Departamento de Ingeniería Civil Industrial
Estadística para la Economía y Gestión - IN3401
Prof.: Constanza Contreras Piña
Aux's: Paz Montaño - Rubén Ortega - Camila J. Pulgar
Carolina Salgado - Florencia Signorini

INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

Tarea #1

Probabilidades - Test de Hipótesis

En la presente tarea intentaremos dar respuesta a la pregunta **¿Por qué renuncian los empleados?**. Algunas de las causas que normalmente se plantean son la búsqueda de nuevas oportunidades, mal ambiente o cultura organizacional, exceso de horas extras no remuneradas, mal o falta de liderazgo de los superiores, entre otras. Para realizar este análisis contamos con datos de 1470 trabajadores de una misma empresa en el dataset *employee_attrition.csv*. En estos datos la variable de interés se llama *attrition*, la cual toma dos valores: *yes* o *no*. Para los empleados que renunciaron a la compañia (variable *attrition* igual a *yes*), los valores que se presentan en las otras variables corresponden a los atributos que presentaba la persona en el momento que dejó la empresa.

Existen varias dimensiones para analizar. Puede escoger alguna de las siguientes o proponer otra:

- **Socio-demográfica:** ¿Existen diferencias entre hombres y mujeres? ¿Afecta la Edad en el hecho de renunciar? ¿Son iguales los niveles de renuncia en todos los niveles de educación? ¿Influye el hecho de vivir cerca o lejos del trabajo?
- **Ambiente organizacional:** ¿Hay diferencias en las renuncias a lo largo de los distintos roles y/o en una misma división? ¿Cambia la satisfacción con el trabajo la cantidad de años con el mismo gerente? ¿Trabajadores con gerentes nuevos presentan mayores niveles de satisfacción? ¿Cómo afecta el balance entre trabajo y vida social?
- **Desempeño:** ¿Cómo afecta la cantidad de años desde la última promoción en las renuncias? ¿Hay diferencias según salarios en la satisfacción por el trabajo y/o las renuncias? ¿Existen diferencias según la cantidad de años en la compañia, horas extras, nivel educacional? ¿Los empleados mejor evaluados presentan menor proporción de renuncias?

Pregunta 1

1. Seleccione 3 variables que considere le permiten caracterizar a los empleados del conjunto de datos y grafíquelas.
2. Seleccione 3 variables que considere que pueden explicar el comportamiento de la variable *attrition*. Explique el por qué de su elección y que comportamiento espera de ellas. Grafique cada variable contrastandola con la variable*attrition*.
3. Realice dos gráficos en los cuales contraste dos variables del dataset que usted crea puedan estar correlacionadas o poseer un patrón interesante entre ellas. Explique su elección y comente los gráficos.

Ahora, vamos a analizar el salario mensual de los trabajadores de esta empresa. Para esto debe trabajar con la variable *MonthlyIncome*.

4. Calcule la probabilidad de que un empleado en la empresa tenga un salario mensual mayor a 9.000 USD. Calcule la probabilidad de que un empleado en la empresa tenga un salario mensual menor a 2.000 USD.

5. Supongamos ahora que el salario mensual de los empleados es una variable aleatoria, que denotaremos W, que distribuye de la siguiente forma:

$$W \sim Lognormal(\bar{X}_{MI}, \hat{\sigma}^2_{MI})$$

   Donde $\bar{X}_{MI}$ es la media muestral de la variable *MonthlyIncome* y $\hat{\sigma}^2_{MI}$ corresponde a la varianza muestral de dicha variable. ¿Cuál es la probabilidad de que un empleado gane menos de 2.000 USD en este caso? ¿Y de que gane más de 9.000 USD?

6. Compare lo obtenido en la parte 4 y 5. ¿Se parecen estas probabilidades? Utilice un gráfico qqplot para comparar la distribución de la variable *MonthlyIncome* contra una distribución lognormal.

Pregunta 2

De las dimensiones mencionadas en este enunciado, o alguna otra de su preferencia:

1. Plantee su objetivo de estudio (pregunta de investigación). Indicando que variables cree que afectan y cuáles son sus resultados esperados. Utilice los gráficos realizados en la pregunta 1 para argumentar. Puede buscar investigaciones al respecto para respaldar también su objetivo.

2. Realice 4 test de hipótesis, señalando explícitamente su hipótesis nula y alternativa, el nivel de significancia y la conclusión del test. Debe utilizar **al menos dos tipos de test diferentes** vistos en cátedra (es decir, no puede realizar 4 test de diferencia de medias).

3. Analice sus resultados y concluya.

Reglas de la Tarea:

- **Fecha de entrega:** Martes 01 de junio de 2021 hasta las 23:59. No se aceptan atrasos.
- **Entregables:** Debe entregar un reporte con sus resultados y los códigos utilizados. El código debe correr sin errores. Además deberá entregar completado el archivo *Evaluacion_Tarea1.xlsx* agregandole su nombre al nombre del archivo. Ejemplo: *Evaluacion_Tarea1_ConstanzaContreras.xlsx*.
- **Lenguajes permitidos:** R y Python.
- **Formato de entrega:** Puede entregar un archivo PDF con el reporte y un archivo .R o .py con los códigos, o puede entregar un archivo .ipynb, .rmd (con el .html respectivo) que contenga reporte y códigos.
- Esta Tarea puede realizarla de manera individual o en parejas.
- **Auxiliares Encargados:** Carolina Salgado y Paz Montaño.
- Favor realizar las preguntas sobre la Tarea en el foro de U-cursos para que todo el curso posea la misma información.
- **Nota Tarea:** La nota de la presente Tarea será ponderada de acuerdo a la pauta de evaluación que entregue usted y su pareja. Se evaluarán 3 dimensiones en cada pregunta: diseño de la tarea, análisis de resultados/conclusiones y código. Para más detalles revisar el archivo *Evaluacion_Tarea1.xlsx*.