

Trabajo practico - ESTADISTICA APLICADA I

Franco Gentile - Pablo Zanetti

- **Docente a cargo: Pedro Cosatto**
- **1º Cuatrimestre 2022**
- Franco Gentile - PADRÓN : 105226
- Pablo Zanetti - PADRÓN : 103006



- **1. Simule 1000 valores de una variable aleatoria normal con los parámetros asignados a su grupo (variable 1) y simule además 1000 valores de la variable aleatoria 2 asignada a su grupo. Asuma que estas dos variables simuladas representan a las dos poblaciones que estudia.**

Para realizar este enunciado hicimos uso de las librerías proporcionadas por el software Rstudio

- **2. Explique cómo tomaría una muestra aleatoria de estas poblaciones en una situación en la que no tuviese acceso a todos los datos digitalizados.**

Si no tuvieramos acceso a los datos digitalizados, una manera prudente sería utilizar el metodo de *Encuestas por muestreo* la idea principal es definir una población finita de N unidades, subpoblaciones de un tamaño fijado de antemano, donde el número de muestras de tamaño n viene dado por:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = C_{N,n}$$

Este es un metodo de seleccion de n unidades sacadas de N , de tal manera que cada una de las muestras tiene la misma probabilidad de ser elegida.

Posteriormente, una vez identificada nuestra población, lo que haremos será dividirla en grupos de 6.

Tomaremos un dado, lo arrojaremos y el lado que quede boca arriba será el que defina a que grupo tendremos que ir a extraer un dato. Este proceso se repetirá dependiendo de la cantidad de datos que deseemos que tenga cada muestra. Lo ideal será que el tamaño de nuestra población finita sea multiplo de 3, para que el tamaño de cada grupo sea igual.

Por ejemplo: si tenemos una población que contiene 12 unidades y queremos extraer muestras de tamaño 3. Primero vemos que podemos tener 220 combinaciones posibles de manera que los datos no se repitan, no le damos importancia al orden de los elementos. Luego separamos a las 12 unidades entre 6, de manera que cada grupo tenga 2 unidades. Posteriormente, arrojaremos el dado 3 veces, si sale 1, vamos al grupo 1, donde tendremos dos datos, para elegir el dato buscado, optamos por un sistema de bolillas, en donde cada bolilla representa un dato. Como no hay reposición de los datos, estos no se pueden repetir, de manera que también puede suceder que tengamos que ir a buscar un dato al grupo 1 y no tengamos mas información para sustraer, en estos casos, debemos volver a arrojar el dado.

- 3. Utilizando la aplicación estadística de su preferencia obtenga una muestra de tamaño 50 en cada población y obtenga / calcule:
- 3.1. Las principales métricas de tendencia central y de dispersión, así como los coeficientes de asimetría y curtosis.

Para este ítem, decidimos directamente, dejar en evidencia los resultados obtenidos por las métricas de las distribuciones realizadas 50 simulaciones y 1000 simulaciones

Comparación normal con 50 ensayos y normal con 1000 ensayos, respectivamente

Coeficiente de asimetría (*Skewness*)

```
## [1] 0.07542979 -0.04229538
```

Coeficiente de curtosis (α_4)

```
## [1] 2.991476 2.821904
```

Comparación gamma con 50 ensayos y gamma con 1000 ensayos, respectivamente

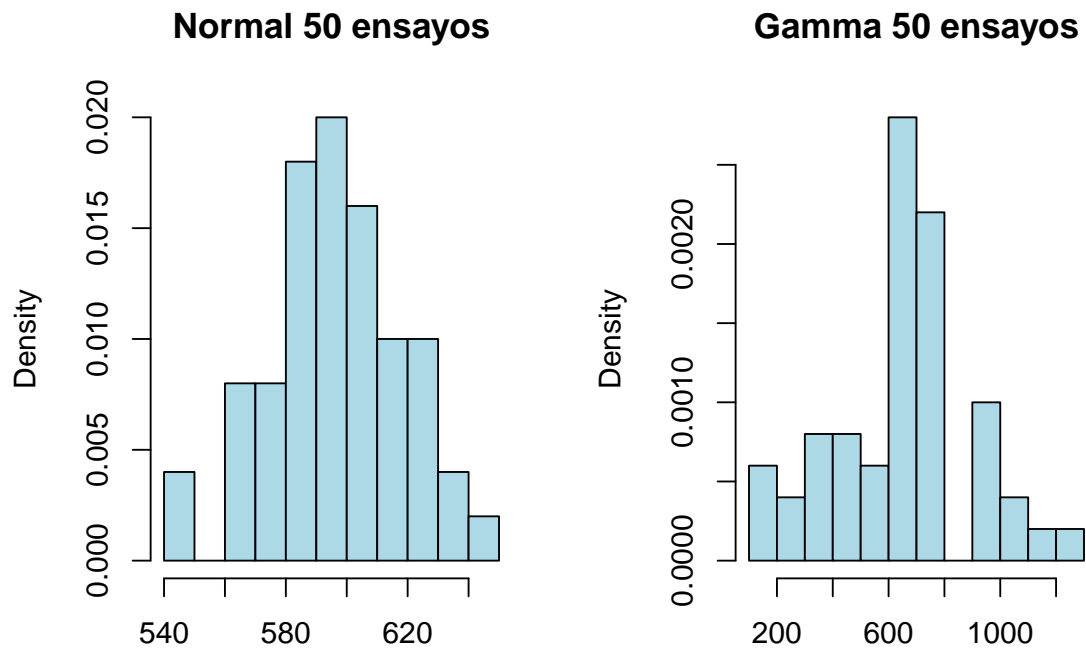
Coeficiente de asimetría (*Skewness*)

```
## [1] 0.62570031 0.08964726
```

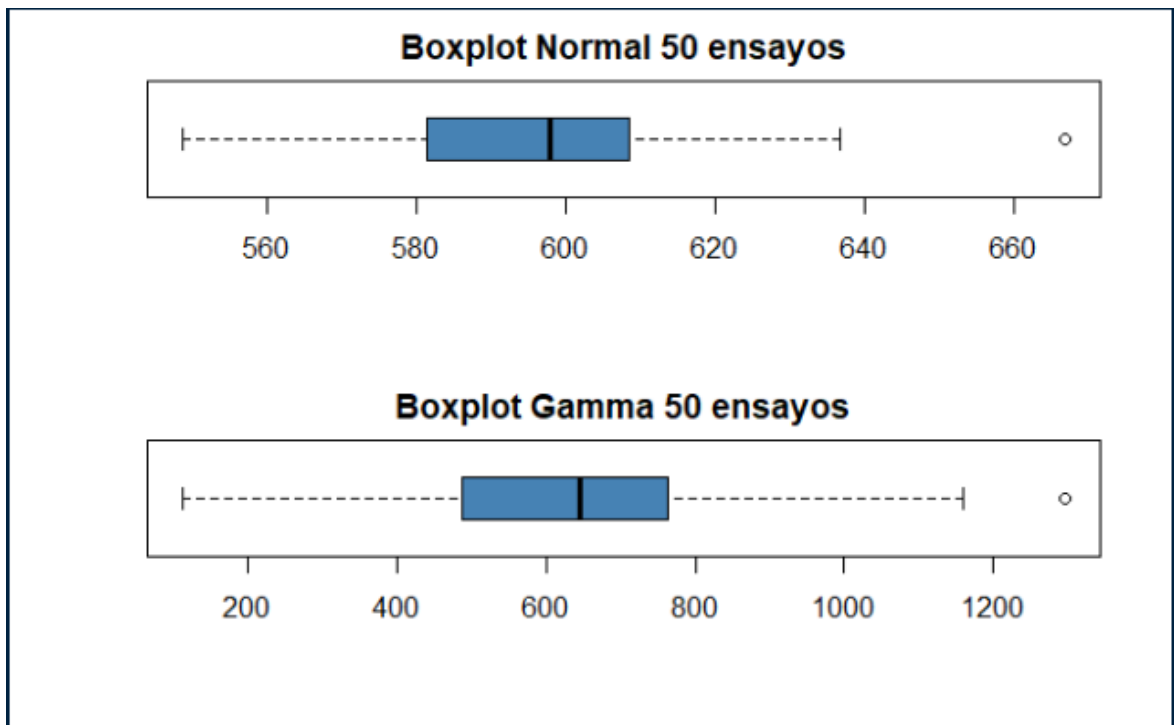
– Coeficiente de curtosis (α_4)

```
## [1] 3.031486 2.928466
```

- 3.2. El histograma de densidad de frecuencias



- 3.3. El box plot, identifique los outliers



Se puede observar que para cada simulación obtuvimos un outlier, son los puntos blancos situados aproximadamente a 670 y 1300 respectivamente.

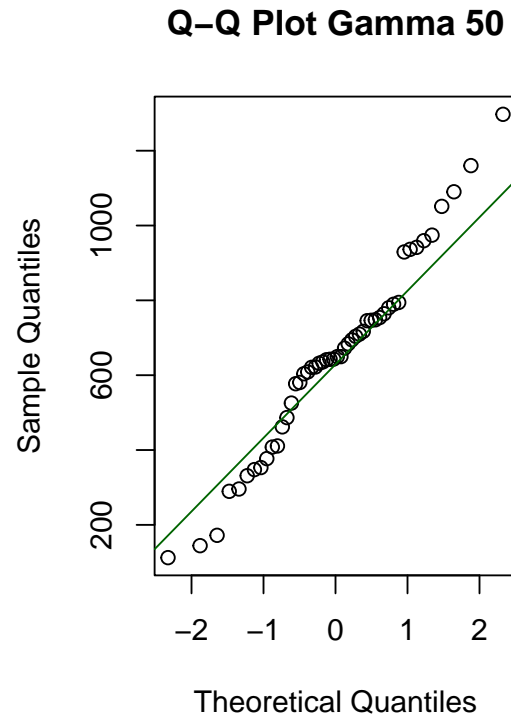
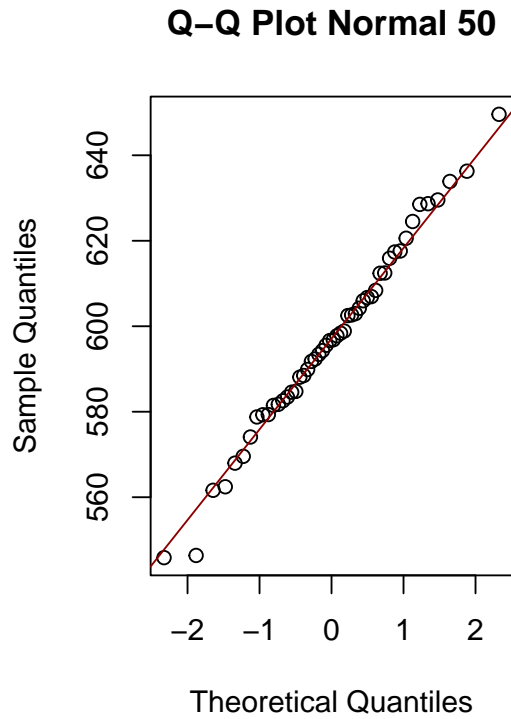
- **3.4. Explique en una situación real cómo trataría a los outliers. En una situación real, primero buscaría la manera de encontrar los outliers, y esto podría realizarse con el método del boxplot, para así poder reconocer con mayor facilidad los valores atípicos.**

Una vez reconocidos estos valores, intentaríamos analizarlos y categorizarlos para así luego poder tomar una decisión a partir de la evaluación del riesgo a los que estos someten a nuestro estudio.

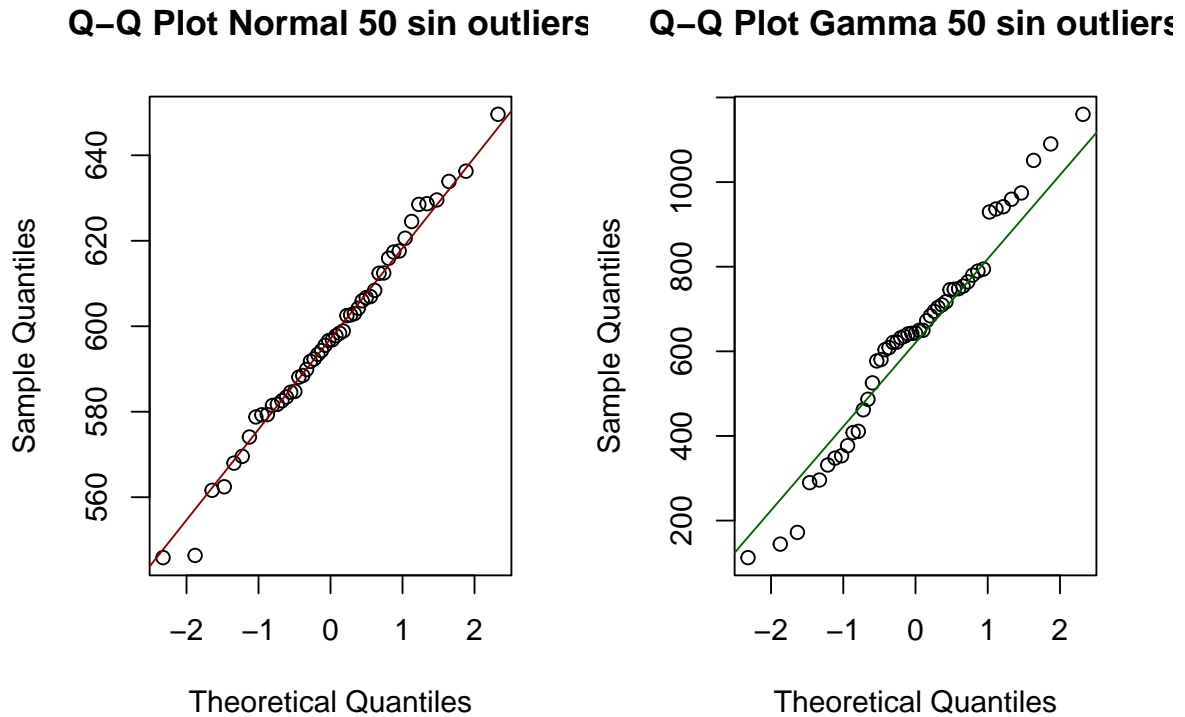
Consideramos que estos casos atípicos se pueden clasificar en algunas categorías, como por ejemplo aquellos que surgen de un error de procedimiento, que pueden ser eliminados del análisis. También se deben tener en cuenta aquellas observaciones que ocurren como consecuencia de algún acontecimiento extraordinario, donde se optará por reconocer que no representa ningún riesgo y también puede ser eliminado. Cuando se trate de un valor cuyo origen no es el de un error de procedimiento o de un caso atípico, se debería analizar aún más en profundidad la influencia del mismo. Una manera de hacerlo es quitarlo del análisis, rehacerlo, y observar cómo se desarrolla luego el análisis. Si este supone un cambio drástico, se deberá tener en cuenta y trabajar sobre ese mismo para poder solucionarlo, pero si el cambio no es notable (por ejemplo, el p-valor sigue siendo significativo), podremos quitarlo con las observaciones y justificaciones adecuadas.

Otra forma de tratar los outliers sería una transformación que minimice el efecto distorsionador, como por ejemplo una transformación logarítmica o una raíz cuadrada que reduzca la distancia entre las observaciones. Otra alternativa sería la de elegir un modelo estadístico con supuestos diferentes.

- 3.5. Obtenga el Q-Q plot de los datos contra una distribución normal.



- 3.6. Repita el punto anterior, pero eliminando los outliers.



- 3.7. A partir de todas las técnicas gráficas y las métricas calculadas verifique la consistencia entre las distribuciones de las poblaciones que se le asignaron y las conclusiones que podrían extraerse a partir de las técnicas de estadística descriptiva aplicadas a estos datos.

Iniciamos calculando el sesgo de las variables para así luego compararlas por medio de la siguiente fórmula correspondiente al momento estandarizado de tercer orden:

$\alpha_3 = E(x - \mu)^3 / (Var(x))^{3/2}$ En cuanto a las variables normales, el valor de α_3 , por definición, debería tender a cero y ser simétrica, es decir, no tener ningún tipo de asimetría. Podemos observar en la simulación correspondiente a las 1000 muestras que el valor de α_3 es aproximadamente **-0.14**, y el de las 50 muestras es de **0.16**, lo cual tiene sentido dado que estamos pasando de trabajar con 1000 muestras a 50, por ende concluimos que a medida que aumentamos la cantidad de repeticiones de la variable normal, α_3 tiende a ser cada vez más parecido a 0 y no toma asimetría de ningún tipo.

La distribución de las variables Gamma tiene asimetría positiva, donde el parámetro α es el que le da la forma y a medida que aumenta atenúa aún más la asimetría mencionada, por ende, el momento estandarizado de tercer orden va a ser mayor a cero. Al realizar las simulaciones, obtuvimos que α_3 es **1.07** en el caso de las 1000 muestras, y **3** para las 50 el resultado fue de **0.77**, por lo que concluimos que a medida que aumentemos las repeticiones, se atenuará más la asimetría positiva de la distribución de la variable.

Tras haber hecho el análisis correspondiente al sesgo, analizamos el caso de la curtosis, donde

nuestro coeficiente será α_4 , es decir, el momento centrado adimensional de cuarto 4 orden:

$$\alpha_4 = E(x - \mu)^4 / (Var(x))^2$$

Entonces α_4 es una medida del grado de apuntamiento de la función de densidad, por ende se 4 considera que si este es mayor a 3 la campana sería leptocúrtica, mientras que si es menor a 3, sería platicúrtica. Iniciamos comparando las dos poblaciones normales y en ambos casos notamos que el valor de α_4 fue siempre mayor a 3 y aumentó cuando tomamos las 50 muestras, por lo que 4 consideramos que si aumentamos el número de repeticiones, cada vez nos acercaremos más a una campana más aplanada.

Al comparar las dos poblaciones de la gamma, observamos que sucedió lo contrario: cuando pasamos de 1000 muestras a 50 el valor de α_4 disminuyó considerablemente, por 4 ende concluimos que a medida que aumentemos las repeticiones, se notará cada vez más la campana leptocúrtica.

- 4. A partir de los datos de la población 1, tome 100 muestras aleatorias diferentes de tamaño 5. Para cada muestra calcule:

A traves de el software utilizado, mediante la función sample, que se encarga de tomar datos aleatoriamente de listas, lo que hicimos fue encargarle que tome 5 valores proporcionados por las distribucion normal que generamos en un principio, y esos valores, que conformarían los datos de una muestra, los guardamos en una matriz, cuyos datos los hicimos visibles en un excel.

En la imagen también se ven los datos que recibimos para resolver el ítem 5.

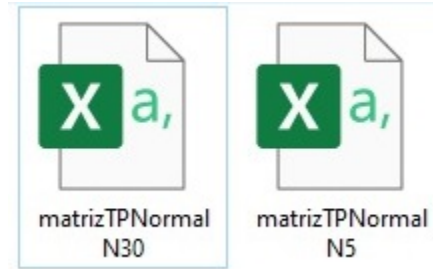


Figure 1: Datos almacenados en excel

- 4.1. El intervalo de confianza del 95% para la media poblacional asumiendo que conoce la varianza. Calcule empíricamente el nivel de confianza.

Nosotros recibimos valores tales que conformamos 100 muestras, a continuación mostramos los resultados obtenidos en solo 20 de ellas.

Datos obtenidos a partir de 1000 simulaciones de la distribución normal, 5 datos : 1 muestra

Normal N=5		IC y Error con sd conocida				IC y Error con sd desconocida			
Numero de n	Media Muestral	LI	LS	Error	Sobs	LI	LS	Error relativo	
1	600.41249249184	578.49942897743	622.32555600625	21.913063514414	15.044720671393	583.11709446760	617.70789051608	17.2953980242353	
2	588.22861910777	566.31555559336	610.14168262219	21.913063514414	37.034992107590	545.65322383685	630.80401437869	42.5753952709227	
3	593.10631870952	571.19325519511	615.01938222394	21.913063514414	20.132616753165	569.96187969860	616.25075772044	23.1444390109197	
4	611.14430274309	589.23123922868	633.05736625751	21.913063514414	14.330691030527	594.66975258367	627.61885290251	16.4745501594185	
5	608.93341661197	587.02035309756	630.84648012638	21.913063514414	10.295908911869	597.09724852740	620.76958469654	11.836168084572	
6	615.59521605127	593.68215253685	637.50827956568	21.913063514414	14.438665247112	598.99653882476	632.19389327778	16.5986772265134	
7	609.39538590144	587.48232238703	631.30844941585	21.913063514414	10.599611833745	597.21008119747	621.58069060541	12.1853047039691	
8	616.89352986114	594.98046634672	638.80659337555	21.913063514414	30.755100677691	581.53749241024	652.24956731204	35.3560374509	
9	589.13052945685	567.21746594244	611.04359297127	21.913063514414	14.806297503428	572.10922250272	606.15183641099	17.0213069541374	
10	611.58986140097	589.67679788656	633.50292491539	21.913063514414	24.359490789901	583.58621161038	639.59351119157	28.0036497905958	
11	612.93729110587	591.02422759145	634.85035462028	21.913063514414	30.155971775831	578.27001172843	647.60457048330	34.6672793774314	
12	583.19171905961	561.27865554520	605.10478257403	21.913063514414	17.485628835931	563.09025509581	603.29318302342	20.1014639638022	
13	609.42984532861	587.51678181419	631.34290884302	21.913063514414	23.220037603610	582.73610994702	636.12358071020	26.6937353815929	
14	611.71061921339	589.79755569897	633.62368272780	21.913063514414	16.338635989713	592.9277372452	630.49350118158	18.7828819681909	
15	596.70840394811	574.79534043370	618.62146746253	21.913063514414	26.261591222599	566.51810112592	626.89870677030	30.1903028221899	
16	599.53041640286	577.61735288845	621.44347991728	21.913063514414	25.870555854946	569.78964750508	629.27118530064	29.7407688977762	
17	594.12845951802	572.21539600360	616.04152303243	21.913063514414	30.505417776683	559.05945731670	629.19746171933	35.0690022013188	
18	599.87352167727	577.96045816285	621.78658519168	21.913063514414	14.536213373827	583.16270320767	616.58434014687	16.7108184696039	
19	606.22031776517	584.30725425076	628.13338127959	21.913063514414	24.316443966171	578.26615456634	634.17448096400	27.9541631988308	
20	608.92065012890	587.00758661448	630.83371364331	21.913063514414	10.457472831675	596.89874830021	620.94255195758	12.0219018286829	

Figure 2: 20 muestras aleatorias de tamaño 5

Comparamos

Calculo empirico del intervalo de confianza para varianza conocida

```
## [1] 595.5292 598.6281
```

- **4.2. El intervalo de confianza del 95% para la media poblacional asumiendo que se desconoce la varianza. Calcule empíricamente el nivel de confianza y la longitud promedio de los intervalos y su error relativo promedio.**

Los datos obtenidos sobre las muestras pueden visualizarse en la Figura

Calculo empirico del intervalo de confianza para varianza desconocida

```
## [1] 595.2641 598.8932
```

Calculo empirico del error relativo

```
## [1] 1.814511
```

- **4.3. Compare los resultados obtenidos en 4.1 y 4.2**

Se puede observar que el Error para el caso de el desvío conocido y el Error relativo para el caso de la varianza desconocida es mucho mayor al Error obtenido en el caso empirico, esto es debido a que el tamaño de nuestras muestras es muy pequeño, por lo tanto el sesgo es muy grande.

También vemos que el desvío es mucho mayor al empirico, el motivo es el mismo que mencionamos anteriormente

- 5. A partir de los datos de la población 1, tome 100 muestras aleatorias diferentes de

tamaño 30. Para cada muestra repita los cálculos y análisis efectuados en los puntos 4.1, 4.2 y 4.3.

Datos obtenidos a partir de 1000 simulaciones de la distribución normal, 30 datos : 1 muestra

Normal N=30		IC y Error con sd conocida			Sobs	IC y Error con sd desconocida		
Numero de	Media muestral	LI	LS	Error		LI	LS	Error relativo
1	591.377832751703	582.431862033117	600.323803470288	8.94597071858578	24.9035762276563	582.092130146208	600.663535357198	11.6877933574232
2	596.591007789675	587.645037071089	605.536978508261	8.94597071858578	19.2596991225104	589.409716368392	603.772299210959	9.03899831141774
3	600.276153896482	591.330183177897	609.222124615068	8.94597071858578	24.0373058027572	591.313454279314	609.238853513651	11.2812336880283
4	600.691729681731	591.745758963146	609.637700400317	8.94597071858578	24.6229349385793	591.510668735947	609.872790627516	11.5560822584105
5	600.254385185765	591.308414467179	609.20035590435	8.94597071858578	24.3864916006836	591.161485976218	609.347284395312	11.445114225193
6	608.500516342668	599.554545624082	617.446487061253	8.94597071858578	26.8747941233318	598.479813150006	618.521219535329	12.6129290574728
7	598.581926372948	589.635955654363	607.527897091534	8.94597071858578	27.5332447011927	588.315709195045	608.848143550852	12.9219543243566
8	595.278722303539	586.332751584953	604.224693022124	8.94597071858578	21.1296259559989	587.400198311677	603.157246295401	9.91659589915064
9	602.350494943697	593.404524225112	611.296465662283	8.94597071858578	22.4089122582087	593.994968290009	610.706021597386	10.5169929589354
10	601.616370188804	592.670399470218	610.56234090739	8.94597071858578	20.5548888217478	593.952146265985	609.280594111624	9.64685918348549
11	599.396648305868	590.450677587282	608.342619024454	8.94597071858578	27.3457204271421	589.200352597244	609.592944014492	12.8339450784328
12	599.595951472217	590.649980753631	608.541922190802	8.94597071858578	30.6734656001743	588.15885196332	611.033050981113	14.3957287183812
13	608.078433149209	599.132462430624	617.024403867795	8.94597071858578	25.4164838736266	598.60148460293	617.555381695489	11.9285121410526
14	606.991145049152	598.045174330566	615.937115767738	8.94597071858578	23.4112498862975	598.261880541926	615.720409556378	10.9874119447219
15	597.307743930308	588.361773211722	606.253714648894	8.94597071858578	22.0951548005166	589.069206837305	605.546281023311	10.3697397172189
16	597.779492555321	588.833521836735	606.725463273906	8.94597071858578	23.592403262955	588.982682071314	606.576303039327	11.0724311890673
17	603.88203732682	594.936066608235	612.828008045406	8.94597071858578	24.0302911575147	594.921953233985	612.842121419656	11.2779415614951
18	598.21061208117	589.264641362584	607.156582799756	8.94597071858578	29.784719930003	587.104895820594	609.316328341745	13.9786209244921
19	602.140559096933	593.194588378347	611.086529815519	8.94597071858578	22.0217157741774	593.929404936887	610.351713256979	10.3352731748888
20	598.927208174616	589.98123745603	607.873178893201	8.94597071858578	22.1358020365335	590.67351509992	607.180901249311	10.3888163546775

Figure 3: a partir de 1000 simulaciones de Datos de 20 muestras aleatorias de tamaño 30

Recordando que

Calculo empirico del intervalo de confianza para varianza conocida

```
## [1] 595.5292 598.6281
```

Calculo empirico del intervalo de confianza para varianza desconocida

```
## [1] 595.2641 598.8932
```

Calculo empirico del error relativo

```
## [1] 1.814511
```

- **6. A partir de los análisis efectuados en 4.3 y 4.4 ¿Qué puede concluir?**

Puedo concluir que el Error relativo disminuyo considerablemente, y el desvío esta mas acentuado en torno a los mismos valore, esto es debido a que aumentamos la cantidad de ensayos.

Por lo tanto la aproximación realizada, en donde 30 datos conforman una muestra, es mucho mas precisa, que la utilizada anteriormente

- 7. Repita los puntos 4 y 5 pero ahora para la media de la variable 2. Indique qué

supuestos teóricos está violando. Compare los niveles de confianza empíricos.

El error teórico lo realizamos al momento de calcular el intervalo de confianza, lo estamos haciendo de la misma manera que lo hicimos al calcularlo para la distribución normal.

El supuesto teorico que estamos violando es que para realizar una buena aproximación normal α debería ser >25 y en nuestro caso $\alpha = 6$ por lo que realizar lo dicho anteriormente es sumamente peligroso, a continuación, veremos la consecuencia de hacerlo.

Para comenzar, volvimos a generar datos utilizando la función sample, pero esta vez haciendo uso de la distribución gamma, que generamos al principio, que cuenta con 1000 simulaciones.

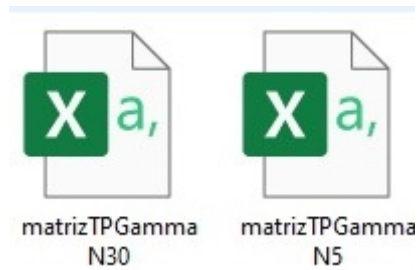


Figure 4: Datos almacenados en excel

Nosotros recibimos valores tales que conformamos 100 muestras, a continuación mostramos los resultados obtenidos en solo 20 de ellas.

Datos obtenidos a partir de 1000 simulaciones de la distribución gamma, 5 datos : 1 muestra

Gamma N=5		IC y Error con sd conocida				IC y Error con sd desconocida			
Numero de muestra	Media Muestral	LI	LS	Error	Sobs	LI	LS	Error relativo	
1	865.79078952709	608.09316259757	1123.4884164566	257.69762692951	407.18829183081	397.68747728592	1333.8941017682	468.10331224117	
2	731.04548177033	473.34785484081	988.74310869984	257.69762692951	359.12394356173	318.19690321589	1143.8940603247	412.848578554437	
3	610.15931720135	352.46169027184	867.85694413087	257.69762692951	326.27591230103	235.07280730709	985.24582709561	375.086509894261	
4	649.72951897470	392.03189204519	907.42714590422	257.69762692951	207.59767509569	411.07540913051	888.38362881890	238.654109844193	
5	503.37169071831	245.67406378880	761.06931764783	257.69762692951	98.078915480121	390.62025331612	616.12312812050	112.751437402191	
6	587.66356874066	329.96594181115	845.36119567018	257.69762692951	202.01586787957	355.42629970102	819.90083778031	232.237269039643	
7	870.43551496796	612.73788803844	1128.1331418974	257.69762692951	284.85447124454	542.96705830692	1197.903971629	327.468456661036	
8	795.64501437252	537.94738744301	1053.3426413020	257.69762692951	395.17965640280	341.34681915491	1249.9432095901	454.298195217611	
9	646.92679495963	389.22916803011	904.62442188914	257.69762692951	284.97098823639	319.32439046435	974.52919945490	327.602404495274	
10	834.06965825258	576.37203132306	1091.7672851821	257.69762692951	362.24251010172	417.63597826939	1250.5033382357	416.433679983189	
11	839.20470716842	581.50708023890	1096.9023340979	257.69762692951	349.85861451283	437.00754296898	1241.4018713678	402.197164199437	
12	924.40800859351	666.71038166400	1182.1056355230	257.69762692951	226.53408535556	663.98461770062	1184.8313994864	260.423390892892	
13	689.55352256739	431.85589563788	947.25114949691	257.69762692951	406.83844374511	221.85239538651	1157.2546497482	467.701127180882	
14	789.37210122455	531.67447429504	1047.0697281540	257.69762692951	247.06987964735	505.34077876696	1073.4034236821	284.03132245759	
15	795.39512284625	537.69749591673	1053.0927497757	257.69762692951	285.61667953628	467.05043218454	1123.7398135079	328.344690661709	
16	882.35210636887	624.65447943936	1140.0497332983	257.69762692951	224.30448182677	624.49186578691	1140.2123469508	257.86024058196	
17	604.36158804394	346.66396111442	862.05921497345	257.69762692951	192.61696337777	382.92929158577	825.79388450210	221.432296458165	
18	997.11904122314	739.42141429363	1254.8166681526	257.69762692951	568.73303467257	343.30403069278	1650.9340517535	653.815010530361	
19	803.63302472254	545.93539779303	1061.3306516520	257.69762692951	302.95882176241	455.35182218101	1151.9142272640	348.281202541535	
20	686.19219363768	428.49456670816	943.88982056719	257.69762692951	266.70942874943	379.58326231940	992.80112495595	306.608931318274	

Figure 5: 20 muestras aleatorias de tamaño 5

Datos obtenidos a partir de 1000 simulaciones de la distribución gamma, 30 datos : 1 muestra

Gamma N=30		IC y Error con sd conocida				IC y Error con sd desconocida		
Numero de muestra	Media muestral	LI	LS	Error	Sobs	LI	LS	Error relativo
1	764.12499545364	658.92037980308	869.32961110421	8.9459707185857	277.51029283041	660.65077851265	867.59921239464	130.241653950001
2	787.43621355934	682.23159790877	892.64082920991	8.9459707185857	340.59433310407	660.44008828762	914.43233883107	159.84837469283
3	848.29239999462	743.08778434405	953.49701564519	8.9459707185857	308.56435113325	733.23917337816	963.34562661108	144.816003153305
4	729.13406760664	623.92945195608	834.33868325721	8.9459707185857	331.90882999268	605.37647314408	852.89166206921	155.772077993783
5	769.97689797319	664.77228232262	875.18151362375	8.9459707185857	264.08850436471	671.50721271017	868.4465832362	123.942514876959
6	688.71738637260	583.51277072203	793.92200202317	8.9459707185857	358.29102929130	555.12276069305	822.31201205214	168.153821519208
7	811.66756239137	706.46294674080	916.87217804194	8.9459707185857	337.99239032891	685.64161371540	937.69351106734	158.627226002943
8	770.02040887652	664.81579322595	875.22502452709	8.9459707185857	327.07366283947	648.06568496645	891.9751327866	153.502828227455
9	700.49866710002	595.29405144945	805.70328275059	8.9459707185857	268.58900653793	600.35089655424	800.64643764580	126.054699043777
10	733.49306703839	628.28845138782	838.69768268896	8.9459707185857	299.52030965145	621.81205808191	845.17407599486	140.571436550141
11	829.96965136473	724.76503571416	935.17426701530	8.9459707185857	355.36422549956	697.46633198378	962.47297074568	166.780208444405
12	640.90439396617	535.69977831560	746.10900961674	8.9459707185857	263.97564380374	542.47679059516	739.33199733719	123.889546945656
13	721.67583606432	616.47122041375	826.88045171489	8.9459707185857	306.20906078008	607.50081768467	835.85085444398	143.710613843236
14	661.84051895236	556.63590330179	767.04513460293	8.9459707185857	281.67699453962	556.81267964951	766.86835825521	132.197178253583
15	775.61435493738	670.40973928681	880.81897058795	8.9459707185857	317.81825929466	657.11065846310	894.11805141166	149.159064782303
16	722.10109070363	616.89647505306	827.30570635420	8.9459707185857	319.89719688130	602.82222860701	841.37995280025	150.134755690852
17	669.43149388988	564.22687823931	774.63610954045	8.9459707185857	313.78102280668	552.43314656837	786.42984121138	147.264301340489
18	830.89997198525	725.69535633468	936.10458763582	8.9459707185857	372.71529999622	691.92702285358	969.87292111692	174.923447447206
19	636.69476340540	531.49014775484	741.89937905597	8.9459707185857	229.98340398738	550.94171837669	722.44780843412	107.936244853711
20	706.05210416013	600.84748850956	811.25671981069	8.9459707185857	318.77935380507	587.19004799815	824.91416032210	149.610127470327

Figure 6: 20 muestras aleatorias de tamaño 30

Calculo empirico del intervalo de confianza para varianza conocida

[1] 631.1255 667.5694

Calculo empirico del intervalo de confianza para varianza desconocida

[1] 628.5826 670.1123

1. Calculo empirico del error relativo

[1] 20.76484

Se puede ver que tenemos un desvío muy grande, por lo tanto esta aproximación no es buena, esto mejora un poco al aumentar la cantidad a 30 datos en una muestra, en donde el Error, con desvío conocido, se reduce de 257 a 8.

También visualizamos que los datos se acentúan, al igual que lo sucedido en el ensayo con la distribución normal, marcan una tendencia y los intervalos de confianza se acercan a los valores calculados empíricamente.

- **8. A partir de los análisis efectuados elabore sus conclusiones generales.**

Concluyo que para hacer una buena aproximación de la distribución normal no necesitamos tener un tamaño de muestra tan grande, como si lo vamos a necesitar para realizar una aproximación de una distribución gamma.

A partir de los análisis efectuados también concluimos que cuanto mayor sea la cantidad de datos que tengamos en nuestra muestra mejor va a ser la aproximación.

Si queremos realizar una buena aproximación a partir de una distribución gamma, debemos tener especial cuidado en el valor de sus parametros, por que estaremos rompiendo un supuesto teórico al realizar el calculo de su intervalo de confianza.