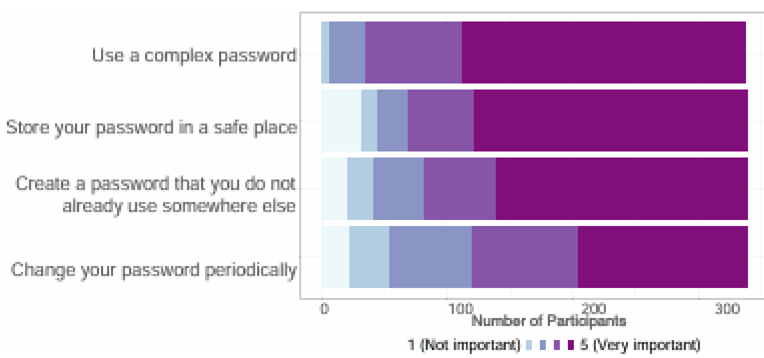
Prueba de hipótesis.

Para comprender el grado de homogeneidad o dispersión en las percepciones de los usuarios respecto a las prácticas recomendadas de seguridad en contraseñas, es relevante calcular medidas de variabilidad como la varianza. En particular, el estudio (archivo adjunto en el repositorio de GitHub) de Habib et al. (2018) presenta datos categorizados mediante una escala de Likert, en la cual se evaluó la importancia asignada por los participantes a diversas acciones relacionadas con la gestión de contraseñas, tales como cambiarla periódicamente, usar contraseñas complejas o almacenarlas de forma segura.

Este tipo de datos permite realizar un análisis cuantitativo complementario, al asignar valores numéricos a las categorías y calcular promedios y dispersiones. La varianza, en este contexto, permite identificar qué tan consistentes o divergentes son las opiniones de los usuarios respecto a una práctica específica, como el cambio periódico de contraseña. Dicha información es clave para interpretar el nivel de concienciación, la aceptación de políticas de seguridad y para establecer referencias válidas para contrastes estadísticos.

A partir de la distribución observada en las respuestas, es posible estimar una varianza poblacional que sirva como base para pruebas de hipótesis y análisis comparativos con otras muestras o poblaciones.




A partir del conteo visual de la gráfica, se estimaron aproximadamente los siguientes valores para la afirmación “Change your password periodically”.

| **Valor Likert ($x_i$)** | **Frecuencia ($f_i$)** | **$f_i * x_i$** | **$f_i * (x_i - \bar{x})^2$** |
|---|---|---|---|
| 1 | 20 | 20 | 143.93 |
| 2 | 35 | 70 | 99.12 |
| 3 | 60 | 180 | 28.04 |
| 4 | 90 | 360 | 9 |
| 5 | 95 | 475 | 164.73 |
| **Total** | **300** | **1105** | 444.82 |

Entonces,

$$x = \frac{\sum x_i f_i}{\sum f_i} = \frac{1105}{300} = 3.6833$$

$$s^2 = \frac{1}{n-1}\sum f_i (x_i - x)^2$$

$$s^2 = \frac{444.82}{299} = 1.48$$

Ahora con el conocimiento de este dato, proseguimos a lo siguiente:

CALCULO DE VARIANZA Y DESVIACION ESTANDAR

X = {Frecuencia de cambio de contraseñas}

| Clase | FA | Media de clase ($X_i$) | $(Xi - X)^2$ | $FA \cdot (Xi - X)^2$ |
|---|---|---|---|---|
| Cada mes | 1 | 30 | 0.8649 | 0.8649 |
| Cada 3-6 meses | 11 | 30.5 | 0.1849 | 2.0341 |
| Rara vez | 39 | 31 | 0.0049 | 0.1911 |
| Nunca | 8 | 31.5 | 0.3249 | 2.5992 |
| **Totales** | **59** | | | **5.6893** |

| Parámetro | Valor |
|---|---|
| Media general ($\bar{X}$) | 30.93 |
| Varianza muestral ($s^2$) | 0.0981 |
| Desviación estándar (s) | 0.3132 |

| Clase | FA | Media de Clase ($X_{1i}$) |
|---|---|---|
| Cada mes | 1 | 30 |
| Cada 3-6 meses | 11 | 30.5 |
| Rara vez | 39 | 31 |
| Nunca | 8 | 31.5 |
| **Total** | 59 | — |

Con base en los datos recolectados en la frecuencia con la que los usuarios cambian sus contraseñas. Ahora se desea determinar si la varianza es igual a un valor de referencia esperado. Este valor de referencia está determinado en la aceptación de políticas de seguridad y de estudios previos realizados. Para este análisis, se considera como valor de la varianza igual a 1.48 con un nivel de significancia igual a 0.05.

Entonces:

$$H_0: \sigma^2 = 1.48$$
$$H_A: \sigma^2 \neq 1.48$$

Datos:

$$n = 59$$
$$nivel\ de\ significancia = 0.05$$
$$s^2 = 0.0981$$
$$grado\ de\ libertad = 59 - 1 = 58$$

Calculando el estadístico de la distribución chi-cuadrada:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(59-1)(0.0981)}{1.48} = 3.8444$$

Calculando $\frac{\alpha}{2}$

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025;\ \frac{\alpha}{2} = 1 - \frac{0.05}{2} = 0.975$$

Calculando valores críticos

$$\chi^2_{58,0.025} = 79.082$$
$$\chi^2_{58,0.975} = 39.363$$

Ahora:

$$H_0: 39.363 \; < 3.8444 < 79.082$$

Con un nivel de significancia del 5%, se rechaza la hipótesis nula. Esto indica que existe una diferencia estadísticamente significativa entre la varianza muestral observada (0.0981) y la varianza poblacional conocida (1.48).

Es decir, los datos muestran una variabilidad mucho menor que la esperada según el valor poblacional, lo cual sugiere una posible discrepancia estructural entre la muestra y la población.