Suffolk County Community College

Michael J. Grant Campus

Department of Mathematics

Spring 2017

MAT 103
Statistics I

Final Exam

Instructor:

Name: Alexander Kasiukov
Office: Suffolk Federal Credit Union Arena, Room A-109
Phone: (631) 851-6484
Email: kasiuka@sunysuffolk.edu
Web Site: `http://kasiukov.com`

Please print the requested information in the spaces provided:

Student:

Name: _____

Student Id: _____

Email: _____ *include to receive the final grade via email* ONLY *if you are not getting email updates*

- *Notes and books are permitted on this exam.*
- *Graphing calculators, smartwatches, computers, cell phones and any other communication-capable devices are prohibited. Their mere presence in the open (even without use) is a sufficient reason for an immediate dismissal from this exam with a failing grade.*
- *You will not receive full credit if there is no work shown, even if you have the right answer. Please don't attach additional pieces of paper: if you run out of space, please ask for another blank final.*

Problem 1. According to the Himalayan Database [1] which contains 1921–2023 data on Mount Everest expeditions [2] two routes are most popular among those who want to reach the summit. Namely, among 2,306 such expeditions recorded in the database, 1,276 used the Southeast Ridge/South Corridor from Nepal, and 780 followed the Northeast Ridge/North Corridor from China (with the next most popular route, the South Pillar/South East Ridge, having been used by only 13 expeditions) [3].

(1). Assuming that the popularity of routes does not change each season, where would you look for a friend who is trying to reach the summit of Mount Everest, if you know nothing else about their location? What are the chances you would be looking for them in the right place? Round the answer to the nearest whole percent.

Space for your solution:

(2). Would the answer change if you happen to know that your friend uses the Southeast Ridge/South Corridor or the Northeast Ridge/North Corridor?

Space for your solution:

[1] `http://www.himalayandatabase.com/downloads.html` assessed May 20, 2024

[2] and covers all known attempts on 406 Himalayan peaks — including Mount Everest — from 1905 to 2023

[3] Consider these counts — as well as all other numbers appearing in this problem — to be (very close) approximations rather than the exact representation of reality, since we are routinely ignoring some border cases, like the fact that some — very few – expeditions used different routes on ascend and descent.

(3). Among the individual climbers who used the Southeast Ridge/South Corridor, $8,096$ succeeded and $6,323$ failed to reach the summit. Among those who followed the Northeast Ridge/North Corridor, the numbers were 3505 for successes and 3652 for the failed attempts. Find the probabilities of success for an individual climber taking each of these routes.

Space for your solution:

(4). During the year 2000 season, 55 climbers attempted to reach the summit of Mount Everest via the Northeast Ridge/North Corridor. If their success rate was independent from each other, how many of them, give or take, do you expect to reach the summit?

Space for your solution:

(5). In the same situation as before, find the 95% confidence interval for the number of successful climbs.

Space for your solution:

(6). With all assumptions of the previous two sub-problems in effect, what is the probability that 37 or more among them reach the summit?

Space for your solution:

(7). Knowing that the actual number of climbers who succeeded to reach the summit of Mount Everest via the Northeast Ridge/North Corridor in the year 2000 was 55, how can you explain this apparent discrepancy with your result in the previous sub-problem?

Space for your solution:

(8). Two teams attempt to climb Mount Everest: one using the South, and another — the North root. Assuming that success of the team is as likely as that of an individual, and the success of one team is independent from the success of the other, what is the probability that both teams reach the summit?

Space for your solution:

(9). What is the probability that at least one of the teams will reach the summit?

Space for your solution:

Problem 2. This problem will introduce you to the *Simpson's Paradox*.

1314 women took part in a study [4] of thyroid disease that was conducted in 1972-1974 in Newcastle, United Kingdom. A follow-up study of the same subjects [5] took place nearly thirty years later.

(1). The subjects of the study were classified according to their smoking habits (current smokers at the time of the original 1970's study or those who never smoked) and according to their survival status 20 years after the original study. The outcomes are summarized in the following table:

	Smoker	**Non-smoker**
Dead	139	230
Alive	443	502

Based on this table, determine if smoking has positive or negative effect on survival.
Hint: compute and compare the conditional probabilities:

$$P(\text{Alive}|\text{Smoker})$$

$$P(\text{Alive}|\text{Non-smoker})$$

Space for your solution:

[4] W. M. G. Tunbridge, D. C. Evered, D. Appleton, M. Brewis, F. Clark
"The Spectrum of Thyroid Disease in a Community: The Whickham Survey",
Clinical Endocrinology, Volume 7, Issue 6, December 1977, Pages 481-493
http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2265.1977.tb01340.x/abstract

[5] David R. Appleton, Joyce M. French and Mark P. J. Vanderpump
"Ignoring a Covariate: An Example of Simpson's Paradox",
The American Statistician, Volume 50, Number 4, November 1996, Pages 340-341
http://www.jstor.org/stable/2684931?seq=1#page_scan_tab_contents

(2). The subjects were further classified according to their age at the time of the original study. The outcomes for women aged 18 to 64 are summarized in this table:

Age 18 to 64	**Smoker**	**Non-smoker**
Dead	97	65
Alive	436	474

Determine if smoking has positive effect on survival of women in this age group.

Space for your solution:

(3). The outcomes for women aged 65 and above are summarized in this table:

Age 65 and above	**Smoker**	**Non-smoker**
Dead	42	165
Alive	7	28

Determine if smoking has positive effect on survival of women in this age group.

Space for your solution:

(4). What conclusion can you draw from this consideration: does smoking improve or harm survival chances? If smoking is beneficial, why it is not shown by the analysis of age groups? If smoking is harmful, why does it contradict the outcome for the combined analysis (that ignores age)?

Space for your solution: