

# Análisis de variables cuantitativas y cualitativas en la Investigación Científica.

Una perspectiva desde la Ciencia de Datos.

Guillermo Pizarro, [gpizarro@ups.edu.ec](mailto:gpizarro@ups.edu.ec)

Universidad Politécnica Salesiana

15 de diciembre de 2020

# CITIC 2020

## VIII Congreso Internacional en Tecnologías de Información y Comunicación

# Proceso de Diseño de una Investigación Científica

Una investigación requiere de un diseño previo su implementación donde se contemple:

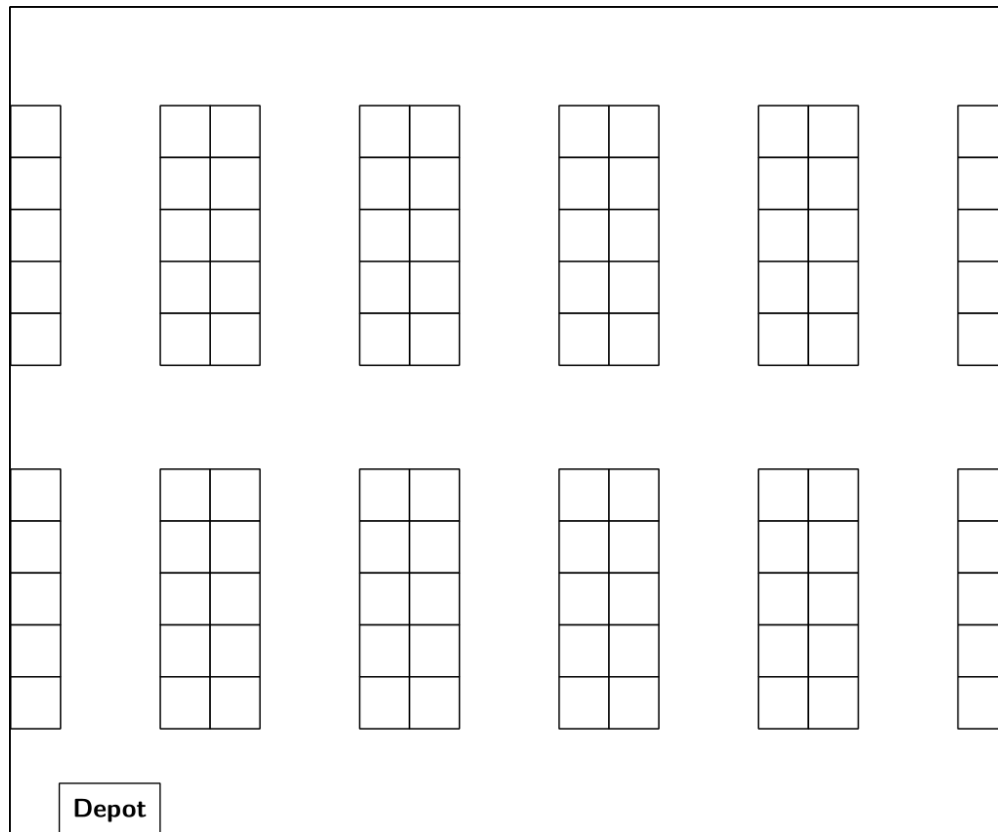
- Enfoque: cuantitativo, cualitativo y mixto.
- Alcance: exploratorio, descriptivo, correlacional y explicativo.
- Diseño: experimentales y no experimentales.
- Planteamiento de Hipótesis.
- Selección de la Muestra: por tipo o por proceso.
- Técnicas de Recolección de datos.
- Análisis de los Datos.

# Variables Cuantitativas y Cualitativas

Diseño Experimental Puro Aleatorio con un Alcance Correlacional

# Problema de Investigación

- Optimización del agrupamiento y recogida de pedidos en un almacén rectangular con un pasillo transversal.



# Preguntas de Investigación

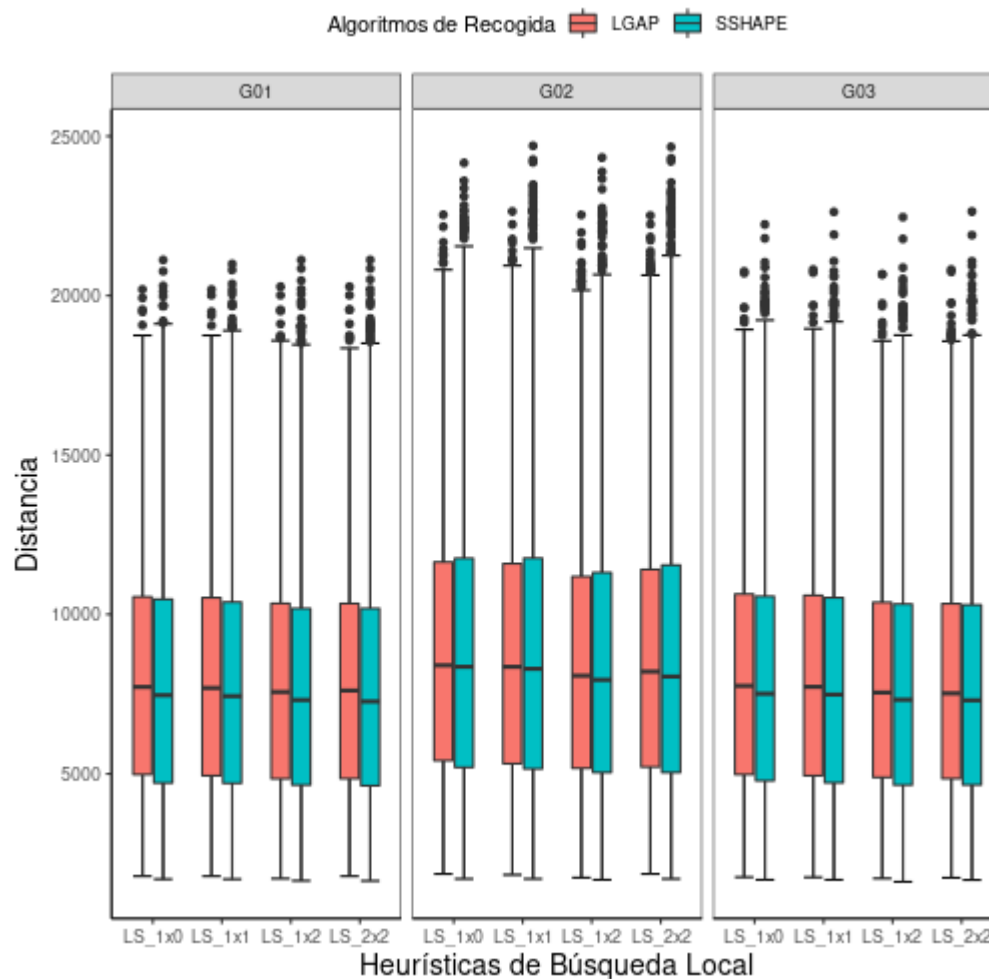
- ¿Existen diferencias entre las medias de las distancias con respecto a los algoritmos de agrupamiento de pedidos?
- ¿Existen diferencias entre las medias de las distancias con respecto a los algoritmos de recogida de pedidos?
- ¿Existen diferencias entre las medias de las distancias con respecto a los algoritmos de búsqueda local y sus variaciones?
- ¿La aplicación de los algoritmos heurísticos de búsqueda local inciden en la minimización de la distancia total recorrida con respecto a la distancia total obtenida mediante los algoritmos de agrupamiento y recogida de pedidos?

# Análisis Exploratorio de los Datos

```
summary(datos)
```

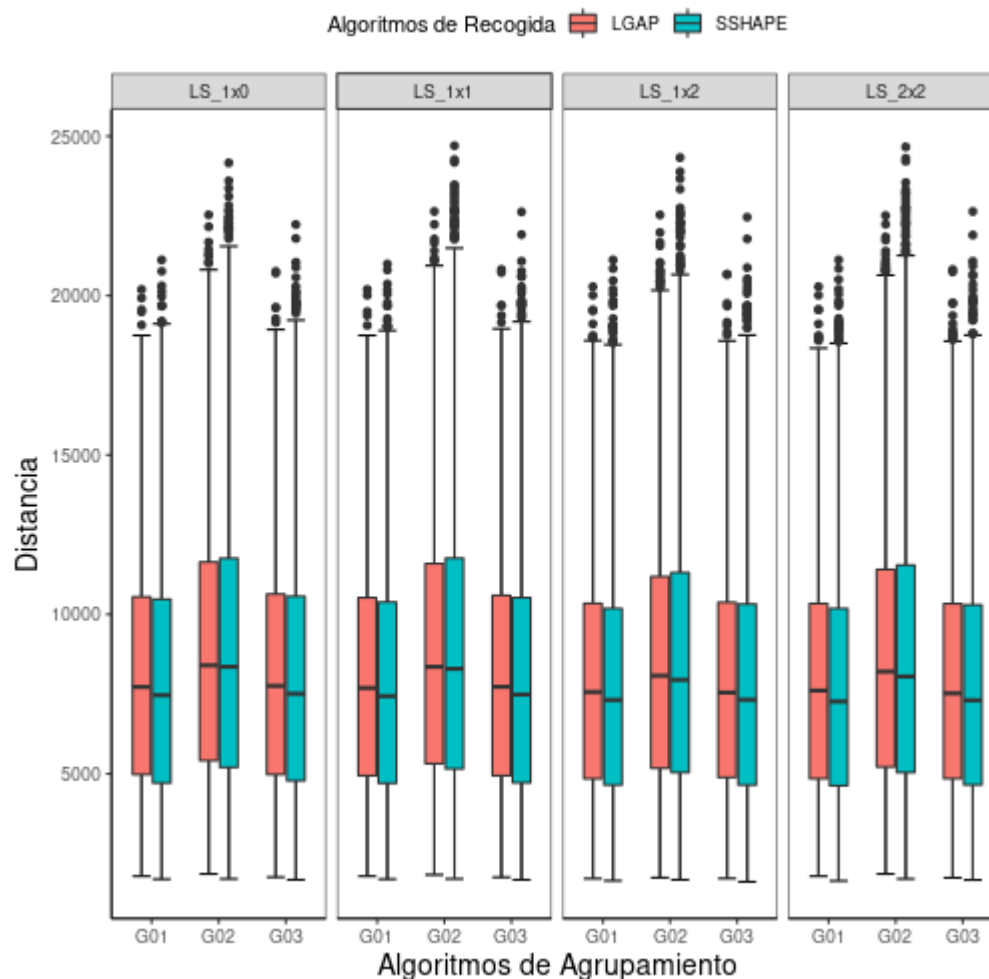
```
##  obp_algorithm prp_algorithm  ls_algorithm      time      capacity_c
##  G01:11520      LGAP   :17280    LS_1x0:8640    Min.       :2.937e+05    30:8640
##  G02:11520      SSHAPE:17280    LS_1x1:8640    1st Qu.:3.448e+06    45:8640
##  G03:11520                      LS_1x2:8640    Median  :9.093e+06    60:8640
##                               LS_2x2:8640    Mean    :2.194e+07    75:8640
##                               3rd Qu.:2.519e+07
##                               Max.      :2.168e+09
##
##      num_orders  before_distance after_distance
##  20      : 3840    Min.       : 1676    Min.       : 1612
##  30      : 3840    1st Qu.: 5005    1st Qu.: 4926
##  40      : 3840    Median  : 7871    Median  : 7749
##  50      : 3840    Mean    : 8422    Mean    : 8297
##  60      : 3840    3rd Qu.:10976    3rd Qu.:10789
##  70      : 3840    Max.      :24698    Max.      :24698
##  (Other):11520
```

# Análisis Exploratorio de los Datos

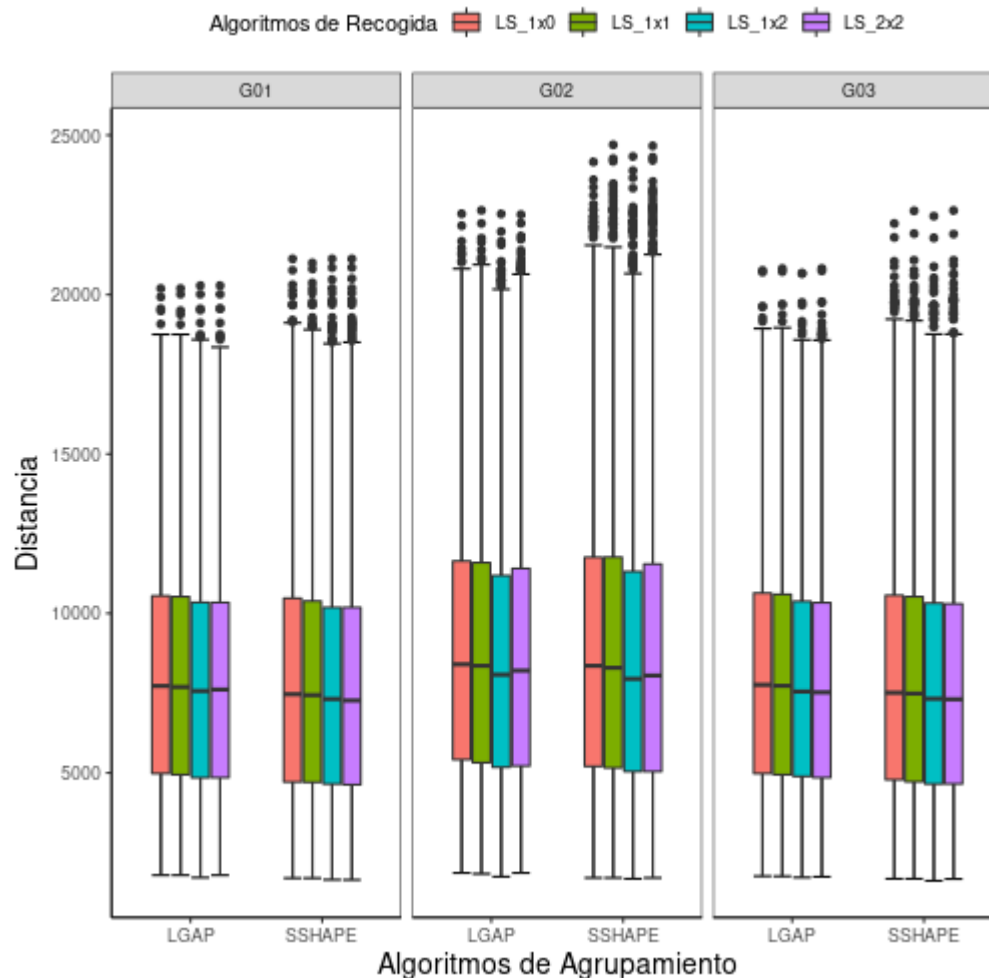




# Análisis Exploratorio de los Datos



# Análisis Exploratorio de los Datos



# Análisis Exploratorio de los Datos

Aquí se puede conocer que el algoritmo que minimiza la distancia, sería la siguiente combinación: Greedy-03 + SShape + Local Search 1x2.

```
## # A tibble: 10 x 6
## # Groups:   obp_algorithm, prp_algorithm [3]
##   obp_algorithm prp_algorithm ls_algorithm PROMEDIO DESVIACION MINIMO
##   <fct>         <fct>         <fct>         <dbl>         <dbl>         <dbl>
## 1 G03          SSHAPE        LS_1x2         7957.         4175.         1612
## 2 G01          SSHAPE        LS_2x2         7875.         4088.         1642
## 3 G01          SSHAPE        LS_1x2         7875.         4072.         1646
## 4 G03          SSHAPE        LS_2x2         7942.         4221.         1670
## 5 G02          SSHAPE        LS_1x2         8745.         4782.         1676
## 6 G03          SSHAPE        LS_1x0         8125.         4229.         1676
## 7 G03          SSHAPE        LS_1x1         8092.         4234.         1676
## 8 G01          SSHAPE        LS_1x0         8008.         4104.         1694
## 9 G01          SSHAPE        LS_1x1         7975.         4100.         1694
## 10 G02         SSHAPE        LS_1x1         9032.         4910.         1704
```

# Análisis Inferencial

Dado que se desea conocer si la aplicación de los algoritmos heurísticos de búsqueda local en alguna de sus variaciones (1x1, 1x2, 2x2 o 2x3) inciden en minimizar la distancia recorrida; se necesita verificar esta hipótesis estadísticamente.

Para realizar este procedimiento, es necesario verificar algunos supuestos; tales como:

- **Supuesto de independencia**, este supuesto se cumple, dado que en la simulación computacional se obtuvieron observaciones independientes;
- **Supuesto de Normalidad**, los datos de las distancias tendrían que tener una distribución normal.
- **Supuesto de Homocedastecidad**, sería la homogeneidad de la varianza de la variable dependiente entre los grupos.

# Verificación del supuesto de Normalidad

Se puede verificar gráficamente si existe normalidad en los datos, mediante el gráfico de la densidad o el histograma de las distancias.

# Verificación del supuesto de Normalidad

Verificación de normalidad mediante la prueba de Lilliefors.

```
lillie.test(datos$after_distance)
```

```
##  
##      Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data:  datos$after_distance  
## D = 0.061849, p-value < 2.2e-16
```

En este caso, la  $H_0$  es que los datos son normales; sin embargo, el valor de  $p$  es muy cercano a cero lo que permite concluir que se rechaza la  $H_0$ ; por tanto, los datos **no son normales**.

# Otra alternativa: Pruebas No Paramétricas

Dado que no se puede utilizar una prueba paramétrica, se puede utilizar una prueba no paramétrica; sin embargo, para esto es necesario realizar la verificación del **supuesto de la homocedastecidad**.

# Supuesto de Homocedasticidad

A continuación, la realiza la prueba de Bartlett.

```
bartlett.test(data=datos, before_distance ~ obp_algorithm)
```

```
##  
##      Bartlett test of homogeneity of variances  
##  
## data:  before_distance by obp_algorithm  
## Bartlett's K-squared = 366.1, df = 2, p-value < 2.2e-16
```

La  $H_0$  es que existe homogeneidad en la varianza; sin embargo, el valor de  $p$  es muy cercano a cero; por tanto, existe evidencia estadística para rechazar la  $H_0$ ; es decir, no se cumple este supuesto tampoco.



# Prueba ANOVA Robusta

Dado que no se cumplieron los supuestos de normalidad ni de homocedastecidad, se debe utilizar una prueba robusta.

# Prueba ANOVA Robusta

Primero utilizaremos la alternativa ANOVA Robusta (con la media recortada en un 20%), donde la  $H_0$  define que las medias de cada grupo son iguales.

```
tlway(before_distance ~ obp_algorithm, data = datos)
```

```
## Call:
## tlway(formula = before_distance ~ obp_algorithm, data = datos)
##
## Test statistic: F = 101.493
## Degrees of freedom 1: 2
## Degrees of freedom 2: 13776.19
## p-value: 0
##
## Explanatory measure of effect size: 0.1
## Bootstrap CI: [0.09; 0.12]
```

Dado que el valor de  $p$  es muy cercano a cero, se rechaza la  $H_0$ ; por lo tanto, las medias de los diferentes grupos son diferentes.

# Prueba ANOVA Robusta

Si deseamos revisar de manera detallada las similitudes entre cada grupo, se detalla lo siguiente:

```
lincon(before_distance ~ obp_algorithm, data = datos)
```

```
## Call:
```

```
## lincon(formula = before_distance ~ obp_algorithm, data = datos)
```

```
##
```

```
##               psihat  ci.lower  ci.upper p.value
## G01 vs. G02 -827.61690 -978.9798 -676.25400 0.00000
## G01 vs. G03  -59.69676 -201.5794   82.18588 0.31463
## G02 vs. G03  767.92014  615.6355  920.20480 0.00000
```

Se puede ver que las medias entre los algoritmos de agrupamiento Greedy-01 y Greedy-03 son similares, como se verificó visualmente.

# Prueba ANOVA Robusta

En esta prueba se rechaza la  $H_0$ ; por lo tanto, las medias de los grupos no son iguales.

```
t1way(before_distance ~ prp_algorithm, data = datos)
```

```
## Call:
## t1way(formula = before_distance ~ prp_algorithm, data = datos)
##
## Test statistic: F = 10.4167
## Degrees of freedom 1: 1
## Degrees of freedom 2: 20725.73
## p-value: 0.00125
##
## Explanatory measure of effect size: 0.03
## Bootstrap CI: [0.01; 0.04]
```

# Prueba ANOVA Robusta

En esta prueba se acepta la  $H_0$ ; por lo tanto, las medias de los grupos son iguales.

```
t1way(before_distance ~ ls_algorithm, data = datos)
```

```
## Call:
## t1way(formula = before_distance ~ ls_algorithm, data = datos)
##
## Test statistic: F = 0
## Degrees of freedom 1: 3
## Degrees of freedom 2: 11517.78
## p-value: 1
##
## Explanatory measure of effect size: 0
## Bootstrap CI: [0; 0.02]
```

# Prueba ANOVA Robusta

Si deseamos revisar de manera detallada las similitudes entre cada grupo, se realiza la siguiente prueba:

```
lincon(before_distance ~ ls_algorithm, data = datos)
```

```
## Call:
```

```
## lincon(formula = before_distance ~ ls_algorithm, data = datos)
```

```
##
```

##		psihat	ci.lower	ci.upper	p.value
##	LS_1x0 vs. LS_1x1	0	-189.0181	189.0181	1
##	LS_1x0 vs. LS_1x2	0	-189.0181	189.0181	1
##	LS_1x0 vs. LS_2x2	0	-189.0181	189.0181	1
##	LS_1x1 vs. LS_1x2	0	-189.0181	189.0181	1
##	LS_1x1 vs. LS_2x2	0	-189.0181	189.0181	1
##	LS_1x2 vs. LS_2x2	0	-189.0181	189.0181	1

Se verifica la similitud entre todos los grupos.

# Prueba YUEND Robusta

Esta prueba se realiza para dos muestras dependientes:

```
yuend(datos$before_distance, datos$after_distance)
```

```
## Call:
## yuend(x = datos$before_distance, y = datos$after_distance)
##
## Test statistic: 110.9393 (df = 20735), p-value = 0
##
## Trimmed mean difference: 136.0534
## 95 percent confidence interval:
## 133.6496      138.4572
##
## Explanatory measure of effect size: 0.02
```

Se rechaza la  $H_0$ ; por lo tanto, las medias no son iguales; además, se puede inferir que la media de la distancia anterior es mayor a la distancia donde se han aplicado los algoritmos de búsqueda local; por lo tanto, si incide la aplicación de los algoritmos de Búsqueda Local para encontrar una distancia mínima.

**Muchas gracias**



