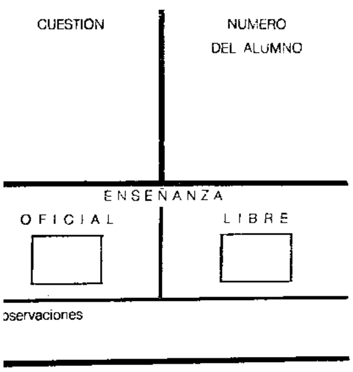
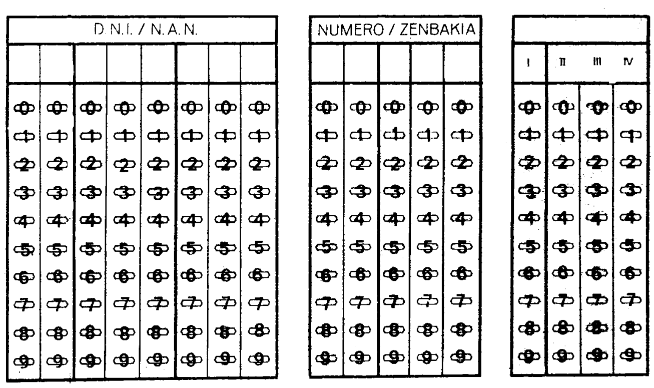
INSTRUCTIONS

1. The exam contains multiple choice questions that must be answered in the orange code sheet we have provided you with, together with three exercises that must be answered in detail in separate sheets of paper.
2. To select an answer, all you need to do is mark in the orange code sheet, **filling the rectangle over which the selected answer is located appropriately**. Please make sure you know the answer you wish to mark before doing it. Even though you can always erase your mark if you have used a pencil (number 2 or similar), any mark that has not been completely erased could be read by the machine. Therefore, we advice you to first mark your selected answers in the exam and to use only the last ten minutes or so from the time assigned to the multiple choice questions-part of the exam to copy them into the code sheet.
3. In the multiple choice questions-part of the exam there is always **only one correct answer** for every question. Every question correctly answered is worth 1 point, while each question incorrectly answered will penalize your grade by 0.2 points. Questions that have not been answered do not penalize your grade in any form.
4. Each one of the exercises, A, B and C, must be answered in a separate sheet of paper. We will collect the different parts of the exam at the indicated times and in this order: first, we will collect the code sheet for the multiple choice questions part of the exam and, then, and in this order, exercises A, B and C will be collected.
5. The exam has six numbered sheets, going from 0.1 to 0.6. Please make sure that you have all sheets and contact your professor if this is not the case. There are different exam types. This exam is of type 0. Mark a 0 in the column labelled with I in your code sheet, just as is illustrated in the example.
6. The maximum final grade for each of the parts of the exam is 30 points. You will need to obtain 15 points in each part of the exam to pass it. However, exams having a multiple choice question part with grades greater than or equal to 14 could, under special circumstances, be compensated with a good grade in the exercises part of the exam.
7. Please fill in your personal information in the appropriate places both in the code sheet and in the sheets provided for the exercises. In "Resit" (column II) you will write the number of times you have registered for an exam in this course *not including this one.*

Example:

$\underbrace{12545}$ PEREZ, Ernesto Exam type $\underbrace{0}$ Resit





MULTIPLE CHOICE QUESTIONS (Time: 1 hour and 45 minutes)

1. FREE-QUESTION. The capital of Spain is:

   (A) Paris (B) Sebastopol (C) Madrid (D) London (E) Pekin

2. If X is a r.v. having a binomial distribution, then we have that $P(1 \leq X < 3)$ is equal to:

   (A) $F(3) - F(1)$

   (B) $P(X \leq 2) + P(X \leq 1)$

   (C) All false

   (D) $P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

   (E) $F(2) - F(0)$

Questions 3 to 5 refer to the following exercise:

In a specific Civil Service examination, the probability that a candidate passes the exam is equal to 0.2. We assume independence between candidates.

3. If 15 of those examinations are graded, the probability that more than 2 candidates pass the exam is:

   (A) 0.6020 (B) 0.3980 (C) 0.2309 (D) 0.8647 (E) 1

4. In the same set of 15 of those examinations, what will be the mean passing number of examinations?

   (A) 4 (B) 2 (C) 1 (D) 5 (E) 3

5. If we have a total of 300 of those examinations, what will be the approximate probability that as much as 53 candidates pass it?

   (A) 0.8264 (B) 0.1587 (C) 0.8413 (D) 0.1736 (E) 1

Questions 6 to 8 refer to the following exercise:

Let X_1, X_2 and X_3 be three independent random variables variables having normal distributions with means equal to -2, 2 and 0, and variances equal to 4, 9 and 1, respectively.

6. The probability that the random variable $W = \dfrac{(X_1+2)^2}{4} + (X_3)^2$ takes on values smaller than 2.77 is:

   (A) 0.05 (B) 0.90 (C) 0.25 (D) 0.75 (E) 0.10

7. The probability that the random variable $Y = \dfrac{\sqrt{2}(X_2 - 2)}{3\sqrt{W}}$ takes on values smaller than 2.92 is:

   (A) 0.20 (B) 0.90 (C) 0.05 (D) 0.95 (E) 0.10

8. The probability that the random variable Y^2 takes on values smaller than 8.53 is:

   (A) 0.05 (B) 0.95 (C) 0.99 (D) 0.90 (E) 0.01

9. If we have a Poisson distribution with parameter λ, where we know that $P(1) = 0.334695$ and that $P(2) = 0.251021$, then the value of the parameter λ is:

   (A) 1.5 (B) 2 (C) 2.5 (D) 1 (E) 0.5

10. Let X and Y be independent r.v. having each an exponential distribution with parameter $\lambda = \frac{1}{2}$. The distribution of the r.v. $Z = X + Y$ is:

(A) $\gamma(1,1)$ (B) $\exp(\lambda = 1)$ (C) All false (D) $\exp(\lambda = \frac{1}{2})$ (E) χ^2_4

Questions 11 and 12 refer to the following exercise:

Let X be a r.v. with probability mass function given by:

$$P(X = -2) = \frac{3\theta}{2} \qquad P(X = 0) = \frac{3\theta}{2} \qquad P(X = 2) = 1 - 3\theta$$

In order to estimate the parameter θ, a r.s. of size n, $X_1, \ldots, X_n$ has been taken.

11. The method of moments estimator of θ is:

(A) $\overline{X}$ (B) $\dfrac{1 - \overline{X}}{2}$ (C) $\dfrac{2 - \overline{X}}{9}$ (D) $\dfrac{9}{2 - \overline{X}}$ (E) $\dfrac{1}{\overline{X}}$

12. In order to be able to obtain an estimate of the parameter θ, a r.s. of size $n = 8$ has been taken providing the following results: -2, -2, -2, -2, 0, 0, 2, 2. The maximum likelihood estimate of θ is equal to:

(A) 0.25 (B) 0.42 (C) 0.28 (D) 2.40 (E) 2.00

Questions 13 and 14 refer to the following exercise:

Let X be a r.v. with probability density function given by:

$$f(x) = (\theta + 2)\, x^{-(\theta+3)}, \qquad x > 1, \quad \theta > 0,$$

and $X_1, \ldots, X_n$ be a r.s. of size n from this distribution. We know that the mean of this r.v. is $m = \frac{\theta+2}{\theta+1}$.

13. The method of moments estimator of the parameter θ is:

(A) $\left(\overline{X} - 2\right)$ (B) $\left(2 - \overline{X}\right)$ (C) $\left(\frac{\overline{X}-2}{1-\overline{X}}\right)$ (D) $\left(\frac{\overline{X}-2}{1+\overline{X}}\right)$ (E) All false

14. The maximum likelihood estimator of the parameter θ is:

(A) $\frac{n-1}{\ln(\prod X_i)}$ (B) $\frac{n}{\ln(\prod X_i)}$ (C) All false (D) $\frac{-n}{\ln(\prod X_i)}$ (E) $\frac{n}{\ln(\prod X_i)} - 2$

Questions 15 to 18 refer to the following exercise:

Let $X_1, \ldots, X_n$ $(n > 3)$ be a r.s. from a population having a Poisson distribution with parameter λ. In order to estimate the parameter λ, we consider the two estimators:

$$\widehat{\lambda}_1 = \frac{X_1 + X_2 + \ldots + X_n}{n} = \overline{X} \qquad \text{and} \qquad \widehat{\lambda}_2 = \frac{3X_1 + X_2 + \ldots + X_n}{n+2}$$

15. Are they unbiased estimators of λ?

(A) Only $\hat{\lambda}_2$ (B) Both of them (C) Only $\hat{\lambda}_1$ (D) None of them (E) -

16. Are they consistent estimators of λ?

(A) Only $\hat{\lambda}_1$ (B) Both of them (C) - (D) None of them (E) Only $\hat{\lambda}_2$

17. If we know that the Cramer-Rao lower bound for a regular and unbiased estimator of λ is $L_c = \frac{\lambda}{n}$, is any of the aforementioned estimators an efficient one?

(A) Only $\hat{\lambda}_2$ (B) Both of them (C) Only $\hat{\lambda}_1$ (D) None of them (E) -

18. Which one of them has smaller variance?

(A) $\hat{\lambda}_1$ (B) Both of them have the same variance (C) -
(D) It depends on the sample values (E) $\hat{\lambda}_2$

Questions 19 to 21 refer to the following exercise:

Let X be a random variable having a normal distribution with variance equal to 49. To test the null hypothesis that the mean is equal to 5, a r.s. of 16 observaciones has been taken. In order to be able to carry out the test, we consider the following critical region for the sample mean, $\text{CR} = (1.29, 8.71)^C$

19. The significance level for this test is:

(A) 0.068 (B) 0.017 (C) 0.034 (D) 0.966 (E) 0.983

20. The power of the test for a mean value $m_1 = 5.315$ is:

(A) 0.9631 (B) 0.0262 (C) 0.9893 (D) 0.0369 (E) 0.9738

21. If we wish to have a larger significance level for the test, the width of the critical region will have to be:

(A) Smaller (B) The same (C) Larger (D) More information is required (E) All false

Questions 22 and 23 refer to the following exercise:

A publicist claims that the mean sales of a given product in a supermarket can be increased if its location is changed. In order to verify his/her claim, a r.s. of the sales of that product that occurred during a 31-day period was recorded, providing sample mean and standard deviation sales values of 3480 and 482 euros, respectively. Once the location of the product was changed, a second r.s. of the sales of the product that occurred during a 41-day period was recorded, providing sample mean and standard deviation sales values of 3300 and 504 euros, respectively.

We assume normality and independence for the sales occurring in different days.

22. A 90% confidence interval for the ratio of the "variance before the change of location" over the "variance after the change of location" is:

(A) $(0.554, 1.726)$ (B) $(0.515, 1.604)$ (C) $(0.626, 1.485)$
(D) $(0.530, 1.650)$ (E) $(0.344, 1.420)$

23. At the 10% significance level, the decision about the increase of the **mean sales** will be:

(A) They have increased (B) - (C) They have not increased
(D) - (E) More information is required

Questions 24 and 25 refer to the following exercise:

A firm wished to test if the probability of buying a new product is equal to 0.2, $p = 0.2$, against the alternative hypothesis that it is greater than 0.2, $p > 0.2$. In order to do so, a r.s. of 20 possible consumers is selected, obtaining that 6 of them would eventually buy the product.

24. If $Z = \sum_{i=1}^{n} X_i$ is used as test statistic, and at the $\alpha = 0.05$ significance level, the most powerful critical region for this test will be:

(A) $[7, \infty)$ (B) $[0, 7]$ (C) $[0, 1]$ (D) $[0, 8]$ (E) $[8, 20]$

25. At the $\alpha = 0.05$ significance level, the decision of the test will be

(A) More information is required (B) - (C) Reject H_0 (D) Do not reject H_0 (E) -

Questions 26 and 27 refer to the following exercise:

We wish to test the null hypothesis that the distribution of the clients, stratified by age, five large phone firms have is the same. In order to do so, five r.s. of 500, 600, 350, 800 and 700 clients from those phone firms were taken. These clients were accordingly classified as a function of their age in four classes: younger than 30, between 30 and 44, between 45 and 60, and older than 60.

26. The test to be carried out will be:

(A) Homogeneity

(B) Independence

(C) Ratio of Variances

(D) Goodness-of-fit to a completely specified distribution

(E) All false

27. The distribution of the test statistic to be used for this test is, under H_0:

(A) χ^2_{20} (B) $F_{5,4}$ (C) χ^2_{12} (D) t_{20} (E) N(0,1)

Questions 28 to 30 refer to the following exercise:

The major of a location with 10000 inhabitants decides to take a poll to be able to estimate the proportion p of people in favor of eliminating bullfights from the local festivities program. We wish to estimate p so that the 95% confidence interval has an estimation error equal to 0.01.

28. What is the required sample size if the sampling is taken with replacement?

(A) 9604 (B) 49 (C) 6724 (D) 960 (E) 4900

29. What is the required sample size if the sampling is taken without replacement?

(A) 4900 (B) 49 (C) 9604 (D) 960 (E) 6724

30. The same poll is taken in 10 different locations of the same province, using the same sample size in each of them. If the sampling is taken without replacement and the true probability in favor of eliminating bullfights is the same for all of the sampled locations, in which one of these locations the estimate of p will be more precise?

(A) In the location with the smallest number of inhabitants

(B) More information is required

(C) It will have the same precision in all locations

(D) In the location with the largest number of inhabitants

(E) All false

EXERCISES (Time: 75 minutes)

A. (10 points, 25 minutes)

Let X be a r.v. having probability density function $f(x) = 1 - \theta^2 \left(x - \frac{1}{2}\right),\ x \in (0,1),\ -1 < \theta < 1.$

We wish to test the null hypothesis $H_0 : \theta = 0$ against the alternative hypothesis $H_1 : \theta = 0.5$. In order to do so, a random sample of one observation, X, is taken.

i) Obtain the form of the most powerful critical region for this test.

ii) At the 5% significance level, find the specific most powerful critical region for this test.

iii) Obtain the power for this test.

B. (10 points, 25 minutes)

Let X be a random variable with probability mass function given by:

$$P(x) = e^{-(\lambda/2)} \left(\frac{\lambda}{2}\right)^x \left(\frac{1}{x!}\right) \qquad x = 0, 1, \cdots, \qquad \lambda > 0$$

In order to be able to estimate the parameter λ, a random sample of size n, $X_1, X_2, \cdots, X_n$, has been taken. It is known that both the mean and the variance of this distribution are equal to $\dfrac{\lambda}{2}$.

i) Find, **providing all relevant details**, the method of moments estimator of the parameter λ.

ii) Find, **providing all relevant details**, the maximum likelihood estimator of the parameter λ.

iii) If we use the estimator $\hat{\lambda} = 2\overline{X}$. Is this an unbiased estimator of λ? Is it consistent? Is it efficient?

Remark: The Cramer-Rao lower bound for a regular and unbiased estimator obtained from a r.s. is:

$$L_c = \frac{1}{nE\left[\frac{\partial \ln f(X,\theta)}{\partial \theta}\right]^2}$$

C. (10 points, 25 minutes)

The number of people that request assistance in a given emergency service at a hospital each day follows a Poisson distribution with parameter $\lambda = 8$. The number of available beds in that emergency service is of only six beds. We assume independence between the different patient arrivals for the different days.

i) What is the probability that, in a given day, more than two patients need to be moved to another hospital because of the lack of available beds?

ii) What would be the required number of beds so that all emergencies can be handled with a minimum probability of 85%?

iii) If we keep the number of beds from the aforementioned statement fixed; that is, six available beds each day, what is the probability that, in two days, it will be necessary to move exactly 2 people to another hospital?

iv) What is the approximate probability that, in a five-day period, more than 30 people request assistance in that hospital's emergency service?

SOLUTIONS TO MULTIPLE CHOICE QUESTIONS (exam type 0)

1: C	11: C	21: C
2: E	12: A	22: D
3: A	13: C	23: C
4: E	14: E	24: E
5: D	15: B	25: D
6: D	16: B	26: A
7: D	17: C	27: C
8: D	18: A	28: A
9: A	19: C	29: A
10: E	20: D	30: A

SOLUTIONS TO EXERCISES

Exercise A

We have a r.v. X with probability density function given by $f(x) = 1 - \theta^2 \left(x - \frac{1}{2}\right), \; x \in (0,1), \; -1 < \theta < 1.$

We wish to test the null hypothesis $H_0 : \theta = 0$ against the alternative hypothesis $H_1 : \theta = 0.5$. In order to do so, a r.s. of one element, X, is taken.

i) To be able to obtain the most powerful critical region for this test, we use the Neyman Pearson Theorem, so that the critical region will be defined by:

$$\begin{aligned}
&\frac{f(x;\theta=0)}{f(x;\theta=0.5)} \leq K \\
&\frac{1}{1-\frac{1}{4}\left(x-\frac{1}{2}\right)} \leq K \\
&\frac{1}{\frac{9}{8}-\frac{x}{4}} \leq K \\
&\frac{9}{8}-\frac{x}{4} \geq K_1 \\
&\frac{x}{4} \leq K_2 \\
&X \leq C
\end{aligned}$$

Therefore, the most powerful critical region for this test will reject the null hypothesis H_0 if $X \leq C, \; C > 0 \implies$ CR = (0, C).

ii) Given that the significance level is $\alpha = 0.05$, we have that:

$$\begin{aligned}
\alpha = 0.05 = P(X \leq C|\theta = 0) &= \\
&= \int_0^C f(x;\theta=0)\; dx = \\
&= \int_0^C 1 \; dx = \\
&= x\,\Big|_0^C = \\
&= C
\end{aligned}$$

Therefore, the most powerful critical region for this test will reject the null hypothesis H_0 if $X \leq 0.05 \implies$ CR = (0, 0.05).

iii) Power

$$\begin{aligned}
\text{Power} = P(X \leq 0.05|\theta = 0.5) &= \\
&= \int_0^{0.05} f(x;\theta=0.05)dx = \\
&= \int_0^{0.05} \left(\frac{9}{8}-\frac{x}{4}\right) dx = \\
&= \left(\frac{9x}{8}-\frac{x^2}{8}\,\Big|_0^{0.05} = \\
&= \frac{1}{8}\left[9 \cdot 0.05 - 0.05^2\right] = \\
&= 0.05594
\end{aligned}$$

Exercise B

$$P(x,\lambda) = e^{-(\lambda/2)} \left(\frac{\lambda}{2}\right)^x \left(\frac{1}{x!}\right) \qquad x = 0,1,2,\ldots, \qquad \lambda > 0$$

$$\mathrm{E}(X) = \mathrm{Var}(X) = \frac{\lambda}{2}$$

i) **Method of moments estimator**

In order to be able to obtain the method of moments estimator of the parameter λ, we need to equate the first population moment $\alpha_1 = \mathrm{E}(X) = m$ to the first sample moment $a_1 = \overline{X}$. That is, $a_1 = \alpha_1$. Given that we have that $\alpha_1 = \frac{\lambda}{2}$ and that $a_1 = \overline{X}$, we obtain that: $\dfrac{\lambda}{2} = \overline{X}$, so that,

$$\hat{\lambda}_{\mathrm{MM}} = 2\overline{X}$$

ii) **Maximum likelihood estimator**

$$L(\vec{x};\lambda) = P(x_1;\lambda)\ldots P(x_n;\lambda) = \left[e^{-\frac{\lambda}{2}}\left(\frac{\lambda}{2}\right)^{x_1}\frac{1}{x_1!}\right]\ \ldots\ \left[e^{-\frac{\lambda}{2}}\left(\frac{\lambda}{2}\right)^{x_n}\frac{1}{x_n!}\right] =$$

$$= e^{-\frac{n\lambda}{2}}\left(\frac{\lambda}{2}\right)^{x_1+\ldots+x_n}\frac{1}{x_1!\,x_2!\ldots x_n!} = e^{-\frac{n\lambda}{2}}\left(\frac{\lambda}{2}\right)^{\sum_{i=i}^n x_i}\frac{1}{\prod_{i=1}^n x_i!}$$

$$\ln L(\vec{x};\lambda) = -\frac{n\lambda}{2} + \left(\sum_{i=1}^n x_i\right)\ln\lambda - \left(\sum_{i=1}^n x_i\right)\ln 2 - \ln\left(\prod_{i=1}^n x_i!\right)$$

$$\frac{\partial \ln L(\vec{x},\lambda)}{\partial\lambda} = -\frac{n}{2} + \left(\sum_{i=1}^n x_i\right)\frac{1}{\lambda} = 0$$

$$\hat{\lambda}_{\mathrm{ML}} = \frac{2\sum_{i=1}^n X_i}{n} = 2\overline{X}$$

iii) **Unbiasedness**

The proposed estimator is unbiased because,

$$\mathrm{E}\left(\hat{\lambda}\right) = \mathrm{E}\left(2\overline{X}\right) = 2\mathrm{E}(\overline{X}) = 2\left(\frac{\lambda}{2}\right) = \lambda$$

Consistency

It is a consistent estimator because the two sufficient conditions required for it hold:

1) $\hat{\lambda}$ is an unbiased estimator and

2) $\lim\limits_{n\to\infty}\left(\mathrm{var}\left(\hat{\lambda}\right)\right) = \lim\limits_{n\to\infty}\mathrm{var}\left(2\overline{X}\right) = \lim\limits_{n\to\infty} 4\dfrac{\mathrm{var}(X)}{n} = \lim\limits_{n\to\infty} 4\left(\dfrac{\lambda}{2n}\right) = \lim\limits_{n\to\infty}\left(\dfrac{2\lambda}{n}\right) = 0$

Efficiency

To verify if the estimator is efficient, we compute the Cramer-Rao lower bound:

$$Lc = \frac{1}{n\mathrm{E}\left[\dfrac{\partial \ln P(X,\lambda)}{\partial\lambda}\right]^2}$$

$$\ln P(x, \lambda) = -\frac{\lambda}{2} + x \ln\left(\frac{\lambda}{2}\right) - \ln x!$$

$$\frac{\partial \ln P(x, \lambda)}{\partial \lambda} = -\frac{1}{2} + \frac{x}{\lambda} = \frac{1}{\lambda}\left(x - \frac{\lambda}{2}\right)$$

$$\mathrm{E}\left[\frac{\partial \ln P(X, \lambda)}{\partial \lambda}\right]^2 = E\left[\frac{1}{\lambda}\left(X - \frac{\lambda}{2}\right)\right]^2 = \frac{1}{\lambda^2}\ \mathrm{var}(X) = \left(\frac{1}{\lambda^2}\right)\ \left(\frac{\lambda}{2}\right) = \frac{1}{2\lambda}$$

Thus, we have that:

$$Lc = \frac{1}{n\left(\frac{1}{2\lambda}\right)} = \frac{2\lambda}{n}$$

As the estimator's variance is equal to the Cramer-Rao lower bound, we conclude that the estimator is efficient.

Exercise C

Let X be the number of people requesting assistance in a given emergency service at a hospital each day, $X \in \mathcal{P}(8)$.

i) More than two patients will be moved to another hospital in a given day if we have that more than 8 patients request assistance at the specific service (there are only 6 available beds):

$$P(X > 8) = 1 - F(8) = 1 - 0.592548 = 0.4074$$

ii) Let k be the number of required beds so that all emergencies can be handled with a minimum probability of 85%:

$$P(X \leq k) \geq 0.85 \quad \Rightarrow \quad F(k) \geq 0.85 \quad \Rightarrow \quad k = 11$$

iii) If there are six beds, in order to have that, in two days, exactly two patients need to be moved to another hospital, we should have that: (a) in one day 6 or fewer patients arrive (we can then have empty beds) and in the other day 8 patients arrive; or (b) 7 patients arrive in each one of the two days.

Let X_1 be the number of patients arriving the first day ($X_1 \in \mathcal{P}(8)$) and X_2 be the number of patients arriving the second day ($X_2 \in \mathcal{P}(8)$):

$$\begin{aligned} P(\text{exactly 2 moves}) &= P(X_1 \leq 6)\, P(X_2 = 8) + P(X_1 = 7)\, P(X_2 = 7) + P(X_1 = 8)\, P(X_2 \leq 6) = \\ &= 2\, F(6)\ [F(8) - F(7)] + [F(7) - F(6)]^2 = \\ &= 2 \cdot 0.313374\ [0.592548 - 0.452961] + [0.452961 - 0.313374]^2 = \\ &= 0.10697 \end{aligned}$$

iv) Let $Z = X_1 + \ldots + X_5$ be the number of patients arriving in a 5-day period.

$Z \in \mathcal{P}(40) \in AN(m = 40, \sigma^2 = 40)$.

The probability that more than 30 patients request assistance in that hospital's emergency service is:

$$\begin{aligned} P(Z > 30 \mid Z \in \mathcal{P}(40)) &\simeq P\left(Z > 30.5 \mid Z \in N(m = 40, \sigma^2 = 40)\right) \\ &= P\left[T > \frac{30.5 - 40}{\sqrt{40}} \mid T \in N(0,1)\right] \\ &= P\left(T > -1.50208\right) \simeq \\ &\simeq \Phi(1.50) = \\ &= 0.9332 \end{aligned}$$