

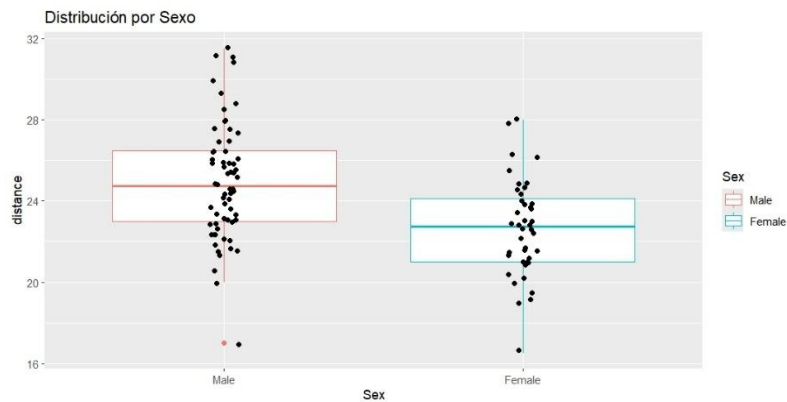
Tarea 2

Pregunta 1

Realice un análisis exploratorio de los perfiles por sujeto indicando si encuentra diferencias entre las distancias de varones y mujeres.

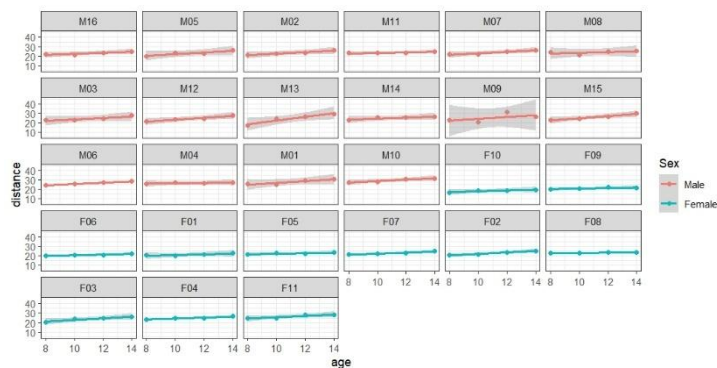
RESPUESTA:

Para contestar esta pregunta, podemos ayudarnos de herramientas visuales que nos permitan comparar la distribución de los datos. En este caso, comparamos la distancia entre varones y mujeres usando un boxplot:



Podemos ver que las distancias entre hombres y mujeres son distintas, siendo la de los hombre, en general, mayor a la de las mujeres. En este caso, los hombres tienen una distancia mediana de 25 y las mujeres de 23.

Además, gracias al gráfico de perfil podemos identificar que la tendencia entre la edad y la distancia es lineal y que las pendientes entre los individuos no son del todo homogéneas, por lo que un modelo de intercepto y/o pendiente aleatoria es prometedor.



Pregunta 2

Analice los datos considerando los siguientes modelos.

$$distance_{ij} = \beta_0 + b_{i0} + \beta_1 age_{ij} + \epsilon_{ij}$$

donde $b_{i0} \sim N(0, \sigma_{b_0}^2)$ y $\epsilon \sim N(0, \sigma^2)$, y

$$distance_{ij} = \beta_0 + b_{i0} + (\beta_1 + b_{i1})age_{ij} + \epsilon_{ij}$$

donde $b_{i0} \sim N(0, \sigma_{b_0}^2)$ es independiente de $b_{i1} \sim N(0, \sigma_{b_1}^2)$ y $\epsilon \sim N(0, \sigma^2)$.

RESPUESTA:

Para el primero modelo, usamos un modelo mixto con intercepto aleatorio. Si nos basamos en el gráfico anterior, podemos esperar que el intercepto aleatorio no sea significativo pues los primeros valores registrados para cada individuo tienen a ser alrededor del mismo valor de 22, presentando una baja varianza.

A continuación, tenemos los resultados del modelo mixto. Podemos ver que el intercepto aleatorio tiene una baja varianza (2.11) por lo que podemos no incluir un intercepto aleatorio en el modelo final.

Linear mixed-effects model fit by REML

Data: datos

	AIC	BIC	logLik
	455.0025	465.6563	-223.5013

Random effects:

Formula: ~1 | Subject
(Intercept) Residual

StdDev: 2.114724 1.431592

Fixed effects: distance ~ age

	Value	Std.Error	DF	t-value	p-value
(Intercept)	16.761111	0.8023952	80	20.88885	0
age	0.660185	0.0616059	80	10.71626	0

Para el segundo modelo con intercepto y, además, pendiente aleatoria los resultados son los siguientes. Acá corroboramos que tanto una pendiente como un intercepto aleatorio no son necesario para ajustar un modelo

Random effects:

Formula: ~1 + age | Subject

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	2.3270341	(Intr)
age	0.2264278	-0.609
Residual	1.3100397	

Fixed effects: distance ~ age

	Value	Std.Error	DF	t-value	p-value
(Intercept)	16.761111	0.7752460	80	21.620377	0
age	0.660185	0.0712533	80	9.265333	0

Pregunta 3

Proponga una interpretación de cada uno de los parámetros del modelo seleccionado.

RESPUESTA

Modelo con Intercepto Aleatorio

El valor del intercepto aleatorio presenta una baja varianza, lo que es un indicativo de que no es necesario su uso para este modelo. Vemos que tanto el intercepto general como el coeficiente de Age son significativos para explicar la distancia y que, a mayor Edad, mayor distancia se tiene (cada año, la distancia crece en, aproximadamente, 0.66 milímetros), empezando en un nivel base de 16.

Modelo con Intercepto y Pendiente Aleatoria

El valor del intercepto y la pendiente aleatoria presenta una baja varianza (2.11 y 0.22, respectivamente), lo que es un indicativo de que no es necesario su uso para este modelo. De igual manera que en el Modelo con Intercepto Aleatorio, cuando observamos los efectos fijos de la regresión, vemos que tanto el intercepto general como el coeficiente de Age son significativos para explicar la distancia, y que, a mayor Edad, mayor distancia se tiene (cada año, la distancia crece en, aproximadamente, 0.66 milímetros), empezando en un nivel base de 16.

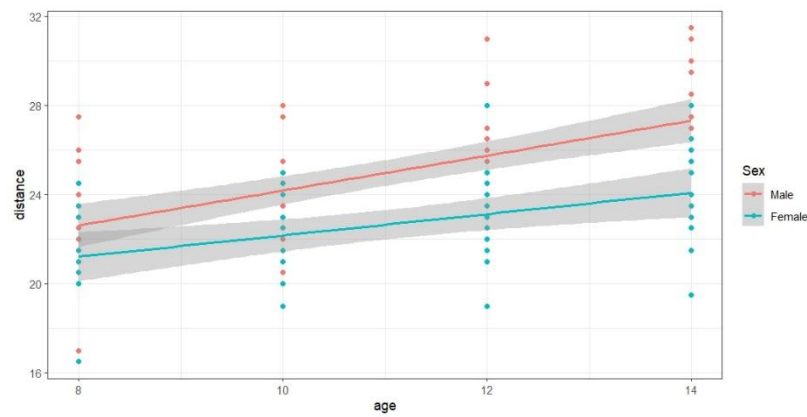
Al tomar en cuenta el valor la varianza de los componentes aleatorios, podemos dejar el intercepto aleatorio, pero botar el coeficiente de Edad pues es una varianza pequeña.

Pregunta 4

Suponga que los investigadores están interesados en una segunda pregunta de interés, a saber, si existen diferencias en el crecimiento entre varones y mujeres. Realice un análisis por separado para varones y mujeres y compare los resultados.

RESPUESTA

Para empezar, podemos hacer una comparativa entre las tendencias iniciales entre varones y mujeres para determinar que los varones tienden a presentar un mayor crecimiento en la variable de distancia (un coeficiente mayor)



Podemos hacer una prueba que nos ayude a determinar si un modelo mixto con intercepto aleatorio es útil.

Al ajustar una regresión lineal simple (sin componentes aleatorios) solo con la variable edad, obtenemos un R^2 de 0.25 con intercepto y coeficiente de edad significativos.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	16.7611	1.2256	13.676	< 2e-16 ***
age	0.6602	0.1092	6.047	2.25e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.537 on 106 degrees of freedom

Multiple R-squared: 0.2565, Adjusted R-squared: 0.2495

Al agregar la variable dummy de Sexo, mejoramos el R^2 a 0.4. Cuando estamos evaluando a una mujer, la distancia baja -2.32 con respecto al valor calculado para un varón para un mismo valor de edad.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	17.70671	1.11221	15.920	< 2e-16 ***
age	0.66019	0.09776	6.753	8.25e-10 ***
SexFemale	-2.32102	0.44489	-5.217	9.20e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.272 on 105 degrees of freedom

Multiple R-squared: 0.4095, Adjusted R-squared: 0.3983

Al ajustar un modelo mixto con coeficientes e interceptos aleatorios podemos concluir que tanto el intercepto como el coeficiente aleatorio de ser hembra ayudan a ajustar un mejor modelo pues su varianza es mayor a la del coeficiente de la edad.

Haciendo el análisis por separado, vemos que para ambas muestras (varones y mujeres), la desviación estándar del intercepto aleatorio ronda valores de entre 1.6 y 2, lo que nos indica que, de agregar un componente aleatorio, tendría que sea el intercepto.

Sí hay una diferencia entre el crecimiento de la distancia entre hombres y mujeres. Cuando evaluamos el modelo mixto ajustado de manera independiente a cada muestras, vemos que el coeficiente de edad para varones es de 0.78 y el de mujeres es de 0.48; esto nos indica que el aumento de distancia cuando aumenta la edad es un 62% mayor en varones con respecto a las mujeres.

Linear mixed-effects model fit by REML

Data: mujeres

	AIC	BIC	logLik
	149.4287	159.8547	-68.71435

Random effects:

Formula: ~1 + age | Subject

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	1.8841866	(Intr)
age	0.1609278	-0.354
Residual	0.6682746	

Fixed effects: distance ~ age

	Value	Std.Error	DF	t-value	p-value
(Intercept)	17.372727	0.7606027	32	22.840737	0

age	0.479545	0.0662140	32	7.242353	0
-----	----------	-----------	----	----------	---

Linear mixed-effects model fit by REML

Data: varones

AIC	BIC	logLik
285.1374	297.9002	-136.5687

Random effects:

Formula: ~1 + age | Subject

Structure: General positive-definite, Log-Cholesky parametrization

	StdDev	Corr
(Intercept)	2.6665775	(Intr)
age	0.1887314	-0.788
Residual	1.6090569	

Fixed effects: distance ~ age

	Value	Std.Error	DF	t-value	p-value
(Intercept)	16.340625	1.2099003	47	13.505761	0
age	0.784375	0.1015729	47	7.722289	0