Homework 4

PPHA 31002 | Statistics for Data Analysis I

Jay Ballesteros

December 05, 2025

Pen & paper exercises

Is Hand Sanitizer Effective? Evidence from One Study

(1) Calculate the proportion of the treatment group who had an infection.

$$p_{ti} = \frac{49}{116} = 0.422$$

(2) Calculate the proportion of the control group who had an infection.

$$p_{ci} = \frac{49}{96} = 0.510$$

(3) Calculate the difference between the proportion of people who had an infection in the treatment group and the control group.

$$p_{ti} - p_{ci} = 0.42 - 0.51 = -0.088$$

(4) Conduct a formal hypothesis test examining whether the treatment group had a lower infection rate than the control group. Be sure to report all of the steps of the test.

First of all, we state the hypothesis in statistical terms (Null and alternative hypothesis). Thus,

$$H_0 : p_{ti} = p_{ci}$$

$$H_A : p_{ti} < p_{ci}$$

Next, the significance level. For simplicity $\alpha = 0.05$.

Considering that we are trying to compare the difference between two proportions, we are going to use the two-proportion z-test. To get this started, we need to get several values from the following equation:

$$Z = \frac{\hat{p_A} - \hat{p_B}}{\sqrt{p^*(1 - p^*)(\frac{1}{n_A} + \frac{1}{n_B})}}$$

Where the numerator corresponds to the difference between my two sample proportions $(p_{ti} - p_{ci})$ and the denominator is the estimated standard error.

Translated into our data it might look like this:

$$Z = \frac{\frac{49}{116} - \frac{49}{96}}{\sqrt{\frac{49+49}{212}(1 - \frac{49+49}{212})(\frac{1}{116} + \frac{1}{96})}}$$

So the result of the test its:

$$Z = \frac{-0.088}{\sqrt{0.462(1 - 0.462)(0.019)}} \approx -1.280542$$

Next we have to convert the test statistic into a p-value, which results in 0.2003546. To get that result I executed the following code in R:

```
2*pnorm(-1.280542)
```

Given that $p - value = 0.20$ is greater than $\alpha = 0.05$, we fail to reject H_0.

(5) Suppose you are a public health official in the middle of a pandemic. You have a sworn duty to only recommend evidence-backed treatments to reduce the spread of viruses. A reporter asks you if the evidence supports alcohol-based hand sanitizers as being effective at preventing disease. What do you say?

Given the sizes of the samples and the information that we have right now, we don't have enough evidence to say that alcohol-based hand sanitizers are effective at preventing disease.

(6) Suppose you learn that there was noncompliance in the study. Of the 116 individuals assigned to the treatment group, 93 used hand sanitizer. Of the 96 individuals assigned to the control group, 20 used hand sanitizer. Estimate the intent-to-treat effect as well as the complier average treatment effect based on this information. If it is not possible to estimate a particular quantity, briefly explain why that is so.

Following the ITT equation:
$$ITT = \bar{Y}_{Z_i=1} - \bar{Y}_{Z_i=0}$$

We simply:
$$ITT = \frac{49}{116} - \frac{49}{96} = 0.42 - 0.51 = -0.088$$

So now, for CATE, we have to:

$$CATE = \frac{\bar{Y_{Z=1}} - \bar{Y_{Z=1}}}{\bar{T_{Z=1}} - \bar{T_{Z=0}}}$$

Based in our data,

$$CATE = \frac{-0.088}{\frac{93}{116} - \frac{20}{96}} = \frac{-0.088}{0.59} = -0.15$$

Does Tylenol Cause Autism?

(7) Calculate the proportion of children diagnosed with autism who were exposed to acetaminophen in utero and the proportion of children diagnosed with autism not exposed to acetaminophen in utero. Based only on those two proportions (i.e., only comparing those proportions without further analysis), can we reach a conclusion as to whether there is a true difference in the incidence of autism diagnoses between children exposed and not exposed in utero to acetaminophen? You can assume that these data represent a sample from a broader population. Briefly explain in 1-2 sentences.

Given the data in table 1, the proportions are:

$$P_{ewa} = \frac{5,912}{185,909} = 0.031$$

$$P_{newa} = \frac{62,672}{2,294,888} = 0.027$$

Just to go further, the differences are:

$$P_{ewa} - P_{newa} = 0.031 - 0.027 = 0.004$$

3

Following these results only, we can't directly assume if being expose to Acetaminophen influences the autism diagnosis. Nonetheless, the small difference of 0.004 between the two proportions can provide us guidance about our hypothesis, where we can state for the null hypothesis (H_0) that there's not difference whatsoever.

(8) Formally conduct a two-sided hypothesis test that there is no difference in the proportion of children diagnosed with autism between the group exposed and not exposed to acetaminophen. Be sure to report all of the steps of the hypothesis test.

First, the null and alternative hypothesis:

$$H_0 : P_{ewa} = P_{newa}$$

$$H_A : P_{ewa} \neq P_{newa}$$

Significance level at: $\alpha = 0.05$

Considering the data and the equation for a two-sided z-test:

$$Z = \frac{\hat{p_A} - \hat{p_B}}{\sqrt{p^*(1-p^*)(\frac{1}{n_A} + \frac{1}{n_B})}}$$

Now with our data:

$$Z = \frac{\frac{5,912}{185,909} - \frac{62,672}{2,294,888}}{\sqrt{\frac{5,912+62,672}{2,480,797}(1 - \frac{5,912+62,672}{2,480,797})(\frac{1}{185,909} + \frac{1}{2,294,888})}}$$

$$Z = \frac{0.00449}{0.00039536} = 11.35674$$

Now, to get a $p-value$, given that we got a positive test statistic result, we run in R `2 * (1-pnorm(11.35674))`. Our resulting $p-value = 0$

Considering that $\alpha = 0.05$ and $p-value = 0$, we reject the H_0 of $P_{ewa} = P_{newa}$

(9) Please elaborate on your finding using the concepts of practical significance (sometimes, also referred to as economic significance) and statistical significance.

Given that, based in our results, $p-value < \alpha$, we have statistical significance that the proportion of children diagnosed with autism differs between those exposed and non-exposed to Tylenol. Nonetheless, the practical significance of the difference found is relatively low (difference of about 0.00449), so whether this is relevant in practice depends on how weighted is this small difference against the benefits and costs on Tylenol use.

(10) We have thus far examined the association between acetaminophen exposure in utero and autism. As we have emphasized throughout the course, the presence of a correlation does not imply a causal relationship. This STAT News article summarizes the evidence on the possibility of a causal link between acetaminophen and autism. After reading the article, describe one possible source of selection into acetaminophen usage by pregnant women, which would make the comparison between the acetaminophen-exposed group and the non-exposed group invalid.

Based on the article, one possible source of selection into Tylenol use is that women with higher genetic risk for ADHD and autism, are more likely to experience more pregnancy pain and migraines. Thus they are more prone to take pain medications like Tylenol. Given this, the possibility of exposed mothers with higher genetic risk having a higher baseline that the non-exposed might already a potential reason to make the comparison invalid.

(11) Construct a 95% confidence interval (you can use 1.96 as the critical value given the enormous sample size, as the appropriate critical value from t distribution would closely approximate 1.96) for $\hat{\beta}$ reported above, and determine based on the confidence interval whether you would reject the null hypothesis that $\beta = 0$ for $\alpha = 0.05$.

Given the information, for the margin of error,

$$1.96 \times 0.00054 = 0.0010584$$

And for the confidence interval,

$$Lower = 0.0009 - 0.0010584 = -0.00016$$

$$Upper = 0.0009 + 0.0010584 = 0.00196$$

Therefore, the 95% confidence interval $(-0.00016, 0.00196)$ includes 0, so we do not reject the null hypothesis of $\beta = 0$ at $\alpha = 0.05$.

Considering this results, the estimated effect of Tylenol exposure is not statistically significantly different from 0 at the 5% level, because the 95% confidence interval for β includes 0.

(12) Interpret the regression coefficient in words. Is the coefficient statistically significant from zero? What is your conclusion.

Assuming a 95% confidence interval:

$$0.002 \pm 1.96(0.0008) = (-0.0014, 0.0018)$$

The coefficient is not statistically significantly different from 0.

For instance, the regression of 0.0002 estimate is not statistically different from 0. Thus, within families, the data do not provide evidence that Tylenol exposure in utero affects autism risk.

Has Banning Neonicotinoids in the European Union Lowered Agricultural Yields?

(13) State the null and alternative hypotheses for a test that seeks to evaluate whether the prediction for the crop index changes after the 2018 EU neonicotinoids ban.

$$H_0 : \hat{\beta} = 0$$

$$H_A : \hat{\beta} \neq 0$$

(14) Using the critical value approach, can you reject the null hypothesis at the 5% significance level? Throughout the question, use $n - 2$ as the number of degrees of freedom.

Considering the degrees of freedom of $n - 2$,

$$756 - 2 = 754$$

And our

$$\alpha = 0.05$$

then our critical value is 1.96

So given that out t in absolute value is 3.24 and greather than 1.96, we reject H_0 at the 5% level. So there is statistically significant evidence that the predicted change of $\hat{\beta}$ is different than 0.

(15) Calculate the p-value of observing the t-statistic of -3.24. Based on this p-value alone (as in do not use the critical threshold), can you reject the null hypothesis at the 5% significance level?

Running the code in R

```
2 * pnorm(-3.24)
```

we get that $p - value = 0.0012$.

And since the $p - value < 5\%$ significance level, we reject the H_0.

(16) Construct a 95% confidence interval for the coefficient. Using only this 95% confidence interval alone, determine whether you (fail to) reject the null hypothesis at the 5% significance level?

Give the data, for our margin error,
$$1.96 \times 0.028 = 0.0549$$

$$Lower = -0.091 - 0.0549 = -0.146$$

$$Lower = -0.091 + 0.0549 = -0.036$$

So the 95% confidence interval is:
$$(-0.146, -0.036)$$

Thus considering that the 0 from the H_0 is not inside the 95% confidence interval, we reject H_0.

How Do Minimum Wage Increases Affect Employment?

Data exercise

(19) Read the first three pages of the paper and provide a high-level summary in 5-7 sentences. Your summary should address the following points: what a "Ban the Box" (BTB) policy does, one reason why the policy could succeed in achieving its goal, and one reason why it might fail. You do not need to include all of the details; focus on the main ideas

BTB policy aims to restrict employers for asking to applicants about its criminal history in order to reduce discrimination and unemployment among black men –because this population group disproportionately have criminal records.

One of the arguments about the potential success of its goal is that BTB might reduced bias of employers by reducing the weight of criminal history of applicants down to 0, so applicants have a better chance to be assessed based in their skills, experience, among other aptitudes. Nonetheless, the drawbacks might be that employer could move their assessments using race of the applicants as a proxy, enhancing discrimination even further.

(20) Each row is a different fictitious resume. How many resumes are there in the sample? What share of these resumes received a positive response (i.e. a value =1 in the response variable)?

In this database there are 3,712 resumes, from which 11.8% received a positive response.

(21) We are going to start our analysis with two different two-sample t-tests. Each test assesses whether there is a difference in the response outcome between fictitious Black and white applicants. One test compares the difference in response for resumes sent before the BTB policy, and the other test uses the resumes sent after the BTB policy. For each test, conduct a two-sided t-test comparing using the response variable as the outcome.

Can you reject the null that the proportion of responses is the same across the Black and white resumes before the BTB policy? What about after the BTB policy? Write down the null and alternative hypotheses (you only need to do this once). Then for each test, report the p-value, the decision to reject the null or not, and your interpretation of the result of the test.

First, the hypothesis:

$$H_0 : \mu_{black} = \mu_{white}$$

$$H_A : \mu_{black} \neq \mu_{white}$$

```
# filter groups per period
before <- df %>% filter(post == 0)
after  <- df %>% filter(post == 1)

#t test for pre policy
t_pre <- t.test(response ~ black,
                data = before,
                alternative = "two.sided")
t_pre
```

```
##
##  Welch Two Sample t-test
##
## data:  response by black
## t = 0.51812, df = 1852.3, p-value = 0.6044
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.02100966  0.03609587
## sample estimates:
## mean in group 0 mean in group 1
##       0.1142241       0.1066810
```

```
#t test for pre policy
t_post <- t.test(response ~ black,
                 data = after,
                 alternative = "two.sided")
t_post
```

```
##
##  Welch Two Sample t-test
##
## data:  response by black
## t = 2.8772, df = 1812.5, p-value = 0.004059
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  0.01406449 0.07429758
## sample estimates:
## mean in group 0 mean in group 1
##       0.1476293       0.1034483
```

Based in the t-test results and $\alpha = 0.05$, we fail to reject the H_0 in the pre-policy period ($p-value = 0.6044$), but we do reject it in the post-policy period ($p-value = 0.004059$).

These results mean that in the pre-policy period, the difference in callback rates between black and white applicants is not statistically significant. In contrast, in the post-policy period, the difference is statistically significant with more than 4 pp difference in callback rates, favoring non-black applicants.

(22) We will make a small adjustment to the exact regression specification we ask you to run, but it is capturing the most important features of the main analysis presented in the paper. First, you will need to create a new variable which is the interaction term of post and black. You simply need to create a variable that multiplies post by black. Then run a linear regression model of the following form

$$response_i = \alpha + \beta_1 Black_i + \beta_2 post_i + \beta_3 post_i \times Black_i + \epsilon_i$$

Report the estimated coefficients from the regression. Note that you should obtain results that are close, but not identical to those in Table 4, column 2 above (the main difference is with the estimated standard errors, but you can ignore those differences). You should arrive at a nearly identical value to the one reported in the first row of estimated coefficients reported in the second column.

```
regression <- lm(response ~ black + post + post_black, data = df_regression)

summary(regression)
```

```
##
## Call:
## lm(formula = response ~ black + post + post_black, data = df_regression)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.1476 -0.1142 -0.1067 -0.1035  0.8965
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.114224   0.010580  10.796   <2e-16 ***
## black       -0.007543   0.014962  -0.504   0.6142
## post         0.033405   0.014962   2.233   0.0256 *
## post_black  -0.036638   0.021160  -1.731   0.0835 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3223 on 3708 degrees of freedom
## Multiple R-squared:  0.00296,    Adjusted R-squared:  0.002153
## F-statistic: 3.669 on 3 and 3708 DF,  p-value: 0.01177
```

(23) This study used data from an experiment that sent out fictitious resumes to employers in New Jersey and New York City. Do you think we can easily generalize the results from this study when discussing how a similar policy would perform in San Francisco or Washington DC? Discuss briefly in 1-2 sentences.

Not really. We can't generalize the results from the study because of different populations and unknown means and baselines.

Simulation exercise

Sample size and rejecting the null hypothesis

(25) Run the code provided in simulation treatment effects.R. We will first use the baseline values of the true effect and the sample size, as provided in the code (0.5 and 50, respectively). Based on the distribution of p-values (use the hist() function at the end of the code), if you were to collect just one sample, should you feel confident that you will reject the null hypothesis of no difference in the population means at a significance level of 5%? (Including the graph of the distribution of p-values is optional)

Given the sample size of 50 and looking at its distribution after the simulation, the histogram looks almost uniform. And taking into the consideration the frequency of simulations with p-values under 0.05 (mean around 5% of observations) we can't use this sample to confidently reject the null hypothesis of no difference in the population of means.

(26) Increase the sample size to 500, and then to 5,000, and re-run the simulation for each sample size. How does your answer to the previous question change for each new sample size value?

Radically. For instance, when running the simulation at 500 about 8.3% of simulation are under 0.05. But when increasing to 5000, this proportion increases to 42%. So the more I increase the sample size, more confident I am to potentially use a sample to reject the H_0.

(27) Reset the sample size value to 50. Increase the size of the true effect from 0.5 to 1 and then to 5, and re-run the simulation for each value. How does your answer to the original question change?

Very similar case to the previous ones. By each increase in the `true.effect` (while holding sample size at 50) the proportion of p-values under 0.05 increase. So by each increase in `true.effect`the more confident I am to use that sample size to reject the H_0.