Tamar M. Shaw

Omega Graduate School

Richard Nti

Submission Date: July 1, 2023

100 Day - Essay

The following essay questions review the basic principles and concepts of inferential statistics.

The order of the questions approximates that of the five-step model.

1. Each response is to include a minimum of two descriptive paragraphs, examples, and

charts, tables, or graphics as appropriate.

 a. Hypothesis testing or significance testing can be conducted only with a

random sample. Why?

 b. Under what specific conditions can it be assumed that the sampling

distribution is normal in shape?

 c. Explain the role of the sampling distribution in the test of the hypothesis.

 d. The null hypothesis is an assumption about reality that makes it possible

to test sample outcomes for their significance. Explain.

 e. What is the relationship between or among significance level, p-value, and

confidence level?

2. Structure (Paper Evaluation includes the following structure below).

a. Download the “OGS APA Course Assignments Template 7th Ed 2021” template

from the General Helps folder in the AA-101 The Gathering Place Course on

DIAL. Using the template create the following pages.

b. Title Page (Not included in page count).

c. Copy and paste the assignment instructions from the syllabus starting on a new

page after the title page, adhering to APA 7th edition style (APA 7 Workshop,

Formatting, and Style Guide, APA 7 Quick Guide).

d. Start the assignment on a new page after the copied assignment instructions.

3. Be sure to meet the following expectations.

a. Document all sources in APA style, 7th edition (APA 7 Reference Example, APA

7 Quick Guide)

b. Include a separate Works Cited page, formatted according to APA style, 7th

edition

c. Use a minimum of fifteen scholarly research sources. Three to four books and

the remaining scholarly peer-reviewed journal articles, ideally from your

developmental reading.

4. Submit the completed paper to DIAL.

**Statistics Examples for Affordable Housing**

**a. Hypothesis testing or significance testing can be conducted only with a**

**random sample. Why?**

A hypothesis is a hunch, proposal, or explanation of an observation with little to no evidence that has to be tested. In order for the hypothesis to be genuinely tested, and to end up with a verified result without bias, the sample has to be random and not predetermined, preset or obvious. In significance testing there is a threshold of usually 𝛂 0.05. One will calculate if the probability value supports a null hypothesis which is usually assumed to be true or is the status quo or does the probability value support the alternative hypothesis.

If the probability value is (<) less than 0.05 then there is low chance that the null hypothesis is true, and we reject the (Ho) null hypothesis. If the probability value is (>) greater than 0.05 then we cannot reject the Ho (null hypothesis). Hypothetically if I owned a mixed income building with many residents at 50% of the (AMI) average median income of $95,000, I may want to foster greater community and social capital within the complex. If one of the common spaces was an art gallery with installations provided by the residents, I could use this space to foster greater community. In this hypothetical scenario, the average time spent inside the gallery is thirty minutes. If I created fun interactive activities within the art gallery, what would be the probability that the sample size of 50 guests visiting the gallery in a day would stay for more than sixty minutes?

Ho: *u* = 30 minutes, Ha: u > 30 minutes after interactive activities are created.

n = 50 residents, ‾x = 60 minute. If my p-value < α 0.05 then I can reject the null hypothesis that the change of installing interactive activities is only going to be 30 minutes. This is significant in that I know my change in the art gallery is making a difference and causing residents to intermingle with each other, thereby perhaps fostering social interaction and creating social capital.

**b. Under what specific conditions can it be assumed that the sampling**

**distribution is normal in shape?**

If the sampling is a fair or true representation of the general population or group and a high number of the sample represents the average, then the distribution is normal with little standard deviation. A normal distribution is therefore not skewed too far to the right or left. A low sample size can result in a skewed distribution. A large sample size tends to result in a normal distribution and small standard deviation.

If the values are displayed on a bell curve, and the values fall down the sides of the curve equally on each side then the distribution is normal. Each normal distribution has its own ‾X.

30 45 60 75 90 105 120

In the above chart, the (µ) score is 75 and the standard deviation (σ) is 15. If I wanted to know the self-esteem levels in a housing complex of both males and females, in this example, I find that the average self-esteem score is 75 and these only deviate, higher or lower by 15 for the females. In this example, there are two means to represent male and female.

**c. Explain the role of the sampling distribution in the test of the hypothesis.**

A hypothesis will state that there is an average number of a particular fixed variable in a given situation. The average number or mean is in the center of the data distribution. This distribution is usually displayed or expressed in a bell curve with the average number in the middle. If the top of the bell curve or if most of the participants in the sampling are not toward the middle but skewed to the left or to the right, then the sample is skewed or there is deviation from the standard average.

Out of a low-income housing complex of 300 residents, it may be hard to test a hypothesis with each and every resident. It is best in statistical calculations to get a large enough sample population out of the 300 residents. A good number may be more than half such as one hundred and eighty or at least one hundred and twenty-five. Too small of a sample of the population within the housing complex would probably be thirty (30). The sampling distribution in this case must truly represent the complex as a whole as far as age, average number of children in household, ethnicity and income, etc..

The below example (Pilkauskas, N., & Michelmore, K., 2019, p. 1310) is from an article discussing how the EITC or Earned Income Tax Credit benefit for low-income families has an effect on homelessness, evictions and foreclosures; (2) reduces housing costs burdens and (3) reduces the need for doubling up or shared living arrangements.

|  |  |  |
| --- | --- | --- |
| Table 1 Sample descriptive statistics: CPS, ACS, and FFCWS |  |  |
|  | CPS | ACS | FFCWS |
| Maternal CharacteristicsAge | 33.62 | 33.78 | 29.65 |
|  | (6.82) | (6.83) | (7.34) |
| Number of children | 1.78 | 1.79 | 2.02 |
|  | (0.78) | (0.78) | (0.82) |
| Race/Ethnicity Non-Hispanic white | 0.51 | 0.51 | 0.14 |
| Non-Hispanic black | 0.23 | 0.27 | 0.57 |
| Hispanic | 0.20 | 0.19 | 0.26 |
| Other race/ethnicity | 0.05 | 0.03 | 0.03 |
| EducationLess than high school | 0.20 | 0.16 | 0.44 |
| High school diploma | 0.40 | 0.48 | 0.30 |
| Some college | 0.41 | 0.36 | 0.26 |
| Earned Income Tax Credit (EITC)Average simulated EITC (in $1,000s, 2011 dollars) | 1.54 | 1.62 | 1.87 |
|  | (0.59) | (0.60) | (0.51) |
| Eligible for the EITC | 0.58 | 0.65 | 0.69 |
| EITC credit (in $1,000s, 2011 dollars) | 1.44 | 1.44 | 1.67 |
|  | (1.67) | (1.66) | (1.75) |
| Number of Observations | 85,089 | 757,977 | 9,946 |

Notes: Standard deviations are shown in parentheses. The sample is restricted to single mothers with less than a college degree and with at least one coresident child under the age of 19. The CPS and ACS are also restricted to mothers aged 19–45.

Sources: Current Population Survey (CPS) 1990–2016. Census 1990/American Community Survey (ACS) 2000–2016. Fragile Families and Child Wellbeing Study (FFCWS) 1998–2016.

We can see in the above chart that the average age of the mother in the Current Population Survey (CPS) is 33.62, almost 34 years old. The average age of the mother in the American Community Survey (ACS) is 33.78, even closer 34 years old. In these two surveys, the standard deviation was about the same (6.82) and (6.83) respectively. This means that some of the mothers were almost 7 years older. For the Fragile Families survey (FFCWS), the average age found was lower at 29.65, almost 30 years old. In the (FFCWS) survey, the standard deviation was higher at 7.34 years. For the number of children, the women had, the average was below two children for the CPS and ACS surveys and at least 2 children for the FFCWS survey. The standard deviation was .78, .78 and .82 respectively. Standard deviations are also listed for the amount of earned tax credit received.

**d. The null hypothesis is an assumption about reality that makes it possible**

**to test sample outcomes for their significance. Explain.**

The null hypothesis is considered the status quo. The below table (Kramer, K. Z., et al., (2019) is from an article that discusses the effects of providing the (EITC) Earned Income Tax Credit in advance of four payments. The table describes the results of testing three out of five hypothesis. The base reality before testing any of the hypothesis was that distributing the EITC as the normal lump sum payment has null effect on the recipient family. Model 1 tests hypothesis 3 stating that residents who receive advance periodic EITC payments will experience lower levels of perceived financial stress than those who receive the lump sum EITC payment. Model 2 tests hypothesis 4 stating that financial stressors such as unpaid bills will partially mediate the relationship between the EITC distribution method (advance/periodic or lump sum) and perceived financial stress over time. Model 3 tests hypothesis 5 which states financial resources (savings) will partially mediate the relationship between the EITC distribution method (advance/periodic or lump sum) and perceived financial stress over time. Hypotheses 3 and 4 were supported but hypothesis 5 was not.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variables | Model 1b | se | Model 2b | se | Model 3 |  |
| b | se |
| Time 2 | 0.15\* | 0.07 | 0.01 | 0.07 | 0.02 | 0.07 |
| Time 3 | 0.15\* | 0.08 | 0.02 | 0.08 | 0.03 | 0.08 |
| Time 4 | − 0.08 | 0.08 | − 0.15† | 0.08 | − 0.14† | 0.08 |
| Female | − 0.13 | 0.12 | − 0.09 | 0.11 | − 0.07 | 0.11 |
| In relationship | 0.08 | 0.07 | 0.02 | 0.06 | 0.02 | 0.06 |
| Age | − 0.00 | 0.00 | − 0.00 | 0.00 | − 0.00 | 0.00 |
| Less than HS | 0.17\* | 0.08 | 0.16\* | 0.07 | 0.14\* | 0.07 |
| Periodic EITC payment | 0.20\*\* | 0.08 | 0.10 | 0.07 | 0.11 | 0.07 |
| Periodic\*Time 2 | − 0.31\*\* | 0.08 | − 0.20\* | 0.09 | − 0.21\* | 0.09 |
| Periodic\*Time 3 | − 0.43\*\* | 0.09 | − 0.24\* | 0.09 | − 0.26\*\* | 0.09 |
| Periodic\*Time 4 | − 0.39\*\* | 0.09 | − 0.20\* | 0.10 | − 0.21\* | 0.10 |
| Borrowed money | – | – | 0.27\*\* | 0.04 | 0.26\*\* | 0.04 |
| Food insecurity | – | – | 0.19\*\* | 0.04 | 0.18\*\* | 0.04 |
| Unpaid bills | – | – | 0.32\*\* | 0.04 | 0.30\*\* | 0.04 |
| Savings | – | – | – | – | − 0.0004\*\* | 0.0001 |
| Disposable income | – | – | – | – | − 0.00007\* | 0.00004 |
| Constant | 3.67\*\* | 0.20 | 3.39\*\* | 0.17 | 3.47\*\* | 0.17 |
| N (observations) | 1685 (498) |  | 1545 (492) |  | 1545 (492) |  |

As you can see in the chart for Hypothesis 3 for the variable Periodic\* Time 4 there is an inverse (less than 0) correlation because the result is -0.30. No strong correlations are seen and many weak correlations such as in Hypothesis 4 (model 2) for the Periodic EITC payment at 0.10. This chart represents the strength of each Hypothesis tested.

**e. What is the relationship between or among significance level, p-value, and**

**confidence level?**

The significance level is usually 0.05. The probability value or p-value is measured against the significance level to assure a greater confidence level. If it is less than the standard 0.05 then one can reject their null hypothesis. You must calculate the p-value in order to reject the null hypothesis or assume that it is true. The lower the p-value, the higher the statistical significance.

(Rumsey, 2019) states when estimating using a confidence interval, ultimately you want to have a small margin of error. More precise results will have a narrower interval.

When figuring out the mean or average of a population (when what is measured is numerical), the formula to use for the confidence interval is as follows: 1.96 presents *z* \*

 (Rumsey, 2019, p. 76). ‾x is the sample mean. The z\* is the appropriate value from the Z distribution. The population standard deviation is divided by the square root of the sample size.

By taking a random sample of 175 tenants in a low-income building of 500 , I may want to estimate the average income of each household within the complex. Because of the likelihood of variation, it is best to calculate a margin of error in order to capture a true representation of the whole population . This is done by adding plus or minus to the result of my calculation. If I come up with an average income of $39,000, the resulting (CI) confidence interval could be $39,000 + or – 3.5k which would be a range of 35,500 to 42,500. Everyone in the building would most likely fall into this income range.

(Rumsey, 2019) gives the following more commonly used confidence levels:

|  |  |
| --- | --- |
| Percentage Confidence | *z*\*- value |
| 80 | 1.28 |
| 90 | 1.64 |
| 95 | 1.96 |
| 98 | 2.33 |
| 99 | 2.58 |

“The confidence level of a confidence interval corresponds to the percentage of the time your result would be correct if you took numerous random samples” (Rumsey, 2019, p. 72).

As stated earlier, the larger the sample provides the best conditions for the most accurate results. A large sample allows for a small margin of error.

Many times, what we want to measure is not numerical but categorical, such as how often residents in a multi-family go to church during the month. In this case we may want to calculate the percentage of people going to church in one month. The following formula is appropriate for the confidence interval of a population proportion:  p +/- z\*√p (1-p) / n (Rumsey, 2019, p. 77).

|  |
| --- |
| TABLE 2—Associations Between All-Cause and Cause-Specific Mortality and HousingTypology: EPIPorto Cohort Study, Porto, Portugal, 1999–2019 |
|  Model 1, RR (95% CI) Model 2, RR (95% CI) | Model 3, RR (95% CI) |
| All-cause mortalityConventional housing (Ref) | 1 | 1 | 1 |
| Affordable housing | 1.43 (0.91, 2.14) | 1.37 (0.86, 2.08) | 1.38 (0.83, 2.16) |
| Social housing | 1.68 (1.30, 2.15) | 1.59 (1.22, 2.06) | 1.52 (1.14, 1.99) |
| Ilhas | 1.74 (1.19, 2.45) | 1.64 (1.12, 2.33) | 1.45 (0.96, 2.12) |
| Cardiovascular mortality Conventional housing (Ref) | 1 | 1 | 1 |
| Affordable housing | 1.14 (0.40, 2.55) | 0.94 (0.29, 2.28) | 0.83 (0.20, 2.29) |
| Social housing | 1.64 (0.98, 2.64) | 1.64 (0.96, 2.70) | 1.29 (0.69, 2.28) |
| Ilhas | 2.46 (1.30, 4.30) | 2.48 (1.29, 4.43) | 2.36 (1.18, 4.35) |
| Cancer mortalityConventional housing (Ref) | 1 | 1 | 1 |
| Affordable housing | 1.64 (0.63, 3.50) | 1.63 (0.63, 3.49) | 1.87 (0.72, 4.03) |
| Social housing | 1.55 (0.90, 2.53) | 1.51 (0.86, 2.54) | 1.53 (0.86, 2.62) |
| Ilhas | 1.95 (0.90, 3.73) | 1.86 (0.85, 3.62) | 1.63 (0.71, 3.30) |

Note. CI = confidence interval; RR = rate ratio. Ilha is the local word for substandard housing. The sample size was n = 2485. Model 1 = adjusted for sex and age. Model 2 = additionally adjusted for residence in Porto for 20 years or more, education, occupation, and marital status. Model 3 = additionally adjusted for 25 · 25 risk factors (current smoking, harmful use of alcohol, insufficient physical activity, obesity, hypertension, and diabetes).

The above chart (Ribeiro & Barros, 2020) displays the correlation between types of housing and mortality. Each model represents different variables. Model 1 is adjusted for sex and age, Model 2 is adjusted for demographics such as age, and education and Model 3 is adjusted for certain risk factors such as obesity. Each model has a confidence level of 95% and for each model in each type of housing, there is a margin of error.

Although the significance threshold is usually 0.005, there are some highly significant results at p < 0.001. Pomery, et al (2021, p. 1423) gives us the following chart as an example:

|  |  |
| --- | --- |
| **Table 2.****CHRONIC CONDITIONS AND DISABILITIES** |  |
| **Condition**  | **Fairfax Housing Health Survey****All HCV RAD**  | **Behavioral Risk Factor Surveillance System** |
|  **All Virginia** |
| High blood pressure | 52 (57.14) | 32 (58.18) | 20 (55.56) | 181,098 (41.04) | 3,895 (41.34) |
| Heart attack | 3 (3.26) | 3 (5.56) | 0 (.00) | 26,004 (5.81) | 546 (5.70) |
| Angina | 7 (7.53) | 5 (9.09) | 2 (5.26) | 25,389 (5.69) | 509 (5.33) |
| Stroke\*\* | 9 (9.89) | 9 (16.98) | 0 (.00) | 18,956 (4.22) | 392 (4.08) |
| Asthma | 22 (23.66) | 14 (25.45) | 8 (21.05) | 62,411 (13.91) | 1,301 (13.56) |
| Cancer (other than skin cancer) | 9 (9.78) | 7 (12.96) | 2 (5.26) | 44,289 (9.87) | 994 (10.36) |
| Depression | 42 (46.15) | 27 (50.94) | 15 (39.47) | 89,209 (19.92) | 1,877 (19.59) |
| Diabetes | 16 (17.78) | 11 (20.75) | 5 (13.51) | 60,440 (13.81) | 1,354 (14.47) |
| Overweight/ Obesity\* | 49 (52.69) | 35 (62.50) | 14 (37.84) | –  | –  |
| Requires special equipment | 22 (24.18) | 13 (24.53) | 9 (23.68) | –  | –  |

*Note:*

FHHS: Fairfax Housing Health Survey

BRFSS: Behavioral Risk Factor Surveillance System.

Values indicate frequency followed by (percent).

Missing, “refused,” and “don’t know” responses are not included in the denominator for percentage calculations.

N=106 for all FHHS participants, N=65 for HCV program participants, and N=41 for RAD residents.

N=450,016 for all BRFSS participants and N=9,630 for Virginia.

– No comparison data collected for the Behavioral Risk Factor Surveillance System.

\* Difference between HCV and RAD significant at p < .05.

\*\* Difference between HCV and RAD significant at p < .01.

,

Here we see that in the category of Obesity, 63% of voucher residents self-reported and 38% of RAD residents self-reported. For this result the proportion value is less than, p < .05 the standard significance value and we can reject the null hypothesis. For voucher tenants, 17% reported a stroke, with an even greater significance of p < .01.

Harford, T. (2021). “Social scientists have long understood that statistical metrics are at their most pernicious when they are being used to control the world, rather than try to understand it” (p. 59).

**WORKS CITED**

Brucker, D. L., & Helms, V. (2019). Health, health behaviors, and health care utilization among adults with serious psychological distress who receive federal housing assistance.*The Journal of Behavioral Health Services & Research, 46*(4), 586-606. https://doi.org/10.1007/s11414-018-09647-z

Kramer, K. Z., Flávia Cristina, D. A., Greenlee, A. J., Mendenhall, R., Bellisle, D., & Renee, L. B. (2019). Periodic earned income tax credit (EITC) payment, financial stress and wellbeing: A longitudinal study.*Journal of Family and Economic Issues, 40*(3), 511-523. https://doi.org/10.1007/s10834-019-09618-2

Khan Academy. (2018) Normal conditions for sampling distributions of sample proportions| AP Statistics [Video].

 [Normal conditions for sampling distributions of sample proportions | AP Statistics | Khan Academy - YouTube](https://www.youtube.com/watch?v=L7AX2RcbqCg)

Khan Academy. (2018). *P-values and Significance tests*| AP Stasticstics[Video].

 [P-values and significance tests | AP Statistics | Khan Academy - YouTube](https://www.youtube.com/watch?v=KS6KEWaoOOE)

Pilkauskas, N., & Michelmore, K. (2019). The effect of the earned income tax credit on housing and living arrangements.*Demography, 56*(4), 1303-1326. <https://doi.org/10.1007/s13524-019-00791-5>

Pomeroy, J. M., Johnson, E., & Weinstein, A. A. (2021). Subsidized housing and health: An exploratory study examining resident perspectives on community health and access to care.*Journal of Health Care for the Poor and Underserved, 32*(3), 1415-1432. https://www.proquest.com/scholarly-journals/subsidized-housing-health-exploratory-study/docview/2566490496/se-2

Ribeiro, Ana Isabel, PhD., M.P.H., & Barros, H.,M.D.PhD. (2020). Affordable, social, and substandard housing and mortality: The EPIPorto cohort study, 1999–2019.*American Journal of Public Health, 110*(7), 1060-1067. https://doi.org/10.2105/AJPH.2020.305661

Rumsey, J. D., Ph.D., (2019). *Statistics essentials for dummies*. John Wiley & Sons, Inc.