

# Final Paper

## *PUBH 512 Biostatistical Modeling & Sampling*

**Due: 23:59 pm, Monday, May 3, 2021**

Final paper is homework-like activity that will include statistical analysis, interpretation of your findings and write-up of the final paper. Final paper will be graded based on a clear explanation and concise presentation of statistical findings. A unique data set and questions related to your dataset (use your last four digits on your ID card to navigate your final paper assignment) will be available on the Moodle. Please submit your final paper on the Moodle or send directly to my email address (alpamys.issanov@nu.edu.kz).

Please read carefully the following instructions:

Final paper ought to contain two parts: *technical description* (describing your “kitchen”) and *write-up* part. Technical description is a part where you can describe what you have done to reach the final model in your statistical analysis. Make sure to clearly outline how you chose your final model, and a reason for including/excluding each variable. Write-up summary is a part where you should present your findings from statistical analysis as a manuscript, which is a “fit for publication”. In addition, your final paper will be assessed against your academic writing skills.

### **Technical Description (15 points)**

Here you have an opportunity to describe step-by-step your analysis to reach the final model. Firstly, describe whether you have found any peculiarities in variables, such as errors, problems or inaccuracies in univariate analysis. If there are any issues, how you addressed them. Secondly, describe in brief relationships (bivariable analysis) between the outcome variable and exposure variables, as well as, relationships among exposure variables. You should clearly explain why you have chosen a certain statistical test in doing bivariable analysis. Next, you should clearly outline how you chose your final model, and a reason for including/excluding each variable. Discuss whether you have checked for multicollinearity, confounding and interaction. Lastly, you should summarize whether assumptions for the selected model were checked and how well the model fit.

### **Write-Up Part (25 points).**

This part is familiar for you from Assignment #1 and #2. Few changes were introduced, such as, introduction and methods were expanded to 3-4 paragraphs. Added specifications for write-up in each part of the manuscript.

Based on the research question presented in your final assignment, and the analysis that was carried out, present your findings in a short paper written in a format for publication. The goal is to have you think about how this would be presented/organized for a manuscript as “fit

for publication”. Also, note that some of the relevant info that would go into a paper is not given in the description of the data. Please limit your write up to at most 10 pages, excluding any plots and summary tables presented. Below is a guideline for preparation.

**Introduction (6 points)** – In three paragraphs, please describe importance and background information related to this study objective. State the study objective.

- i. First paragraph should contain following info: Public health importance (for example, morbidity, mortality, socio-economic burden, health care burden) of the outcome. In general, how this health problem affects the general population. Is this question worth investigating?
- ii. Describe exposure variable (s), what is known about it/them. How the outcome and exposure variables are connected. Previous studies conducted on this research question.
- iii. What are the gaps, and what could be investigated? Why investigating these gaps is important? State your research aim (objectives) and state the importance of your study for public health.

**Data/Methods (6 points)** – 3-4 paragraphs describe:

- i. How many participants were in the study? How they were selected: inclusion/exclusion criteria. Describe in brief study population (where they are selected, how they were selected). Study design, sampling technique, and data collection.
- ii. Describe/define what is your outcome variable.
- iii. Describe what exposure variables were collected. How they were measured or categorized. If definitions for each category is needed, please do so.
- iv. Statistical analysis plan: all steps utilized to perform analysis, such as, type of numeric summaries in univariate analysis, types of bivariate analysis tests used and what was done in multivariate analysis. You should also mention whether you checked for assumptions, interaction, multicollinearity and goodness of fit of the final model.
- v. DAG. Draw Direct Acyclic graph to get better understanding of relationships between variables and selecting important variables in the model. Give a brief description why variables were chosen for the model building, from epidemiological, clinical or public health perspectives.

**Results (7 points)**– About 1 page summarizing the analyses and findings. Better to look at each of variables separately, then investigate associations of an outcome variable with response variables. So, you start with describing overall the study sample in terms of demographical characteristics, then move to clinical/epidemiological or other characteristics. In general, flow should be logical from describing important findings from tables starting from univariate to bivariable to multivariable analysis. You do not need to describe all study results here, only the most important ones.

Graphs or tables should be included. They help describe the results. Make sure that tables and figures have been appropriately titled and axes have labels; legends, notations or explanations/footnotes are provided for ease of interpreting graphical results.

Example Table for univariate and bivariable analysis (Linear):

Variable	Total, # (%) or mean±sd	Mean±sd for each group or correlation coefficient	p-value

Example Table for univariate and bivariable analysis (Logistic, Survival analysis):

Variable	Total, N=	Y=1 group, n=	Y=0 group, n=	p-value
Numeric variable	Mean±sd	Mean±sd	Mean±sd	
Categorical variable	n(%)	n(%)	n(%)	

Figure. Kaplan-Meier Survival Curves (Survival Analysis)

Example Table for multivariable analysis (Linear regression):

Variable	Unadjusted coefficient (95% CI)	p-value	Adjusted coefficient (95% CI)	p-value

Example Table for multivariable analysis (Logistic regression):

Variable	Unadjusted OR (95% CI)	p-value	Adjusted OR (95% CI)	p-value

Example Table for multivariable analysis (Cox regression):

Variable	Unadjusted HR (95% CI)	p-value	Adjusted HR (95% CI)	p-value

Example Table for effect modification/interaction (if you needed)

Variable	Unadjusted/Adjusted coefficient	p-value
Var1=0 & Var2=0		
Var1=1 & Var2=0		
Var1=0 & Var2=1		
Var1=1 & Var2=1		

**Discussion (6 points)** – 3-4 paragraphs interpreting the results and where they sit in the bigger picture. Make sure it was enough discussion of the study findings: previous studies, underlying explanation for the results, conclusion and recommendations based on the study findings.

- i. What you have found interesting – main finding(s).
- ii. Are your findings similar to previous studies. Why are they different? Do you have any explanation for your findings (biological, clinical social, public health)
- iii. Limitations/strengths of the study
- iv. Bigger picture – reflect what you found in a broader prospective.

Conclusions: What you can conclude? Recommendations. Public health importance.

### **Academic writing skills (no points, but deductions may apply).**

No points are given for your academic writing. However, the overall score for a poorly written final paper will be deduced based on the list of criteria indicated below. In assessing the academic writing, the instructor will examine your work based on relevance of the content, use of source materials, organization, cohesion and coherence, language accuracy, and presentation.

- *The relevance of the content* of your manuscript will be assessed against your writing ability to present your paper in well-focused manner, relevant to the main topic, comprehensive coverage of the topic, well supported arguments with scientific evidence.
- *Use of source materials* - references are provided. Citation is appropriate and consistent citation style is used throughout of the paper.
- *Organization* – sections and paragraphs are easy to navigate. The paper follows the logical flow from the introduction to conclusion(s).
- *Cohesion and coherence* – smooth flow of the text, sentences are logically connected. Writing is cohesive, avoiding redundancy and repetitions of ideas, sentences and results.
- *Language accuracy* – avoiding grammatical and stylistic errors, appropriate use of vocabulary and writing style. No incomplete, confusing and too long sentences.

- *Presentation* – clear presentation of both text and any tables and figures; avoiding a use of technical (coded) names for variables and categories in interpretations and discussion.

Here are some useful resources on how to write a research paper:

<https://pubmed.ncbi.nlm.nih.gov/15071152/>

<http://rc.rcjournal.com/content/49/10/1222/tab-pdf>

<https://adc.bmj.com/content/103/9/820.long>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3178846/>

You may look how authors presented their findings in publications (I did not critically appraise these two papers, but methodology and tables looked good to me. So, I decided to share with you as potential examples. Feel free to search for other papers which suit your preferences.):

[https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667\(21\)00001-3/fulltext](https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(21)00001-3/fulltext)

<https://www.frontiersin.org/articles/10.3389/fpubh.2021.649974/full>