0.3085375.$$

This is the answer if you use R. If you read the table you get $1 - 0.6915 = 0.3085$. This should give you answer b).

Q19. The thickness of a plastic film (in mils) on a substrate material is thought to be influenced by the temperature at which the coating is applied. A completely randomized experiment is carried out. Eleven substrates are coated at 125^oF, resulting in a sample mean coating thickness of $\bar{x}_1 = 103.5$ and a sample standard deviation of $s_1 = 10.2$. Another 11 substrates are coated at 150^oF, for which $\bar{x}_2 = 99.7$ and $s_2 = 11.7$ are observed. We want to test equality of means against the two-sided alternative. The value of the appropriate test statistics and the decision are ($\alpha = 0.05$):

(a)	0.81; Reject H_0.	(b)	0.81; Do not reject H_0.	(c)	1.81; Reject H_0.
(d)	1.81; Do not reject H_0.	(e)	None of the preceding		

Hint: You may consider that population variances are unknown but equal.

Solution to Q19:

This is two-sample test. $s_p^2 = \dfrac{10 * 10.2^2 + 10 * 11.7^2}{20} = 120.465$, $s_p = 10.97$. The value of the test statistics

$$\frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}} = \frac{103.5 - 99.7}{10.97 * \sqrt{1/11 + 1/11}} = 0.81.$$

The p-value is

$$2P(t_{20} > 0.81) \in 2 \times (0.2, 0, 3) = (0.4, 0.6)$$

Do not reject. Answer b).

Q20. Ten individuals have participated in a diet modification program to stimulate weight loss. Their weight both before and after participation in the program is shown below:

Before	$195, 213, 247, 201, 187, 210, 215, 246, 294, 310$
After	$187, 195, 221, 190, 175, 197, 199, 221, 278, 285$

Is there evidence to support the claim that this particular diet-modification program is effective in producing mean weight reduction? Compute p-value and state your conclusion using $\alpha = 0.05$.

(a) p-value< 0.0005; reject H_0

(b) p-value< 0.0005; do not reject H_0

(c) $0.25 <$ p-value< 0.40; reject H_0

(d) $0.25 <$ p-value< 0.40; do not reject H_0

(e) None of the preceding

Solution to Q20:

This is the paired t-test. Compute difference (After-Before):

Before	$195, 213, 247, 201, 187, 210, 215, 246, 294, 310$
After	$187, 195, 221, 190, 175, 197, 199, 221, 278, 285$
D_i	$-8, -18, -26, -11, -12, -13, -16, -25, -16, -25$

To test $H_0 : \mu_D = 0$ vs. $H_0 : \mu_D < 0$. We have $\bar{D} = -17$, $s_D^2 = 41.11$. The observed p-value is

$$P\left(\bar{D} < -17\right) = P\left(\frac{\bar{D}}{s_D/\sqrt{10}} < -8.38\right) = P\left(t_9 < -8.38\right).$$

The P-value is

$$P(t_9 < -8.38) = P(t_9 > 8.38) < 0.0005,$$

reject H_0 - evidence that the diet reduces a weight.

Q21. The following output was produced with `t.test` command in `R`

```
One Sample t-test

data:  x

t = 2.0128,df = 99, p-value = 0.02342

alternative hypothesis: true mean is greater than 0
```

Based on this output, which statement is correct:

(a) If the type I error is 0.05, then we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu > 0$;

(b) If the type I error is 0.05, then we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu \neq 0$;

(c) If the type I error is 0.01, then we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu > 0$;

(d) If the type I error is 0.01, then we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu < 0$;

(e) Type I error is 0.02342.

Q22. A pharmaceutical company claims that a drug decreases a blood pressure. A physician doubts this claim. He tests 10 patients and records results before and after the drug treatment:

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

He is a big fan of `R`, so that he types

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

and obtains

```
data:  Before and After
 t = 0.5499, p-value = 0.2946
alternative hypothesis: true difference in means is greater than 0

estimates: mean of x mean of y
    136.5     133.8
```

His assistant claims that the command should be

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

He obtains

```
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
```

(a) The assistant uses the correct command. There is not enough evidence to justify that the new drug decreases blood pressure;

(b) The assistant uses the correct command. There is enough evidence to justify that the new drug decreases blood pressure for any reasonable choice of α;

(c) The physician uses the correct command. There is enough evidence to justify that the new drug decreases blood pressure for any reasonable choice of α;

(d) Nobody is correct, t-test should not be used here.

For Questions 23, 24 use the data below:

	1	2	3	4	5	6	7	8	9	10
x	825	215	1070	550	480	920	1350	325	670	1215
y	3.5	1.0	4.0	2.0	1.0	3.0	4.5	1.5	3.0	5.0

Hint: $\sum_{i=1}^{n} x_i^2 = 7104300$, $\sum_{i=1}^{n} y_i^2 = 99.75$, $\sum_{i=1}^{n} x_i y_i = 26370$.

Q23. A company employs 10 part-time drivers for its fleet of trucks. Its manager wants to find a relationship between number of kilometers driven (X) and number of working days (Y) in a typical week. (The drivers are hired to drive half-day shifts, so that 3.5 stands for 7 half-day shifts). The manager wants to use the linear regression model

$$Y = \beta_0 + \beta_1 x + \epsilon.$$

The fitted regression line is:

(a) $\hat{y} = 0.1181 + 0.0036x$

(b) $\hat{y} = 117.15 - 0.15x$

(c) $\hat{y} = 5.59 - 0.0036x$

(d) $\hat{y} = 2.85 - 0.1521x$

(e) None of the preceding.

Solution to Q23:

$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = 1297860$, $S_{xy} = 4653$, so that $\hat{\beta}_1 = S_{xy}/S_{xx} = 0.003585$, $\hat{\beta}_1 = \sum_{i=1}^{n} y_i/n - \hat{\beta}_1 \sum_{i=1}^{n} x_i = 0.11823$. Answer a)

There is a rounding problem here. If you round first $\hat{\beta}_1 = 0.0036$, then you get $\hat{\beta}_0 = 0.1068$ that may tempt you to choose e). I will accept both.

Q24. The correlation coefficient is (round to 3 digits)

(a) 0.437

(b) 0.949

(c) 0.113

(d) 1.123

(e) None of the preceding.

Solution to Q24:

$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = 1297860$, $S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = 18.525$, $S_{xy} = 4653$, so that the correlation coefficient is

$$S_{xy}/\sqrt{S_{xx} * S_{yy}} = 0.949$$

This is the last question

Solutions to multiple choice questions:

Q1 $\longrightarrow$ b

Q2 $\longrightarrow$ b

Q3 $\longrightarrow$ b

Q4 $\longrightarrow$ e

Q5 $\longrightarrow$ c

Q6 $\longrightarrow$ e

Q7 $\longrightarrow$ c

Q8 $\longrightarrow$ b

Q9 $\longrightarrow$ b

Q10 $\longrightarrow$ d

Q11 $\longrightarrow$ b

Q12 $\longrightarrow$ a

Q13 $\longrightarrow$ c

Q14 $\longrightarrow$ b

Q15 $\longrightarrow$ b

Q16 $\longrightarrow$ d

Q17 $\longrightarrow$ b

Q18 $\longrightarrow$ b

Q19 $\longrightarrow$ b

Q20 $\longrightarrow$ a

Q21 $\longrightarrow$ a

Q22 $\longrightarrow$ b

Q23 $\longrightarrow$ a

Q24 $\longrightarrow$ b