

STATISTICS MASTERS/Ph.D.-QUALIFYING EXAM: TAKE HOME

August, 2024

General directions

Complete both problems in this exam. Your report is to be typed, double spaced, no smaller than ten-point font with one-inch margins, and should be identified by your “CODE WORD” in the header on each page; *do not include your name or UNM ID number*. Each problem is to be no longer than four pages, and an additional four-page appendix is allowed for each problem but will be examined only at the discretion of the graders; the better constructed your appendix with cross-references from the text, the more likely it is to get examined. In your data analysis, raw uninterpreted computer output will be graded as the dress it is.

Write your answers completely, but concisely. Insert tables and figures to support your points. Tables and figures should be well-labelled and cross-referenced from text, such as, “in Table 1 ...”, or if in the appendix, “in Table A1 ...” and each should have a caption that describes it and briefly tells the reader why it is of value. Figures should include appropriate symbols suitable for black-and-white reproduction (that is, avoid use of color if possible; consider symbols, line types, and distinct shades of gray to distinguish categories or values).

Organize your sections to justify the validity of what you uncovered and the methods you used to uncover it. We want a summary of what you think is important, not a diary of how you spent your time. Remember that even the best data analysis is worthless if your reader does not understand it, so you are being graded on presentation as well as statistical content.

As necessary:

1. Plot and describe the data (that is, plot all the individual observations, in addition to summaries of data you might present with the results, such as the mean and confidence intervals).
2. Clearly define population parameters and sample statistics.
3. Clearly specify hypotheses tested and explicitly state the associated model at least once (i.e., write the model equation).
4. Define and assess method assumptions.
5. Write a coherent evidence-based conclusion that a layperson can understand.

You may use your course notes as well as any available books or web resources on general statistical methods for the exam. You may not consult any other person when working on this exam or discuss your exam with anyone else, regardless of whether or not the person is taking the exam nor are you allowed to use the internet to find analyses of these data.

Any points of clarification can be directed to Prof. Erik Erhardt, erike@stat.unm.edu.

Email solutions as a pdf file by **3 PM, Fri Aug 16, 2024** to Ana Parra Lombard, mathstatgradprogram@unm.edu or alombard@unm.edu Department of Mathematics and Statistics, University of New Mexico. Please do not turn in a physical copy of your solutions.

Problem 1, Circuit board shorts

In inner-layer fabrication of circuit boards, copper-clad glass epoxy laminate panels are cleaned. Dry-film photoresist is applied to the panels under lamination temperature and pressure using a hot-roll laminator. The circuitry is plotted on film, placed on the panel, and exposed to ultraviolet radiation. The photoresist below the opaque area of the film is not affected, the rest is polymerized.

The experiment involves one response and three factors:

Shorts: a measure of short circuits in the board (fewer is better).

Prep: surface preparation of the panels which is Scrub, Pumice, or Chemical.

Heat: Preheating of the panels, Yes or No.

Pres: Lamination pressure, 20, 40, or 60 psi.

The current operating levels are Scrub, No, and 40.

Analyze the data assuming no three-factor interaction.

Should the process be changed to another set of operating levels?

Write a succinct, coherent, and complete summary of your analysis that addresses the original goal of the experiment and discuss anything else of interest.

Download the data from

https://math.unm.edu/sites/default/files/files/qual-exams/stat/unm_exam_202408_stat_qual-takehome_dat1.csv.

Problem 2, WHO Life Expectancy

The Global Health Observatory (GHO) data repository, managed by the World Health Organization (WHO), tracks health status and related factors for all countries. Publicly available datasets, including life expectancy and health factors for 193 countries, were sourced from the WHO and corresponding economic data from the United Nations. We selected critical health-related factors for analysis. Over the past 15 years, significant health sector developments have improved mortality rates, particularly in developing nations. Consequently, this project analyzes data from 2000-2015 for 193 countries.

1. **Country:** Country
2. **Year:** Year
3. **Status:** Developed or Developing status
4. **Life expectancy:** Life Expectancy in age
5. **Adult Mortality:** Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population)
6. **infant deaths:** Number of Infant Deaths per 1000 population
7. **Alcohol:** Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)
8. **percentage expenditure:** Expenditure on health as a percentage of Gross Domestic Product per capita (%)
9. **Hepatitis B:** Hepatitis B (HepB) immunization coverage among 1-year-olds (%)
10. **Measles:** Measles - number of reported cases per 1000 population
11. **BMI:** Average Body Mass Index of entire population
12. **under-five deaths:** Number of under-five deaths per 1000 population
13. **Polio:** Polio (Pol3) immunization coverage among 1-year-olds (%)
14. **Total expenditure:** General government expenditure on health as a percentage of total government expenditure (%)
15. **Diphtheria:** Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)
16. **HIV/AIDS:** Deaths per 1000 live births HIV/AIDS (0-4 years)
17. **GDP:** Gross Domestic Product per capita (in USD)
18. **Population:** Population of the country
19. **thinness 1-19 years:** Prevalence of thinness among children and adolescents for Age 10 to 19 (%)
20. **thinness 5-9 years:** Prevalence of thinness among children for Age 5 to 9 (%)
21. **Income composition of resources:** Human Development Index in terms of income composition of resources (index ranging from 0 to 1)
22. **Schooling:** Number of years of Schooling (years)

Exclude from analysis variables with more than 10% missing values.

Model the increase in a country's Life expectancy from year 2000 to 2015 as a function of the other covariates.

Download the data from

https://math.unm.edu/sites/default/files/files/qual-exams/stat/unm_exam_202408_stat_qual-takehome_dat2.csv.

If you're using R, you may use this code to get you started with a dataset, provided that you write a paragraph in the appendix explaining what each line of the code does.

```

dat_life <-
  readr::read_csv(
    file = "unm_exam_202408_stat_qual-takehome_dat2.csv"
  ) |>
  dplyr::mutate(
    Status = Status |> factor()
  ) |>
  dplyr::filter(
    Year %in% c(2000, 2015)
  ) |>
  dplyr::arrange(
    Year
  ) |>
  dplyr::group_by(
    Country
  ) |>
  dplyr::mutate(
    `Life expectancy 2000` = `Life expectancy` |> dplyr::first()
    , `Life expectancy 2015` = `Life expectancy` |> dplyr::last()
    , `Life expectancy Diff` = `Life expectancy 2015` - `Life expectancy 2000`
  ) |>
  dplyr::ungroup() |>
  dplyr::filter(
    Year %in% c(2000)
  ) |>
  dplyr::select(
    -Year
    , -`Life expectancy`
    , -`Life expectancy 2015`
  ) |>
  dplyr::relocate(
    `Life expectancy Diff`
  )

```