Textbook Exercises: Chapters 13, 14, and 16

Due on the 10th of April, 2025.

Most problems here are adapted from your textbook exercises. For Problem 1, you will need to perform the finite population correction, which was introduced in Chapter 9, covered in the previous semester already. Problem 5 is the only problem that is not from your textbook; I set it myself. That being said, it is not particularly challenging, really. I include it because it serves as an exemplar of spurious correlation: no causality.

1. A random sample of undergraduate students were asked whether they had purchased a particular product in the past six months, and then they were asked two survey questions. One was whether they liked or disliked a particular advertisement for the product. The other one was whether they thought the advertisement would be effective in selling the product, or it would be ineffective. The table below summarises the survey results. The first count in each pair is of those who had purchased the product recently, and the second count is of those who had not purchased the product.

	effective	ineffective
like	(85, 95)	(11, 3)
dislike	(2, 0)	(53, 86)

   Below we are comparing those undergraduate students who have purchased the product in the recent past, and those who have not.

   (a) Determine with 90% confidence whether or not those who have purchased the product recently are more likely to not only like the advertisement, but also think it will be effective. (10 marks)

   (b) Estimate the difference between the proportions of undergraduate students who not only like the advertisement, but also think it will be effective. Estimate with 90% confidence. Assume infinite populations. (7 marks)

   (c) Repeat Part (b), but assume finite populations. Suppose this survey is focused on the undergraduate students in a local university. The university has 3,582 undergraduate students. You are not sure how many of them have recently purchased the product, but, in your belief, they should account for 20% of them at least, and 80% of them at most. Make your estimate conservative. (10 marks)

2. An investor recorded the weekly returns of two independent portfolios: A and B. He recorded their returns for one year, i.e., four quarters. The following table shows the standard deviation (Bessel-corrected) of the weekly return of each portfolio in each quarter. The investor uses the standard deviation as a measure of return volatility.

	Q1	Q2	Q3	Q4
A	20.95%	11.77%	17.32%	14.00%
B	24.91%	39.11%	30.25%	24.51%

(a) Determine with 99% confidence whether or not A and B differ in terms of their annual weekly return volatility. (10 marks)

(b) Determine with 99% confidence whether or not B's volatility of weekly return is at least 1.5 times that of A's weekly return during Quarter 2. (7 marks)

3. An experiment was conducted to examine attitude towards overweight job applicants. The experimenter used a 2×2 factorial design, between-participant. The first factor is whether the applicant is overweight. The second factor is a control: whether the applicant is male or female. The experimenter presented participants with a video of the applicant being interviewed. After the presentation, participants scored the applicant in terms of his suitability for the job. A higher score indicates higher suitability. Assume the scores to be interval-scaled. The following table summarises the experiment results. In each pair, the first number is the sum of the scores, and the second one is the sum of squared scores. Each condition involved 80 participants.

	average	overweight
male	(4,774, 290,758)	(3,840, 191,656)
female	(4,237, 233,753)	(2,837, 108,817)

Determine with 99% confidence whether or not there is an interaction between the body weight of the applicant and the sex of the applicant. (15 marks)

4. Review Problem 1 in Homework 3. Determine for each commercial style, with 99% confidence, whether or not there is a positive linear relationship between the commercial length and one's memory test score. (7 marks for each)

t	y_t	x_t	t	y_t	x_t
2004	969	17,449	2014	355	5,845
2005	776	16,988	2015	340	5,175
2006	669	15,548	2016	250	4,513
2007	719	14,105	2017	230	3,779
2008	734	12,707	2018	223	3,335
2009	600	11,549	2019	214	3,020
2010	510	10,456	2020	229	2,795
2011	472	8,850	2021	170	2,441
2012	415	7,705	2022	201	2,260
2013	368	6,658	2023	185	2,064

5. Refer to the table on the previous page.

   - t denotes year;
   - x_t denotes the number of male infants born in the US in year t whose first name is Tyler, according to the Social Security Administration; and
   - y_t denotes the Google Trends Web search interest in *desktop background* in all categories (e.g., arts and entertainment, and games) in the US in year t.

   Below I propose a probabilistic model:

$$\ln(Y_t) = \alpha + \beta \ln(X_{t-1}) + \varepsilon_t$$

   where X_{t-1} and Y_t are random variables corresponding to x_{t-1} and y_t, which are realised values. ε_t is the residual, which is also a random variable.

   (a) Estimate α and β using the method of ordinary least squares. (15 marks)

   (b) Draw a scatter plot where the horizontal axis is x_{t-1}, and the vertical axis is y_t. Include the fitted line. Also include the fitted equation. (5 marks)