

Clases - Semana 4

December 30, 2020

1 Basic Statistical

- Vamos a utilizar conceptos de estadística como prueba de hipótesis:

```
[46]: import pandas as pd
import numpy as np
from scipy import stats
```

```
[49]: df = pd.read_csv('grades.csv')
df.head()
```

```
[49]:
```

	student_id	assignment1_grade	\
0	B73F2C11-70F0-E37D-8B10-1D20AFED50B1	92.733946	
1	98A0FAE0-A19A-13D2-4BB5-CFBFD94031D1	86.790821	
2	D0F62040-CEB0-904C-F563-2F8620916C4E	85.512541	
3	FFDF2B2C-F514-EF7F-6538-A6A53518E9DC	86.030665	
4	5ECBEEB6-F1CE-80AE-3164-E45E99473FB4	64.813800	

	assignment1_submission	assignment2_grade	\
0	2015-11-02 06:55:34.282000000	83.030552	
1	2015-11-29 14:57:44.429000000	86.290821	
2	2016-01-09 05:36:02.389000000	85.512541	
3	2016-04-30 06:50:39.801000000	68.824532	
4	2015-12-13 17:06:10.750000000	51.491040	

	assignment2_submission	assignment3_grade	\
0	2015-11-09 02:22:58.938000000	67.164441	
1	2015-12-06 17:41:18.449000000	69.772657	
2	2016-01-09 06:39:44.416000000	68.410033	
3	2016-04-30 17:20:38.727000000	61.942079	
4	2015-12-14 12:25:12.056000000	41.932832	

	assignment3_submission	assignment4_grade	\
0	2015-11-12 08:58:33.998000000	53.011553	
1	2015-12-10 08:54:55.904000000	55.098125	
2	2016-01-15 20:22:45.882000000	54.728026	
3	2016-05-12 07:47:16.326000000	49.553663	
4	2015-12-29 14:25:22.594000000	36.929549	

	assignment4_submission	assignment5_grade \
0	2015-11-16 01:21:24.663000000	47.710398
1	2015-12-13 17:32:30.941000000	49.588313
2	2016-01-11 12:41:50.749000000	49.255224
3	2016-05-07 16:09:20.485000000	49.553663
4	2015-12-28 01:29:55.901000000	33.236594

	assignment5_submission	assignment6_grade \
0	2015-11-20 13:24:59.692000000	38.168318
1	2015-12-19 23:26:39.285000000	44.629482
2	2016-01-11 17:31:12.489000000	44.329701
3	2016-05-24 12:51:18.016000000	44.598297
4	2015-12-29 14:46:06.628000000	33.236594

	assignment6_submission
0	2015-11-22 18:31:15.934000000
1	2015-12-21 17:07:24.275000000
2	2016-01-17 16:24:42.765000000
3	2016-05-26 08:09:12.058000000
4	2016-01-05 01:06:59.546000000

```
[10]: #vamos a dividir en dos grupos: los que terminaron la materia 1 antes de
      →diciembre de 2015 y los que terminaron despues:

early_finishers = df[pd.to_datetime(df['assignment1_submission']) < '2016']
early_finishers.head()
```

	student_id	assignment1_grade \
0	B73F2C11-70F0-E37D-8B10-1D20AFED50B1	92.733946
1	98A0FAE0-A19A-13D2-4BB5-CFBFD94031D1	86.790821
4	5ECBEEB6-F1CE-80AE-3164-E45E99473FB4	64.813800
5	D09000A0-827B-C0FF-3433-BF8FF286E15B	71.647278
8	C9D51293-BD58-F113-4167-A7C0BAFCB6E5	66.595568

	assignment1_submission	assignment2_grade \
0	2015-11-02 06:55:34.282000000	83.030552
1	2015-11-29 14:57:44.429000000	86.290821
4	2015-12-13 17:06:10.750000000	51.491040
5	2015-12-28 04:35:32.836000000	64.052550
8	2015-12-25 02:29:28.415000000	52.916454

	assignment2_submission	assignment3_grade \
0	2015-11-09 02:22:58.938000000	67.164441
1	2015-12-06 17:41:18.449000000	69.772657
4	2015-12-14 12:25:12.056000000	41.932832
5	2016-01-03 21:05:38.392000000	64.752550
8	2015-12-31 01:42:30.046000000	48.344809

	assignment3_submission	assignment4_grade \
0	2015-11-12 08:58:33.998000000	53.011553
1	2015-12-10 08:54:55.904000000	55.098125
4	2015-12-29 14:25:22.594000000	36.929549
5	2016-01-07 08:55:43.692000000	57.467295
8	2016-01-05 23:34:02.180000000	47.444809

	assignment4_submission	assignment5_grade \
0	2015-11-16 01:21:24.663000000	47.710398
1	2015-12-13 17:32:30.941000000	49.588313
4	2015-12-28 01:29:55.901000000	33.236594
5	2016-01-11 00:45:28.706000000	57.467295
8	2016-01-02 07:48:42.517000000	37.955847

	assignment5_submission	assignment6_grade \
0	2015-11-20 13:24:59.692000000	38.168318
1	2015-12-19 23:26:39.285000000	44.629482
4	2015-12-29 14:46:06.628000000	33.236594
5	2016-01-11 00:54:13.579000000	57.467295
8	2016-01-03 21:27:04.266000000	37.955847

	assignment6_submission
0	2015-11-22 18:31:15.934000000
1	2015-12-21 17:07:24.275000000
4	2016-01-05 01:06:59.546000000
5	2016-01-20 19:54:46.166000000
8	2016-01-19 15:24:31.060000000

```
[11]: late_finishers = df[pd.to_datetime(df['assignment1_submission']) >= '2016']
late_finishers.head()
```

	student_id	assignment1_grade \
2	D0F62040-CEB0-904C-F563-2F8620916C4E	85.512541
3	FFDF2B2C-F514-EF7F-6538-A6A53518E9DC	86.030665
6	3217BE3F-E4B0-C3B6-9F64-462456819CE4	87.498744
7	F1CB5AA1-B3DE-5460-FAFF-BE951FD38B5F	80.576090
9	E2C617C2-4654-622C-AB50-1550C4BE42A0	59.270882

	assignment1_submission	assignment2_grade \
2	2016-01-09 05:36:02.389000000	85.512541
3	2016-04-30 06:50:39.801000000	68.824532
6	2016-03-05 11:05:25.408000000	69.998995
7	2016-01-24 18:24:25.619000000	72.518481
9	2016-03-06 12:06:26.185000000	59.270882

	assignment2_submission	assignment3_grade \
2	2016-01-09 06:39:44.416000000	68.410033

3	2016-04-30 17:20:38.727000000	61.942079
6	2016-03-09 07:29:52.405000000	55.999196
7	2016-01-27 13:37:12.943000000	65.266633
9	2016-03-13 02:07:25.289000000	53.343794

	assignment3_submission	assignment4_grade \
2	2016-01-15 20:22:45.882000000	54.728026
3	2016-05-12 07:47:16.326000000	49.553663
6	2016-03-16 22:31:24.316000000	50.399276
7	2016-01-30 14:34:36.581000000	65.266633
9	2016-03-17 07:30:09.241000000	53.343794

	assignment4_submission	assignment5_grade \
2	2016-01-11 12:41:50.749000000	49.255224
3	2016-05-07 16:09:20.485000000	49.553663
6	2016-03-18 07:19:26.032000000	45.359349
7	2016-02-03 22:08:49.002000000	65.266633
9	2016-03-20 21:45:56.229000000	42.675035

	assignment5_submission	assignment6_grade \
2	2016-01-11 17:31:12.489000000	44.329701
3	2016-05-24 12:51:18.016000000	44.598297
6	2016-03-19 10:35:41.869000000	45.359349
7	2016-02-16 14:22:23.664000000	65.266633
9	2016-03-27 15:55:04.414000000	38.407532

	assignment6_submission
2	2016-01-17 16:24:42.765000000
3	2016-05-26 08:09:12.058000000
6	2016-03-23 14:02:00.987000000
7	2016-02-18 08:35:04.796000000
9	2016-03-30 20:33:13.554000000

[12]: *#otra manera de hacer late_finishers:*

```
late_finishers = df[~df.index.isin(early_finishers.index)] #le estoy diciendo
→que quiero todo lo que no esta en early
late_finishers.head()
```

	student_id	assignment1_grade \
2	D0F62040-CEB0-904C-F563-2F8620916C4E	85.512541
3	FFDF2B2C-F514-EF7F-6538-A6A53518E9DC	86.030665
6	3217BE3F-E4B0-C3B6-9F64-462456819CE4	87.498744
7	F1CB5AA1-B3DE-5460-FAFF-BE951FD38B5F	80.576090
9	E2C617C2-4654-622C-AB50-1550C4BE42A0	59.270882

	assignment1_submission	assignment2_grade \
2	2016-01-09 05:36:02.389000000	85.512541

3	2016-04-30 06:50:39.801000000	68.824532
6	2016-03-05 11:05:25.408000000	69.998995
7	2016-01-24 18:24:25.619000000	72.518481
9	2016-03-06 12:06:26.185000000	59.270882

	assignment2_submission	assignment3_grade \
2	2016-01-09 06:39:44.416000000	68.410033
3	2016-04-30 17:20:38.727000000	61.942079
6	2016-03-09 07:29:52.405000000	55.999196
7	2016-01-27 13:37:12.943000000	65.266633
9	2016-03-13 02:07:25.289000000	53.343794

	assignment3_submission	assignment4_grade \
2	2016-01-15 20:22:45.882000000	54.728026
3	2016-05-12 07:47:16.326000000	49.553663
6	2016-03-16 22:31:24.316000000	50.399276
7	2016-01-30 14:34:36.581000000	65.266633
9	2016-03-17 07:30:09.241000000	53.343794

	assignment4_submission	assignment5_grade \
2	2016-01-11 12:41:50.749000000	49.255224
3	2016-05-07 16:09:20.485000000	49.553663
6	2016-03-18 07:19:26.032000000	45.359349
7	2016-02-03 22:08:49.002000000	65.266633
9	2016-03-20 21:45:56.229000000	42.675035

	assignment5_submission	assignment6_grade \
2	2016-01-11 17:31:12.489000000	44.329701
3	2016-05-24 12:51:18.016000000	44.598297
6	2016-03-19 10:35:41.869000000	45.359349
7	2016-02-16 14:22:23.664000000	65.266633
9	2016-03-27 15:55:04.414000000	38.407532

	assignment6_submission
2	2016-01-17 16:24:42.765000000
3	2016-05-26 08:09:12.058000000
6	2016-03-23 14:02:00.987000000
7	2016-02-18 08:35:04.796000000
9	2016-03-30 20:33:13.554000000

[13]: *#vamos a calcular la media para los grupos formados:*

```
print(early_finishers['assignment1_grade'].mean())
print(late_finishers['assignment1_grade'].mean())
```

74.94728457024303

74.0450648477065

```
[14]: #vamos a utilizar la libreria SciPy para hacer test de hipotesis
#alfa es el nivel de significacion o de confianza (el error que estoy dispuesto
    ↳a aceptar), en este caso adoptamos 0.05
#p-valor: es la probabilidad de que la hipotesis nula sea verdadera
#hipotesis nula: las notas son iguales
#hipotesis alternativa: las notas son distintas

from scipy.stats import ttest_ind

ttest_ind(early_finishers['assignment1_grade'],
    ↳late_finishers['assignment1_grade'])
```

```
[14]: Ttest_indResult(statistic=1.322354085372139, pvalue=0.1861810110171455)
```

```
[15]: #podemos ver que p-valor es 0.18 (mayor que alfa). Por lo tanto esto significa
    ↳que no podemos rechazar la hipotesis nula,
#es decir que segun la evidencia que tenemos las notas son iguales.
#Ahora vamos a chequear con otras materias:
```

```
[16]: print(ttest_ind(early_finishers['assignment2_grade'],
    ↳late_finishers['assignment2_grade']))
print(ttest_ind(early_finishers['assignment3_grade'],
    ↳late_finishers['assignment3_grade']))
print(ttest_ind(early_finishers['assignment4_grade'],
    ↳late_finishers['assignment4_grade']))
print(ttest_ind(early_finishers['assignment5_grade'],
    ↳late_finishers['assignment5_grade']))
print(ttest_ind(early_finishers['assignment6_grade'],
    ↳late_finishers['assignment6_grade']))
```

```
Ttest_indResult(statistic=1.2514717608216366, pvalue=0.2108889627004424)
Ttest_indResult(statistic=1.6133726558705392, pvalue=0.10679998102227865)
Ttest_indResult(statistic=0.049671157386456125, pvalue=0.960388729789337)
Ttest_indResult(statistic=-0.05279315545404755, pvalue=0.9579012739746492)
Ttest_indResult(statistic=-0.11609743352612056, pvalue=0.9075854011989656)
```

```
[17]: #en esta data no hay evidencia para decir que las notas son diferentes. Como
    ↳analista, hay que investigar más sobre la materia3
# ya que tiene un p-valor de 0.1. Esto puede ser por pocos participantes o por
    ↳algo unico del curso.
```

```
[18]: #al ser insuficientes, se estan reemplazando los test de hipotesis por
    ↳intervalos de confianza o analisis de bayes.
#vamos a ver una simulacion de esto:

df1 = pd.DataFrame([np.random.random(100) for x in range (100)]) #creo df1
df1.head()
```

```
[18]:
```

	0	1	2	3	4	5	6	\
0	0.235315	0.865727	0.229445	0.505576	0.497508	0.856427	0.632166	
1	0.323391	0.519747	0.305150	0.399421	0.106721	0.017815	0.657897	
2	0.093247	0.962076	0.064608	0.570445	0.342470	0.012256	0.235565	
3	0.265882	0.611311	0.023636	0.234934	0.843192	0.720463	0.312731	
4	0.651452	0.872515	0.615094	0.894787	0.825892	0.858519	0.833000	

	7	8	9	...	90	91	92	93	\
0	0.397486	0.369785	0.846677	...	0.235239	0.402434	0.997620	0.091872	
1	0.904126	0.521485	0.738789	...	0.713988	0.630668	0.172573	0.541846	
2	0.924496	0.673888	0.566631	...	0.487879	0.746764	0.025636	0.683120	
3	0.283801	0.871774	0.594185	...	0.445531	0.589198	0.708959	0.563728	
4	0.980519	0.225891	0.799765	...	0.930239	0.169294	0.428371	0.693672	

	94	95	96	97	98	99
0	0.007479	0.073533	0.813997	0.174193	0.193754	0.645956
1	0.051981	0.230362	0.062533	0.037113	0.817699	0.426128
2	0.090312	0.188465	0.429843	0.639126	0.883251	0.853001
3	0.911034	0.966791	0.506432	0.575089	0.629220	0.637827
4	0.095203	0.105936	0.621499	0.028379	0.601181	0.660641

[5 rows x 100 columns]

```
[19]: df2 = pd.DataFrame([np.random.random(100) for x in range (100)]) #creo df2
```

```
[20]: #son estos dos df iguales? vamos a crear una funcion para comparar columna por
      ↪columna entre ambos df
```

```
[21]: def test_columns(alpha=0.1):
      num_dif = 0 #contador
      for col in df1.columns: #recorro las columnas del df1
          teststat, pval = ttest_ind(df1[col], df2[col]) #hago test
          if pval <= alpha: #si el p-valor es menor que alfa entonces se debe
      ↪aceptar la hipotesis alternativa
              print('La columna {} es estadisticamente diferente para alfa={} y
      ↪p-valor={}'.format(col, alpha, pval))
              num_dif = num_dif + 1 #aumento contador
              print('El total de columnas diferentes es {}, que es un {}%.'.
      ↪format(num_dif, float(num_dif)/len(df1.columns)*100))

      test_columns()
```

La columna 4 es estadisticamente diferente para alfa=0.1 y
p-valor=0.040804910180119724
La columna 19 es estadisticamente diferente para alfa=0.1 y
p-valor=0.06003893875945721
La columna 35 es estadisticamente diferente para alfa=0.1 y
p-valor=0.03627638663866923
La columna 62 es estadisticamente diferente para alfa=0.1 y

p-valor=0.010588283504909046
 La columna 72 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.05571004246427579
 La columna 78 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.07485714786143179
 El total de columnas diferentes es 6, que es un 6.0%.

[22]: `test_columns(0.05)`

La columna 4 es estadísticamente diferente para $\alpha=0.05$ y
 p-valor=0.040804910180119724
 La columna 35 es estadísticamente diferente para $\alpha=0.05$ y
 p-valor=0.03627638663866923
 La columna 62 es estadísticamente diferente para $\alpha=0.05$ y
 p-valor=0.010588283504909046
 El total de columnas diferentes es 3, que es un 3.0%.

[23]: *#siempre se debe tener en cuenta el valor de alfa para el analisis, y dependerá
 →de cada caso cual tomar.*

[24]: *#ahora vamos a crear otro df, pero usando una distribucion no normal. Vamos a
 →utilizar chi cuadrado:*

```
df2 = pd.DataFrame([np.random.chisquare(df=1, size=100) for x in range (100)])
test_columns()
```

La columna 0 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0003243387492409425
 La columna 1 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=8.201957372826176e-05
 La columna 2 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00453097205611213
 La columna 3 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.03013242702312555
 La columna 4 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0007204861204224606
 La columna 5 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00019034534251848143
 La columna 6 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=8.501900425068969e-06
 La columna 7 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0014066919836154394
 La columna 8 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.000486299132426987
 La columna 9 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00010679659312160572
 La columna 10 es estadísticamente diferente para $\alpha=0.1$ y

p-valor=0.0005813621013618975
 La columna 11 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.025699296596404037
 La columna 12 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0021589790894650383
 La columna 13 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00033996173869926415
 La columna 14 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=2.3651042177707823e-05
 La columna 15 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0005743521143271199
 La columna 16 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00014161713517214913
 La columna 17 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0045923637591393784
 La columna 18 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.008209581676431574
 La columna 19 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00032025479347360026
 La columna 20 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.048648507209462934
 La columna 21 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00011343803634162933
 La columna 22 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0001835667842267419
 La columna 23 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0020501734816585886
 La columna 24 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00013698815655173487
 La columna 25 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=5.260470223873605e-06
 La columna 26 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.004235803799653742
 La columna 27 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0013575364673126995
 La columna 28 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0011909237474932092
 La columna 29 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0004361615673183277
 La columna 30 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0003502413208350381
 La columna 31 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.03109614205086666
 La columna 32 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0020771953639266813
 La columna 33 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.034844844772339324
 La columna 34 es estadísticamente diferente para $\alpha=0.1$ y

p-valor=0.0002509438760305707
 La columna 35 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.07814300233559522
 La columna 36 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0003771586520790147
 La columna 37 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=7.40451908063857e-05
 La columna 38 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00018708254264869593
 La columna 39 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=8.930969519708947e-07
 La columna 40 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0020769886737103875
 La columna 41 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00036080011926563094
 La columna 42 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.000504201897056923
 La columna 43 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.019247876177767378
 La columna 44 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.031791225221398246
 La columna 45 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=1.2330370549380151e-05
 La columna 46 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=3.7980123924798536e-05
 La columna 47 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.02337164000226436
 La columna 48 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.012406954746116914
 La columna 49 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0002620552960631748
 La columna 50 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=2.0951261364337596e-05
 La columna 51 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0014030585213874933
 La columna 52 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=6.810029131001573e-05
 La columna 53 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.003178644329407479
 La columna 54 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=1.0210545400974304e-05
 La columna 55 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0003017841068042615
 La columna 56 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=6.120157079862686e-05
 La columna 57 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0003628287630449215
 La columna 58 es estadísticamente diferente para $\alpha=0.1$ y

p-valor=0.007604574572618228
 La columna 59 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=6.869014635256504e-05
 La columna 60 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.013895280061426006
 La columna 61 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=2.5259005082928402e-05
 La columna 62 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0001786718608694777
 La columna 63 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00020105900419390652
 La columna 64 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0004687876258559973
 La columna 65 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.002424021424328542
 La columna 66 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0012088948397640752
 La columna 67 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.03537260792619573
 La columna 68 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0001780488913016485
 La columna 69 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00043786111950929385
 La columna 70 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0006951658225582343
 La columna 71 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00022242432786259854
 La columna 72 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.015436996672373282
 La columna 73 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=9.443272833664962e-05
 La columna 74 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0026950471480925407
 La columna 75 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.001460161770837082
 La columna 76 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0012725509146478463
 La columna 77 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0012739139564798291
 La columna 78 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=1.380693032583835e-05
 La columna 79 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.010431088673392415
 La columna 80 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.03574845432397989
 La columna 81 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00038276775130000463
 La columna 82 es estadísticamente diferente para $\alpha=0.1$ y

p-valor=0.019170810090286667
 La columna 83 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0021952181884918793
 La columna 84 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.000743453878703937
 La columna 85 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.04608568391937338
 La columna 86 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.041043552168151164
 La columna 87 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0049721201759872075
 La columna 88 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=4.746335936506911e-05
 La columna 89 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0005323924800928607
 La columna 90 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0018935157079050025
 La columna 91 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=9.150521436767209e-05
 La columna 92 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=5.8021138600771755e-06
 La columna 93 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.00047755023490620867
 La columna 94 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0018233598896186581
 La columna 95 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0038655092583066973
 La columna 96 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.0002534647027815455
 La columna 97 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=1.715842181273524e-05
 La columna 98 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=0.004026280327700301
 La columna 99 es estadísticamente diferente para $\alpha=0.1$ y
 p-valor=9.382676531536942e-05
 El total de columnas diferentes es 100, que es un 100.0%.

[25]: *#ahora vemos que todas las columnas son estadísticamente diferentes, con un*
→ nivel de confianza de 0.1

2 Other Forms of Structured Data

Video explicativo de distintas maneras de almacenar y ver datos.