

Exploración de datos multivariados

Año: 2022

Examen final

“Dataset salarios”



Salim Taleb, Nasim A

Docentes: Fernández Melisa, Friosli Melisa.

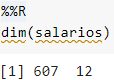
Carrera: Tecnicatura universitaria en procesamiento y explotación de datos

## Introducción

En el presente informe se presenta un análisis de un dataset que contiene datos sobre salarios de trabajadores del rubro de análisis de datos, con el fin de obtener información descriptiva sobre los salarios y encontrar características de los salarios mejores pagos.

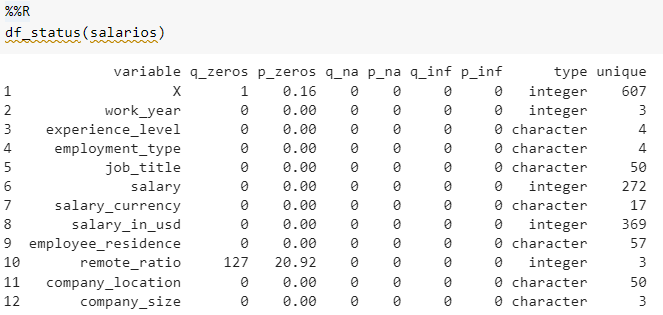
## Descripción del dataset

Luego de la carga del dataset, mediante la siguiente línea se obtiene el tamaño del mismo:



Esto nos dice que el dataset está compuesto por 607 observaciones con un total de 12 variables

Una descripción detallada de los datos de cada variable del dataset:



En cuanto a la calidad del dataset, se observa que no contiene valores incorrectos y en cuanto a valores nulos estos sólo aparecen en la variable "remote\_ratio" la cuál tiene como posible valor al cero, por lo cual no hay ningún inconveniente.

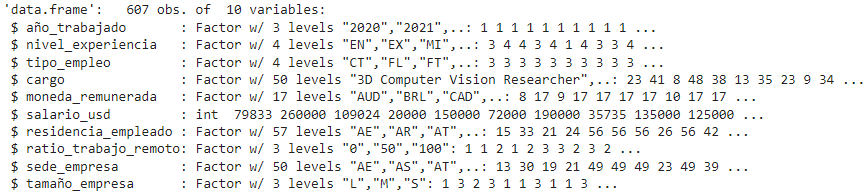
## Limpieza del dataset

La limpieza realizada al dataset en cuanto a los datos fue mínima ya que no se encontraron valores inválidos en el mismo, sin embargo pueden ocurrir cambios más adelante posterior a realizar análisis más detallados en la distribución de las variables.

Se quitó la primera columna por ser, al parecer, una ID que no aporta información relevante, y la columna del salario en moneda recibida, ya que se tiene el valor estandarizado en dólares, lo que permite una mejor comparación y deja inútil a esta columna.

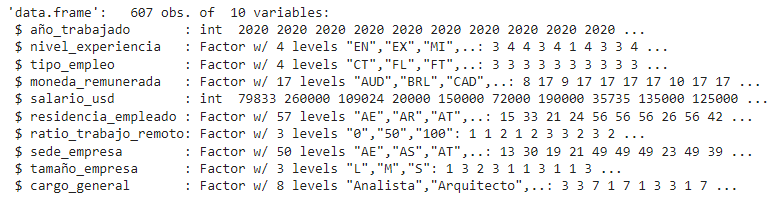
También se modificaron los nombres de las columnas para ser interpretadas con mayor facilidad.

Quedando finalmente el dataset con la siguiente estructura:



Durante el proceso de análisis univariado de la variable “cargo” se determinó que era más útil reemplazarla por una variable “cargo\_general”, agrupando los distintos cargos en los factores “Científico", "Ingeniero", "Analista", "Gerente", "Arquitecto", "Desarrollador", "Consultor" y "Jefe".

Quedando finalmente la siguiente estructura.



## Hipótesis de variables pre análisis

Antes de comenzar con el análisis univariado, bivariado y multivariado se plantean las siguientes hipótesis de cómo cada variable afectará a los salarios, siendo esta la más importante, a ser verificadas o refutadas mediante el análisis.

Año trabajado: Al estar los datos correspondidos entre 2020, 2021 y 2022 se espera que dado el contexto de pandemia que ocurrió por esas épocas, está variable afecte a los salarios en cierta medida.

Nivel de experiencia: Está variable puede tener relación fuerte con el salario ya que por lo general los trabajadores más experimentados suelen recibir una mayor remuneración por sus servicios.

Tipo de empleo: En cuanto al tipo de empleo, puede estar relacionado ligeramente al salario por ejemplo en el caso de los trabajadores a contrato o freelance con respecto a los trabajadores permanentes ya que realizan trabajos por periodos de tiempo cortos y deben cobrar más para asegurar su bienestar económico.

Cargo: Se espera que está variable tenga una fuerte relación respecto al salario ya que la importancia de la posición ocupada dentro de una empresa suele ser un gran indicador en las diferencias salariales.

Moneda remunerada: La variable puede tener alta relación con el salario aunque no necesariamente de manera directa, sino más bien indirectamente debido a que el tipo de moneda es afectado por el lugar de residencia del empleado donde los costos de vida pueden variar en los distintos países. Sin embargo, no puede ser descartada ya que puede ocurrir que la moneda remunerada por un empleado no sea la misma que se utiliza en su país de residencia.

Salario en USD: La variable más importante del dataset, la cuál indica el equivalente del salario de cada trabajador unificado en dólares para permitir la comparación entre trabajadores que reciben su remuneración en monedas diferentes.

Residencia del empleado: La variable puede estar moderadamente relacionada al salario ya que, como se mencionó previamente, en cada país los costos de vida pueden variar y el trabajador puede exigir un mayor salario.

Ratio de trabajo remoto: Es posible que la variable esté ligeramente relacionada al salario ya que puede ocurrir que un trabajador que realiza su trabajo de manera presencial requiere un mayor salario para costear, por ejemplo, transporte.

Sede de la empresa: Está variable debería impactar moderadamente al salario ya que dependiendo la sede de la empresa los mercados de los distintos países pueden ser diferentes.

Tamaño de la empresa: La cantidad de trabajadores de la empresa puede afectar en cierta medida al salario de los mismos ya que puede ocurrir que una empresa pequeña tenga pocos trabajadores mejores pagados en un área, o en una empresa grande muchos trabajadores con un salario promedio.

Estableciendo un valor de prioridad a cada variable según cómo podrían afectar el sueldo se confecciona la siguiente gráfica de torta:

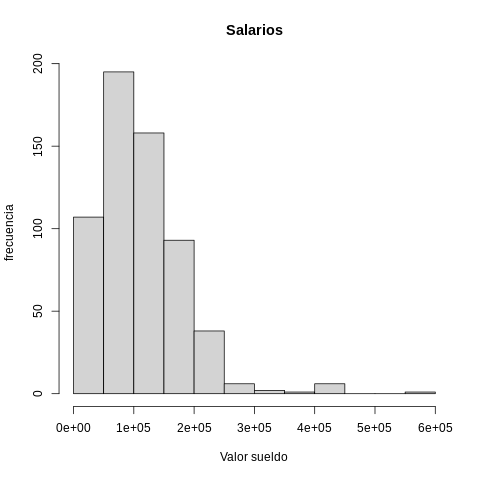


## 

## Análisis univariado

**Variables cuantitativas**

Salario en USD



Se observa que están distribuidos mayoritariamente entre 0 y 300000, con algunos valores mayores a está cantidad.

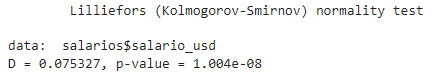
En cuanto a medidas cuantitativas:



Observando el coeficiente de variación que es mayor a 0,3 se puede decir que la variable representa un conjunto heterogéneo por lo que la media no representa una buena medida de tendencia central, en su lugar es preferible la mediana.

En base a los coeficientes de asimetría y curtosis se puede decir que el salario sigue una distribución asimétrica a la derecha y muy leptocúrtica, lo que puede visualizarse gráficamente en el histograma.

En cuanto a la verificación de la normalidad en la distribución se utiliza el test de Lilliefors (Kolmogorov-Smirnov) por tratarse de un dataset de más de 50 muestras

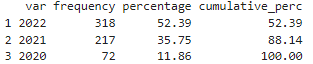
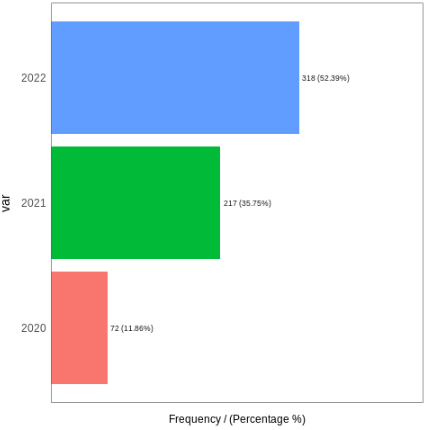


El p-value es menor a 0.05 por lo que puede rechazarse la hipótesis nula y concluir que la distribución no es normal.

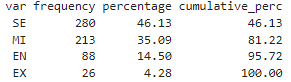
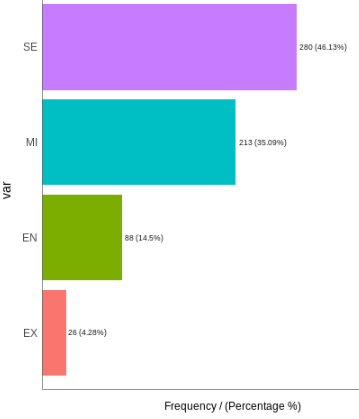
Posteriormente se analizará en conjunto con otras variables para determinar si puede estudiarse los salarios en distintos grupos que sí siguen una distribución normal

**Variables cualitativas**

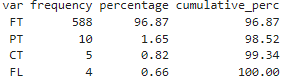
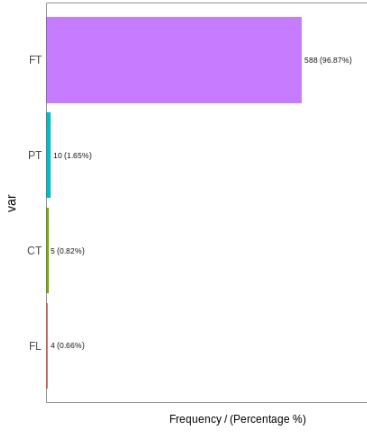
Año trabajado: Observando los años de los datos, aproximadamente la mitad corresponden al 2022, y del resto una gran parte corresponde al 2021 mientras que el resto son del 2020.



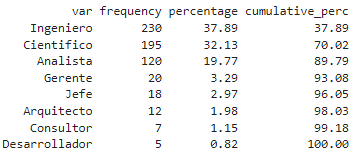
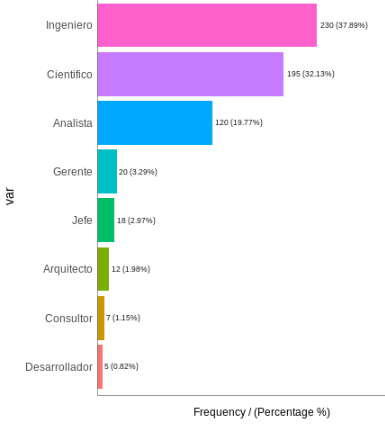
Nivel de experiencia: Para el caso de los distintos niveles de experiencia se observa que la mayor cantidad de trabajadores son de nivel Senior y un poco menos de nivel intermedio, mientras que hay pocos de nivel Junior y la menor cantidad es de nivel ejecutivo.



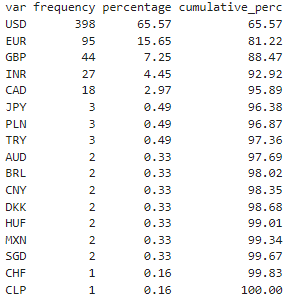
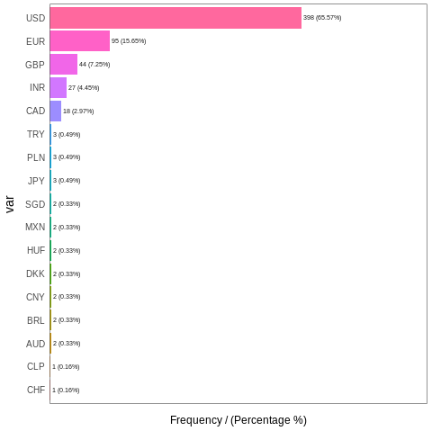
Tipo de empleo: Para el caso del tipo de empleo se ve que casi en su totalidad los trabajadores son full time, con muy pocos datos de los demás tipos.Para este dataset, la variable está muy desbalanceada, por lo que al utilizar está variable para análisis se debe tener cuidado al tener poca cantidad de muestras de algunos valores.



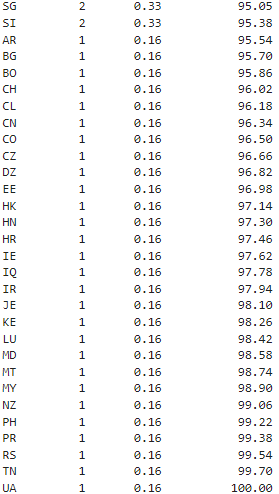
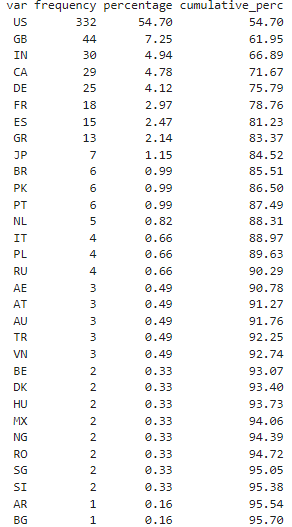
Cargo: Las conclusiones que pueden extraerse es que la gran mayoría, casi el 90%, de los trabajadores tienen cargos como ingenieros, científicos o analistas y el resto están distribuidos en otros cargos como gerentes o jefes, etc.

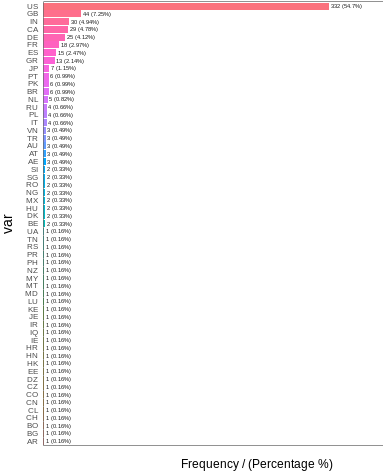


Moneda remunerada: Se ve que la moneda predominante en las remuneraciones es en dólares, seguido por el euro, la libra, la rupia india y el dólar canadiense, luego otras monedas en una cantidad muy pequeña

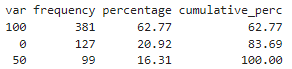
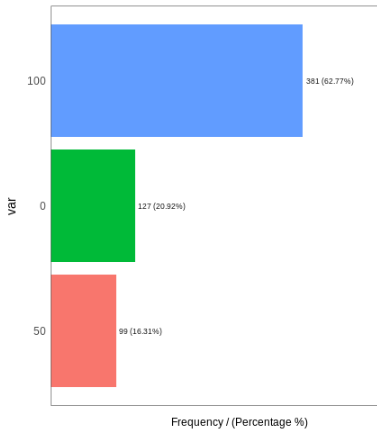


Residencia del empleado: Como cabe esperar, los porcentajes son similares a los de las monedas remuneradas, ya que por lo general una persona cobrará en la moneda del país donde vive. Sin embargo, hay mayor variedad ya que hay casos dónde los trabajadores prefieren cobrar en dólares por ser una moneda más fuerte que la local.

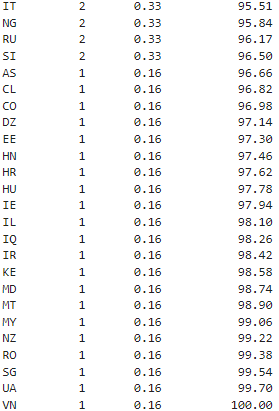
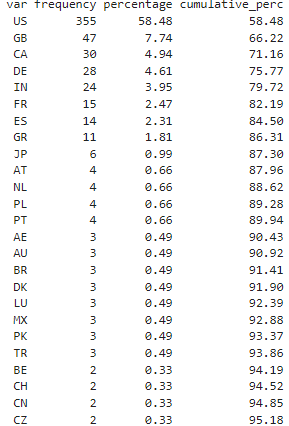


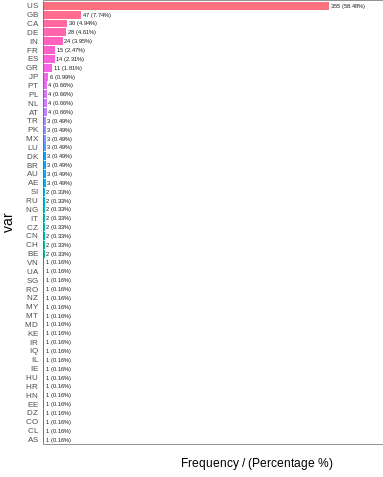


Ratio de trabajo remoto: En el caso del ratio de trabajo remoto se puede observar que casi dos tercios de los trabajadores trabajan de manera totalmente remota, mientras que el porcentaje que lo hace de manera presencial o semi-remota ocupan casi un tercio y son similares entre ellos.

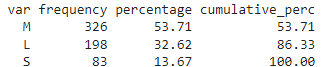
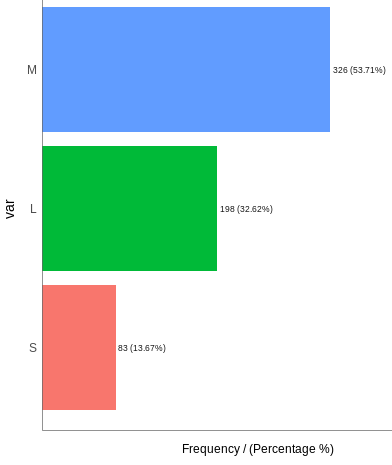


Sede de la empresa: Nuevamente una variación similar a la de la residencia de los empleados.





Tamaño de la empresa: En cuanto al tamaño de la empresa la mayoría de los trabajadores pertenecen a una empresa mediana, un tercio a empresa grandes y unos pocos a empresas pequeñas.



## Análisis bivariado

**Relaciones cuantitativas - cuantitativas**

En este caso no pueden analizarse interacciones entre variables cuantitativas ya que solo hay una variable que es el salario.

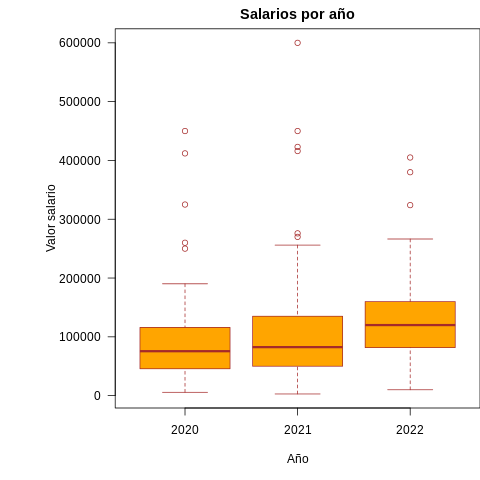
**Relaciones cuantitativas - cualitativas**

Se analiza la variable salario con respecto a cada una de las otras variables para encontrar posibles efectos de las mismas con el salario.

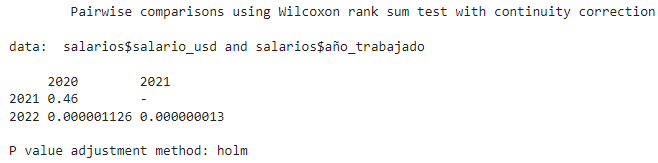
En todos los casos se realizaron tests no paramétricos ya que en todos los análisis al menos un grupo no cumplía el supuesto de normalidad.

En todos los casos se encontraron diferencias significativas entre al menos dos grupos.

Salario - Año

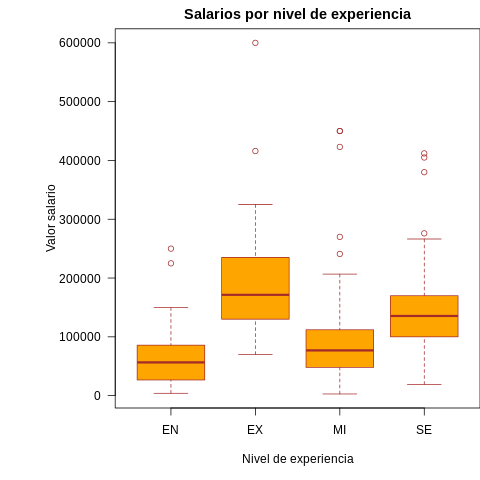


Gráficamente pareciese que los salarios de 2020 y 2021 son similares pero que ambos difieren de los pagados en 2022.

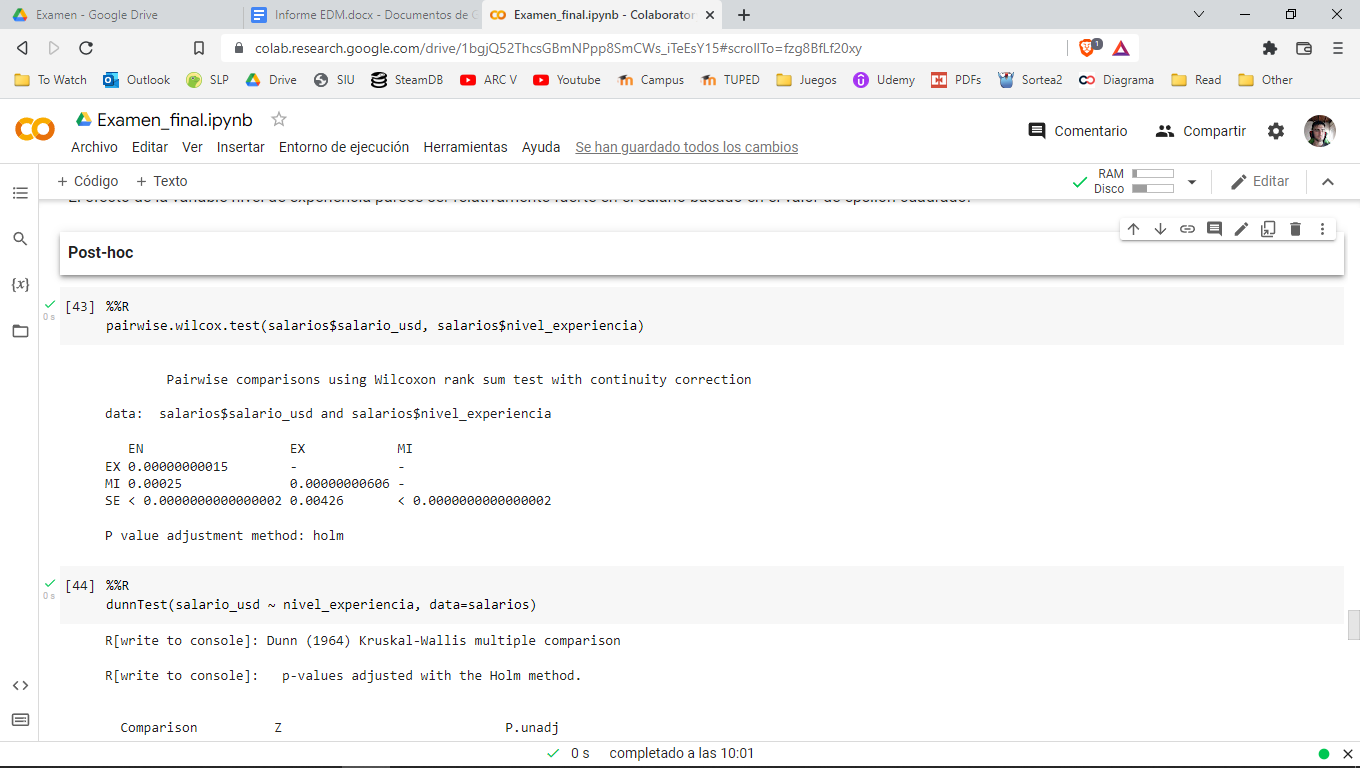


En conclusión, no existe evidencia para decir que los salarios de 2020 y 2021 sean diferentes entre sí, pero si se diferencian ambos con respecto a los salarios de 2022, lo cual indica que los salarios pagados en 2022 son más altos. Esto puede explicarse, tal vez, por el contexto de pandemia que se vivió en 2020/2021, donde su efecto ya fue muy leve durante 2022.

Salario - Nivel de experiencia

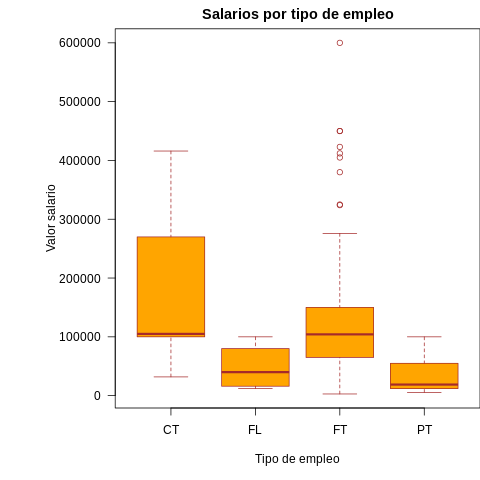


Gráficamente se puede observar que parecieran ser similares los salarios entre trabajadores de experiencia junior e intermedio pero que según los otros niveles de experiencia el salario difiere significativamente.

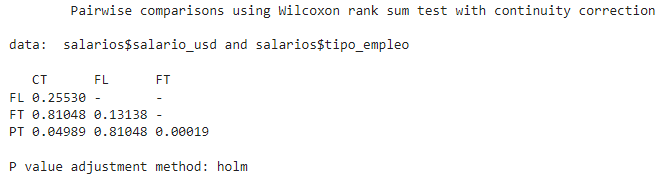


En conclusión, se puede decir que hay diferencia significativa entre todos los niveles de experiencia. Los trabajadores de nivel ejecutivo reciben mejores salarios, luego los trabajadores senior, seguidos por los intermedios y finalmente los junior.

Salario - Tipo de empleo



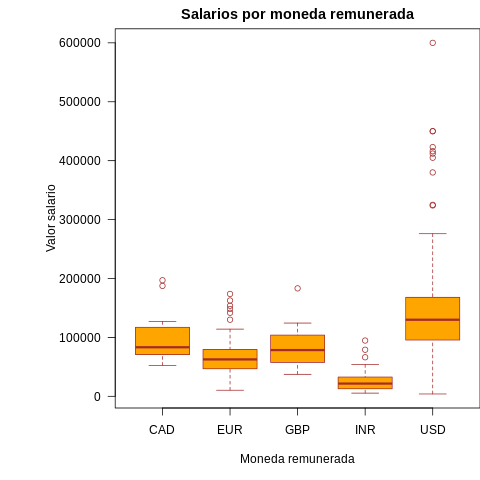
Gráficamente pareciese que la distribución de los salarios según tipo de empleo son diferentes.



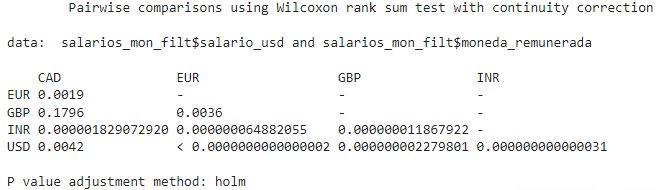
En conclusión, existe evidencia para afirmar que los salarios entre trabajadores de tipo part-time con los contract-time y full-time, pero no para el caso de las demás comparaciones. Esto indicaría que está variable no es tan significativa para el caso de los salarios, también pueden deberse los resultados a la baja representación de trabajadores de tipo de empleo diferente al full-time como se vió previamente.

Salario - Moneda remunerada

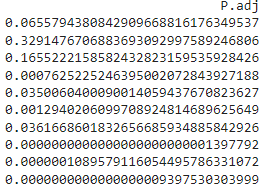
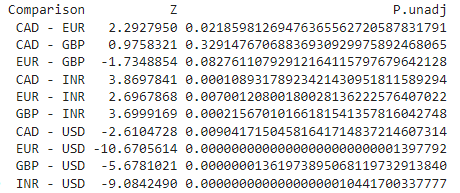
Para este caso se analizarán solo monedas las cuales haya más de 3 observaciones para simplificar el análisis y evitar sacar conclusiones de un grupo del cuál se posee pocas observaciones.



Gráficamente es difícil establecer diferencias o no, pareciese que los salarios pagados en rupias indias son considerablemente más bajos al resto y los pagados en dólares mayores.



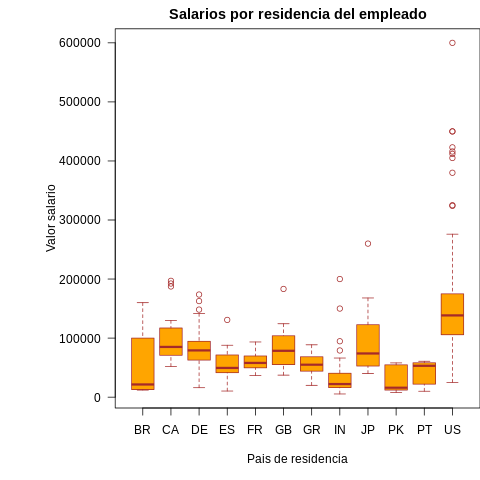
Dunnet test



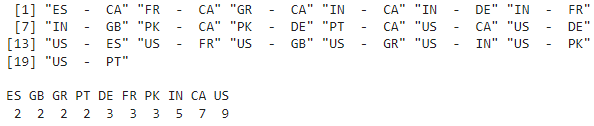
En conclusión, en ambos test se ve que los salarios remunerados en rupia india y en dólar son diferentes significativamente con los demás, lo que indica que los salarios pagados en rupia india son menores y los pagados en dólares mayores. También se puede decir que no existe evidencia para decir que los salarios remunerados en dólar canadiense y libras son diferentes. Finalmente, según los valores obtenidos en el test de dunnett, en los salarios remunerados en euros comparados con los remunerados en dólar canadiense y libras no existe evidencia para decir que son diferentes, sin embargo, el test de Kruskal-Wallis dice que los salarios en euros son diferentes a los remunerados en dólar canadiense y libras.

Salario - Residencia del empleado

Para este caso se analizarán sólo países los cuales haya más de 5 observaciones para simplificar el análisis y evitar sacar conclusiones de un grupo del cuál se posee pocas observaciones.



Como los valores al comparar son bastantes, se seleccionan las combinaciones que hayan presentado un p-value menor a 0.05 para encontrar las que son significativamente diferentes



No se detectaron diferencias entre trabajadores brasileños y japoneses, esto puede deberse a la gran variación que existe entre los salarios, lo que puede visualizarse por la gran amplitud de las cajas en el gráfico.

Si nos concentramos en los 10 países restantes:

Se observa que los salarios de trabajadores de Estados Unidos poseen diferencias significativas con todos los demás, se puede decir que sus salarios son los más altos.

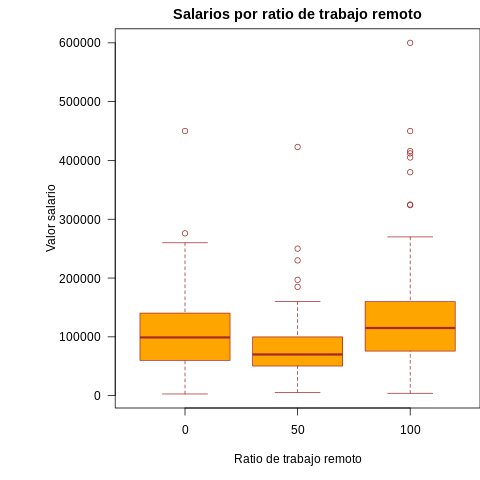
Los salarios de canadienses fueron diferentes significativamente con casi todos los demás países excepto Dinamarca y Gran Bretaña, esto indica que los salarios de canadienses son también muy altos pero inferiores a los de estadounidenses.

Los salarios de los Indios parecen diferenciarse significativamente con los de los 4 países antes mencionados y Francia, lo que indica que son bastante inferiores.

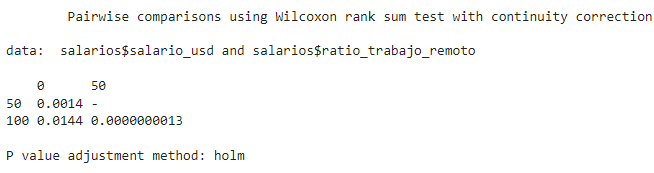
Para el caso de los salarios de Pakistaníes presentan diferencias significativas con los de estadounidenses, canadienses y daneses, indicando que son bastante bajos pero no presentan diferencias con los de indios.

El resto de salarios, franceses, daneses, portugueses, griegos, británicos y españoles, de manera general presentan diferencias con alguno de los países fuera de su continente.

Salario - Ratio de trabajo remoto



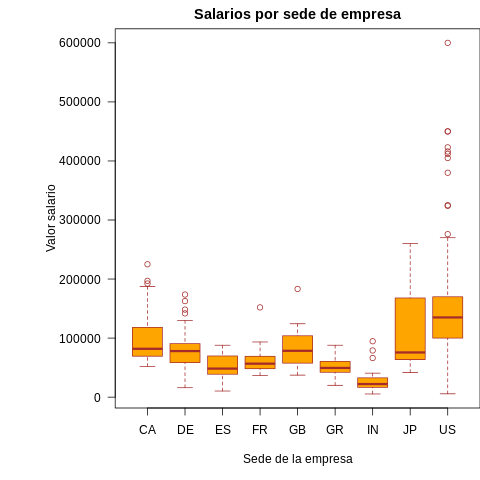
Gráficamente pareciese existir diferencia entre los salarios de trabajadores que realicen su trabajo de manera semi-remota con respecto a las otras modalidades que parecen similares.



En conclusión, existe evidencia para afirmar que los salarios entre trabajadores con diferentes ratios remotos tienen diferentes sueldos. Siendo mayor el de los que realizan trabajos remotos y menor el de los que lo hacen de manera semi presencial.

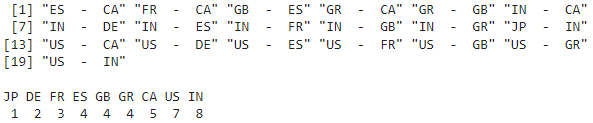
Salario - Sede de la empresa

Para este caso se analizarán sólo países los cuales haya más de 5 observaciones para simplificar el análisis y evitar sacar conclusiones de un grupo del cuál se posee pocas observaciones.



Gráficamente se observan mayores diferencias entre las empresas de India, Japón y Estados Unidos

Como los valores son bastantes se seleccionan las combinaciones que hayan presentado un p-value menor a 0.05 para encontrar las que son significativamente diferentes



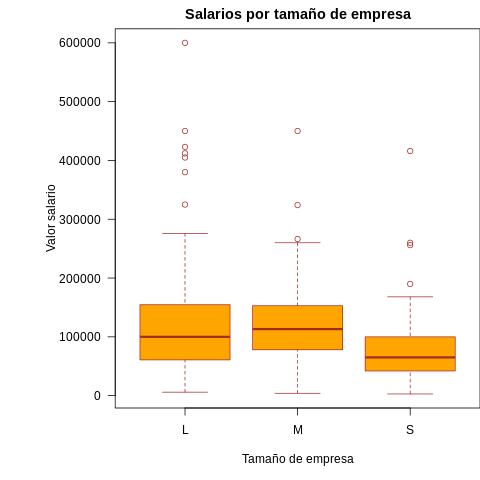
Solo se encontró diferencia significativa en los salarios de las empresas con sede en Japón con las de India, la falta de diferencia con las demás observaciones puede deberse a la gran variedad en los sueldos de trabajadores de empresas japonesas, lo que puede visualizarse gráficamente en la gran amplitud de la caja en la gráfica

Se observa que los salarios de trabajadores de empresas con sede en India poseen diferencias significativas con todos los demás, se puede decir que sus salarios son los más bajos.

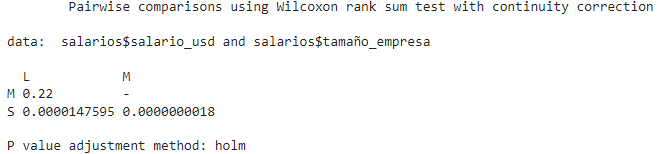
De los salarios de empresas con sede en Estados Unidos se encontraron diferencias significativas con casi todas las otras observaciones excepto con Japón. Se puede decir que las empresas con esta sede tienen trabajadores mejor remunerados. Algo similar sucede con las empresas con sede en Canadá

Para este caso sucede algo similar a lo ocurrido en la residencia del empleado, las empresas con sede en América del norte tienen trabajadores mejores remunerados que en Europa. Esto puede explicarse con el hecho de que es probable que la gran mayoría de gente que trabaja en una empresa tenga residencia igual a la sede de la empresa, esto se comprobará más adelante.

Salario - Tamaño de empresa

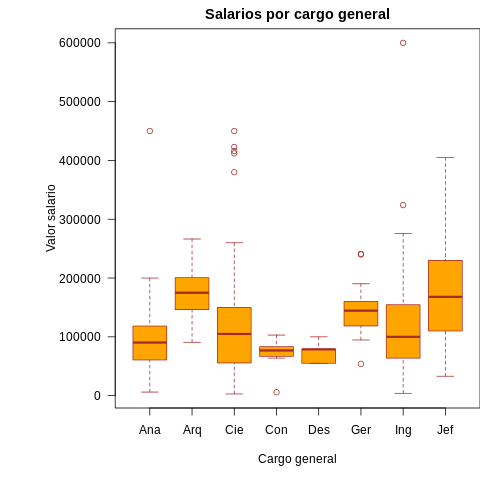


Gráficamente parece existir diferencia entre los salarios de empresas pequeñas con los demás



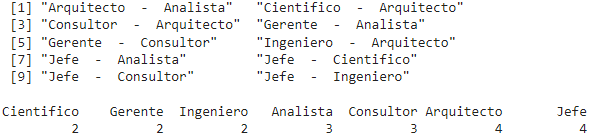
En conclusión, existen diferencias entre los salarios de empresas pequeñas con las medianas y grandes, pero no entre las grandes y medianas.

Salario - Cargo general



Gráficamente parece existir diferencia entre la mayoría de salarios por cargo

Como los valores son bastantes se seleccionan las combinaciones que hayan presentado un p-value menor a 0.05 para encontrar las que son significativamente diferentes

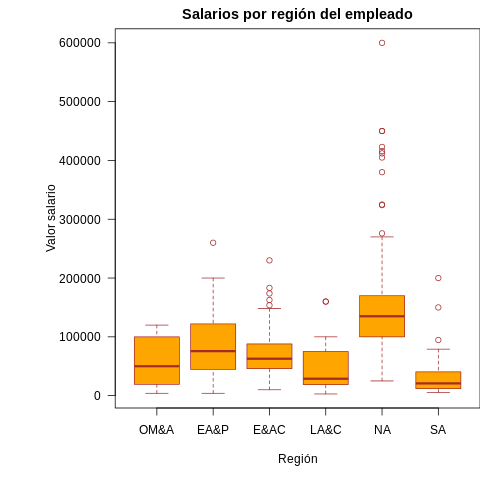


En conclusión se observa que los cargos que presentan mayor diferencias son los de arquitecto y jefe lo que indicaría que los trabajadores que ocupan esos cargos son los mejores remunerados.

Los salarios de trabajadores con cargos de científicos se diferencian en solamente con los de arquitectos y jefes lo que podría indicar que no son de los mejores remunerados.

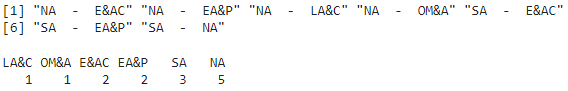
Salario - Región de residencia del empleado

En pasos posteriores se repitió el análisis de la residencia de empleados y sede de las empresas con el salario pero agrupando estás variables por continente.



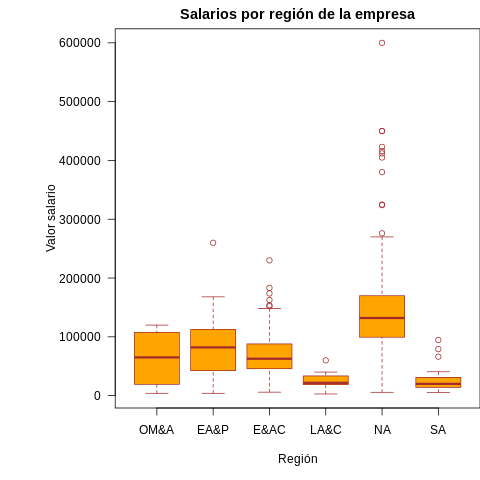
Gráficamente parece haber una diferencia significativa entre los salarios por región, principalmente con la región de norteamérica y el sur de Asia. Se puede observar también que los outliers generales observados al analizar de manera univariada al salario que tenían valor mayor a 300000 se encuentran en la región de norteamérica.

Resultados de test suma de rangos de Wilcoxon filtrados por ser presentar p-value significativo:



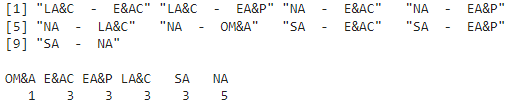
En conclusión, existe evidencia para afirmar diferencias entre Norteamérica y todas las demás regiones, indicando que los trabajadores con residencia allí tienen los salarios más altos. También hay evidencia de diferencia de la región del sur asiatico con la de Europa & Asia central y el este de Asia & Asia-pacifico lo que indica que los salarios allí son bajos en comparación con los de estas regiones.

Salario - Región de la sede de empresa



Gráficamente se observa algo similar a lo ocurrido anteriormente pero aún más marcada la diferencia entre las empresas con sede en norteamérica y las que están en latinoamerica & el caribe y el sur de Asia

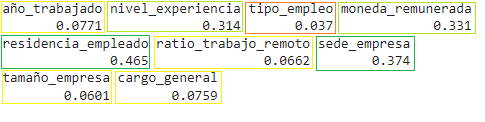
Resultados de test suma de rangos de Wilcoxon filtrados por ser presentar p-value significativo:

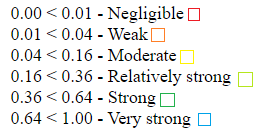


En conclusión, se observan diferencias similares a las de la comparación anterior, existe evidencia para afirmar diferencias entre Norteamérica y todas las demás regiones, indicando que los trabajadores de empresas de allí tienen los salarios más altos. También se ven diferencias entre algunas regiones con las de Latinoamerica & Caribe y el sur Asiatico que parecen ser de allí las empresas con peor remuneración.

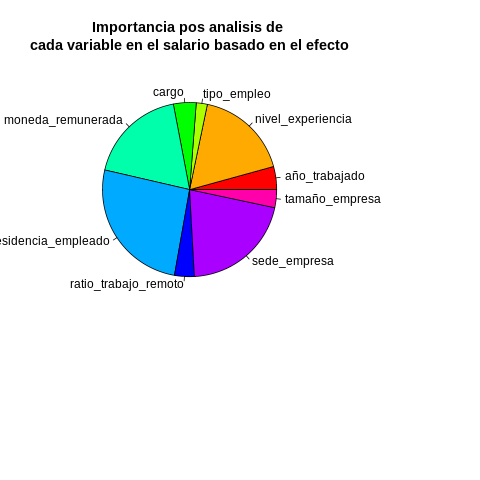
**Análisis de efectos**

Valores de efectos de cada variable con el salario, medido con epsilon cuadrado:

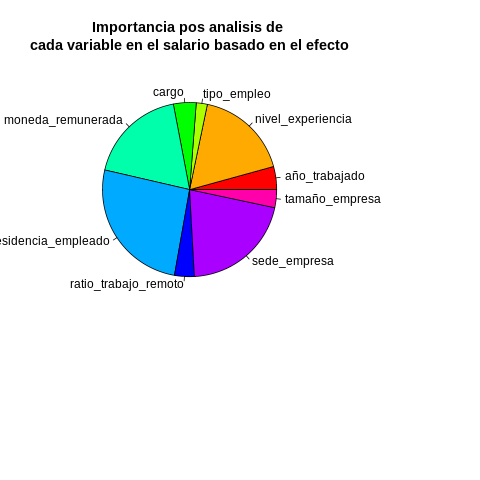




Realizando un gráfico de torta similar al del pre análisis con estos valores se obtiene lo siguiente:

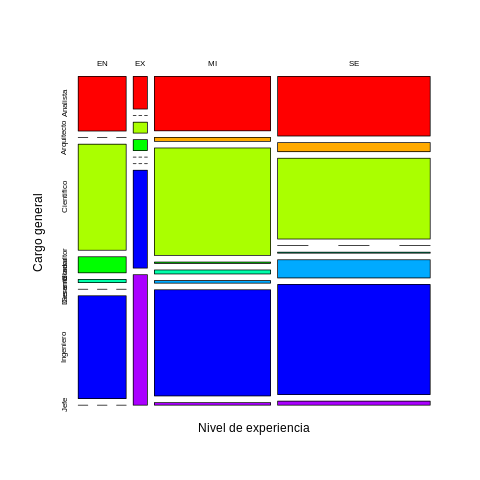


Comparando con la anterior:



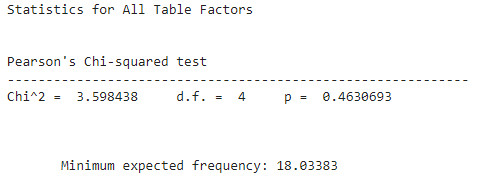
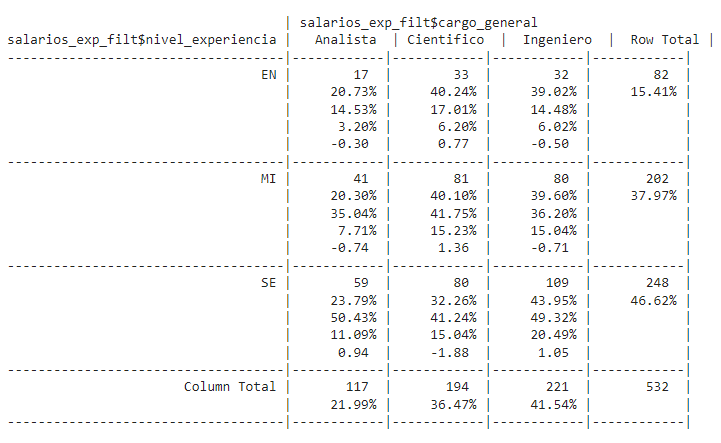
**Relaciones cualitativas - cualitativas**

Nivel de experiencia - Cargo general

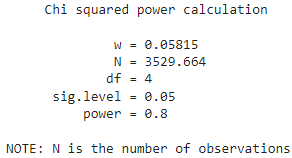
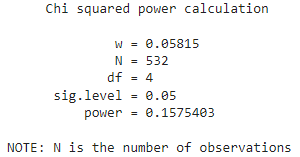


Gráficamente se puede ver que analistas y ingenieros tienen niveles de experiencia muy variados, los jefes son casi en su totalidad de nivel de experiencia ejecutivo, los científicos están distribuidos de manera equitativa en los niveles de experiencia excepto en el nivel ejecutivo y los arquitectos son mayoritariamente de nivel senior e intermedio. Del resto de cargos hay pocas muestras y no puede verse claramente.

Se toman los 3 cargos de mayor cantidad de datos, es decir, científico, ingeniero y analista para intentar realizar un test chi cuadrado ya que pueden presentar frecuencias esperadas mayores a 5, también se eliminan los datos de trabajadores de nivel de experiencia ejecutiva por presentar frecuencias esperadas bajas.



Basándose en el valor del test chi cuadrado se puede decir que no existe asociación entre las variables.

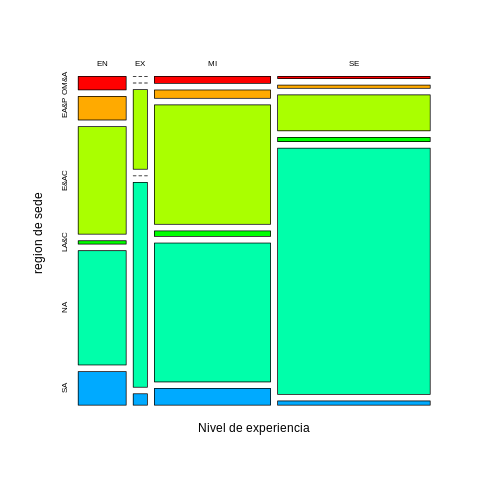


Ejecutando cálculos de potencia se ve que la potencia, es decir, la probabilidad de detectar una diferencia si existiese es de 15%, y para tener una potencia de al menos 80% se necesita una cantidad cercana a 3600 datos.

En conclusión, el nivel de experiencia no parece influir en el cargo general que un trabajador ocupa, algo que resulta contraintuitivo, puede deberse a las simplificaciones que hubo que hacer para poder utilizar el test chi cuadrado debido a las bajas frecuencias esperadas.

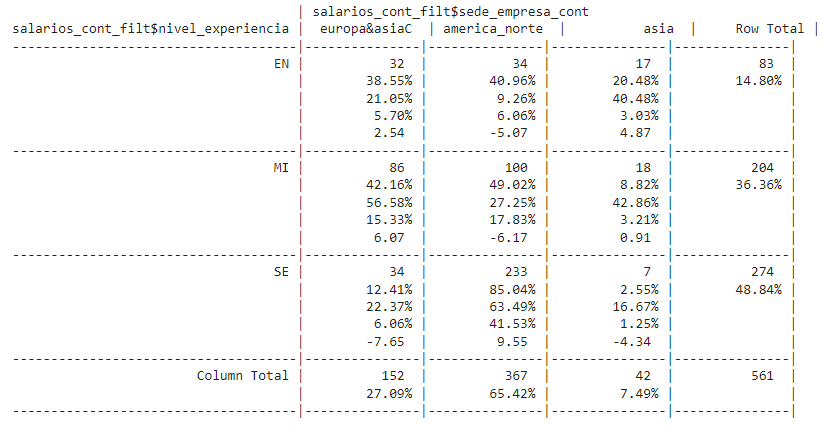
Nivel de experiencia - Región de la sede de empresa

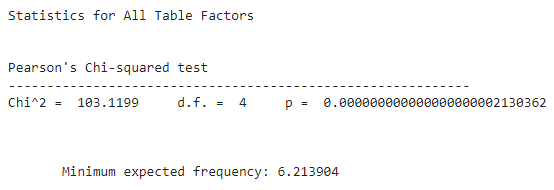
En estos casos se realizaron análisis con las sedes y residencias por país y continente, presentando finalmente las hechas por continente por resultar más interesantes y abarcativas.



Gráficamente se puede ver que en empresas de Estados Unidos predominan trabajadores con niveles de experiencia de mayor jerarquía como senior o ejecutivos. En empresas de Europa & Asia central no parece haber un tipo de experiencia que predomine pero si hay pocos de nivel senior. Los trabajadores de tipo ejecutivo tienen la distribución más desbalanceada, concentrados casi en su totalidad en empresas de norteamérica y Europa & Asia central. Los trabajadores senior parecen concentrados principalmente en norteamérica.

Como hay celdas con una frecuencia esperada menor a 5 se agrupan algunas categorías por continente. Se eliminan los datos de latinoamérica por considerarse muy distintos para agruparlos con América del norte, y los de Oriente medio & África por no poder agruparse de manera lógica. También se borran los datos de trabajadores con nivel de experiencia ejecutiva.



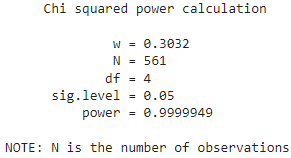


Basándose en el valor del test chi cuadrado se puede decir que existe asociación entre las variables. Se observa que en cada región predomina un nivel de experiencia, en Asia los de nivel junior, el Europa & Asia Central los de nivel intermedio y en América del norte los de nivel senior.





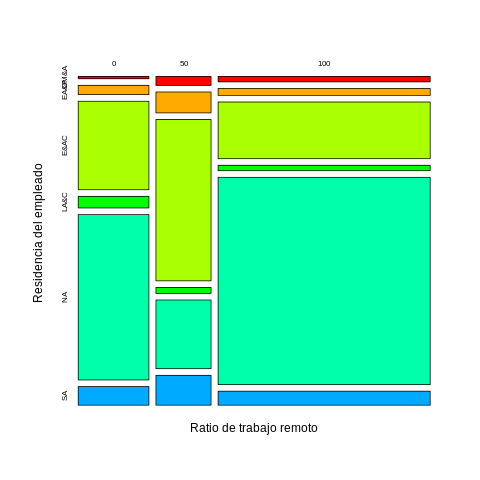
En base al coeficiente de contingencia se puede decir que la dependencia entre las variables es media y según el coeficiente de Cramer con 4 grados de libertad las variables poseen una asociación grande.



Para el caso de la potencia de esta prueba se puede estar casi seguro que la diferencia encontrada no es azarosa.

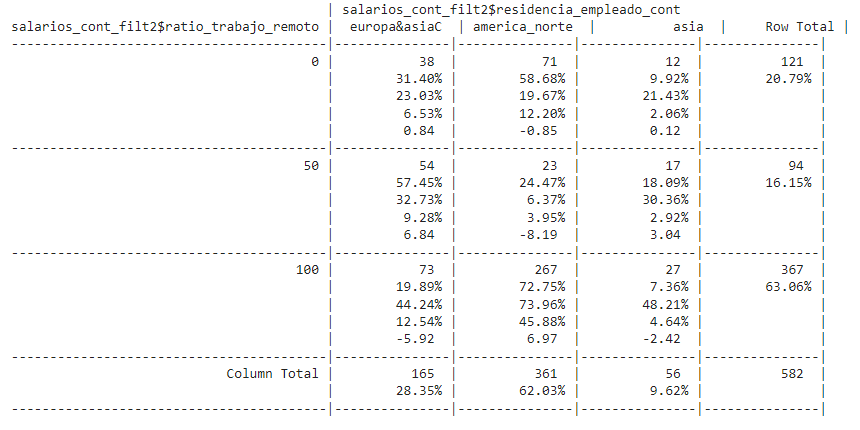
En conclusión, las empresas con sede en América del norte tienen gran cantidad trabajadores de nivel senior y pocos de nivel intermedio y junior en comparación con las de Europa & Asia Central y Asia, donde en Europa & Asia Central se encontró una predominancia por trabajadores de nivel intermedio y una también una frecuencia elevada en trabajadores de nivel junior, finalmente en Asia predominan los trabajadores de junior, también hay una gran cantidad de trabajadores de nivel intermedio y muy pocos de nivel senior.

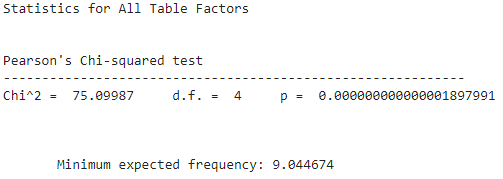
Ratio de trabajo remoto - Región de la residencia del empleado



Gráficamente se puede ver que trabajadores residentes de Estados Unidos preferentemente trabajan de manera remota o presencial y no de manera semi-presencial, todo lo contrario a los trabajadores de Europa & Asia Central. En el resto de países de residencia la distribución del ratio de trabajo remoto es casi uniforme.

Se realiza un agrupamiento de los datos de manera similar al caso anterior



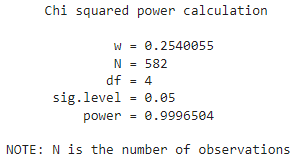


Basándose en el valor del test chi cuadrado se puede decir que existe asociación entre las variables. Se observa que es significativamente superior la cantidad de trabajadores de América del norte que realizan su trabajo de manera remota e inferior la que lo hace de semi remota, en cambio en los de Europa & Asia central ocurre lo contrario, esto también ocurre con los trabajadores de Asia pero en menor medida.





En base al coeficiente de contingencia se puede decir que la dependencia entre las variables es media y según el coeficiente de Cramer con 4 grados de libertad las variables poseen una asociación grande.



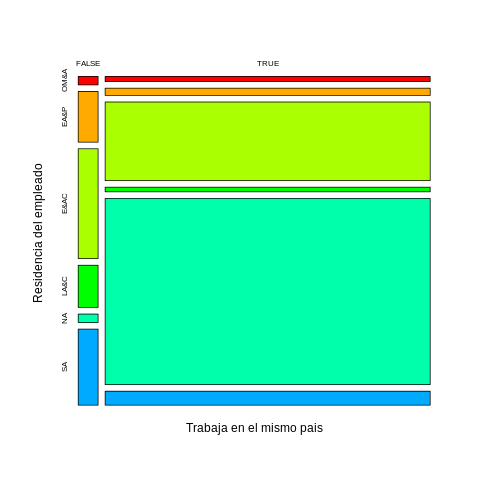
En cuanto a la potencia del test se observa que es casi 1, lo que indica que la probabilidad de cometer un error y que no exista diferencias significativas es casi nula.

En conclusión, los trabajadores de América del norte tienden a trabajar de manera remota por sobre semi remota y los de Europa & Asia central y Asia lo hacen de manera semi remota sobre hacerlo de manera remota.

Región de la sede de empresa - Región de la residencia del empleado

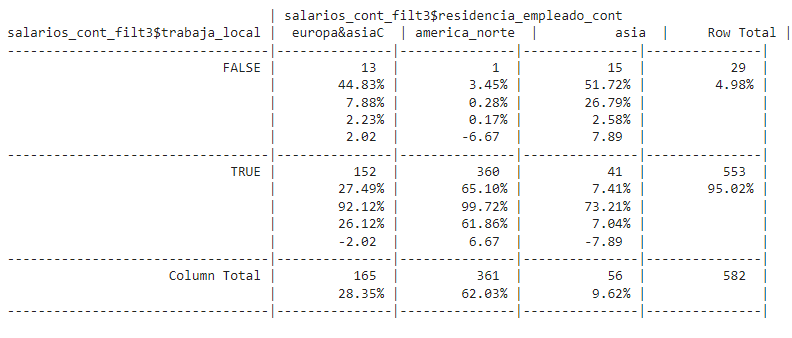
Para este caso se analizará la frecuencia con la que los trabajadores optan por ejercer en una empresa con sede en una región diferente al de su residencia. Se contemplarán todos los trabajadores independientemente de si trabajan de manera remota o no.

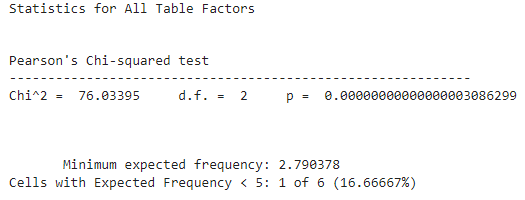
Antes que nada se creará una nueva variable que sea “trabaja\_local” que indicará si la residencia del empleado es igual a la sede de la empresa donde trabaja.



Gráficamente se puede ver que trabajadores que en casi todas las regiones se opta por trabajar fuera de su región, excepto en Norteamérica donde ocurre lo contrario, esta tendencia es menor en las regiones de Europa & Asia Central y Oriente medio & África

Se agrupan nuevamente algunas regiones





No se puede ejecutar el test debido a la baja frecuencia esperada y tampoco es viable la ejecución del test de fisher ni un agrupamiento lógico. Se puede concluir en base a la tabla de contingencias que los trabajadores que viven en Asia suelen ser los que tienden a trabajar para empresas fuera de su país, caso totalmente contrario para los trabajadores de Norteamérica.

## 

## 

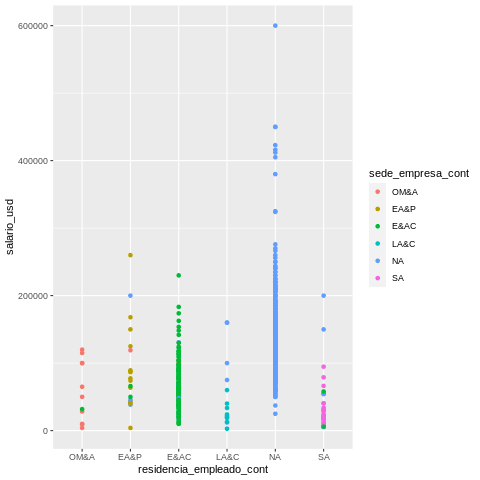
## 

## 

## Análisis multivariado

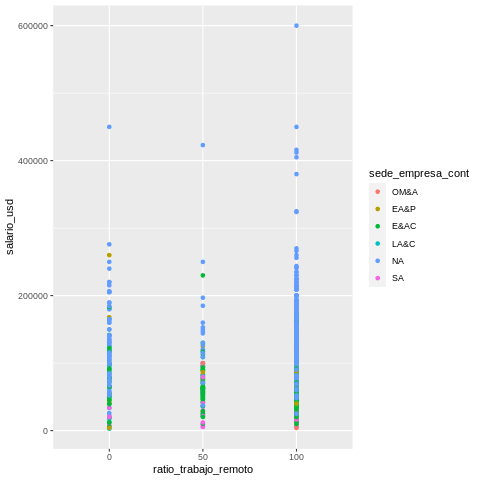
Salario - Región de la sede de empresa - Región de la residencia del empleado

Una opción sería contrastar la variable salario junto con la residencia del empleado y la sede de la empresa, dos variables que presentaron los mayores efectos.



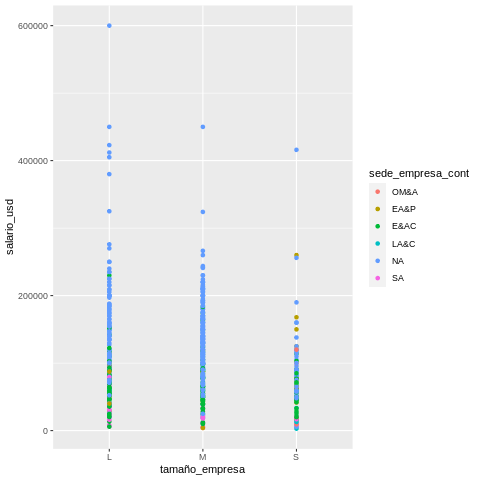
En el gráfico se puede confirmar lo que ya se venía observando en análisis anteriores, los trabajadores de Norteamérica tienden a trabajar en empresas de la misma región donde también los sueldos son de los más altos, también se observan varias observaciones de trabajadores que optan por trabajar en este país aunque no residen allí, sucede algo similar también con la región europea & Asia central, en estos países ningún trabajador opta por trabajar fuera de su región. No sucede esto en las regiones restantes donde siempre hay al menos un trabajador que trabaja fuera de su región en Europa & Asia central o en Norteamérica y en múltiples casos estos salarios son los más altos percibidos por trabajadores de la región.

Salario - Región de la sede de empresa - ratio de trabajo remoto



Recordando que al momento de analizar el ratio de trabajo remoto con el salario resultaron significativos los salarios de trabajadores de modalidad semiremota, aqui puede verse que tal vez la diferencia no se debe a la modalidad sino más bien a que hay pocos salarios en ese grupo que pertenezcan a trabajadores de empresas norteamericanas que son los que más salario perciben.

Salario - Región de la sede de empresa - tamaño de empresa



Algo similar a lo que sucede con el ratio de trabajo remoto sucede también con el tamaño de la empresa, pareciera que la mayoría de empresas en Norteamérica son de tamaño grande y mediano, lo que sesga los salarios a mayores cantidades que en el caso de las pequeñas.

## Conclusión

Recapitulando al objetivo inicial, encontrar características de los salarios mejores pagos, se concluye en base a lo analizado que los salarios mejores pagos son de trabajadores que trabajan para empresas Norteamericanas, esto también está apoyado por la gran cantidad de gente que opta por trabajar en empresas de ese país aunque no residan allí y que los residentes de allí no optan por trabajar en otras regiones.

También se observaron salarios mejores pagos a trabajadores de nivel de experiencia superior, como cabe esperar, está variable resultó mucho más significativa que el cargo que ocupan dentro de la empresa.

El año trabajado tuvo cierto efecto también pero puede deberse al contexto de pandemia, ya que los salarios de 2020 y 2021 no tuvieron diferencias y solo en 2022 subieron un poco que fue el momento en que la pandemia ya casi no tuvo efectos en la economía.

El ratio de trabajo remoto tuvo cierto efecto en el salario aunque fue menos significativo que lo antes mencionado y se observa que es más algo que está relacionado con la preferencia de trabajadores por región, algo similar con el tamaño de la empresa.

Finalmente se observó que la moneda remunerada tuvo un efecto bastante más alto de lo esperado, surge la interrogante de si esto se puede deber a la elección del dólar como la m

moneda estándar para comparar los salarios.