

Spring 2020/21 Statistics Coursework

Coursework submission guidelines

- If you handwrite your answer, then please write legibly on a blank sheet of paper and scan your handwritten answer as a PDF file before submission.
- If you type your answer using Word software, then please use the Insert -> Equation function of Word or professional add-in software (e.g., MathType) to type maths equations. Please Do save your Word file AS a PDF file before submission. Please do NOT submit .DOC or .DOCX files.
- See <https://www.wikihow.com/Create-PDF-Files> for tips on creating PDF files.
- If you have multiple files, make sure that you combine them into a SINGLE PDF file before uploading it.
- Please answer the questions in order, starting from the first question, and clearly indicate which question you are answering by properly numbering your answers.
- Please use the format: Student Number + Student Name as the file name of the PDF file. For example, 19723856+JohnSmith.

Please answer ALL questions and submit your answers as a PDF file attachment (附件) in the 作业 section of the course website before 4:45pm on Monday, May 31st. No extension will be granted. Late submission will be subject to possible punishment judged by a panel.

Points for each question are indicated at the end of each question:

1. A simple random sample of 50 items from a population with $\sigma = 6$ resulted in a sample mean of 32. Provide a 90% confidence interval for the population mean. (Hint: you need to compute the standard error of the mean and also the margin of error.) [10 points]
2. A simple random sample of 800 elements generates a sample proportion $\bar{p} = .70$. Provide a 90% confidence interval for the population proportion. [10 points]
3. Consider the following hypothesis test:
 $H_0 : \mu \leq 12$
 $H_a : \mu > 12$

A sample of 25 provided a sample mean $\bar{x} = 14$ and a sample standard deviation $s = 4.32$.

[to be continued on next page]



- a) Compute the value of the test statistic. [5 points]
 - b) Use the t distribution table (Table 2 in Appendix B) to compute a range for the p -value. [5 points]
 - c) At $\alpha = .05$, what is your conclusion? [5 points]
 - d) What is the rejection rule using the critical value? What is your conclusion? [5 points]
4. The following results come from two independent random samples taken of two populations.

Sample 1	Sample 2
$n_1 = 50$	$n_2 = 35$
$\bar{x}_1 = 13.6$	$\bar{x}_2 = 11.6$
$\sigma_1 = 2.2$	$\sigma_2 = 3.0$

- a) Provide a 95% confidence interval for the difference between the two population means. [10 points]
 - b) The population means of the two populations are μ_1 and μ_2 , respectively. Consider the following hypothesis test:

$$H_0 : \mu_1 - \mu_2 = 1$$

$$H_a : \mu_1 - \mu_2 \neq 1$$

What is the value of the test statistic? What is the p -value? With $\alpha = .05$, what is your hypothesis testing conclusion? (Hint: be careful of how the H_0 is stated.) [15 points]
5. Consider the following results for independent samples taken from two populations.

Sample 1	Sample 2
$n_1 = 400$	$n_2 = 300$
$\bar{p}_1 = .48$	$\bar{p}_2 = .36$

- a) What is the point estimate of the difference between the two population proportions? [5 points]
- b) Develop a 95% confidence interval for the difference between the two population proportions. [10 points]

[to be continued on next page]



6. **Case Problem.** Consumer Research, Inc., is an independent agency that conducts research on consumer attitudes and behaviors for a variety of firms. In one study, a client asked for an investigation of consumer characteristics that can be used to predict the amount charged by credit card users. Data were collected on annual income (measured in thousand dollars) and annual credit card charges (measured in dollars) for a sample of 50 consumers.

Please download the data named “**Data for coursework.csv**” from the “**Resource (资料)**” area of the Course webpage. In this data file, the variables “Income” and “AmountCharged” refer to the annual income and annual credit card charges mentioned above, respectively.

Please do NOT submit your R codes because they have been given in the Hints.

Please do NOT take a screenshot of the results in R console and paste it into the PDF file.

- a) Use classes starting at 1800 and ending at 6000 in increments of 1000 for the variable “AmountCharged” to show its frequency distribution. [10 points]

Hint: After you specify in R where the data file is located on your computer, try the following codes:

```
# import data into R
data <- read.csv("Data for coursework.csv")

# for frequency distribution
group <- cut(data$AmountCharged, seq(1800, 6000, 1000), right=FALSE)
freq <- sapply(split(data, group), nrow)
cbind(freq)
```

- b) Estimate a regression equation:

$$\text{AmountCharged} = \beta_0 + \beta_1 \text{Income} + \varepsilon$$

Obtain the intercept and slope coefficients. Do you expect the sign of the slope to be positive? Explain. [10 points]

Hint: the R codes for this question is

```
summary(lm(AmountCharged ~ Income, data=data))
```