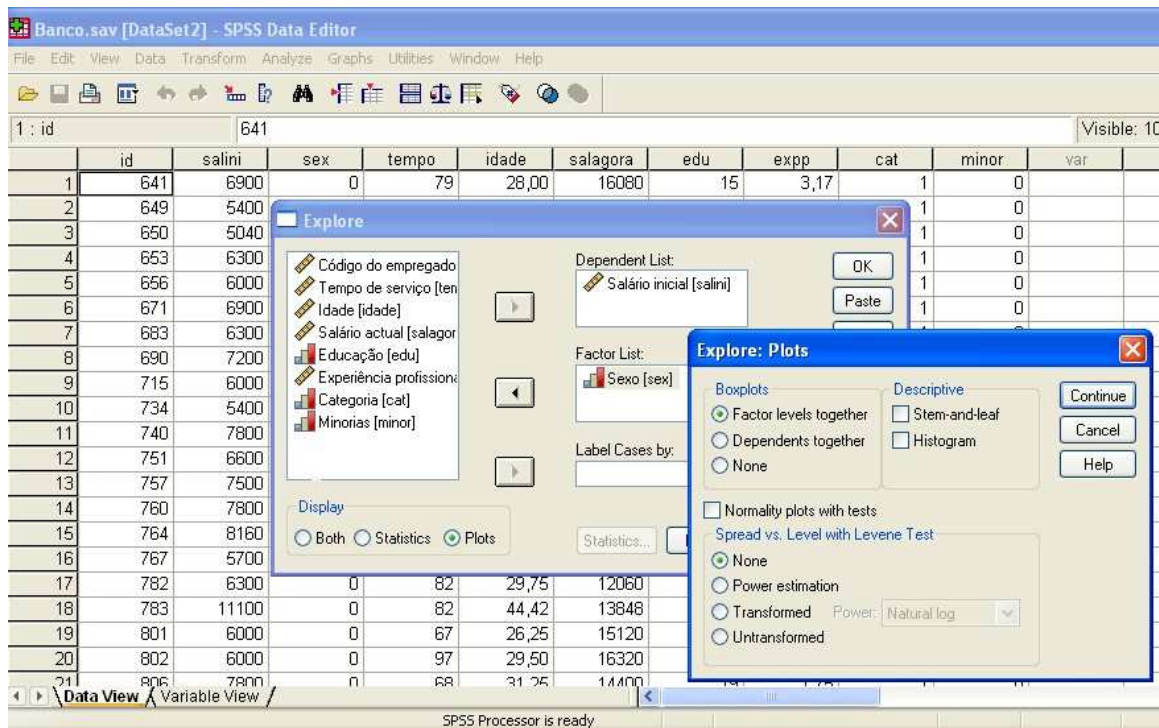


17. O ficheiro Banco.sav encerra informação relativa a 474 empregados contratados por um banco, entre 1969 e 1971. Este banco esteve envolvido num processo judicial no âmbito da Igualdade de Oportunidade no Emprego.

17.1. Relativamente à variável Salário Inicial apresente, para o sexo feminino e masculino, os box plots paralelos.

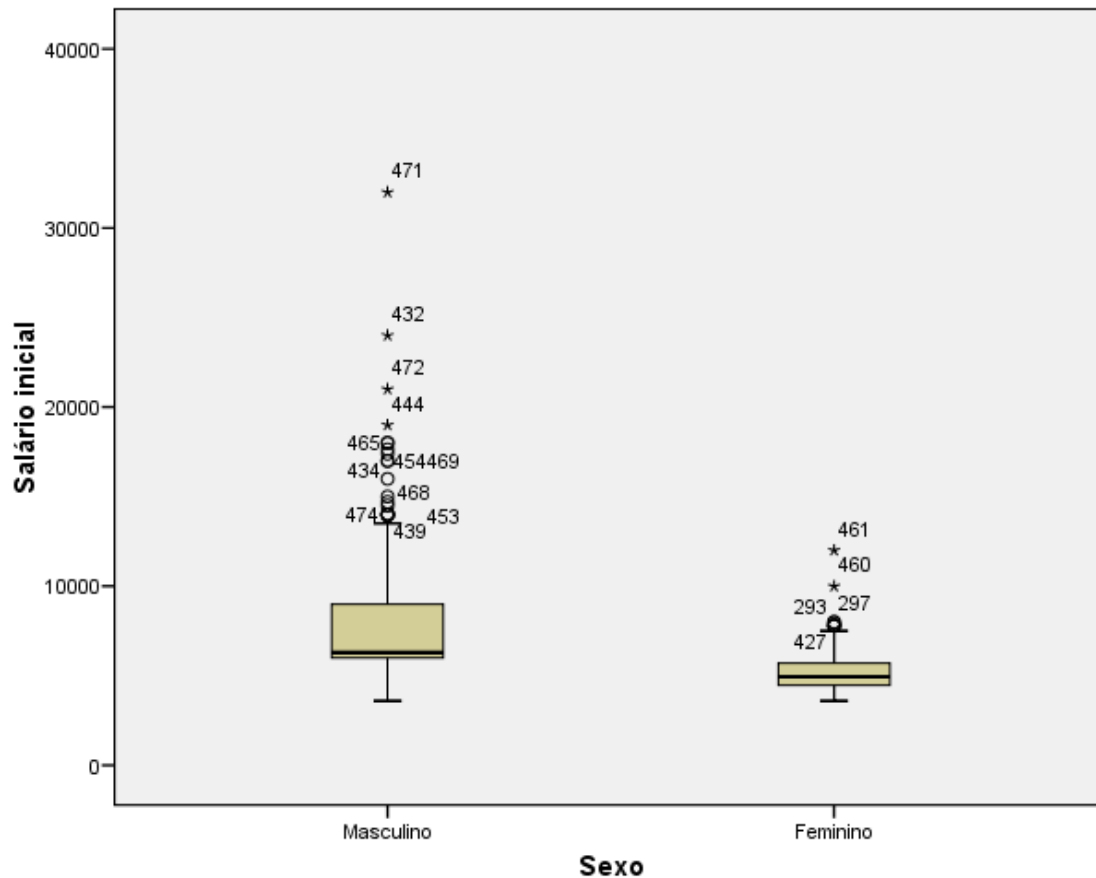
### Opção em SPSS: Analyze / Descriptive Statistics / Explore



### Resultados:

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	Sexo	N	Percent	N	Percent	N	Percent
Salário inicial	Masculino	258	100,0%	0	,0%	258	100,0%
	Feminino	216	100,0%	0	,0%	216	100,0%



Ambas as amostras revelam uma tendência para acumulação de frequências nas observações de mais baixo valor da amostra, isto é, ambas têm assimetria positiva. Mas esta assimetria é muito mais acentuada no caso dos homens uma vez que a dispersão dos valores é muito maior (existem outliers e extremos de valores muito elevados). Podemos também constatar que os valores mínimos de cada amostra devem ser idênticos, mas a localização das medianas é diferente sendo superior no caso dos homens.

### 17.2. Teste, aos níveis usuais de significância, se os dois grupos diferem na localização

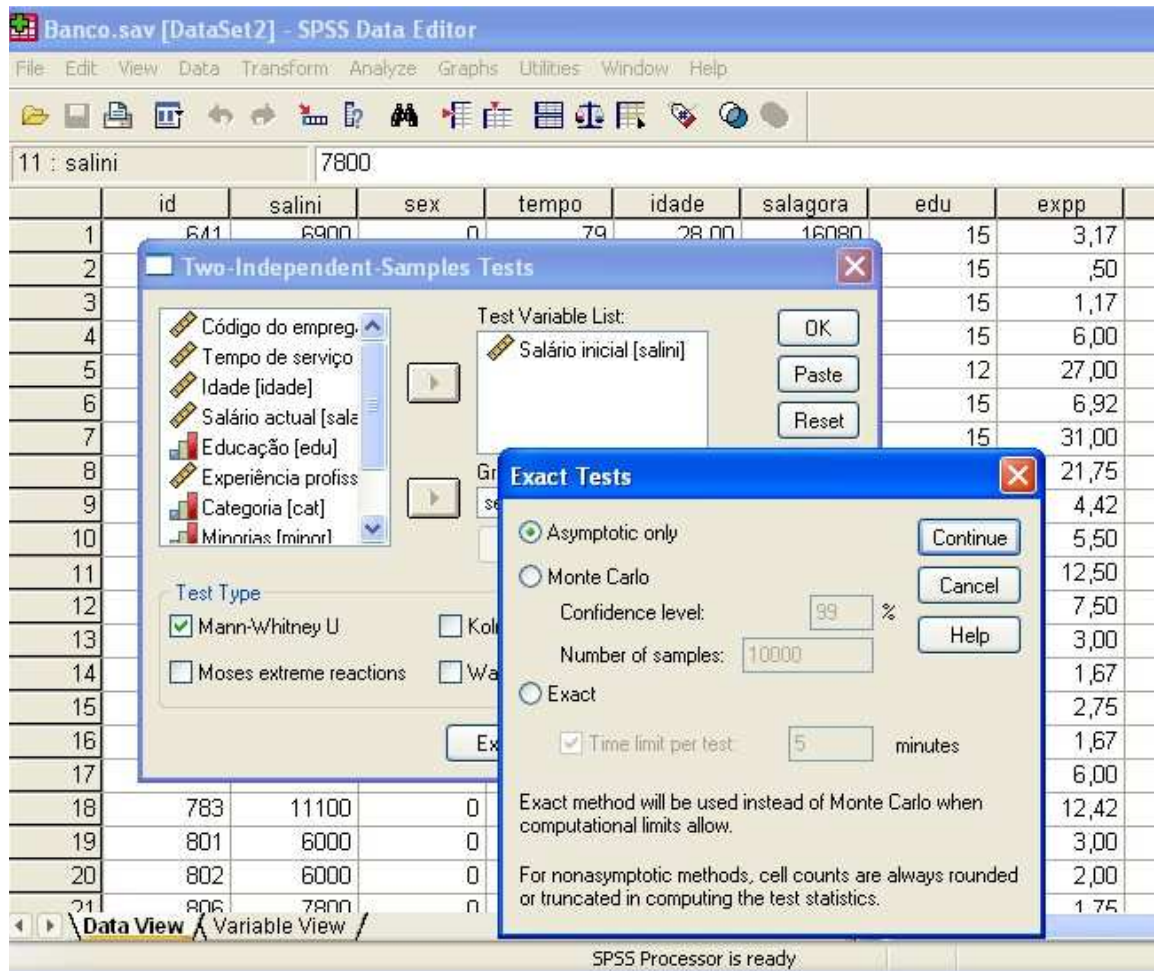
Uma vez que as amostras evidenciam assimetria, não devem ser consideradas como provenientes de populações Normais pelo que o teste a realizar neste contexto será o teste não paramétrico U de Mann-Whitney (ou de Wilcoxon-Mann-Whitney) para comparação de medianas de duas amostras independentes.

O teste a realizar será:

$$H_0 : \mu_{\text{saliniH}} = \mu_{\text{saliniM}} \text{ vs. } H_1 : \mu_{\text{saliniH}} \neq \mu_{\text{saliniM}}$$

onde  $\mu_{\text{saliniH}}$  representa a mediana do salário inicial para os homens e  $\mu_{\text{saliniM}}$  representa a mediana do salário inicial para as mulheres

### Em SPSS: Analyze / Non-Parametric Tests / 2 Independent Samples



### Resultados:

#### Mann-Whitney Test

##### Ranks

	Sexo	N	Mean Rank	Sum of Ranks
Salário inicial	Masculino	258	315,06	81285,00
	Feminino	216	144,86	31290,00
	Total	474		

### Test Statistics(a)

	Salário inicial
Mann-Whitney U	7854,000
Wilcoxon W	31290,000
Z	-13,496
Asymp. Sig. (2-tailed)	,000

a. Grouping Variable: Sexo

**Decisão:** Como  $p\text{-value}_{\text{bilateral}}(\text{assimptótico}) = 0,000 < \alpha$  (para q.q. nível usual de significância)  
 $\Rightarrow$  rejeitar a hipótese nula  $H_0$

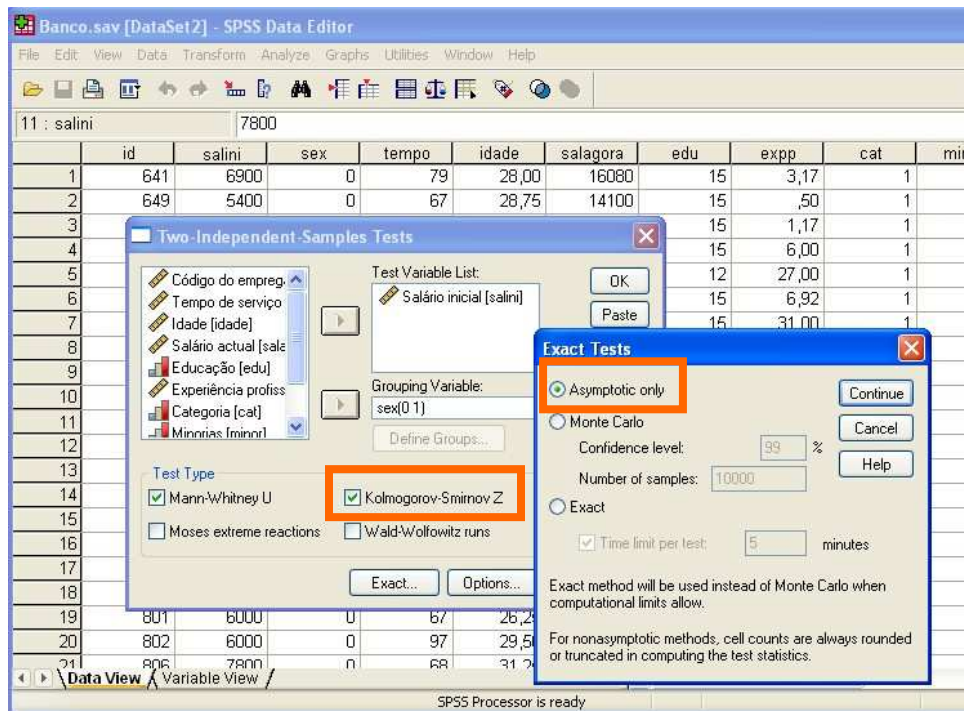
**Conclusão:** Para os níveis usuais de significância existem diferenças significativas entre os salários iniciais no banco dos homens e das mulheres

### 17.3. Será razoável afirmar que os registos correspondentes a ambos os grupos provêm da mesma população?

Para verificar se duas amostras provêm da mesma população teremos que realizar o teste de Kolmogorov-Smirnov para duas amostras independentes cujas hipóteses são:

$H_0$  – as duas amostras provêm da mesma distribuição  
 vs.  $H_1$  – as duas amostras não provêm da mesma distribuição

### Em SPSS: Analyze / Non-Parametric Tests / 2 Independent Samples



## Resultados:

### Two-Sample Kolmogorov-Smirnov Test

#### Frequencies

	Sexo	N
Salário inicial	Masculino	258
	Feminino	216
	Total	474

#### Test Statistics(a)

		Salário inicial
Most Extreme Differences	Absolute	,647
	Positive	,000
	Negative	-,647
Kolmogorov-Smirnov Z		7,015
Asymp. Sig. (2-tailed)		,000

a. Grouping Variable: Sexo

Decisão: Como  $p\text{-value}_{\text{bilateral}}(\text{assimptótico}) = 0,000 < \alpha$  (para q.q. nível usual de significância)  
 $\Rightarrow$  rejeitar a hipótese nula  $H_0$

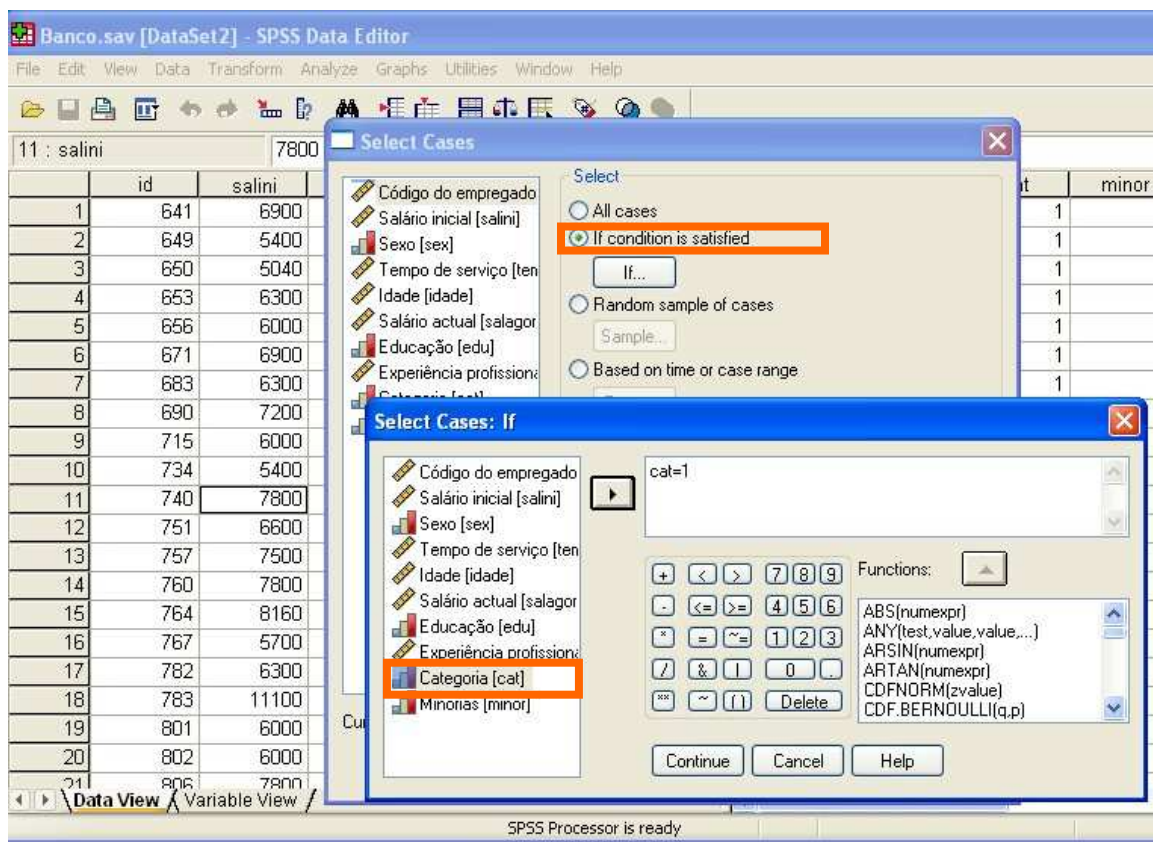
Conclusão: Para os níveis usuais de significância existem evidências para afirmar que as duas amostras provêm de populações diferentes, isto é, provêm de populações com diferentes distribuições de probabilidades.

17.4. Um dos argumentos que o banco poderia utilizar para refutar a acusação será o de que os sectores com maior número de empregados são precisamente aqueles em que as mulheres estão em maioria. No sector administrativo, por exemplo, trabalham 110 homens e 117 mulheres. O banco pode também ter a intenção de fazer salientar o seu empenho numa política contratual sob a égide de “*salário igual para trabalho igual*”. Teste se o salário inicial dos trabalhadores homens do sector administrativo será significativamente diferente do das mulheres do mesmo sector.

1º) Seleccionar apenas os casos de interesse:

### Em SPSS: Data / Select Cases

seleccionar apenas os dados para Categoria=1=”Administrativo”



2º) Realizar o teste de comparação de localizações para o caso de interesse:

O teste a realizar será:

$$H_0 : \mu_{\text{saliniHA}} = \mu_{\text{saliniMA}} \text{ vs. } H_1 : \mu_{\text{saliniHA}} \neq \mu_{\text{saliniMA}}$$

onde  $\mu_{\text{saliniHA}}$  representa a mediana do salário inicial para os homens do sector administrativo e  $\mu_{\text{saliniM}}$  representa a mediana do salário inicial para as mulheres do sector administrativo



## Em SPSS: Analyze / Non-Parametric Tests / 2 Independent Samples

(opções igual a 17.2)

### Resultados:

#### Mann-Whitney Test

Ranks

	Sexo	N	Mean Rank	Sum of Ranks
Salário inicial	Masculino	110	155,12	17063,50
	Feminino	117	75,34	8814,50
	Total	227		

Test Statistics(a)

	Salário inicial
Mann-Whitney U	1911,500
Wilcoxon W	8814,500
Z	-9,171
Asymp. Sig. (2-tailed)	,000

a. Grouping Variable: Sexo

Decisão: Como  $p\text{-value}_{\text{bilateral}}(\text{assimptótico}) = 0,000 < \alpha$  (para q.q. nível usual de significância)  
 $\Rightarrow$  rejeitar a hipótese nula  $H_0$

Conclusão: Para os níveis usuais de significância podemos concluir que os salários iniciais de homens e mulheres na mesma categoria (administrativa) são significativamente diferentes pelo que o banco não se rege pela égide “*salário igual para trabalho igual*”

#### 17.5. Averigue se é razoável afirmar que o tempo de serviço de um empregado deste banco tem distribuição uniforme.

As hipóteses que pretendemos testar são:

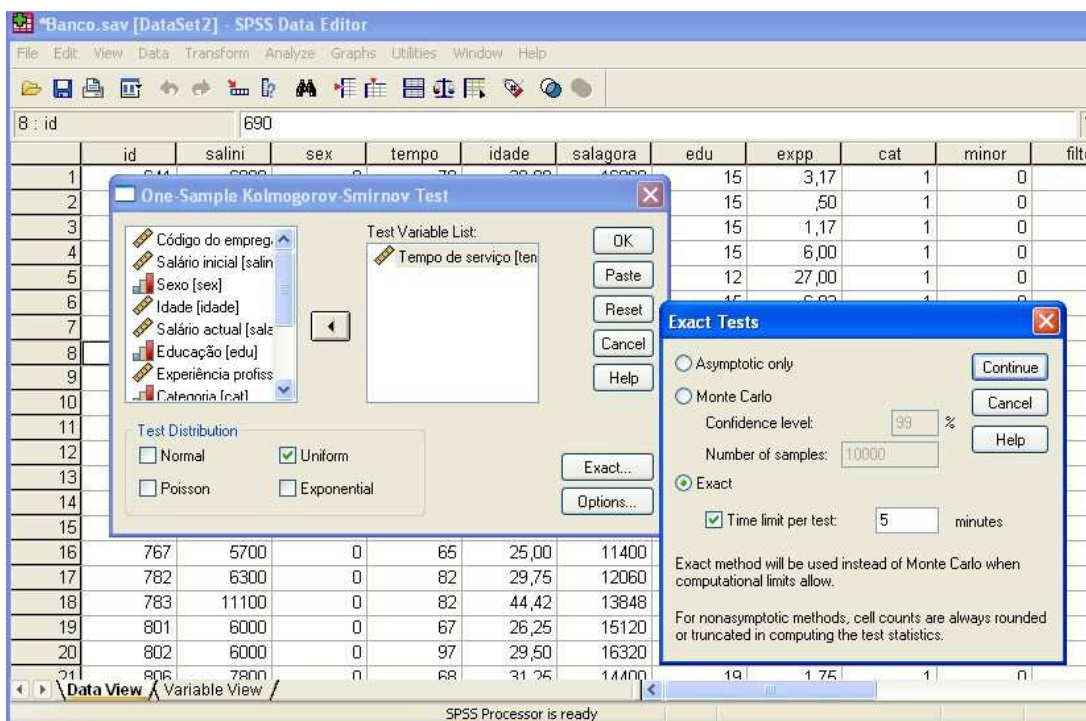
$H_0$ : “tempo de serviço” segue uma distribuição uniforme

$H_1$ : “tempo de serviço” não segue uma distribuição uniforme

O teste a utilizar nesta situação é o teste de ajustamento de Kolmogorov-Smirnov.

ATENÇÃO: não esquecer de anular o SELECT CASE

## Em SPSS: Analyze / Non-Parametric Tests / 1 Sample K-S



## Resultados: NPar Tests

### One-Sample Kolmogorov-Smirnov Test

		Tempo de serviço
N		474
Uniform Parameters(a,b)	Minimum	63
	Maximum	98
	Absolute	,070
	Positive	,018
Most Extreme Differences	Negative	-,070
Kolmogorov-Smirnov Z		1,522
Asymp. Sig. (2-tailed)		,019
<b>Exact Sig. (2-tailed)</b>		<b>,018</b>
Point Probability		,000

a. Test distribution is Uniform.

b. Calculated from data.

Como  $p\text{-value}_{\text{bilateral}}(\text{exacto}) = 0,018 < 0,05 < 0,1 \Rightarrow$  para níveis de significância de 5% ( $\alpha=0,05$ ) e de 10 % ( $\alpha=0,10$ ) temos que rejeitar a hipótese nula do tempo de serviço poder ser modelado por uma distribuição Uniforme. Porém, para o nível de significância de 1 % ( $\alpha=0,01$ ) não podemos rejeitar a hipótese nula, não tendo evidências para refutar que a distribuição do tempo de serviço não seja uniforme.



**17.6. Que teste escolheria se pretendesse verificar se existem diferenças significativas entre o salário inicial e o salário actual. Calcule o p-value do teste que adoptou.**

Como o salário actual de cada trabalhador está relacionado com o salário inicial do mesmo, logo estas amostras são emparelhadas. Para comparação de localização entre amostras emparelhadas temos 3 testes à escolha:

TESTES PARAMÉTRICOS	ALTERNATIVA NÃO PARAMÉTRICA
<b>PARA DUAS AMOSTRAS ALEATÓRIAS EMPARELHADAS</b> ( $X_1, \dots, X_n$ ) e ( $Y_1, \dots, Y_n$ ) formando pares ( $X_i, Y_i$ ), definir amostra das diferenças D: $D_i = X_i - Y_i \forall i$	
<b>Teste para a comparação de <u>médias</u></b> $H_0: \mu_D = 0$ vs. $H_1: \mu_D \neq (<, >) 0$ , $\mu_D$ – média	<b>Teste para a comparação de <u>medianas</u></b> $H_0: \mu_D = 0$ vs. $H_1: \mu_D \neq (<, >) 0$ , $\mu_D$ – mediana
<b>Teste T</b> <b>SPSS:</b> Analyze / Compare Means / Paired-Samples T Test	<b>Teste dos Sinais</b> <b>SPSS:</b> Analyze/Nonparametric Tests/ 2 Related Samples
<b>Pressupostos Exigidos</b>	<b>Pressupostos Exigidos</b>
1. Observações em <u>escala de razões ou intervalos</u>	1. Observações numa escala <u>pelo menos ordinal</u>
2. A f.d. da população de D é <u>Normal</u>	2. A f.d. da população de D é <u>contínua</u>
	<b>Teste de Wilcoxon (signed-ranks)</b> <b>SPSS:</b> Analyze/Nonparametric Tests/ 2 Related Samples
	<b>Pressupostos Exigidos</b>
	1. Observações numa escala <u>pelo menos ordinal</u>
	2. A f.d. da população de D é <u>contínua e simétrica</u> relativamente à sua mediana

O teste T para comparação de médias deve ser usado apenas num contexto paramétrico, o que não é o caso, uma vez que já tínhamos visto na alínea 17.1. que, pelo menos para a variável “salário inicial”, a assimetria é bastante elevada levando a pensar que a amostra não deve ser proveniente de uma população Normal. No caso das alternativas não paramétricas podemos escolher entre o teste de Wilcoxon e o teste dos Sinais.

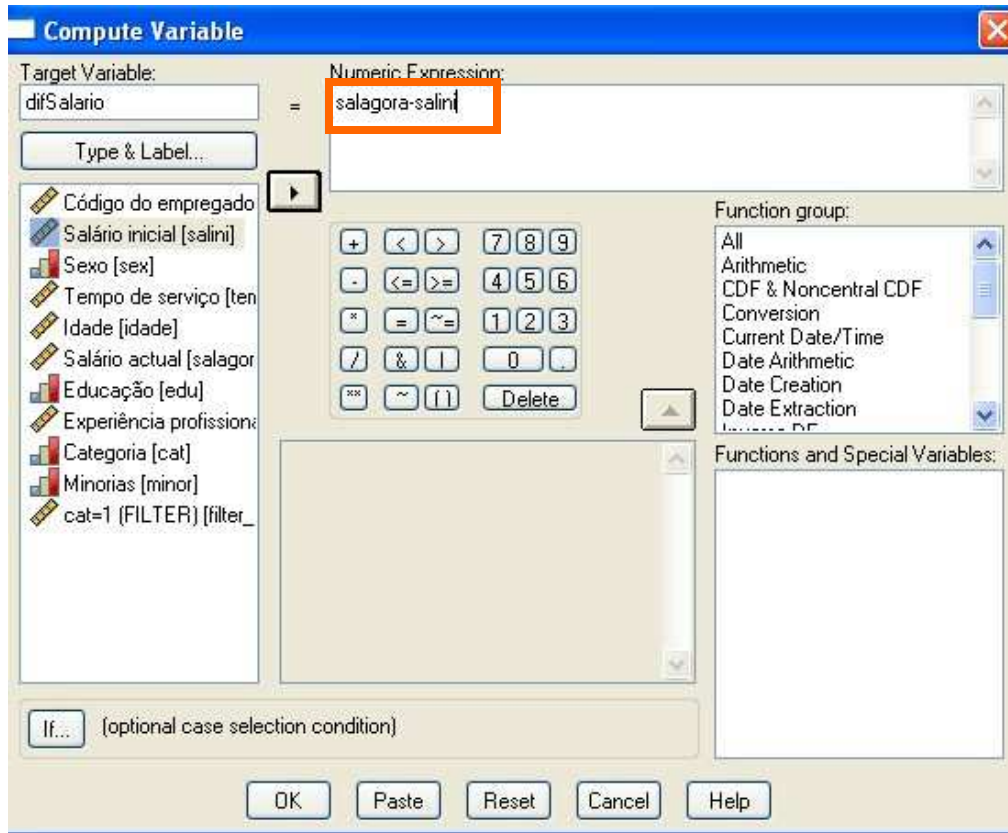
Para poder usar o teste de Wilcoxon temos que verificar se as amostras são simétricas, ou, pelo menos, que a diferença entre as duas amostras emparelhadas tenha uma distribuição de frequência simétrica. No caso de não se verificar a simetria da diferença então podemos escolher o teste dos Sinais. De salientar que dos 3 testes propostos o teste dos Sinais é o menos potente e portanto este só deve ser escolhido quando, num contexto não paramétrico, se verifica que as amostras não são simétricas.

Assim, o primeiro passo para escolher o teste a usar será estudar a simetria da amostra das diferenças:

**1º) Construir a amostra das diferenças:**

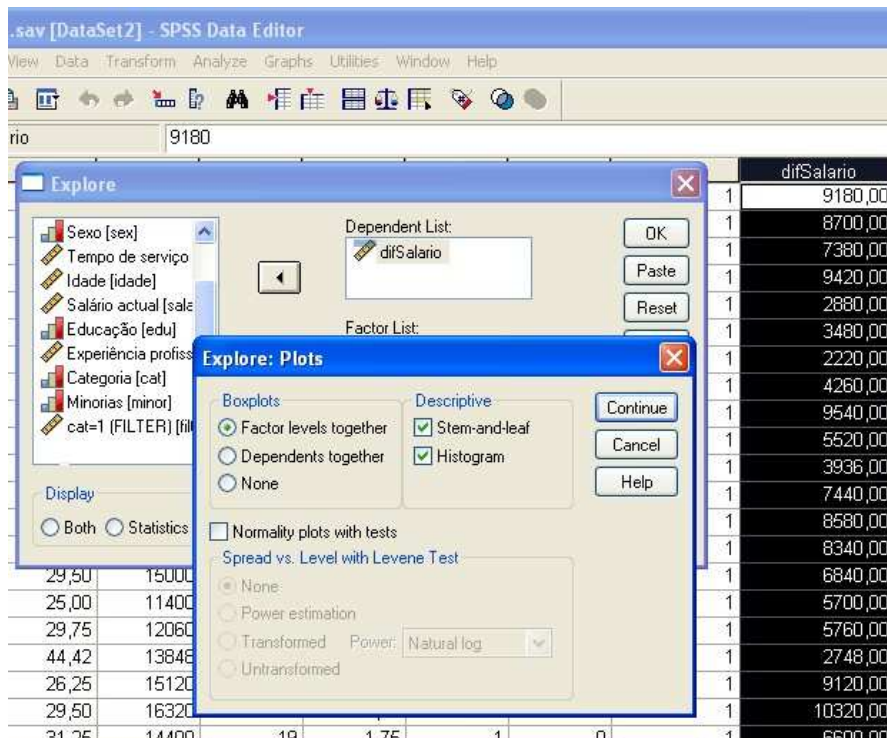
**Em SPSS: Transform/Compute Variable**

Foi criada uma nova variável chamada **difSalario**



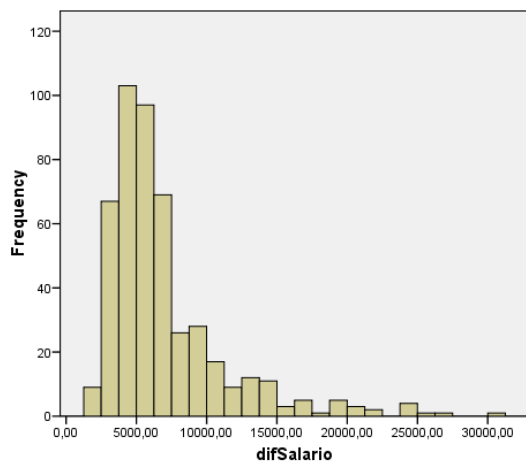
**2º) Estudar a simetria de difSalario:** podemos usar um histograma, um diagrama de caule-e-folhas e/ou uma caixa de bigodes

**Em SPSS: Analyze/Descriptive Statistics / Explore**

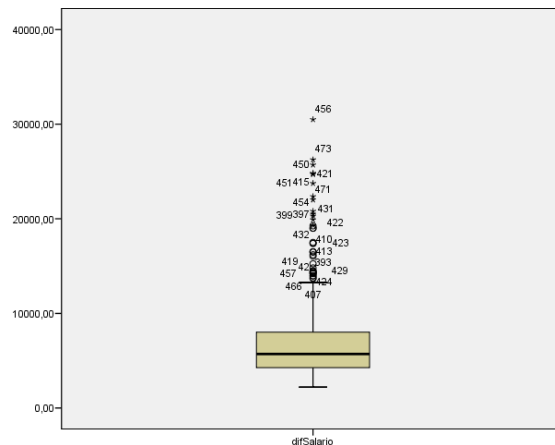


## Resultados:

Histogram



Mean = 6961,39  
Std. Dev. = 4325,8  
N = 474



Em qualquer um dos gráficos podemos observar uma tendência para acumulação de frequências nas observações de mais baixo valor da amostra, isto é, existe assimetria positiva.

## difSalario Stem-and-Leaf Plot

Frequency	Stem &	Leaf
33,00	2 .	22456777889
58,00	3 .	01233344556667778899
88,00	4 .	000001222233444555566667888899
79,00	5 .	011112233444555666777888899
67,00	6 .	00011333334445667999&
30,00	7 .	022334568&
25,00	8 .	012357789&
19,00	9 .	14556&&
12,00	10 .	279&&
11,00	11 .	037&
8,00	12 .	0&&
6,00	13 .	2&
38,00	Extremes	(>=13700)

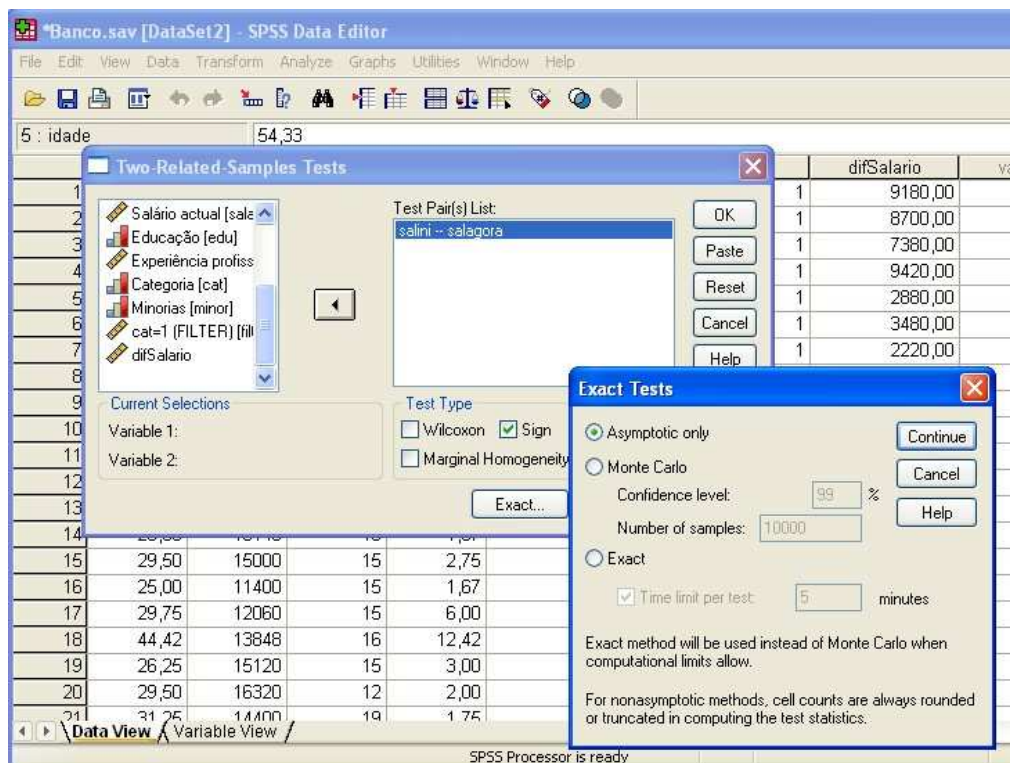
Stem width: 1000,00  
Each leaf: 3 case(s)

& denotes fractional leaves.

Como existe assimetria positiva não podemos usar o teste de Wilcoxon, pelo que nos resta o teste dos Sinais

### 3º) Implementar teste dos Sinais:

## Em SPSS: Analyze/ Nonparametric Tests / 2 Related Samples



As hipóteses a testar são:

$$H_0 : \mu_{\text{salini}} = \mu_{\text{salagora}} \text{ vs. } H_1 : \mu_{\text{salini}} \neq \mu_{\text{salagora}}$$

onde  $\mu_{\text{salini}}$  representa a mediana do salário inicial e  $\mu_{\text{salactual}}$  representa a mediana do salário actual

## Resultados:

### Sign Test

Frequencies

		N
Salário actual - Salário inicial	Negative Differences(a)	0
	Positive Differences(b)	474
	Ties(c)	0
	Total	474

a Salário actual < Salário inicial

b Salário actual > Salário inicial

c Salário actual = Salário inicial

Test Statistics(a)

	Salário actual - Salário inicial
Z	-21,726
Asymp. Sig. (2-tailed)	,000

a Sign Test

Decisão: Como  $p\text{-value}_{\text{bilateral}} (\text{assimptótico}) = 0,000 < \alpha$  para q.q. nível usual de significância  $\Rightarrow$  rejeitar a hipótese nula  $H_0$

Conclusão: Para os níveis usuais de significância podemos concluir que os salários iniciais e actuais dos trabalhadores do banco são significativamente diferentes

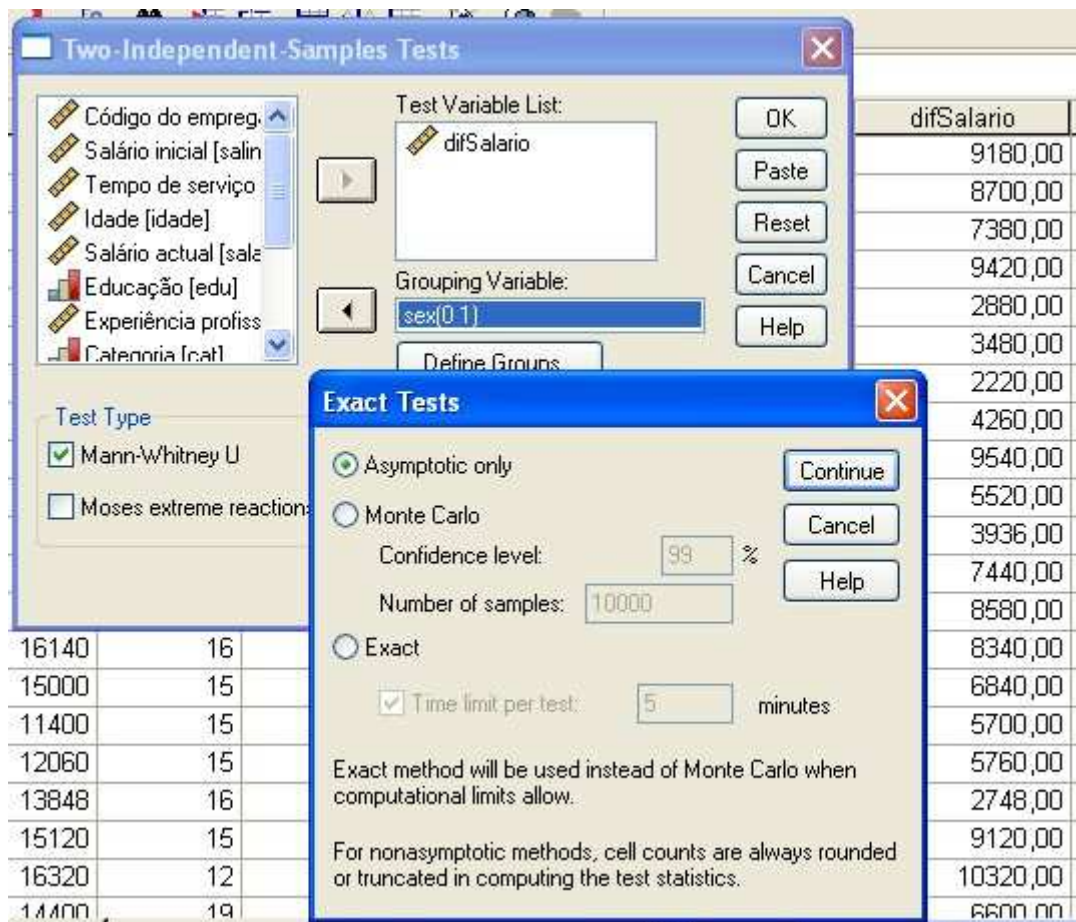
**17.7. Teste, aos níveis usuais de significância, se existem evidências para afirmar que a diferença entre salários inicial e actual é mais acentuada nos homens do que nas mulheres.**

Neste caso estamos perante amostras independentes pois não existe qualquer relação entre a evolução dos salários nos homens e nas mulheres. Alias, o que pretendemos saber é se, ao longo do tempo, os homens vão sendo aumentados mais do que as mulheres. As hipóteses que pretendemos testar são então:

$$H_0 : \mu_{\text{difSalarioH}} = \mu_{\text{difSalarioM}} \text{ vs. } H_1 : \mu_{\text{difSalarioH}} > \mu_{\text{difSalarioM}}$$

O teste a realizar para comparação de localizações entre amostras independentes num contexto não paramétrico é o teste U de Mann-Whitney:

**Em SPSS: Analyze / Non-Parametric Tests / 2 Independent Samples**





## Resultados:

### Mann-Whitney Test

Ranks

	Sexo	N	Mean Rank	Sum of Ranks
difSalario	Masculino	258	290,89	75049,50
	Feminino	216	173,73	37525,50
	Total	474		

Test Statistics(a)

	difSalario
Mann-Whitney U	14089,500
Wilcoxon W	37525,500
Z	-9,275
Asymp. Sig. (2-tailed)	,000

a. Grouping Variable: Sexo

Decisão: O  $p\text{-value}_{\text{bilateral}} (\text{assimptótico}) = 0,000$ . Uma vez que a amostra aponta no sentido da hipótese alternativa (pois o mean rank da variável difSalario é superior nos homens (290,89) do que nas mulheres (173,73), então podemos dividir o p-value bilateral por 2 para obter o p-value unilateral, ou seja:

$p\text{-value}_{\text{unilateral}} = p\text{-value}_{\text{bilateral}} / 2 = 0,000 < \alpha$  para q.q. nível usual de significância  
 $\Rightarrow$  rejeitar a hipótese nula  $H_0$

Conclusão: Aos níveis usuais de significância podemos concluir que a diferença entre os salários inicial e actual é mais acentuada nos homens do que nas mulheres.

Ou seja, neste banco, não só os homens começam por ganhar mais dinheiro à partida como também progridem mais depressa na carreira