HSCI 50 LAB 6: Bivariate Analysis for Continuous Variables

INSTRUCTIONS

The data set that we will be using for the entire course is resampled data from the MIT COVID-19 Beliefs, Behaviors & Norms Survey (https://covidsurvey.mit.edu/api.html). This is a multi-country, online survey that examined different COVID-19 perceptions across time, from July 6, 2020 to March 28, 2021. We will be using data from the Philippines aged 20 to 60. **Assume the data are randomly sampled.**

In this lab, you will learn how to conduct and interpret statistical outputs from an independent samples *t*-test, a paired samples *t*-test, and a one-way analysis of variance (ANOVA).

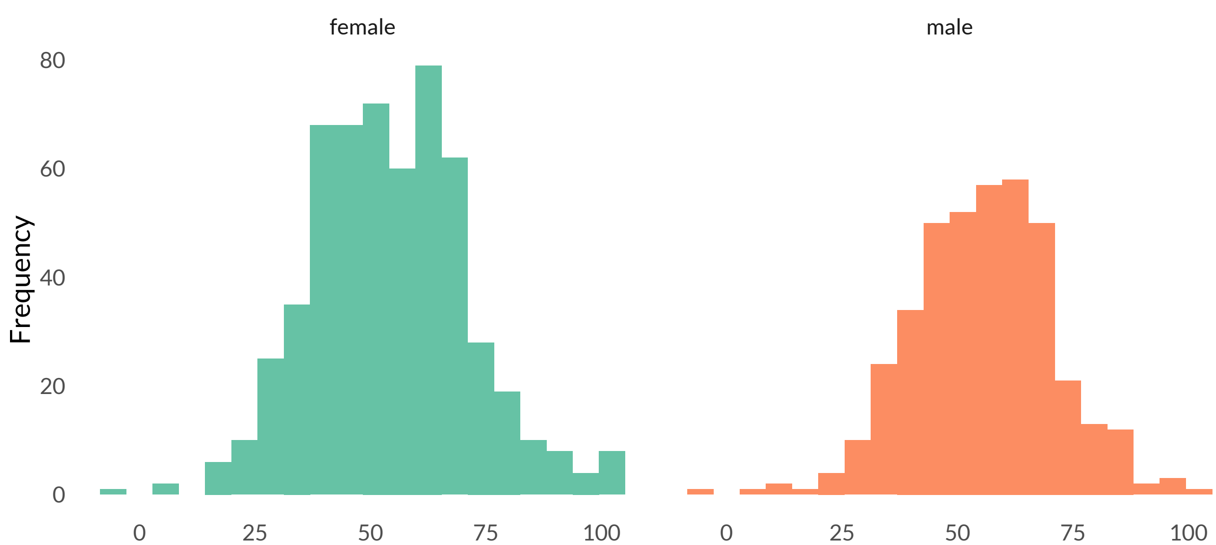
You have many options to submit this worksheet. Either you work on this by hand and scan/take a clear photo of your submission and save as PDF, or type your responses in a Word processor or PowerPoint presentation. You do not have to copy the questions again, but please number them accordingly.

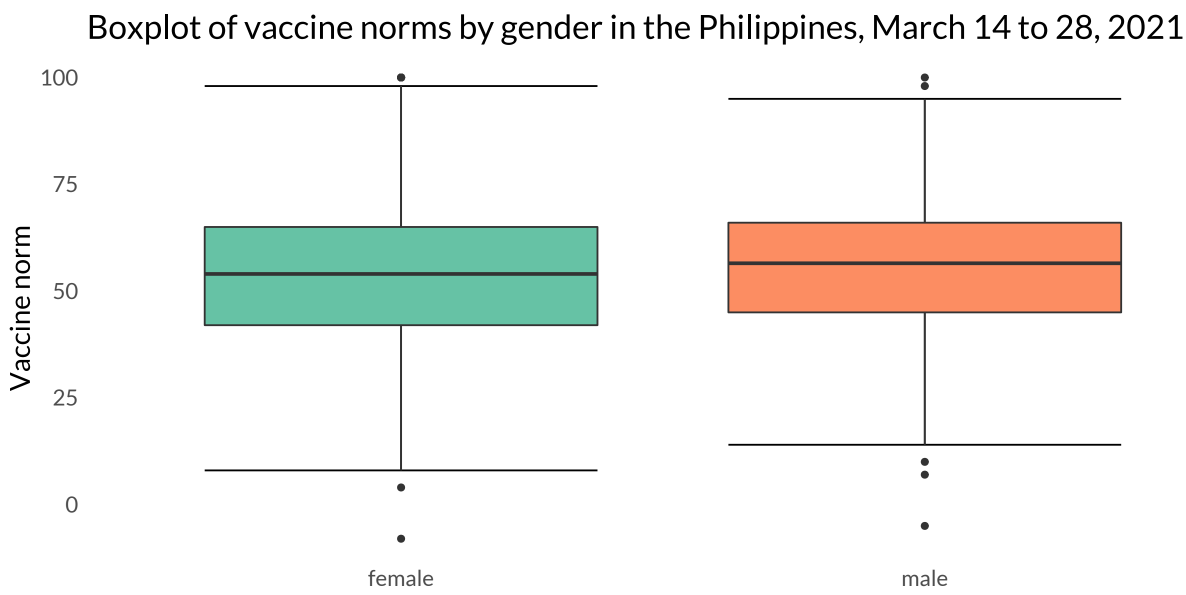
Let's revisit the **vaccine norms** variable we had used in Labs 3 and 4. Individuals were asked about their perceptions of how other people will accept the COVID-19 vaccine, as a measurement of vaccine norms. They were asked, "Out of 100 people in your community, how many do you think would take a COVID-19 vaccine if it were made available?" Again, while this is technically a discrete variable, we will assume for the purposes of this class that this is a continuous variable.

PART A. Independent samples *t*-test

For this part, we want to check if there are any differences in vaccine norm by gender (male vs. female) during the last survey wave (March 14 to 28, 2021). The following are statistical outputs that will help you answer the following questions.

Histogram of vaccine norms by gender in the Philippines, March 14 to 28, 2021





Summary statistics of vaccine norm by gender

Vaccine norm	Male	Female
Count	396	565
Mean	55.6	54.4
Standard deviation (SD)	15.5	17.1
Variance	239.3	292.2
Median	56.5	54
First Quartile (Q1)	45	42
Third Quartile (Q3)	66	65
Interquartile Range (IQR)	21	23
Mode	57	59

```r
# male19 contains all the values of vaccine norms for males during the last survey wave (Wave 19)
# female19 contains all the values of vaccine norms for females during the last survey wave (Wave 19)

# Conduct independent samples t-test, two-sided, equal variances assumed
t.test(x = male19, y = female19, alternative = "two.sided", var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  male19 and female19
## t = 1.1008, df = 959, p-value = 0.2713
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.9286605  3.3013698
## sample estimates:
## mean of x mean of y
##  55.59343  54.40708
```

```r
# Conduct independent samples t-test, two-sided, unequal variances assumed
t.test(x = male19, y = female19, alternative = "two.sided", var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  male19 and female19
## t = 1.1202, df = 899.17, p-value = 0.2629
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8920958  3.2648052
## sample estimates:
## mean of x mean of y
##  55.59343  54.40708
```

1. What are the appropriate null and alternative hypothesis for this question?

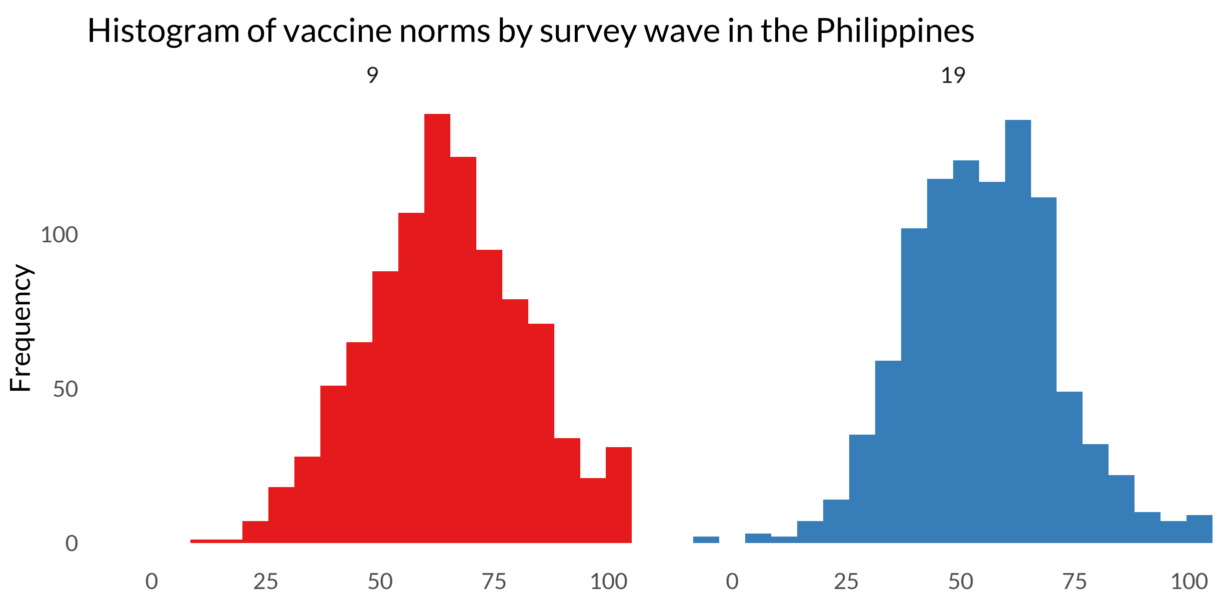
2. What are the assumptions required to conduct an independent samples *t*-test, and how were they fulfilled?

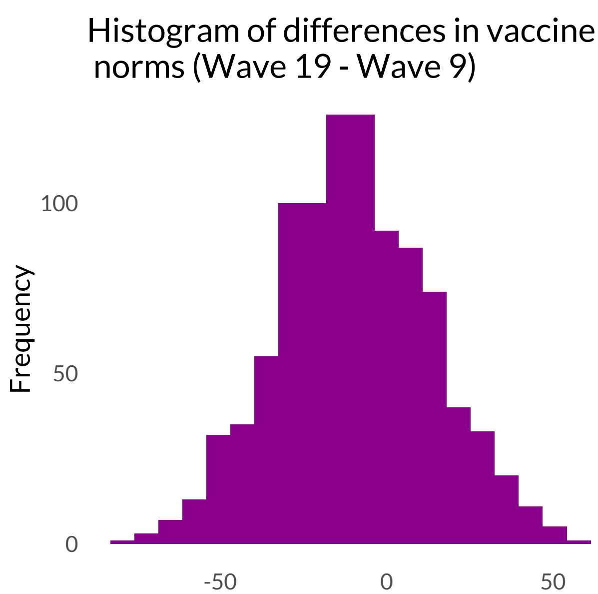
3. Which *t*-test did you use - equal or unequal variances assumed? Explain your response.

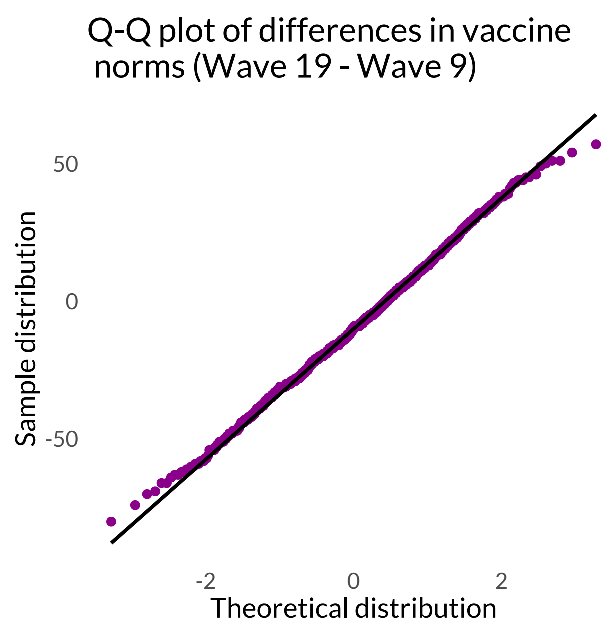
4. What can you conclude? Either use the appropriate test statistic or the 95% confidence interval (CI).

PART B. Paired samples *t*-test

In this part, we will compare the vaccine norms between Wave 9 of the survey (Oct 26 to Nov 10, 2020) to the last wave of the survey (Wave 19, Mar 14 to 28, 2021) and see if there was a significant change in vaccine norms over time. The following are statistical outputs that will help you answer the following questions:







Summary statistics of vaccine norm by survey wave

Vaccine norm	Wave 9	Wave 19	Difference (Wave 19 - Wave 9)
Count	961	961	961
Mean	64.7	54.9	-9.8
Standard deviation (SD)	17.3	16.4	23.1
Variance	298.0	270.5	534.9
Median	64	55	-10
First Quartile (Q1)	54	43	-26
Third Quartile (Q3)	76	66	6
Interquartile Range (IQR)	22	23	32
Mode	100	59	-9

```r
# wave9 contains all the values of vaccine norms for wave 9
# wave19 contains all the values of vaccine norms for wave 19
# data already sorted to align on the same row for the same ID number of survey participant

# Conduct paired samples t-test, two-sided
t.test(x = wave9, y = wave19, alternative = "two.sided", paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  wave9 and wave19
## t = 13.111, df = 960, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   8.317424 11.245532
## sample estimates:
## mean of the differences
##                9.781478
```

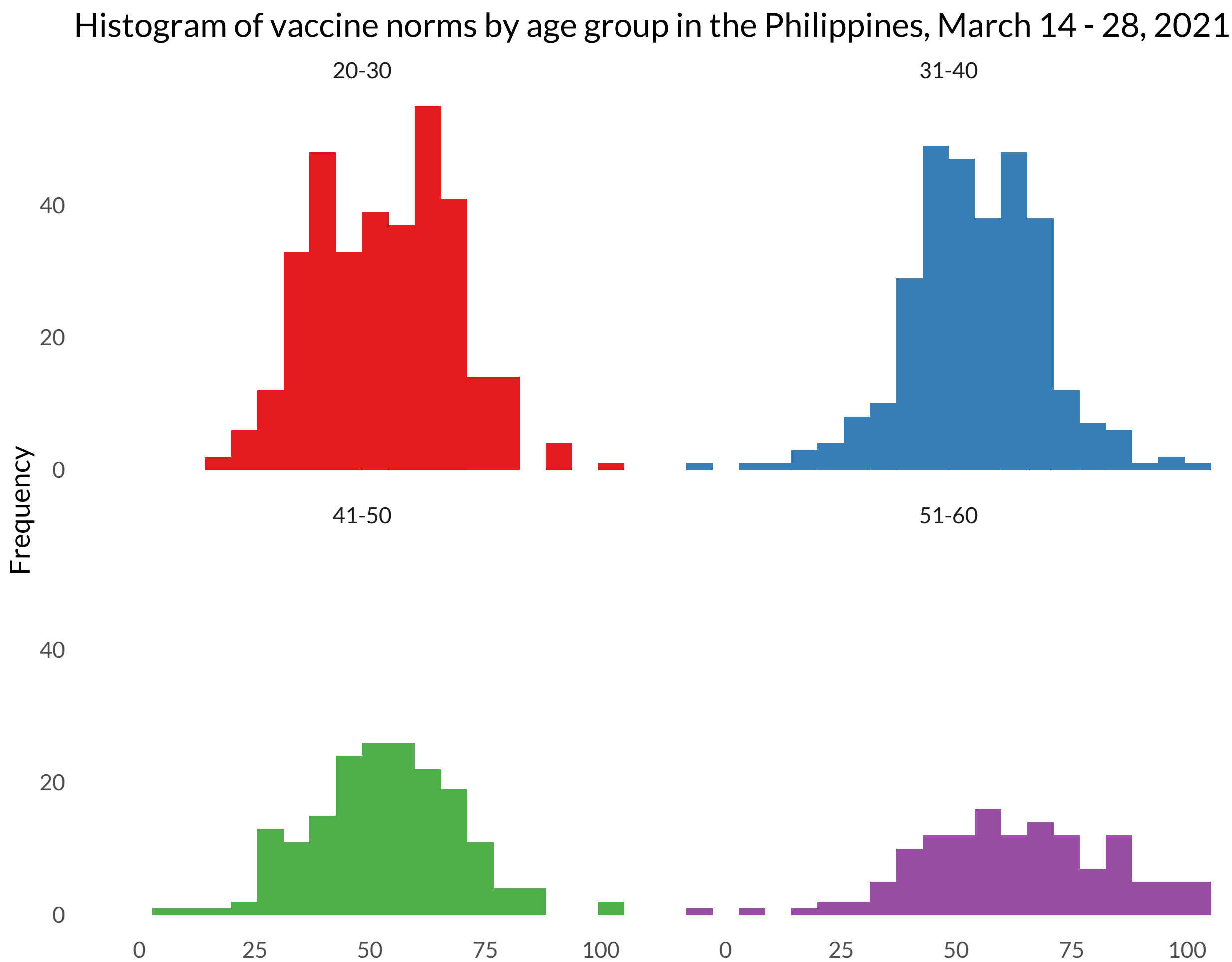
1. What are the appropriate null and alternative hypothesis for this question?

2. What are the assumptions required to conduct a paired samples *t*-test, and how were they fulfilled?

3. What can you conclude? Either use the appropriate test statistic or the 95% confidence interval (CI).

PART C. One-way analysis of variance

For this part, we want to check if there are any differences in vaccine norm by age group during the last survey wave (March 14 to 28, 2021). The following are statistical outputs that will help you answer the following questions.



Summary statistics of vaccine norm by age group

Vaccine norm	20-30 y/o	31-40 y/o	41-50 y/o	51-60 y/o
Count	339	306	182	134
Mean	53.3	54.3	53.2	62.5
SD	15.0	15.1	16.3	20.7
Variance	225.0	227.5	265.5	428.7
Median	54	54.5	54	61.5
Q1	40.5	45.0	43.0	48.2
Q3	64.5	64.0	63.5	77.2
IQR	24.0	19.0	20.5	29.0

```r
# The "vaccine_norm" variable contains the vaccine norm scores
# The "age_grp" variable contains the age group categories

# Conduct one-way ANOVA
summary(aov(vaccine_norm ~ age_grp, data = data19))
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## age_grp       3   9167  3055.7   11.67 1.62e-07 ***
## Residuals   957 250503   261.8                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Conduct post hoc pairwise t-test with Bonferroni correction
# The results shown are the Bonferroni-inflated p-values, meaning the p-values have been inflated
# with the Bonferroni correction factor already. Compare with the original alpha.
pairwise.t.test(data19$vaccine_norm, data19$age_grp, p.adj = "bonf")
```

```
## 
##  Pairwise comparisons using t tests with pooled SD 
## 
## data:  data19$vaccine_norm and data19$age_grp 
## 
##       20-30   31-40   41-50  
## 31-40 1       -       -      
## 41-50 1       1       -      
## 51-60 2.2e-07 8.2e-06 3.5e-06
## 
## P value adjustment method: bonferroni
```

1. What are the appropriate null and alternative hypothesis for this question?

2. What are the assumptions required to conduct a one-way analysis of variance (ANOVA), and how were they fulfilled?

3. What can you conclude from the one-way ANOVA AND the post-hoc pairwise *t*-test with Bonferroni correction (if applicable)? Mention the relevant test statistic and *p*-values, as well as the mean values of vaccine norm in each group

END OF LAB