

Satisfacción de los usuarios en el transporte interurbano

Un enfoque multinivel

*Diego Da Rosa, Lucas Giúdice, Juan
Karawacki y Bruno Pintos*



MOTIVACIÓN

- La calidad del transporte público es clave para la planificación de la movilidad.
- La satisfacción de los usuarios es un indicador central de desempeño del servicio.
- El transporte interurbano presenta desafíos específicos: puntualidad, ocupación, información disponible, heterogeneidad entre líneas.



¿Qué factores explican que algunos usuarios estén más satisfechos que otros?

METODO APLICADO

- 01 Análisis exploratorio y descriptivo
- 02 Ajuste de un modelo lineal simple y verificación de los supuestos
- 03 Ajuste de un modelo multinivel



OBJETIVOS Y PREGUNTAS DE INVESTIGACIÓN

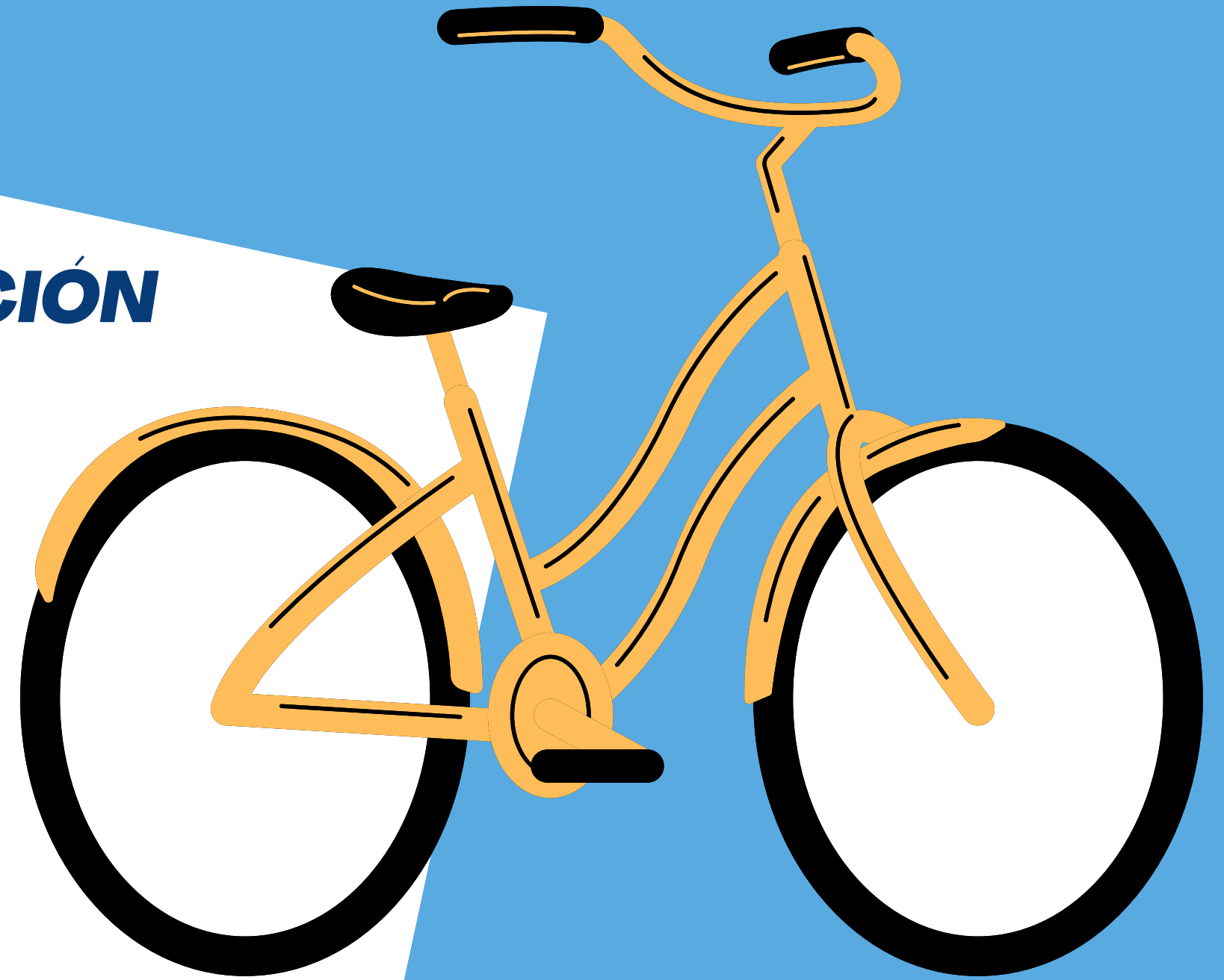
- Objetivo general: estudiar la satisfacción de usuarios del transporte interurbano considerando características individuales y diferencias entre líneas.

Preguntas:

¿Varía la satisfacción sistemáticamente entre líneas?

¿Qué proporción de la variabilidad se debe a diferencias entre líneas?

¿Qué covariables individuales inciden en la satisfacción?



CONTAMOS CON...



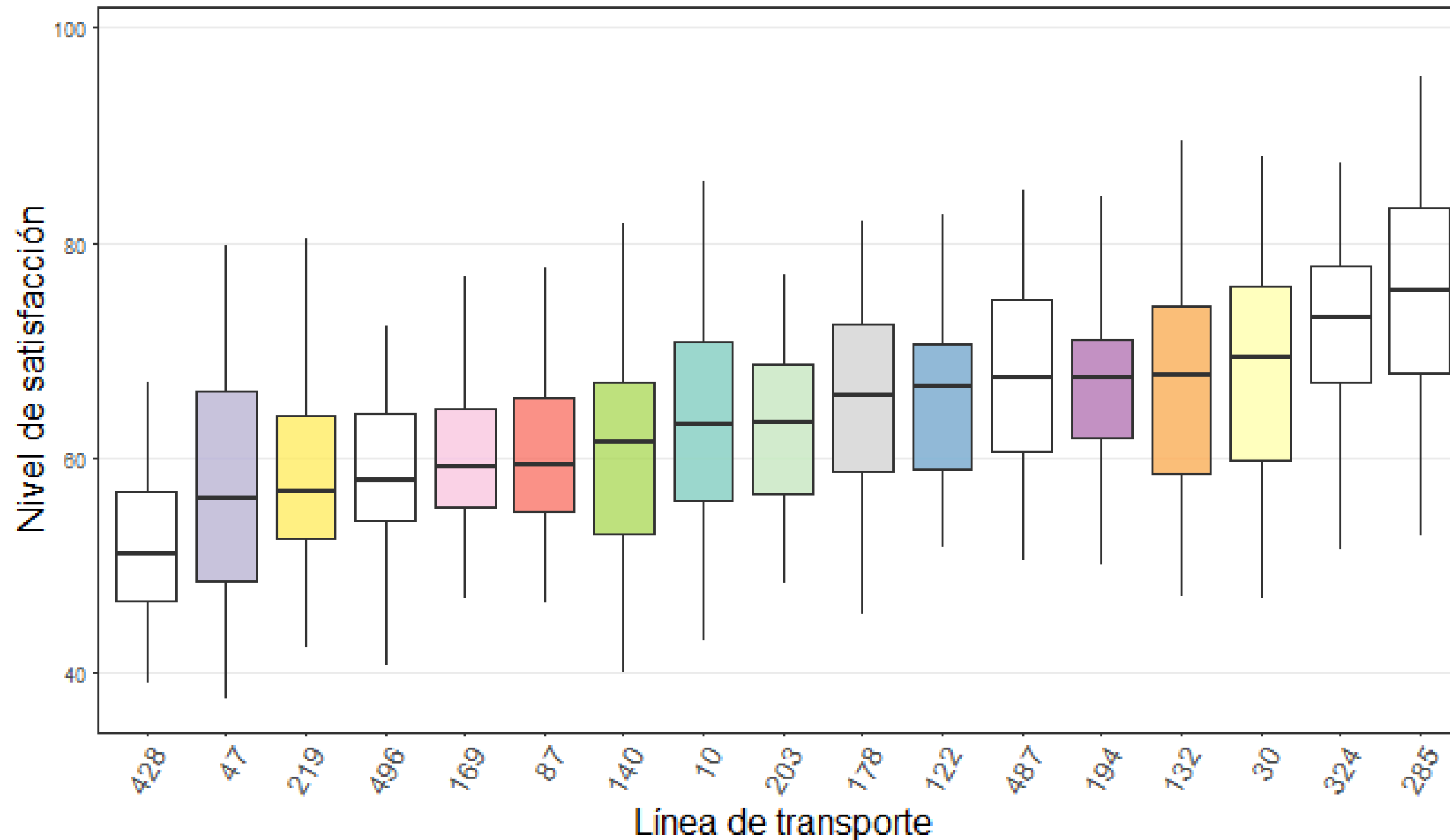
Variable	Tipo_dato	Descripción
linea_id	numeric	Numero de Linea
edad	numeric	Edad del usuario del transporte
genero	character	Genero del usuario del transporte
frecuencia_uso	character	Frecuencia de uso del usuario de esa linea
puntualidad	character	Puntualidad del transporte
p_ocupacion	numeric	Porcentaje de ocupacion del transporte
info_tiempo_real	character	Disponibilidad de Informacion en tiempo real del transporte
satisfaccion	numeric	Satisfaccion del usuario con el transporte

Nota: Tabla con las variables disponibles

- Un total de **780 observaciones** de individuos
- Agrupados en **17 líneas** distintas de transporte interurbano.
- Las lineas cuentan con un entorno de 40-50 observaciones cada una.

PRIMER ACERCAMIENTO: ¿ESTRUCTURA JERÁRQUICA?

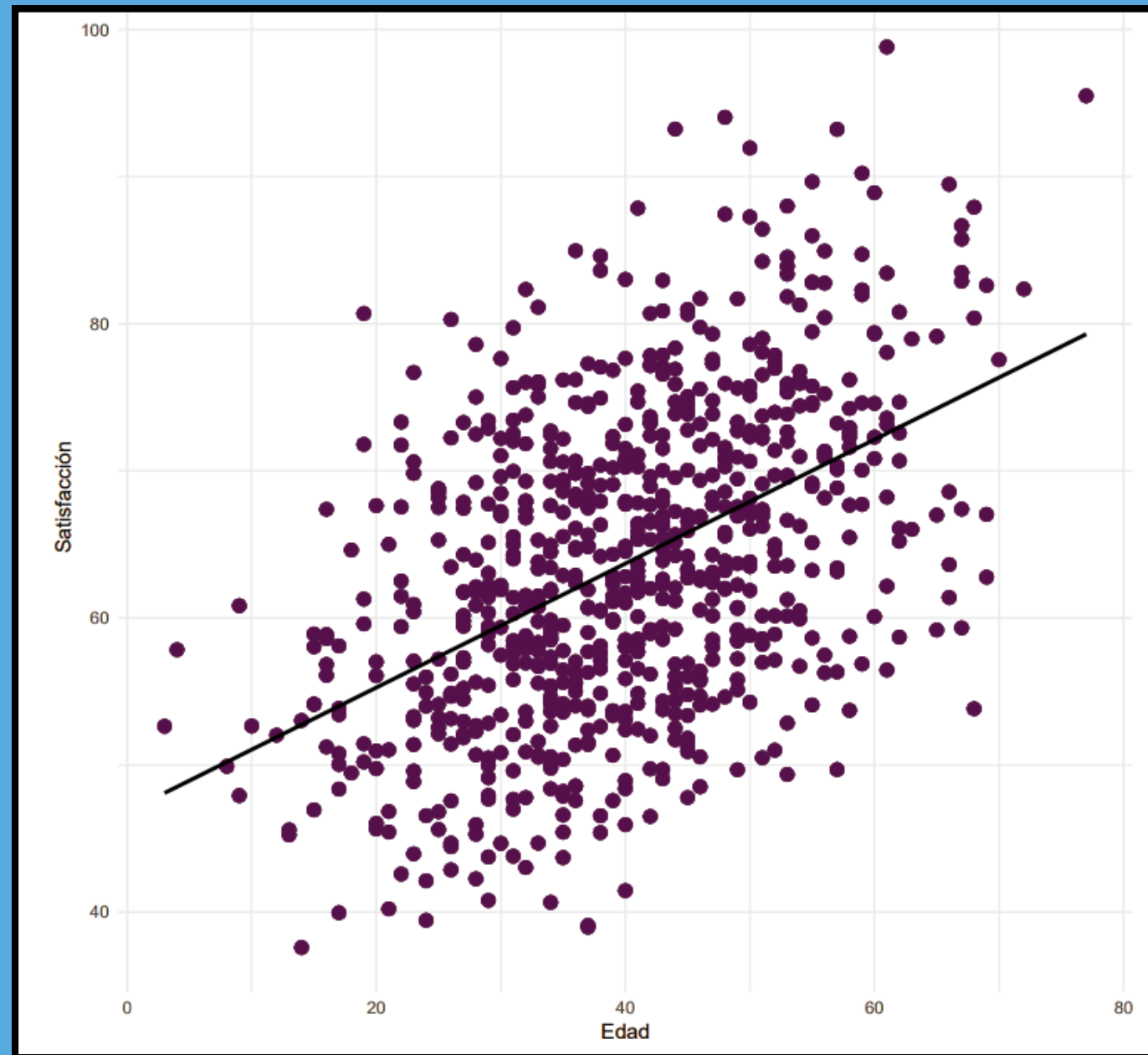
Nivel de satisfacción por línea de transporte



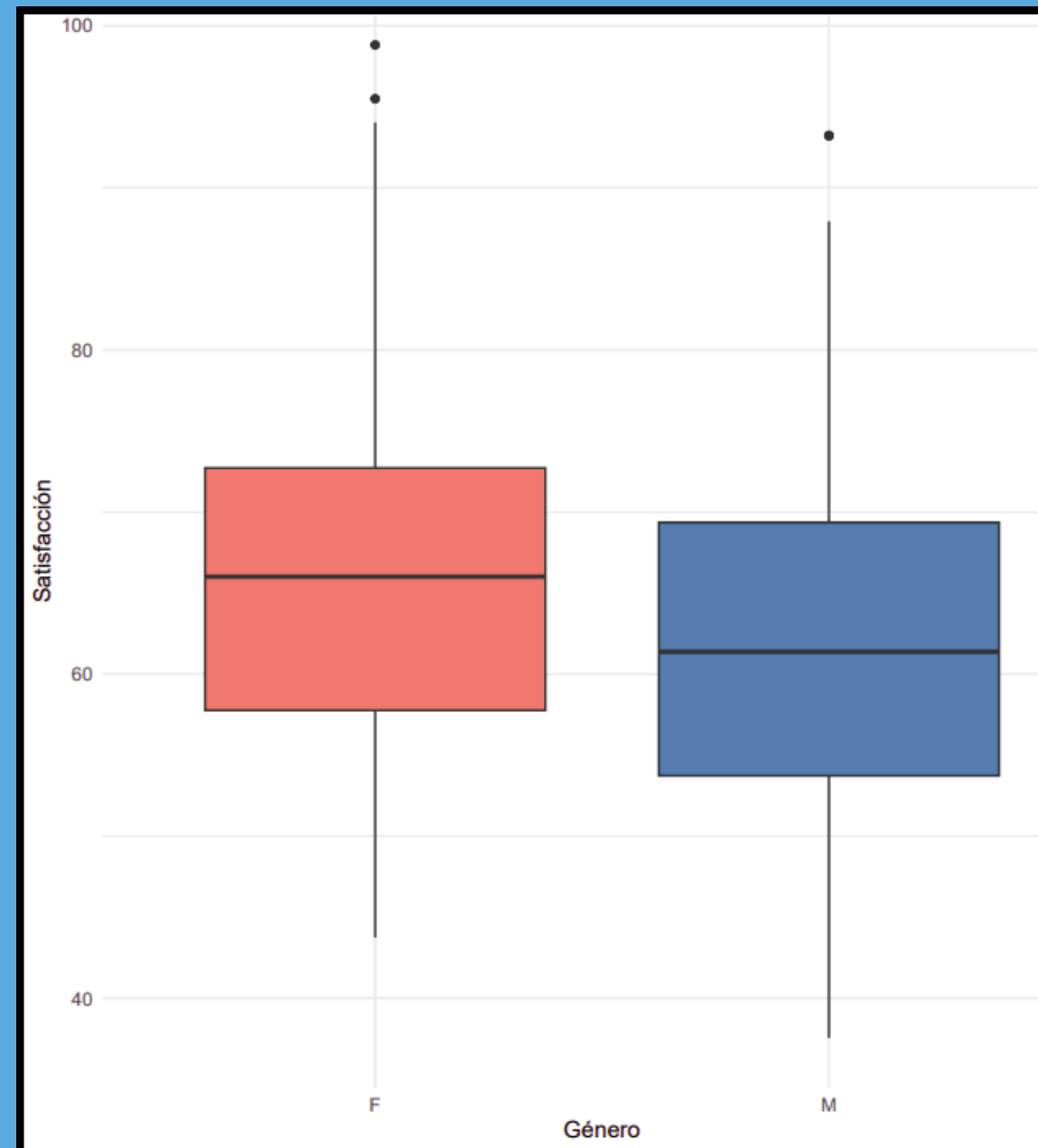
Nota: Diagramas de caja por línea ordenado según la mediana de la misma

- Se observa que la satisfacción de los individuos varía de acuerdo a cada línea.
- Las diferencias son notorias, ¿indicador de algo más?

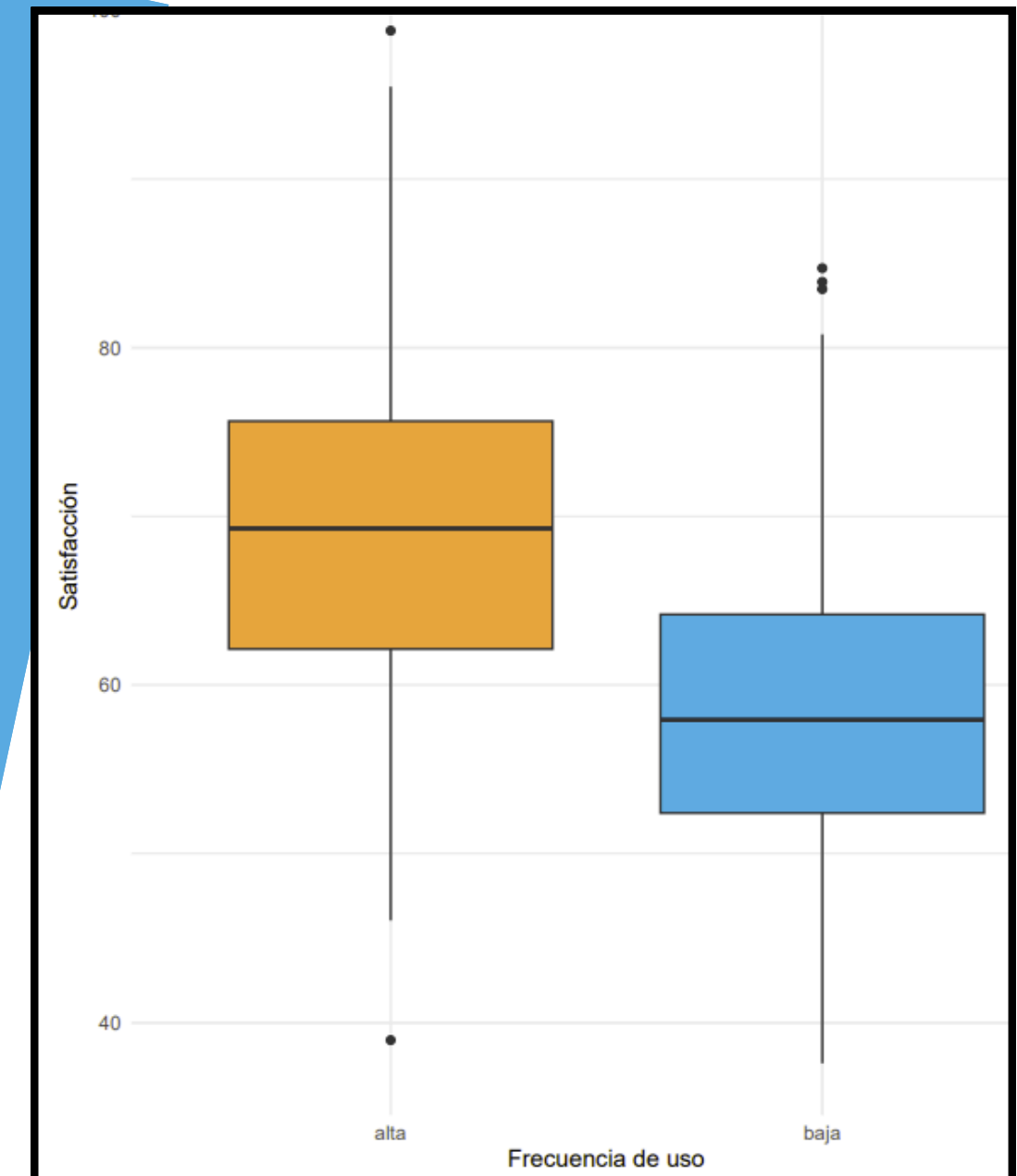
SATISFACCIÓN VS ... : VARIABLES INDIVIDUALES



Edad

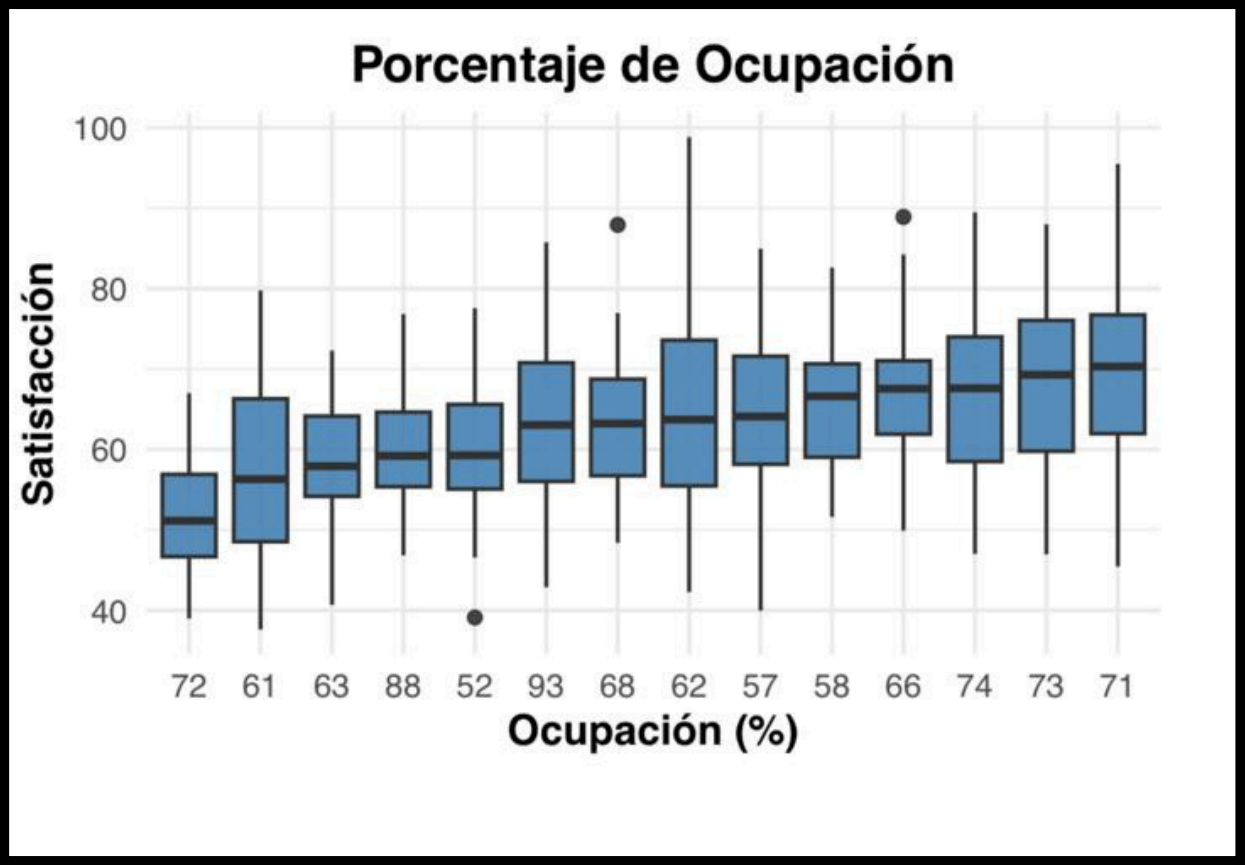


Género

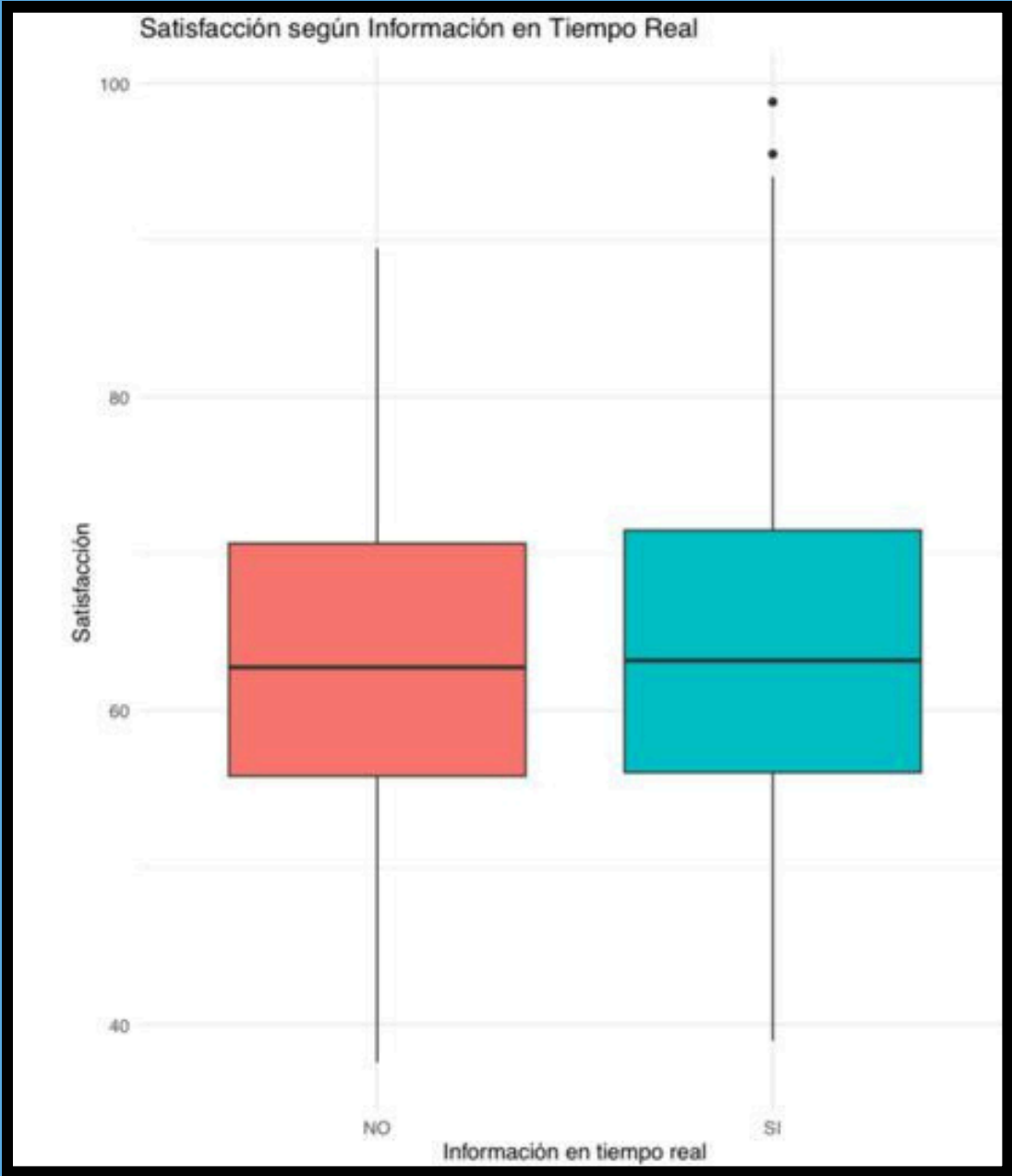


Frecuencia de uso

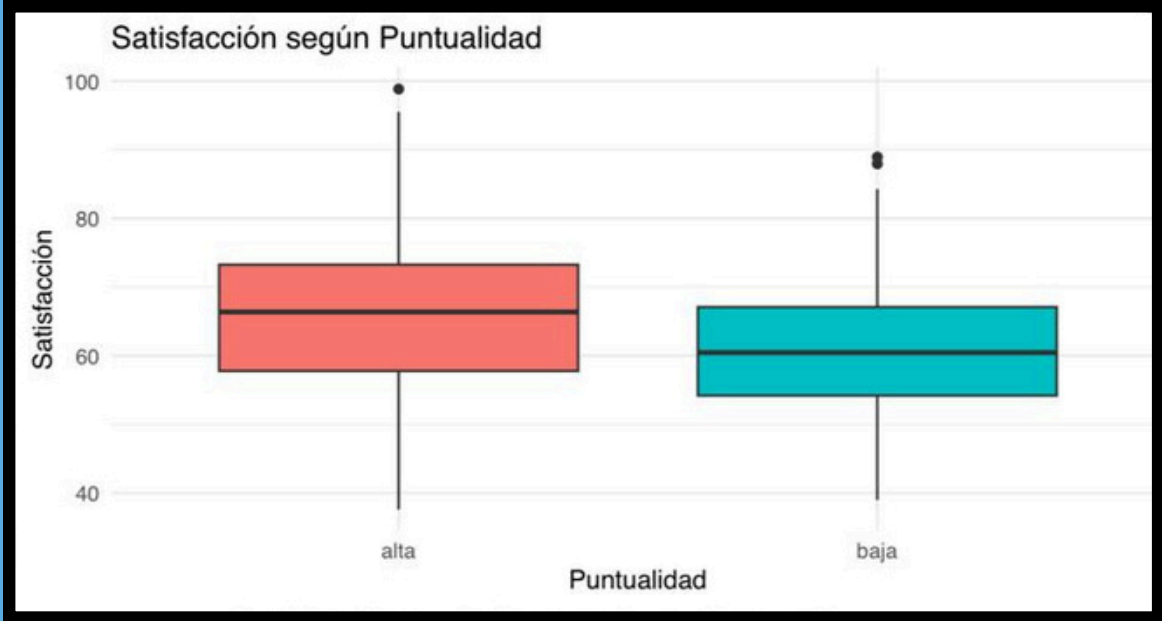
SATISFACCIÓN VS ... : VARIABLES GRUPALES



Ocupación



Tiempo real



Puntualidad

MODELO LINEAL CLÁSICO: ¿ALCANZA?

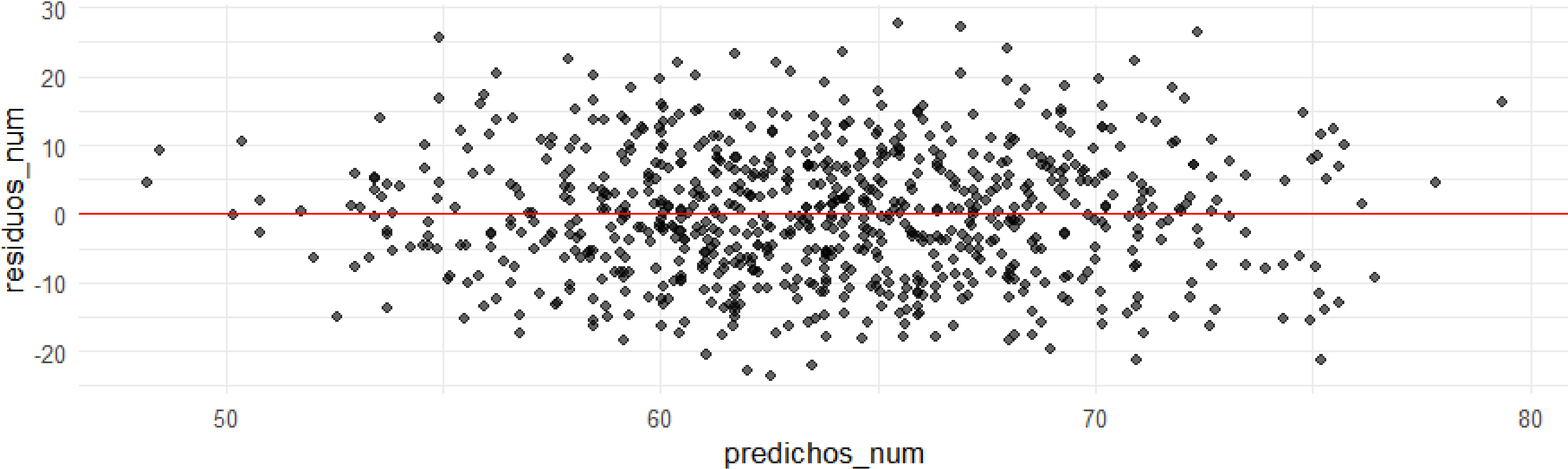


- Ajustamos un modelo lineal simple que contenga todas las variables.

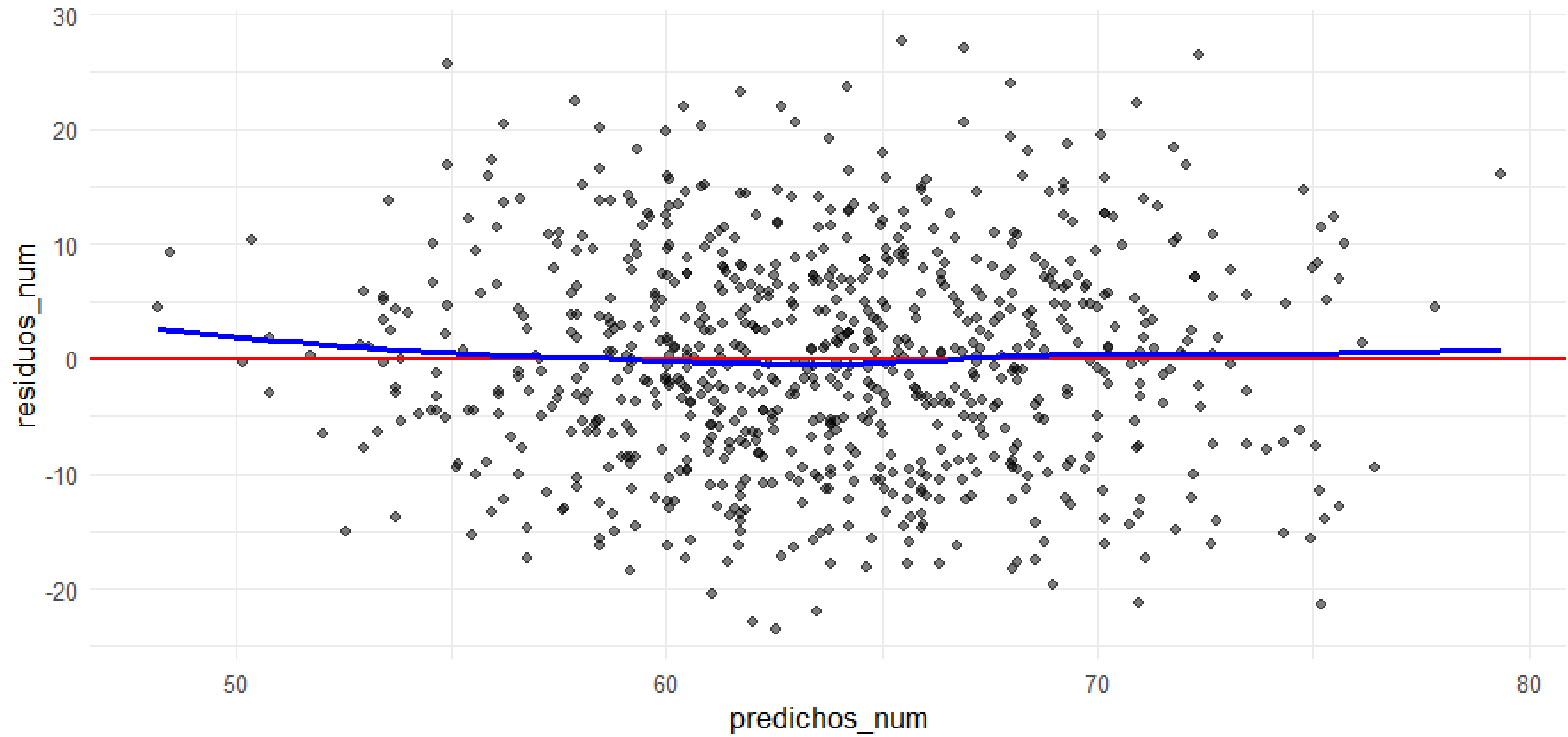
$$satisfaccion_i = \beta_0 + \beta_1 Edad_i + \beta_2 Genero_i + \beta_3 Frecuencia_Uso_i + \beta_4 Porcentaje_Ocupacion_i + \beta_5 Puntualidad_i + \beta_6 Informacion_en_Tiempo_Real_i + \beta_7 Linea_id_i + \epsilon_i$$

- El modelo contendrá muchos niveles, por lo que ya de por si nos resulta inconveniente su manejo...

Residuos vs Valores Predichos

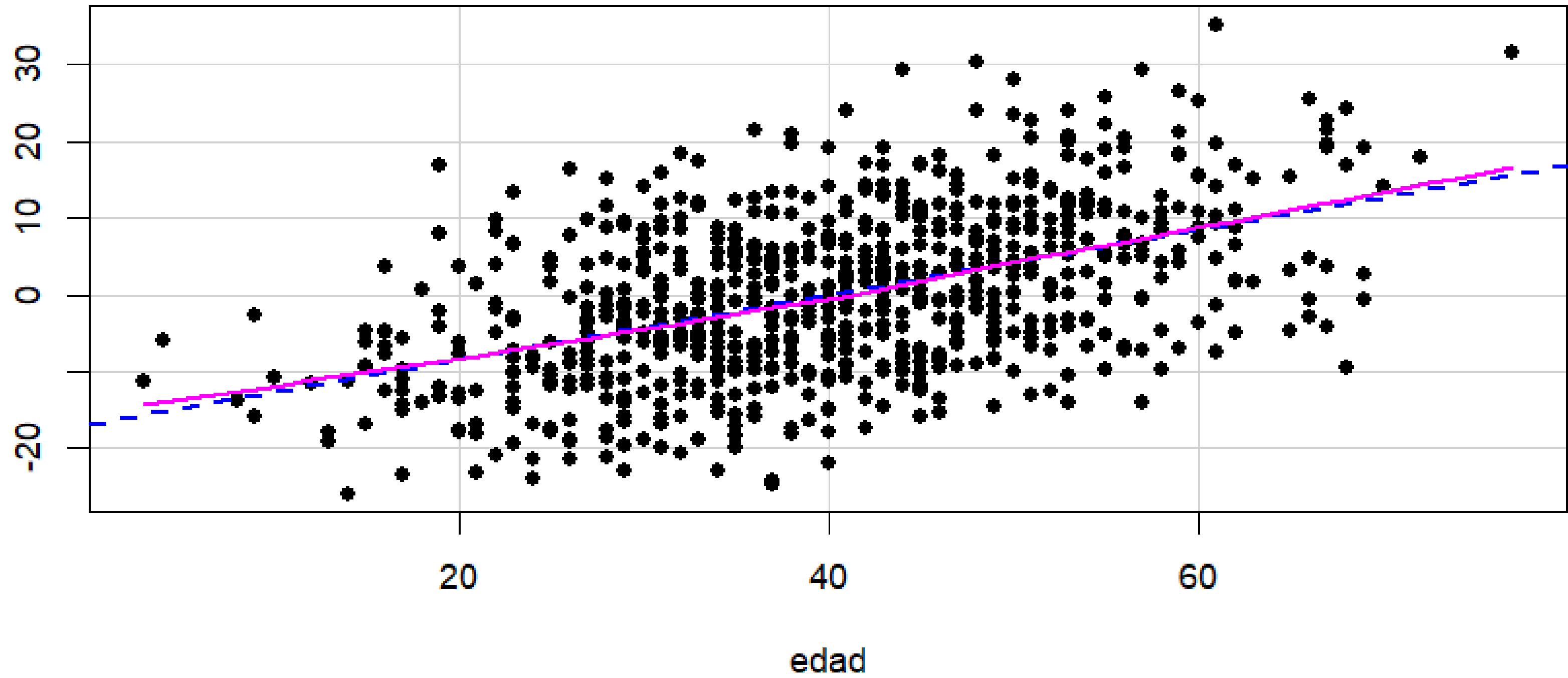


Residuos vs Predichos (LOESS)

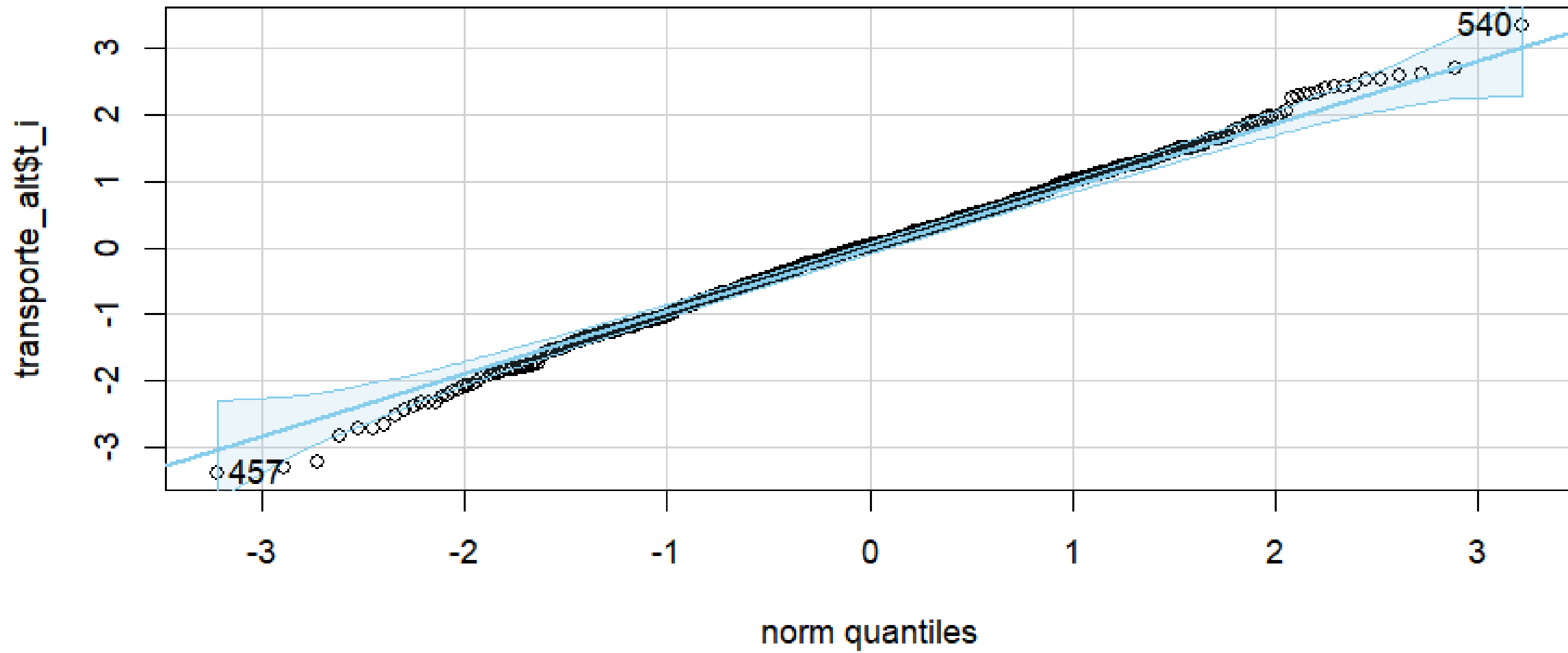


Component + Residual

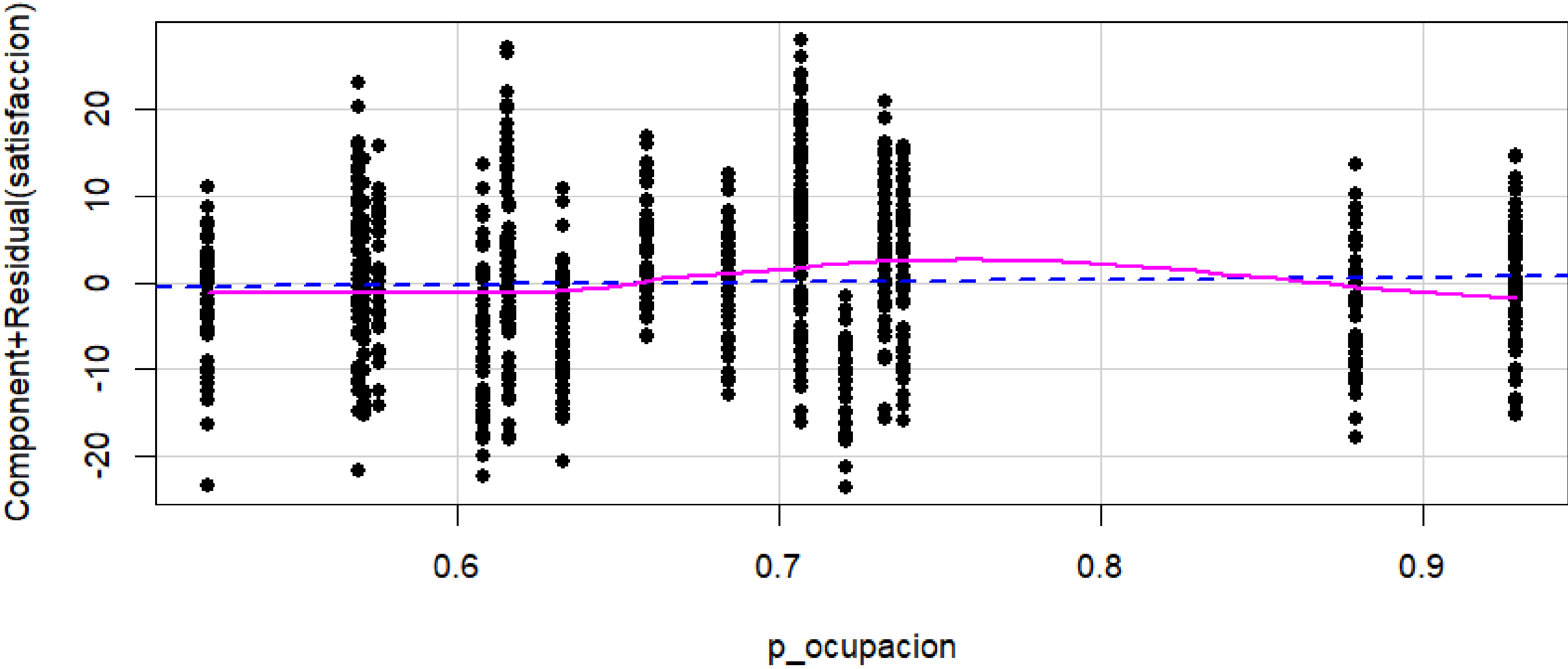
Component+Residual(satisfaccion)



Q-Q Plot con bandas (95%)



Component + Residual



DIAGNÓSTICO

Resultados de Tests Numéricos

	Test	W	p_value
<i>W</i>	Shapiro–Wilk	0.9974381	0.2669876

	Test	BP_stat	df	p_value
<i>BP</i>	Breusch–Pagan	24.65049	19	0.1723711

	Predictor	VIF
<i>edad</i>	edad	1.001412
<i>p_ocupacion</i>	p_ocupacion	1.001412



SIN EMBARGO...

- Observaciones agrupadas por línea, violando independencia.
- Representar 17 líneas como dummies → modelo recargado y difícil de interpretar.

CONCLUSION

Desde los supuestos clásicos el modelo 'funciona', pero no respeta la idea fundamental. Es necesario establecer una estructura jerárquica de los datos.

MODELO MULTINIVEL: ESTRUCTURA

Modelo de efectos mixtos con intercepto aleatorio por línea.

Nivel 1 (pasajeros): edad, género, frecuencia, % ocupación, puntualidad, info tiempo real.

Nivel 2 (líneas): intercepto aleatorio para cada línea



$$satisfaccion_{ij} = \beta_{0j} + \beta_1 Edad_{ij} + \beta_2 Genero_{ij} + \beta_3 Frecuencia_Uso_{ij} + \beta_4 Porcentaje_Ocupacion_{ij} + \beta_5 Puntualidad_{ij} + \beta_6 Informacion_en_Tiempo_Real_{ij} + \beta_7 Linea_id_{ij} + \epsilon_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

ICC y comparación de modelos

ICC es Intraclass Correlation Coefficient, nos sirve para ver qué proporción de la variabilidad total de satisfacción se debe al nivel de línea,

Modelo nulo (solo intercepto aleatorio):

ICC $\approx 0.157 \rightarrow \sim 15\text{--}16\%$ de la variabilidad en satisfacción se debe a diferencias entre líneas.

$$\text{satisfaccion}_{ij} = \beta_{0j} + \varepsilon_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_u^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$\text{Cov}(u_{0j}, \varepsilon_{ij}) = 0$$

ICC y comparación de modelos

Modelo completo (con covariables):

ICC $\approx 0.40 \rightarrow \sim 40\%$ de la variabilidad explicada por diferencias entre líneas.

$$\text{satisfaccion}_{ij} = \beta_{0j} + \beta_1 \text{Edad}_{ij} + \beta_2 \text{Genero}_{ij} + \beta_3 \text{Frecuencia_Uso}_{ij} \\ + \beta_4 \text{Porcentaje_Ocupacion}_{ij} + \beta_5 \text{Puntualidad}_{ij} + \beta_6 \text{Informacion_en_Tiempo_Real}_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_u^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$\text{Cov}(u_{0j}, \varepsilon_{ij}) = 0$$

Elección de modelos

$$\text{satisfaccion}_{ij} = \beta_{0j} + \varepsilon_{ij}$$

VS

$$\begin{aligned} \text{satisfaccion}_{ij} = & \beta_{0j} + \beta_1 \text{Edad}_{ij} + \beta_2 \text{Genero}_{ij} + \beta_3 \text{Frecuencia_Uso}_{ij} \\ & + \beta_4 \text{Porcentaje_Ocupacion}_{ij} + \beta_5 \text{Puntualidad}_{ij} + \beta_6 \text{Informacion_en_Tiempo_Real}_{ij} + \varepsilon_{ij} \end{aligned}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

Criterios de Eleccion:

```
      npar    AIC    BIC  logLik -2*log(L)  Chisq Df Pr(>Chisq)
mod_null      3 5747.6 5761.6 -2870.8    5741.6
mod_mult      9 4876.9 4918.8 -2429.5    4858.9 882.74   6 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Estimación de Efectos Fijos

La siguiente tabla presenta las estimaciones de los efectos fijos del modelo:

Table 2: Tabla 2. Estimaciones de efectos fijos del modelo multinivel.

effect	term	estimate	std.error	statistic	df	p.value
fixed	(Intercept)	64.122	7.916	8.100	13.177	0.000
fixed	generoM	-3.834	0.381	-10.061	760.569	0.000
fixed	edad	0.420	0.015	27.113	760.245	0.000
fixed	frecuencia_usobaja	-10.683	0.381	-28.065	760.633	0.000
fixed	p_ocupacion	-16.319	11.313	-1.443	12.999	0.173
fixed	puntualidadbaja	-10.674	3.062	-3.486	13.079	0.004
fixed	info_tiempo_realSI	8.436	2.945	2.864	13.004	0.013

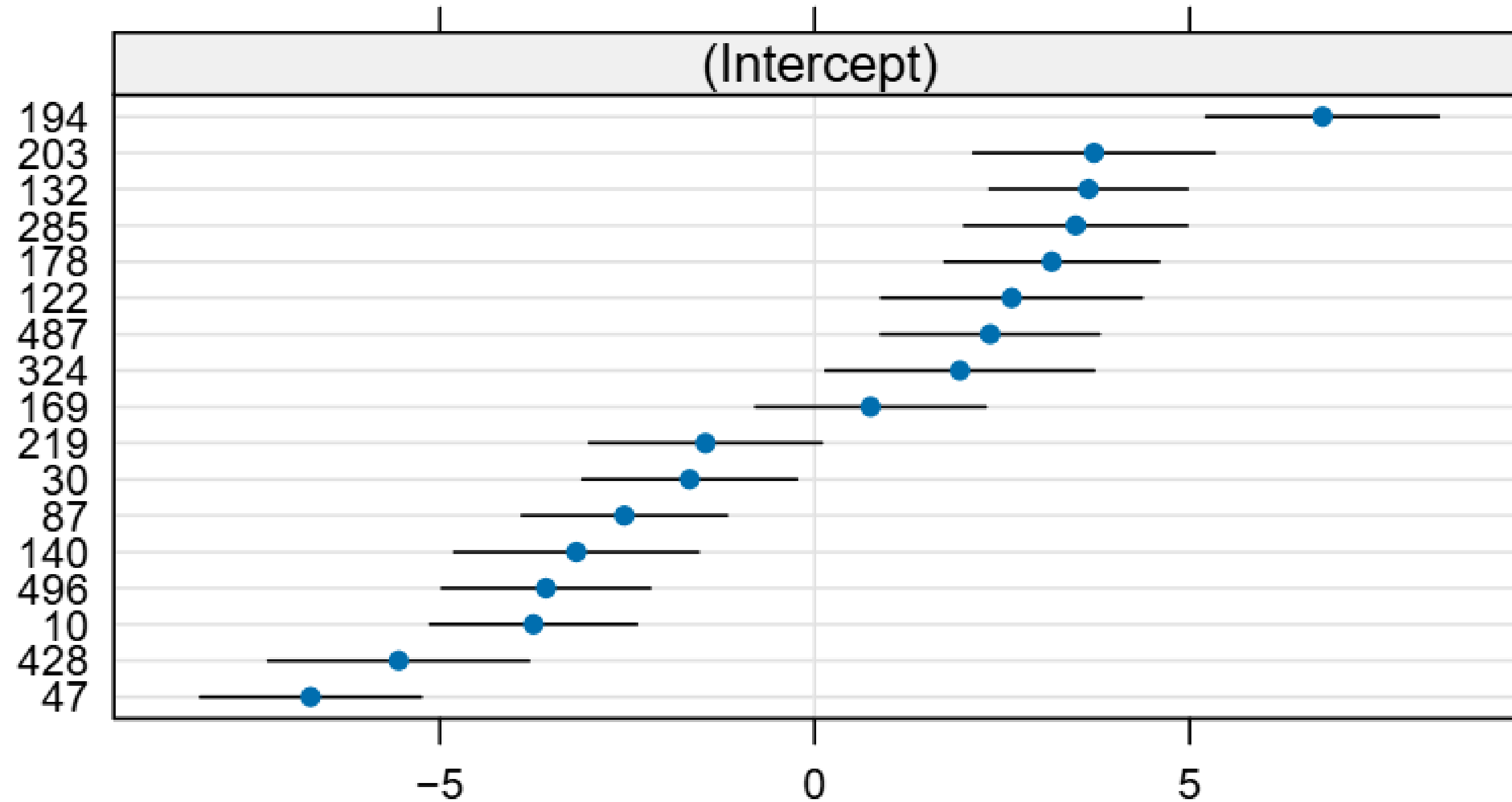
CAPACIDAD EXPLICATIVA DEL MODELO MULTINIVEL

R^2 Marginal (0.611): Los efectos fijos (predictores individuales) explican el 61.1% de la varianza.

R^2 Condicional (0.768): El modelo completo (incluyendo el efecto de las líneas) explica el 76.8% de la variabilidad total en la satisfacción.

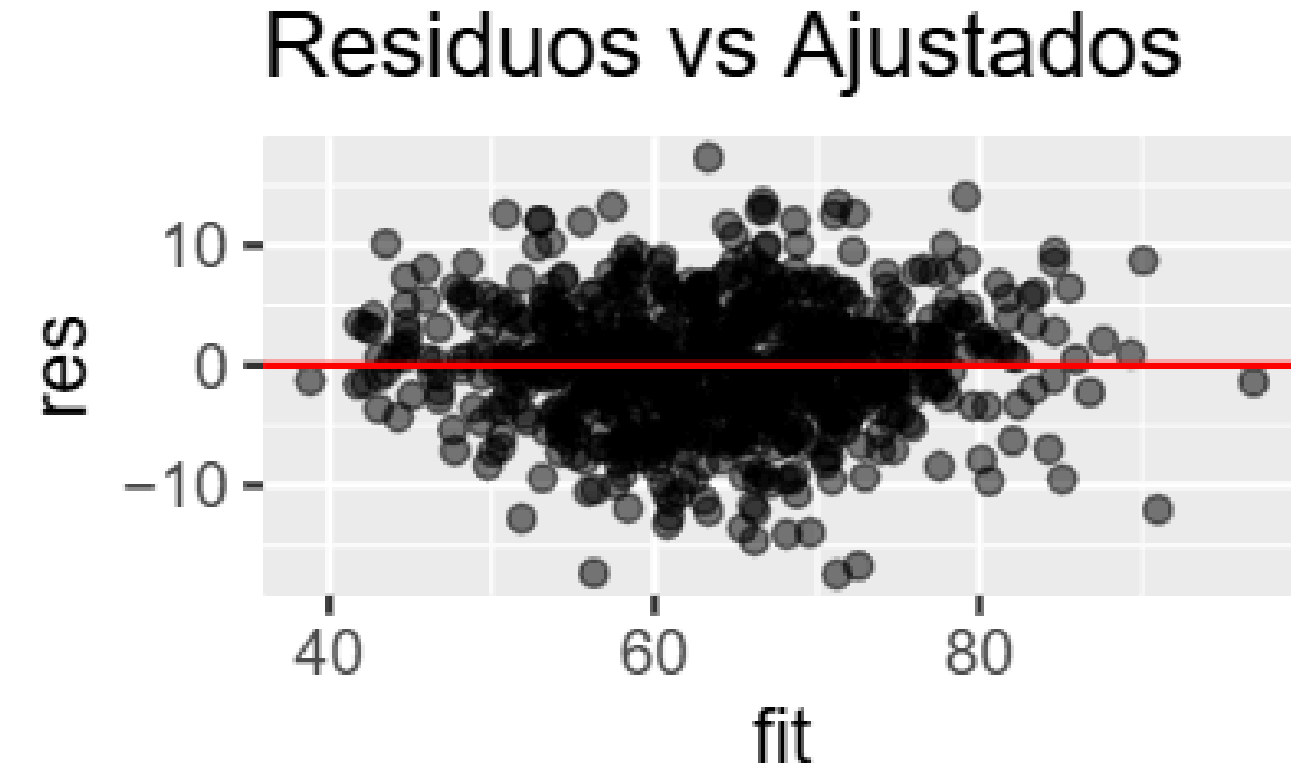
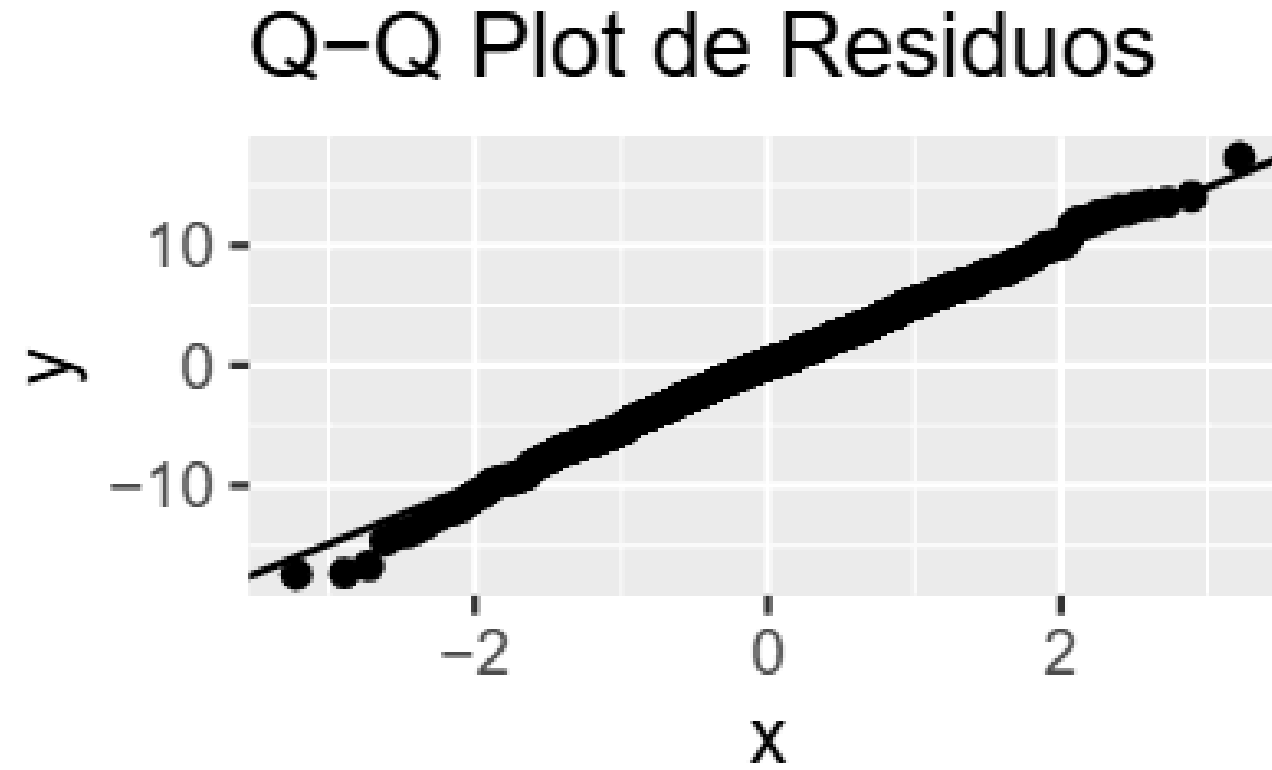


VARIABILIDAD ENTRE LINEAS



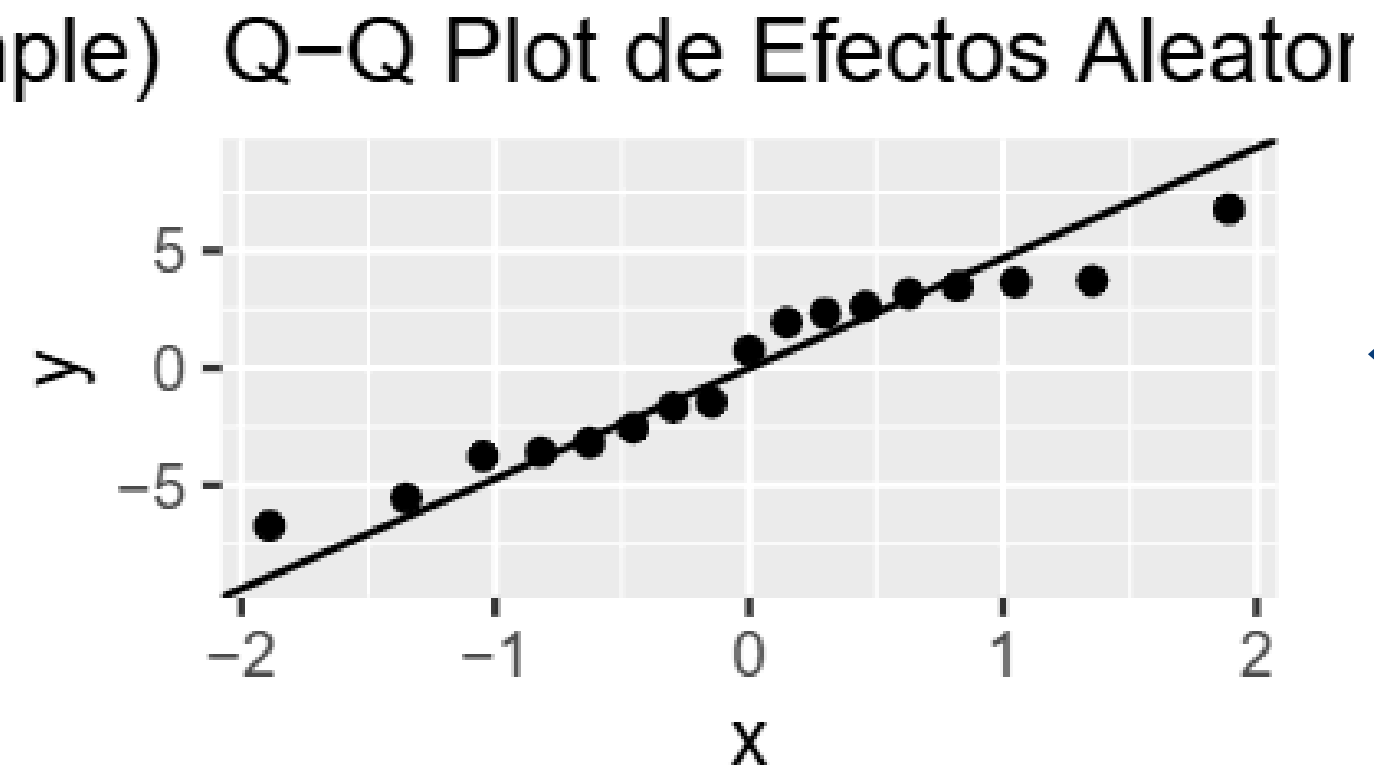
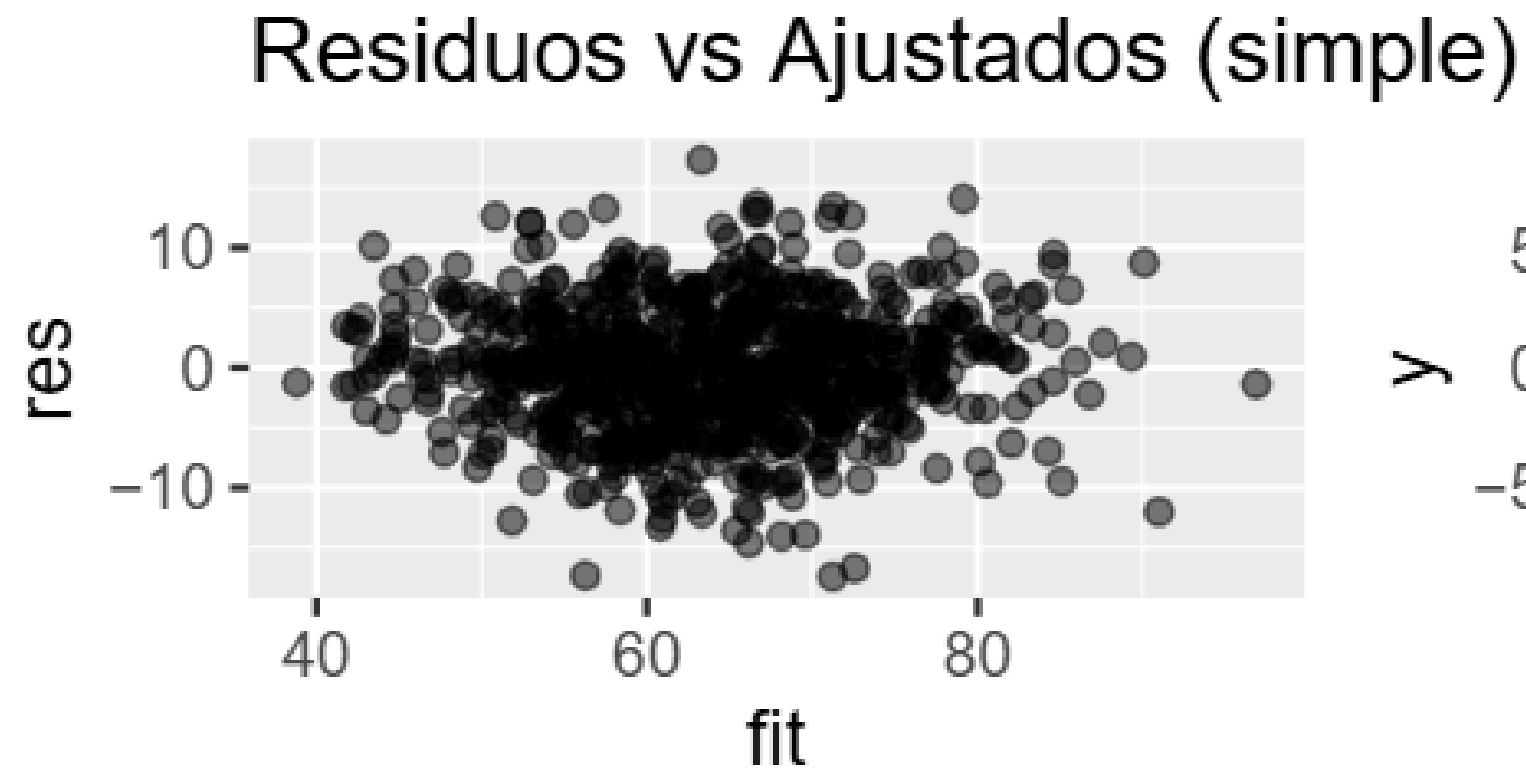
SUPUESTOS DEL MODELO MULTINIVEL

Normalidad □



← Homocedasticidad

Linealidad □



← Normalidad de los efectos aleatorios

VARIANZA EXPLICADA POR NIVELES

```
$var.fixed  
[1] 73.2551
```

```
$var.random  
[1] 18.79614
```

```
$var.residual  
[1] 27.82667
```

```
$var.distribution  
[1] 27.82667
```

```
$var.dispersion  
[1] 0
```

```
$var.intercept  
linea_id  
18.79614
```



MODELOS CON PENDIENTE ALEATORIA

$$\text{Satisfaccion}_{ij} = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j}) \cdot \text{edad}_{ij} + \text{otros efectos fijos} + \varepsilon_{ij}$$

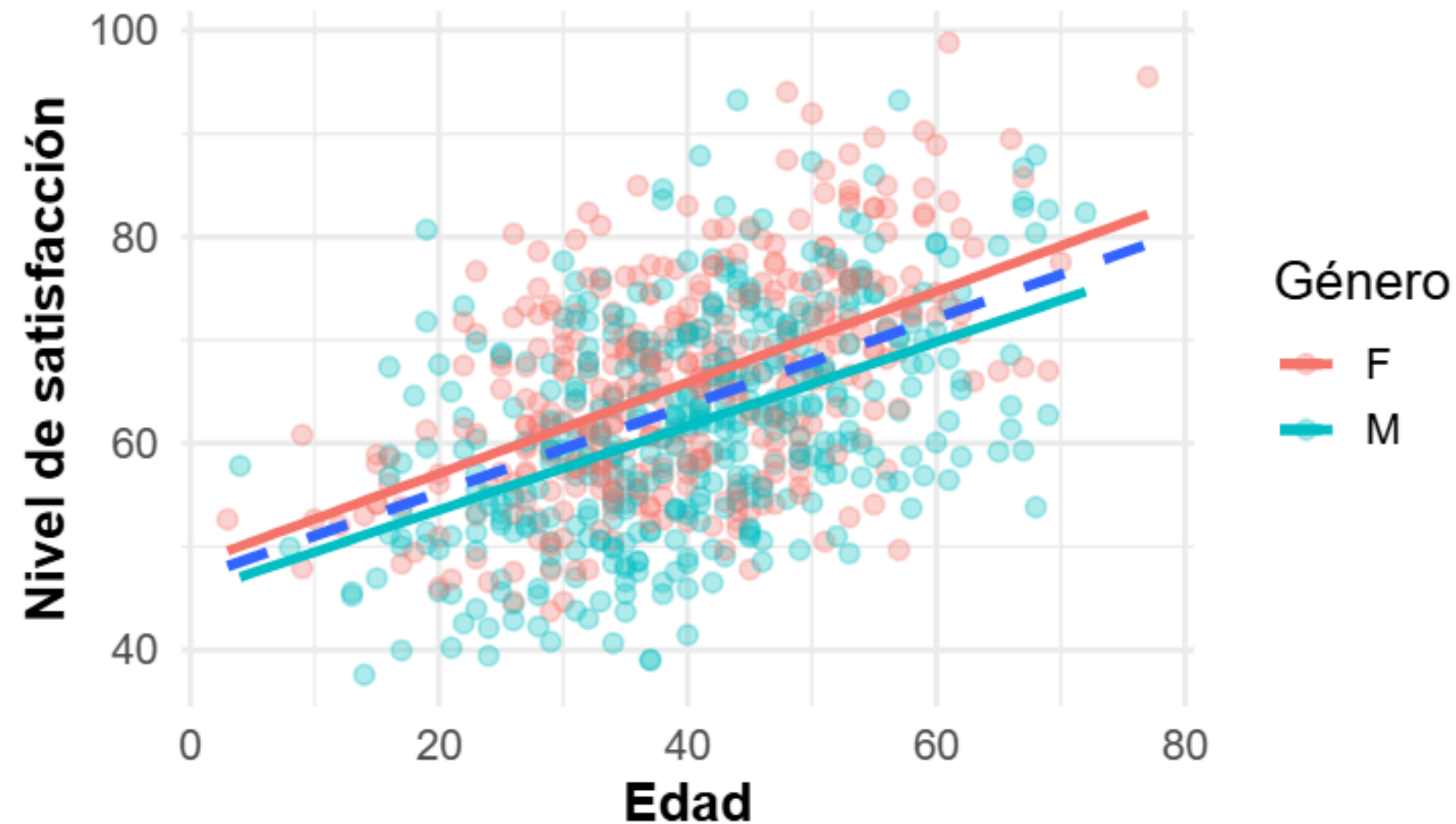
u_{0j} es la desviación del intercepto de la línea j

u_{1j} es la desviación de la pendiente de la edad en la línea j .

	npar	AIC	BIC	logLik	-2*log(L)	Chisq	Df	Pr(>Chisq)
mod_mult	9	4876.9	4918.8	-2429.5	4858.9			
mod_slope	11	4879.8	4931.1	-2428.9	4857.8	1.0876	2	0.5805



MODELOS MULTINIVEL CON INTERACCIONES



$$\text{satisfacción}_i = \beta_0 + \beta_1(\text{género}_i) + \beta_2(\text{edad}_i) + \beta_3(\text{frecuencia}_i) + \beta_4(\text{ocupación}_i) + \beta_5(\text{puntualidad}_i) + \beta_6(\text{infoTR}_i) + \beta_7(\text{género}_i \cdot \text{edad}_i) + \beta_8(\text{género}_i \cdot \text{frecuencia}_i) + \beta_9(\text{género}_i \cdot \text{ocupación}_i) + \beta_{10}(\text{género}_i \cdot \text{puntualidad}_i) + \beta_{11}(\text{género}_i \cdot \text{infoTR}_i)$$

Con un intercepto aleatorio por línea $u_{0j} \sim N(0, \sigma_u^2)$

CONCLUSIONES

- Existen diferencias significativas de satisfacción entre líneas.
- Una fracción importante de la variabilidad (15–40%) se debe al nivel línea.
- A nivel individual, mayor satisfacción se asocia con mayor edad, uso frecuente, puntualidad alta, disponibilidad de información en tiempo real y género femenino (en promedio).
- El porcentaje de ocupación no mostró un efecto claro en el modelo.



DESTINO!!!

