Suffolk County Community College

Michael J. Grant Campus

Department of Mathematics

Spring 2025

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. University of California, Berkeley graduate division admitted 44% of male and 35% of female applicants in the Fall of 1973.

Noticing this apparent discrepancy, Eugene A. Hammel, then the Associate Dean of the Graduate Division,[1] asked Peter Bickel, then a professor of statistics at Berkeley, to analyze the data. The results of that analysis [2] became one of the most widely cited examples of the statistical phenomenon called *Simpson's Paradox*. In this problem, we explore this phenomenon and its ramifications.

The original paper by Bickel at al. does not contain the raw data on the individual departments, but the Data Science Discovery platform [3] has a data set covering all the $12,763$ applicants from the original study. It obscures the specific department names, but identifies the six most popular departments by the department codes A, B, C, D, E and F. In this problem, we will focus only on those six departments, and — in the interest of time — we will further group them into two groups. The departments A and B will form the "easy-to-get-into" group, and departments C, D, E and F will make up the "hard-to-get-into" group. The effect of the Simpson's paradox becomes even more pronounced when only those six departments are considered.

(1). Based on the aggregated six-department data:

	Male	**Female**
Accepted	$1,511$	557
Rejected	$1,493$	$1,278$

compute and compare the conditional probabilities:

$$P(\text{Accepted}|\text{Male}) =$$

$$P(\text{Accepted}|\text{Female}) =$$

and determine if there has been a bias against women in graduate admissions.

Space for your solution:

[1] see Cari Tuna (2009) "When Combined Data Reveal the Flaw of Averages", A Wall Street Journal interview with Peter Bickel, `https://www.wsj.com/articles/SB125970744553071829`,

[2] Bickel, P. J., Hammel, E. A., and O'Connell, J. W. (1975) "Sex bias in graduate admissions: Data from Berkeley", Science, 187, 398–403, `http://brenocon.com/science_1975_sex_bias_graduate_admissions_data_berkeley.pdf`

[3] Berkeley's 1973 Graduate Admissions Dataset, Data Science Discovery, University of Illinois at Urbana-Champaign, `https://discovery.cs.illinois.edu/dataset/berkeley/`

(2). Graduate admission decisions are made by individual departments. In the attempt to "look for the responsible parties", Professor Bickel and his colleagues analyzed data for each of the 101 departments separately. We will use a much more coarse analysis, grouping the six most popular departments into two groups and analyzing the admissions data for those two groups.

Here is the statistics for the easy-to-get-into departments (those labelled as "A" and "B" in the Data Science Discovery dataset):

Easy	**Male**	**Female**
Accepted	1,178	106
Rejected	520	27

and for the hard-to-get-into departments (labelled "C", "D", "E" and "F" in the same dataset):

Hard	**Male**	**Female**
Accepted	333	451
Rejected	973	1,251

Compute and compare the conditional probabilities:

$$P(\text{Accepted}|\text{Male}) =$$

$$P(\text{Accepted}|\text{Female}) =$$

separately for the easy-to-get-into and hard-to-get-into departments.

Space for your solution:

(3). What overall conclusion can you draw from this analysis of admissions data? Did Berkeley discriminate against women in their fall 1973 graduate admissions?

Space for your solution:

Problem 2. The 3rd Nerve Palsy causes the involved eye to deviate in “down and out” direction, and may result in partial or complete ptosis, otherwise know as “lazy eye”. In 98% of cases, this condition is ischaemic (resulting from a restriction to blood supply) and the patients make full recovery without treatment. In 1% of cases, 3rd Nerve Palsy is caused by an aneurysm and in another 1% of cases — by cavernous sinus pathology (CSP). We will assume that these three conditions are always mutually exclusive.

(1). Untreated aneurysms are fatal in 2% of cases, and untreated CSPs — in 50% of cases. Determine the risk of a patient with 3rd Nerve Palsy dying from its cause, if left undiagnosed and untreated.

Space for your solution:

(2). A magnetic resonance angiogaphy (MRA) scan is a non-invasive test for detecting anuerusyms and CSP. However, MRA carries a 5% risk of non-detection.

Determine the risk of a patient with 3rd Nerve Palsy dying from its cause, after having an MRA and the appropriate treatment, if an aneurysm or CSP is detected by the MRA. Assume that the treatment prevents death from aneurysm with certainty, but a treated CSP patient still has 20% risk of death.

Space for your solution:

(3). Assume that magnetic resonance angiogaphy (MRA) has general 5% error rate (meaning both a non-detection of an existing disease, as well as a false detection of a non-existing disease). Suppose a 3rd Nerve Palsy patient has positive MRA result for aneurysm. First guess, and then find using the Bayes formula, the probability that the patient actually has aneurysm. Are you surprised?

Space for your solution:

Problem 3. A grain mill manufactures 100-pound bags of flour for sale in restaurant-supply warehouses. Historically, the weights of bags of flour manufactured at the mill were normally distributed with a mean $\mu = 100$ pounds and a standard deviation $\sigma = 15$ pounds.

(1). What is the probability that the weight of a randomly selected bag of flour falls between 94 and 106 pounds? Use the table of Standard Normal Distribution included at the end of this exam.

Space for your solution:

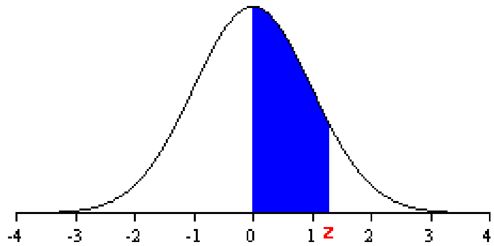
(2). If samples of 36 bags are taken, what is the $\sigma_{\bar{X}}$, the standard error of the mean?

Space for your solution:

(3). What is the probability that a sample of 36 bags of flour has a mean weight between 94 and 106 pounds?

Space for your solution:

Standard Normal Distribution



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.10	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.20	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.30	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.40	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.50	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.60	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.70	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.80	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.90	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.00	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.10	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.20	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.30	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.40	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.50	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.60	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.70	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.80	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.90	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.00	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.10	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.20	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.30	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.40	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.50	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.60	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.70	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.80	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.90	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.00	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990