Excel Analysis ToolPak (Level 2 – Data)

By the QCL

Before Starting

1. Sign-in
2. Locally installed Microsoft Excel
3. Load Analysis ToolPak
4. Download files from: https://github.com/CMC-QCL/Excel-Analysis-ToolPak-L2-Workshop

Load Analysis ToolPak as Add-in

For PC

1.Under the File tab, click Options

2. Under Add-ins, select Analysis ToolPak and click on Go

3. Check Analysis ToolPak and click OK

4. Under the Data tab, in the Analysis group, click on Data Analysis

For Mac: Excel 2016 or later

1. Under the Tools tab, select Excel Add-Ins

2. In the Add-ins window, select Analysis ToolPak and click OK

3. Under the Data tab, click on Data Analysis

If Analysis ToolPak is not available in the list of Add-Ins, click Browse

Agenda

1. What is Analysis ToolPak
2. Common statistical analyses:
   - Histogram
   - Descriptive Statistics
   - F-test
   - t-test
   - ANOVA

   Hands-on time available during workshop

Analysis ToolPak

- Excel add-in
- Good: capability to do analyses without statistical software
- Bad: limited in number of tests and have more manual steps

May 1, 2020

Ashley Rhoades, MBS, RAC
Senior Associate, Regulatory Affairs
Gilead Sciences, Inc.
333 Lakeside Drive
Foster City, CA 94404

Dear Ms. Rhoades:

This letter is in response to your request that the Food and Drug Administration (FDA) issue an Emergency Use Authorization (EUA) for emergency use of remdesivir for the treatment of hospitalized 2019 coronavirus disease (COVID-19) patients, as described in the Scope of Authorization (section II) of this letter, pursuant to Section 564 of the Federal Food, Drug and Cosmetic Act (the Act) (21 U.S.C. 360bbb-3).

https://www.fda.gov/media/137564/download

Data File

id	treatment	radiograph	nose swab	throat swab	rectal swab	score
RM1	remdesivir	0	5.185341	3.916133	0	0
RM2	remdesivir	0	3.638322	3.312539	0	0
RM3	remdesivir	2	7.944003	0	4.314602	3
RM4	remdesivir	1	7.037755	3.407456	3.102061	0
RM5	remdesivir	1	6.421312	3.879554	2.60709	3
RM6	remdesivir	0	7.097592	2.621482	0	0
RM7	vehicle	1	6.328537	4.75539	3.423561	8
RM8	vehicle	1	5.003218	0	5.805897	5
RM9	vehicle	0	6.205009	2.605291	0	10
RM10	vehicle	0	8.243229	5.772659	0	5
RM11	vehicle	3	6.159435	6.4662	0	10
RM12	vehicle	2	7.360754	8.563787	3.734053	10

File: remdesivirCov19list.xlsx

7 fields (column): id, treatment, radiograph, nose swab, throat swab, rectal swab, score

12 records (rows): original data contains 18 tests were downloadable in workbook format

Experimental procedure: 12 rhesus monkeys are divided into two groups, one test and one control. The test group is administered remdesivir while the control group is given vehicle solution over the course of 7 days. Only day 1 and day 7 data used here

https://doi.org/10.1101/2020.04.15.043166

Histogram

Use to summarize the number of data points that fall within a specified range of values called bins

Resembles a vertical bar graph, but contains no gaps in between the bars

Question: what does the score histogram look like?

Histogram

1. Enter bin numbers of 5, 10, 15 (upper range) in H2: H4

2. Under the Data tab, click Data Analysis

3. Select Histogram and click OK

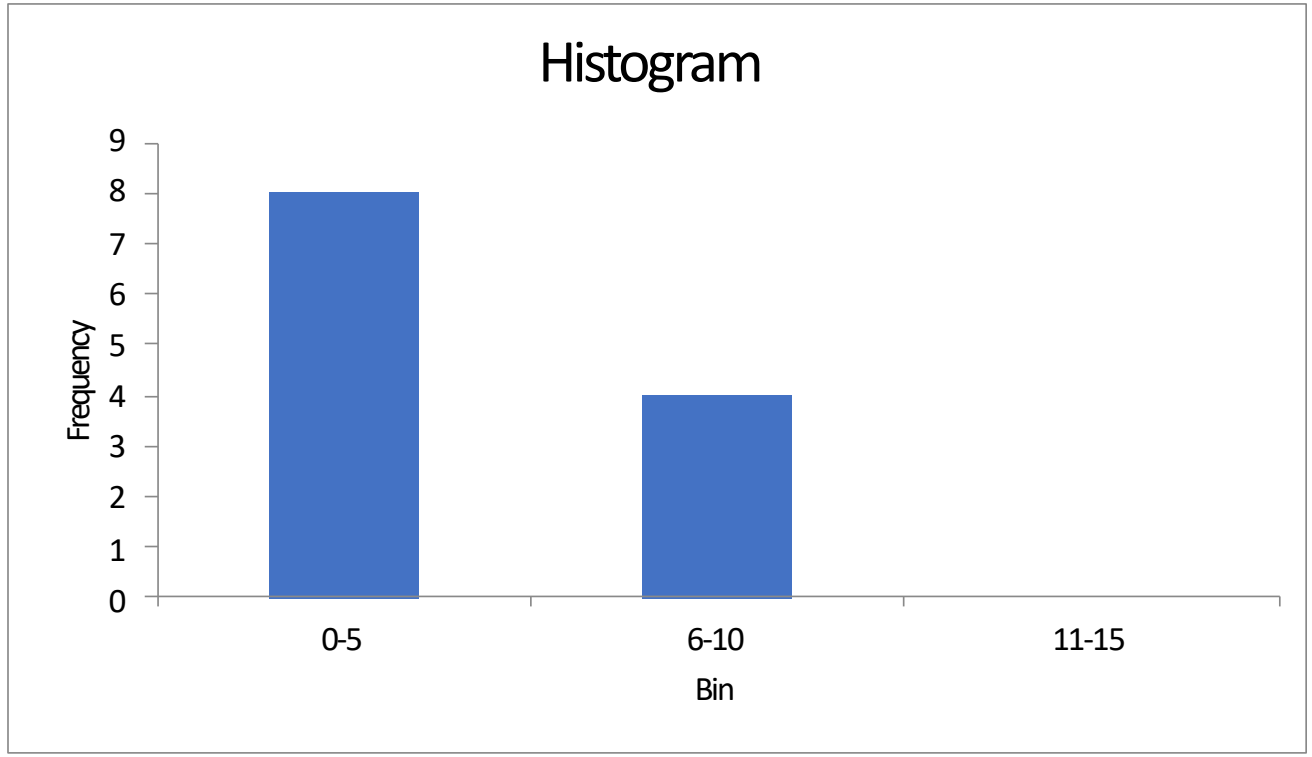
4. Select G3: G14 as Input Range

5. Enter H3: H5 as Bin Range

6. Under Output options, select Cell J1 as Output Range

7. Check Chart Output and click OK

8. Once histogram appears, label range and click on diagram to edit



Descriptive Statistics

Numbers that are used to summarize and describe data:

- Measure of central tendency: mean, median and mode
- Measure of variance: standard deviation, skewness and kurtosis

Question: Using data from day 1, what are the summary statistics for radiograph and nose swab?

Descriptive Statistics Output

Summary statistics output:

1. Under the Data tab, click Data Analysis

2. Select Descriptive Statistics and click OK

3. Select the range C2: D14 as the Input Range. Check Labels in first row

4. Select Cell I1 as the Output Range. Check Summary statistics and click OK

radiograph		nose swab	
Mean	0.916666667	Mean	6.385375583
Standard Error	0.287579589	Standard Error	0.376889203
Median	1	Median	6.3749245
Mode	0	Mode	#N/A
Standard Deviation	0.99620492	Standard Deviation	1.305582496
Sample Variance	0.992424242	Sample Variance	1.704545655
Kurtosis	-0.013845347	Kurtosis	0.404774611
Skewness	0.85362124	Skewness	-0.654436895
Range	3	Range	4.604907
Minimum	0	Minimum	3.638322
Maximum	3	Maximum	8.243229
Sum	11	Sum	76.624507
Count	12	Count	12

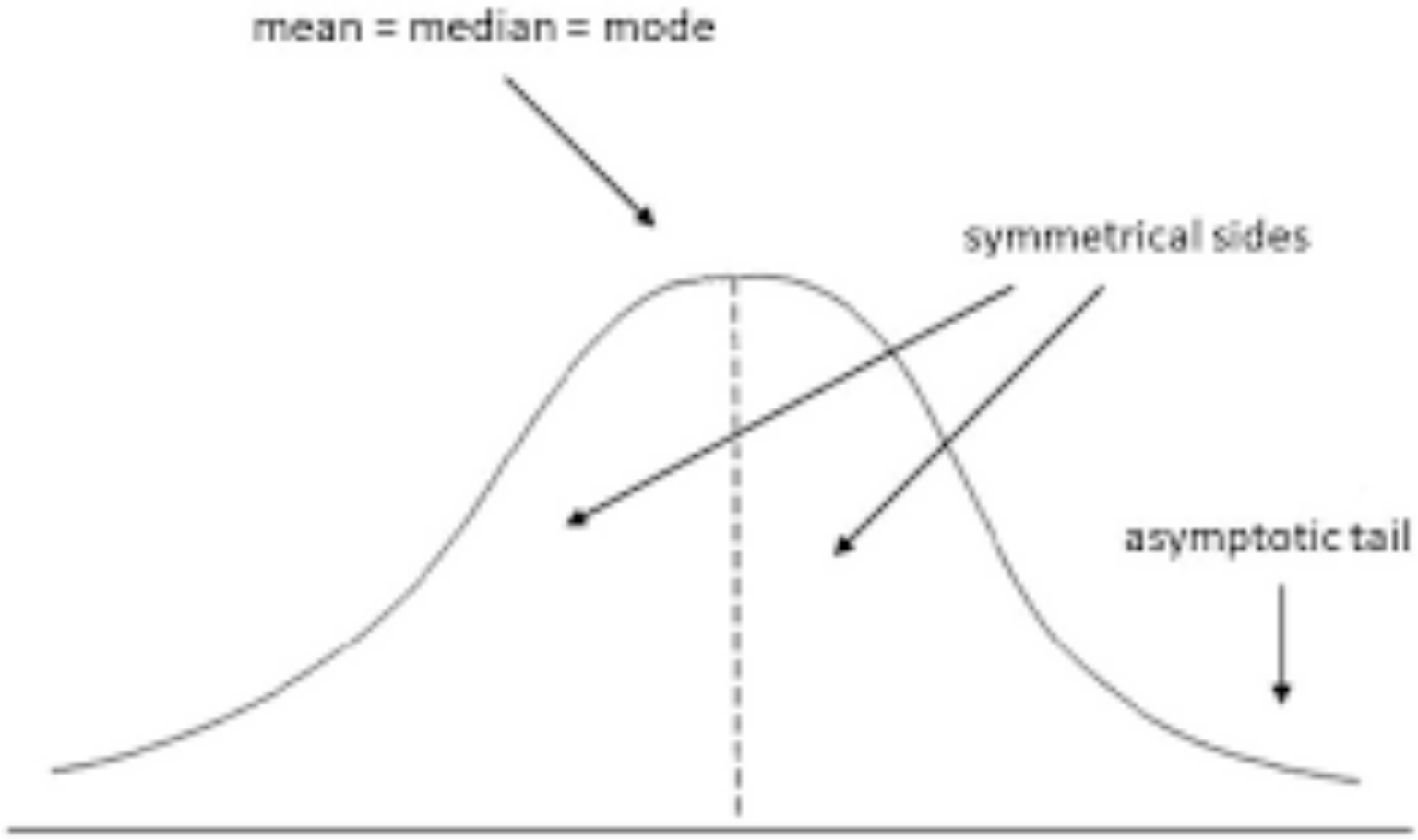
Hypothesis Testing

- Hypothesis: assumption about a phenomenon that is testable either by experiment or observation

- Hypothesis test in statistics allows testing of the results of a study to see if your results are meaningful:
    1. Figure out your null hypothesis (H_0)
    2. State your null hypothesis
    3. Choose the test to perform
    4. Support or reject the null hypothesis (H_1)

- All tests covered assume normal distribution: bell shape curve

Normal Distribution



https://www.simplypsychology.org/normal-distribution.html

F-Test

Used to determine whether the variances of two populations are equal:

Variance: refers to how far the numbers are spread from the center

In the F-test, there are two hypotheses:

H_0 (null hypothesis): $\sigma_1 = \sigma_2$

H_1 (alternative hypothesis): $\sigma_1 \neq \sigma_2$

Question: Using the nose swab data from day 1, are the variances of the control group and test group equal ($\alpha = 0.05$)?

F-Test Output

Nose swab, day 1:

day 1	nose swab	
f-test	remdesivir	vehicle
RM1/RM7	5.185341	6.328537
RM2/RM8	3.638322	5.003218
RM3/RM9	7.944003	6.205009
RM4/RM10	7.037755	8.243229
RM5/RM11	6.421312	6.159435
RM6/RM12	7.097592	7.360754

1. Under Data tab, click Data Analysis, select F-Test and click OK
2. Enter B2: B8 in Variable 1 Range
3. Enter C2: C8 in Variable 2 Range
4. Check Labels
5. Enter Cell E1 in Output Range and click OK

F-Test Two-Sample for Variances		
	remdesivir	vehicle
Mean	6.22072083	6.55003033
Variance	2.43743625	1.24749734
Observations	6	6
df	5	5
F	1.95386087	
P(F<=f) one-tail	0.23996192	
F Critical one-tail	5.05032906	

Check if variances under Variable 1 > Variable 2

Yes: proceed to data interpretation

No: swap data to calculate correct F-value

Result: p-value of 0.2340 > 0.05, H_0 cannot be rejected so assume equal variances

t-Test

Used to determine whether the means of two populations are equal

In the t-test, there are two hypotheses:

H_0(null hypothesis): $\mu_1 = \mu_2$

H_1 (alternative hypothesis): $\mu_1 \neq \mu_2$

Question: Using the nose swab data from day, are the means of the control group and test group equal (α= 0.05)?

t-Test output

Nose swab, day 1:

day 1	nose swab	
f-test	remdesivir	vehicle
RM1/RM7	5.185341	6.328537
RM2/RM8	3.638322	5.003218
RM3/RM9	7.944003	6.205009
RM4/RM10	7.037755	8.243229
RM5/RM11	6.421312	6.159435
RM6/RM12	7.097592	7.360754

1. Under Data tab, click Data Analysis, select t-Test: Two Samples Assuming Equal Variances and click OK
2. Enter B2: B8 in Variable 1 Range
3. Enter C2: C8 in Variable 2 Range
4. Enter 0 for Hypothesized Mean Difference
5. Check Labels
6. Enter Cell E1 in Output Range and click OK

t-Test: Two-Sample Assuming Equal Variances		
	remdesivir	vehicle
Mean	6.220720833	6.550030333
Variance	2.43743625	1.247497343
Observations	6	6
Pooled Variance	1.842466796	
Hypothesized Mean Difference	0	
df	10	
t Stat	-0.420208712	
P(T<=t) one-tail	0.341609663	
t Critical one-tail	1.812461123	
P(T<=t) two-tail	0.683219327	
t Critical two-tail	2.228138852	

The p-value is greater than 0.05 (0.6832 > 0.05) so the null hypothesis cannot be rejected. Therefore, assume that means are equal

Analysis of Variance (ANOVA)

Statistical models and associated estimation procedures used to analyze the differences among group means in a data set

Single factor or one-way ANOVA is used to test the null hypothesis (H_0) that the means of several populations are equal

Two-way ANOVA is used when want to compare the differences between groups that have been split on two independent variables

One-Way ANOVA

In one-way ANOVA:

H_0 (null hypothesis): $\mu_A = \mu_B = \mu_C$

H_1 (alternative hypothesis): at least one mean is different

Question: Using the nose swab data from days 1, 3 and 7, are the means from the control group (RM1 to RM6) equal ($\alpha = 0.05$)?

One-Way ANOVA Output

Using nose swab data on days 1, 3 and 7

1. Under Data tab, click Data Analysis
2. Select ANOVA: Single Factor and click OK
3. Select C3: E8 as Input Range
4. Check Labels in first row
5. Leave alpha as 0.05
6. Under Output options, Cell G1 as Output Range
7. Check Chart Output and click OK

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
day1	6	37.324325	6.22072083	2.43743625
day3	6	28.862184	4.810364	6.46655747
day7	6	17.137164	2.856194	2.17916216

ANOVA

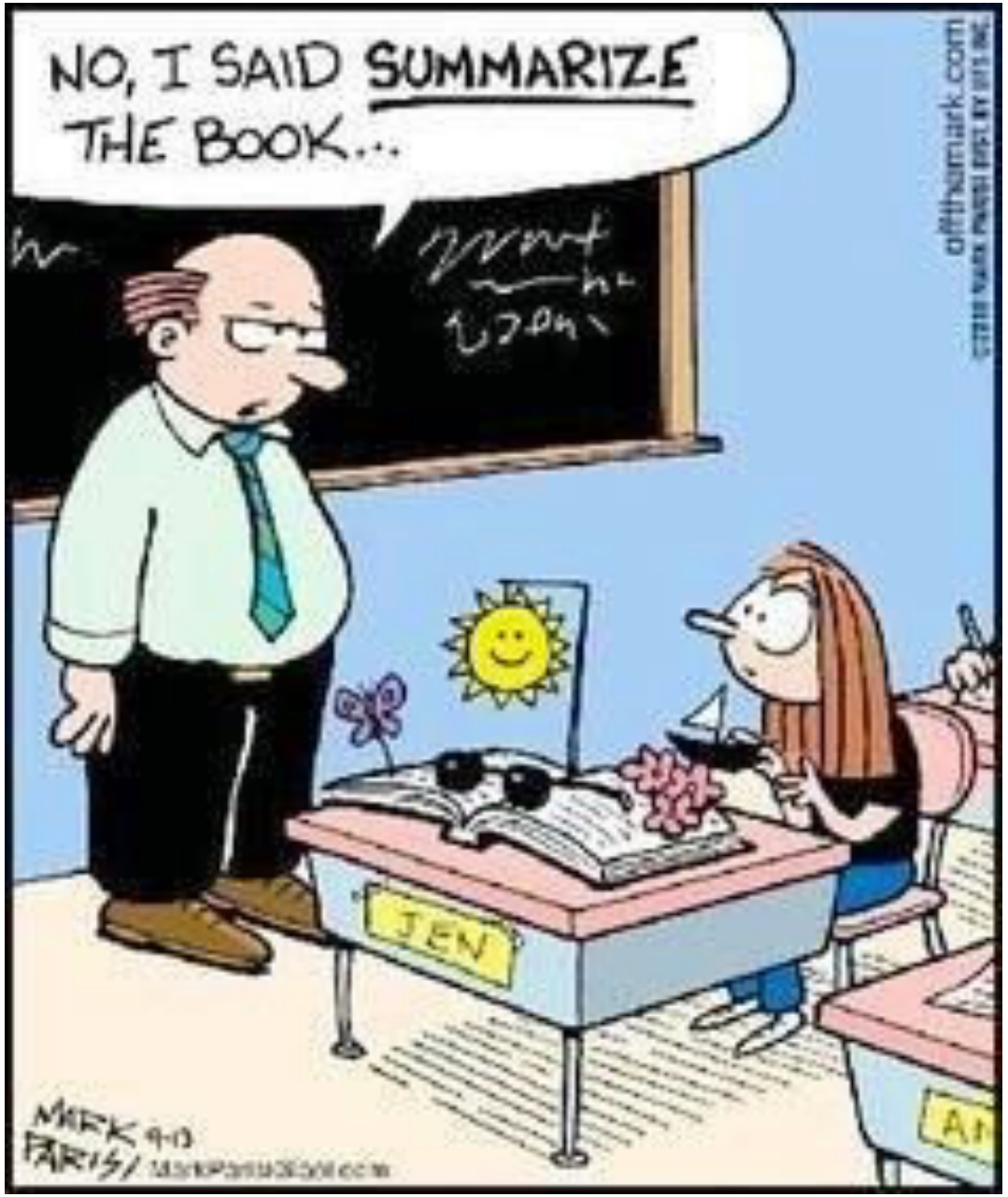
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	34.2558552	2	17.1279276	4.63620501	0.02706062	3.68232034
Within Groups	55.4157794	15	3.69438529			
Total	89.6716346	17				

The p-value is less than 0.05 (0.0271 < 0.05) to show statistical significance, F value is also greater than F crit (4.636 > 3.682). Therefore, the null hypothesis is rejected to suggest that at least one of the means of the three swab is not equal.

Hands-on Exercises

Suppose that you want to own a new startup business and you got a data set (owan04.xls) that contains the startup costs for shops. To evaluate what do you want to go into, a sampling of different amounts of cost are collected. Using $\alpha = 0.05$, determine the following:

1. Generate descriptive statistics for all the variables in the dataset
2. Let's say you set your mind on owning either a pizza parlor or a baker/donut shop, using X1 (cost for pizza) and X2 (cost for baker/goods):
    - The F-test to determine whether the variances of the two variables are equal or not
    - The t-test to determine whether the means of the two variables are equal or not
3. Perform a one-way ANOVA using all variables to determine whether at least one of the means are equal



http://thedumbdatascientist.blogspot.com/2016/12/descriptive-statistics-summarizing-data.html

In the End

1. Excel Analysis ToolPak provides a basic set of tools for data analyses

2. Things to keep in mind:

   What does the output mean?

   Are the results statistically significant or not?

Questions?

Come visit the QCL at Kravis Lower Court or email at qcl@cmc.edu