1.5

2.8

3.5

Los sueldos de 1@s programadores/as en Argentina: análisis y estimadores del 2021

Florencia Molina*, Fernando Cardellino, Nicolás Ferrero, Rodrgo Pizarro and Fernando Carazo

1. Análisis

En la Figura 1 se grafican los valores y las distribuciones estadísticas de las muestras de los sueldos correspondientes a los grupos empleados en el estudio: (i) hombres, (ii) mujeres y (iii) otros. En la Figura 1a se muestran los valores de algunos indicadores estadísticos de los sueldos netos de los tres grupos mencionados. Como se observa, la mayoría las variables aleatorias (tamaño de la muestra) corresponden a hombres, luego mujeres y en menor cantidad a otros. Del análisis de la tabla se desprende que la media (μ) correspondiente a los mayores sueldos netos se da para el grupo otros; no obstante esto, al analizar la desviación y los valores de los cuartiles se observa que esto se debe a que la muestra no es representativa ya que son pocos los datos. Respecto a los dos grupos restantes, como se mencionó los hombres presentan mayor cantidad de datos relevados y una media considerablemente mayor a las mujeres, siendo la desviación estándar (S) mayor en la muestra de hombres que en la del grupo mujeres.

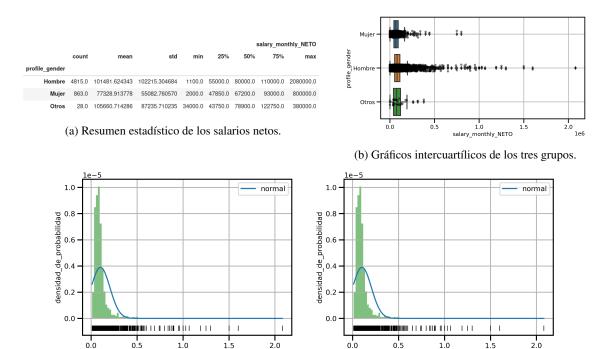


Figura 1. Salarios netos correspondientes al año 2021 extraídos de [1].

(d) Mujeres y otros.

Un desgloce del conjunto de variables de las muestras aleatorias correspondientes a los tres grupos han sidos graficados en la Figura 1b. Al analizar esta figura se aprecia la diferencia de variables en cada muestra y/o grupo (la diferencia de las medias no se distingue debido a un problema de escala). La muestra aleatoria de los sueldos netos de los hombres presenta una gran cantidad de valores mayores al percentil Q3 y outliers o valores anómalos.

(c) Hombres.

^{*}Autor para correspondencia. E-mail: florencia.molina.756@mi.unc.edu.ar.

1.5

En tanto en los grupos mujeres y otros los valores están contenidos entre los percentiles Q1 y Q3 (percentiles 0.25 y 0.75 respectivamente).

1.5

2.8

3.5

En las Figuras 1c y 1d se grafican las distribuciones de los salarios netos correspondientes a los dos grupos seleccionados en los siguientes análisis: hombres (Figura 1c) y mujeres y otros¹ (Figura 1d). Del análisis de estas figuras se desprende que ambas distribuciones presentan asimetría hacia la derecha (lo cual corrobora lo mostrado en la Figura 1b), no correspondiéndose a una distribución normal o gaussiana. El análisis realizado refleja que en los dos grupos hay muchos programadores con sueldas bajos y pocos con sueldos altos, siendo mayor en el grupo hombres que en el grupo mujeres+otros (esto, junto a los valores anómalos de la variable sueldo neto de los programadores en Argentina, también se observa en al Figura 1d). Por otro lado, las medias y las desviaciones (varianzas) son diferentes.

2. Estimadores

Si bien el análisis y las medidas de tendencia central, tendencia y de dispersión hecho en la Sección 1, es una alternativa para comparar la variable sueldos de programadores entre los grupos hombres y mujeres+otros en Argentina, el hecho de que $\mu_X \neq \mu_Y$, con X: grupo hombres e Y: grupo mujeres+otros, no es una evidencia de una diferencia significativa. Esto se debe a que, como se observó, cada muestra tiene su propia variabilidad. Para determinar si la diferencia observada entre μ_X y μ_Y es significativa, a continuación se determinarán estimadores puntuales y por intervalos basados en Z-scores (cuantiles de la distribución normal) o en la distribución T-student (cuantiles de la distribución t-student) para, contemplando la variabilidad de la diferencia de μ_X y μ_Y y un dado nivel de significancia, podamos determinar los intervalos de confianza para conocer entre que valores se encontrará la diferencia real de las medias poblacionales de las muestras analizadas en este artículo².

El estimador puntual basado en $(\mu_X - \mu_Y)$ resultó \$23262.37. Si bien este valor da una idea de la diferencia de los sueldos netos en las poblaciones de los programadorse hombre y mujeres+otros, como se mencionó resulta interesante determinar un intervalo de valores que contenga la mencionada diferencia. Dado que $\mu_X \neq \mu_Y$, $S_X \neq S_Y$ y $n_X \neq n_Y$ (ver Sección 1), la estimación del intervalo de confianza para la diferencia de medias muestrales se hará aplicando el Método de Welch [2]. El Método de Welch se basa en una distribución T-student y el cálculo de los estadísticos que determinan los límites del intervalo de confianza implica conocer los grados de libertad de la distribución mencionada. En el caso de estudio los resultados encontrados fueron:

- estadístico de prueba: 9.69,
- número de grados de libertad de T-student: 2148.05.

Teniendo en cuenta que los grados de libertad son considerablemente mayores a 30, el estadístico a emplear en los cálculos corresponderá a una distribución normal [3]. El valor de $t_{\frac{\alpha}{2},\nu-1} \sim z_{\frac{\alpha}{2}}$ para $\alpha=0.05$ (nivel de significancia) y $\nu\sim 2148$ (grados de libertad) vale $z_{\frac{\alpha}{2}}=1.96$, por lo que los límites del intervalo de confianza estarán dados por $(\bar{X}-\bar{Y})\pm 1.96\sqrt{\frac{S_X^2}{2}+\frac{S_Y^2}{2}}$.

 $(\bar{X}-\bar{Y})\pm 1.96\sqrt{\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}}$. Finalmente el intervalo de confianza hallado corresponde a $1-\alpha=[18560.58,27964.17]$. Teniendo en cuenta que se definió un nivel de significancia del 5% la $\underbrace{P}_{\text{Declebilidad}}\left[(\mu_1-\mu_2)\in(\hat{\mu}_1-\hat{\mu}_2)\right]=1-\alpha=1-0.05=0.95$. Esto

implicará que en el 95% de las realizaciones la diferencia de la media poblacional estará contenido en el intervalo encontrado.

Referencias

- [1] Sysarmy, Resultados de la Encuesta de sueldos 2022, 2022, Último acceso, 25/04/2022. https://sysarmy.com/blog/posts/resultados-de-la-encuesta-de-sueldos-2022-1/.
- [2] R.R. Wilcox, Applying Contemporary Statistical Techniques, in: Comparing two independent groups, Academic Press, 2003, pp. 237–284. ISBN 9780127515410.
- [3] C.B. D. Diez M. Cetinkaya-Rundel, OpenIntro Statistics, 4th edn, OpenIntro, Inc, St. Louis. USA, 2019.

¹Teniendo en cuenta lo mencionado anteriormente, el grupo otros se añadió al grupo mujeres.

²Alternativamente también puede aplicarse un test de hipótesis para determinar cuan significativa es la diferencia entre las variables mencionadas