**Objetivo**

Identificar **qué variables personales y laborales** están **más asociadas** con **altos niveles de absentismo** (medido por Absenteeism\_hours), y usar esa información para **diseñar políticas proactivas**.

**Entendimiento del modelo: Regresión lineal**

La **regresión lineal** estima la relación entre una variable dependiente (Absenteeism\_hours) y una o más variables independientes (edad, hijos, transporte, etc.).

La fórmula general sería:

Absenteeism\_hours = β0 + β1·Edad + β2·Hijos + β3·Distancia + ... + ε

Cada **coeficiente β** nos indica cuánto varía Absenteeism\_hours al cambiar esa variable, **manteniendo el resto constante**.

**Preparación del análisis**

**✅ Variables dependiente**

* Absenteeism\_hours (variable a predecir)

**✅ Variables independientes**

* Personales: Age, Son, Social\_drinker, Social\_smoker, Pet, Weight, Height, Body\_mass\_index
* Laborales: Transportation\_expense, Distance\_Residence\_Work, Service\_time, Education, Disciplinary\_failure, Hit\_target, Work\_load\_Average\_day

También puedes incluir dummies (variables binarias) para categorías como Reason\_absence, Month\_absence, Day\_week, Seasons.

**Interpretación general del modelo**

* **R-squared = 0.278**  
  → El modelo explica el **27.8%** de la variabilidad en las horas de absentismo. No es muy alto, pero en modelos sociales es razonable.
* **Adj. R-squared = 0.223**  
  → Ajustado por el número de variables. También es bajo. Puede haber ruido o demasiadas variables poco útiles.
* **F-statistic p-value = 1.44e-27**  
  → El modelo en su conjunto **sí es estadísticamente significativo**.
* **Nº de observaciones = 840**

→ Tamaño de muestra adecuado

* **Nº de variables = 59**

→ Modelo bastante completo

**Variables más influyentes (significativas)**

**Variables que REDUCEN significativamente el absentismo**

Estas variables tienen coeficientes **negativos** y **p-valor < 0.05**, lo que indica que están **asociadas con menos horas de absentismo**:

| **Variable** | **Coef.** | **p-valor** | **Interpretación práctica** |
| --- | --- | --- | --- |
| Disciplinary\_failure | -8.49 | 0.002 | Como antes, quienes tienen sanciones disciplinarias se ausentan menos, posiblemente por temor o mayor supervisión. |
| Education | -1.90 | 0.020 | A mayor nivel educativo, menos horas de absentismo. |
| Reason\_absence\_23 Consulta medica | -4.66 | 0.029 | Este motivo específico se relaciona con **menos absentismo**. |
| Reason\_absence\_28 Consulta odontológica | -5.84 | 0.009 | Igual que el anterior: este tipo de ausencia es “leve” o de corta duración. |
| Day\_week\_4 (miércoles) | -2.68 | 0.043 | Menos absentismo los miercoles. |
| Day\_week\_5 (jueves) | -3.95 | 0.006 | El jueves también muestra menos absentismo. |
| Day\_week\_6 (viernes) | -3.53 | 0.011 | viernes con menos ausencias, aunque dependerá si es laboral. |

**Variables que AUMENTAN significativamente el absentismo**

Estas variables tienen coeficientes **positivos** y **p-valor < 0.05**, lo que indica que están **asociadas con más horas de absentismo**:

| **Variable** | **Coef.** | **p-valor** | **Interpretación práctica** |
| --- | --- | --- | --- |
| Son | +1.14 | 0.017 | Tener hijos se asocia con más horas de ausencia. Podrían ser por cuidado o emergencias familiares. |
| Reason\_absence\_6 Enfermedad sistema nervioso | +10.83 | 0.019 | Este motivo específico está muy ligado a alto absentismo. |
| Reason\_absence\_9 Enfermedad sistema circulatorio | +32.27 | 0.000 | Muy alta relación. Puede ser una enfermedad seria, accidente, etc. |
| Reason\_absence\_11 Enfermedad del sistema digestivo | +6.54 | 0.030 | También significativo. |
| Reason\_absence\_12 Enfermedad de la piel y tejido subcutaneo | +13.49 | 0.002 | Similar al anterior, representa una razón con gran impacto. |
| Reason\_absence\_13 Enfermedad del sistema musculoesquelético y del tejido conjuntivo | +8.11 | 0.001 | Motivo importante a analizar. |
| Reason\_absence\_19 Lesiones, envenenamiento y consecuencia de causas externas | +9.28 | 0.000 | Otro motivo clave. |

**Variables no significativas (p > 0.05)**

* Variables como edad, transporte, social smoker/drinker, IMC, peso, altura, etc. no son significativas.
* Muchas variables de mes y temporada tampoco lo son.

Esto indica que probablemente se podrían eliminar para simplificar el modelo sin pérdida importante de explicatividad.

**Conclusiones estratégicas**

**1. Intervención en causas específicas**

Los **motivos de ausencia 6, 9, 12, 13 y 19** deben ser prioridad:

* Analiza si están relacionadas con temas médicos, salud mental o conflictos personales.
* Puedes diseñar campañas preventivas, convenios médicos, coaching, etc.
* Planes de apoyo específicos.

**2. Apoyo a empleados con hijos**

* Proveer flexibilidad o apoyo en emergencias familiares puede reducir el absentismo entre quienes tienen hijos.

**3. Formación reduce ausencias**

* Potencia programas de capacitación: los empleados con mayor educación faltan menos.
* Incentivar educación continua o contratar perfiles con formación más alta.

**4. Revisión de sanciones**

* Las personas con fallos disciplinarios se ausentan menos. Cuidado con si esto es por **miedo al castigo** y no por mejora real del compromiso.

**5. Día de la semana influye**

* Miércoles, jueves y viernes muestran menos ausencias. Puede ayudarte a ajustar planificación de turnos o distribuir tareas críticas.