

**Homework 2**

Due, Oct 7th @ 5:00pm

**P8130 Guidelines for Submitting Homework**

Your homework should be submitted only through CourseWorks. No email submissions!

All derivations, graphs, output and interpretations to each section of the problem(s) must be included in the PDF (not the code), otherwise it will not be graded.

Only 1 PDF file should be submitted. When derivations were required and handwriting was allowed, scan the derivations and merge ALL PDF files (<http://www.pdfmerge.com/>) into a single one.

We are encouraged to use R for calculations, but you still have to show the mathematical formulae. Also, make sure to also submit your commented code as a separate R/RMD file.

**DO NOT FORGET:**

You are encouraged to collectively look for answers, explain things to each other, and use questions to test each other knowledge.

*But*

Do NOT hand out answers to someone who has not done any work. Everyone ought to have ideas about the possible answers or at least some thoughts about how to probe the problem further. Write your own solutions!

For all problems below, assume a significance level of 0.05 unless stated otherwise. You can use R to check your calculations, but you still need to type the formulae/calculations – NO PICS !

### Problem 1 (10p)

The incidence of uveal melanoma in the US is approximately 5 per million individuals per year, with a significantly higher incidence in non-Hispanic Whites (6.02 per million), when compared to Blacks and Asians: 0.31 and 0.39 per million, respectively.

- What is the probability that in NYC (population of 8.5 million reported in 2018), exactly 30 cases occur in a given year?
- Compute the same probability in a) by the mentioned racial/ethnic groups and comment on the findings. Demographic data of NYC in 2018: 14.0% Asians, 42.8% non-Hispanic Whites, 24.3% Black.

### Problem 2 (25p)

Remember the exercise study in HW1? Using the same dataset, you will now perform some tests to assess significance for changes in BMI for the two groups (within and between) (see Table 2 for reference). For now, the investigator is only interested in differences, with no implied directionality.

For each question, make sure to state the hypotheses, test-statistics, decision rules/p-values, and provide interpretations in the context of the problem.

Table 2: Pre/Post changes in metabolic parameters

	Intervention N = XX		Control N = XX	
	Baseline	6 month	Baseline	6 month
BMI	31.91 ± 6.58 29.25 (27.35-34.7)	31.21 ± 6.13 29.15 (26.8-33.05)	34.23 ± 6.16 33.4 (29.5-37.65)	34.51 ± 5.97 33.05 (30.35-37.8)
$\Delta$	-0.76 ± 1.44		0.28 ± 0.97	

- For each group, perform appropriate tests to compare the BMI absolute changes at 6 months follow-up vs baseline.
- Now perform a test to compare the BMI absolute changes between the two groups.
- What are the main underlying assumptions for the tests performed in parts a) and b)?
  - Use graphical displays to check the normality assumption and discuss the findings.
  - If normality is questionable, how does this affect the tests validity and what are some possible remedies?

### Problem 3 (10p)

According to recent studies, around 60% of the new restaurants opened in NYC fail/close within the first year and most of the time, location is to blame. Out of 20 restaurants that opened this year (random sampling), calculate the probability that 10 or more will close by the end of 2019. Use exact and approximation (if the case) methods and **comment on your findings**.

### Problem 4 (20p)

Use the built-in R data 'sleep' to assess if there is a statistically significant effect of a particular drug on sleep (Drug 2 compared to control - Drug 1) for 10 patients. The investigator is only interested in improving the efficacy of the drug, i.e., increase the hours of sleep.

For more information about the data, type '? sleep'.

- a) State the hypotheses, test-statistics, decision rule/p-value, and provide interpretation in the context of the problem.
- b) Compute a 95% confidence interval for the mean increase of hours of sleep.
- c) Using the **observed effect size and variance**, calculate the *a posteriori* power associated with the test performed in part a). Comment on your **findings**.
- d) State PROs and CONs of using *a posteriori/post-hoc power analysis* in research studies.