

Significancia e Interacciones

José Eduardo Valencia Espinosa

29/9/2022

Entorno

```
library(pacman)
p.load(data.table, fixest, magrittr, wooldridge, multcomp, ggplot2, gridExtra)
```

1. Significancia estadística

```
## [1] 3.28861e-25
```

a. Demuestre que el estadístico t de educ=10.94

$$t = \frac{\hat{\beta}_{educ}}{se(\hat{\beta}_{educ})}$$

```
round(est_0$coefficients["educ"]/est_0sse["educ"], 2)
```

```
## educ
## 19.94
```

b. ¿Se rechaza la $H_0 : \beta = 0$?

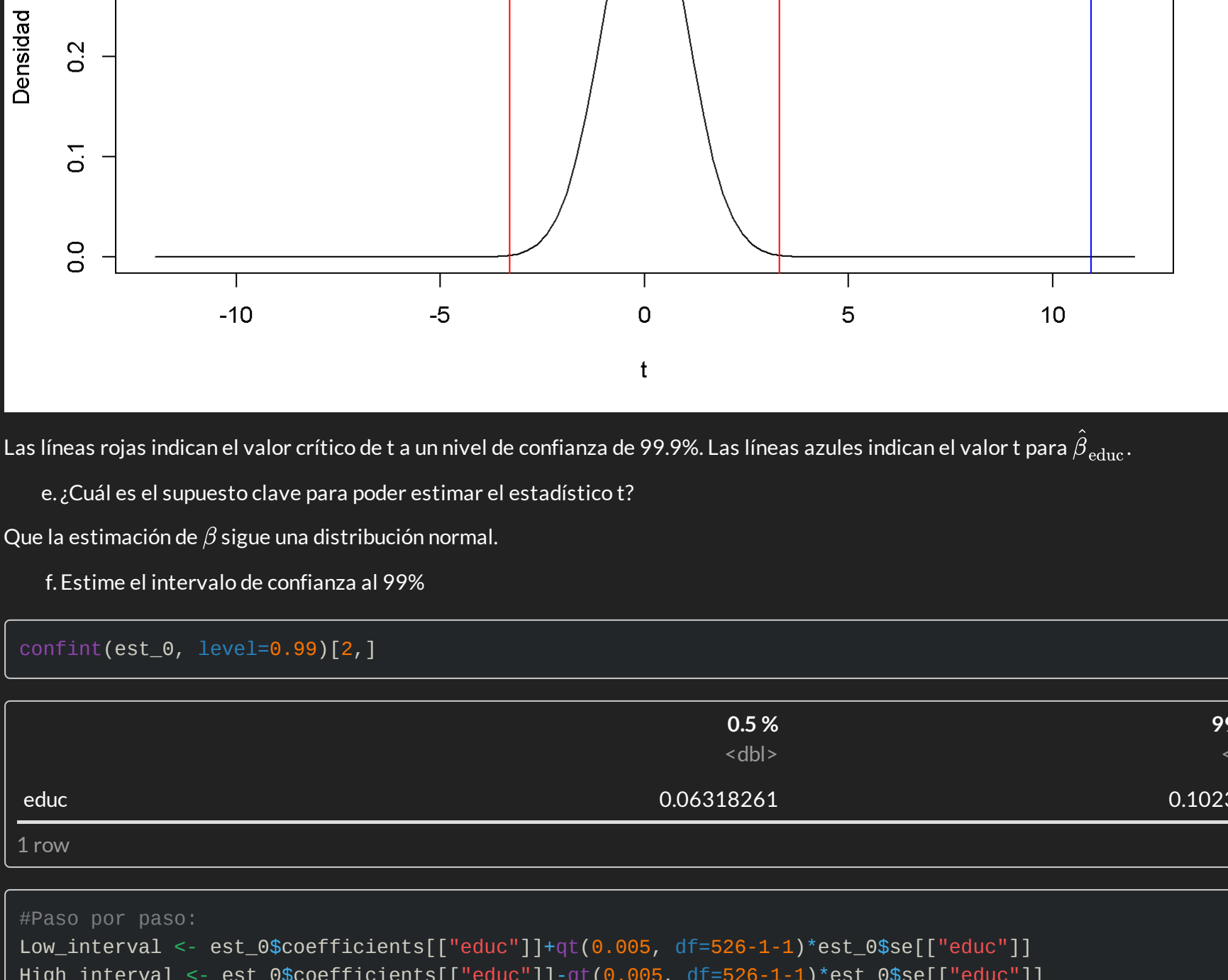
Si, el valor t está en la zona de rechazo (>2).

c. ¿Por lo menos con qué nivel de confianza estadística es significativo el coeficiente de educación? ¿con qué valor crítico?

Por lo menos, con 99.9% de nivel de confianza. El valor crítico es ± 3.309 (de dos colas).

d. Dibuje (a mano) el gráfico de la distribución t que surge de c. donde aproximadamente identifique el área que equivale al p-value.

```
curve(dt(x, df=524), -12, 12, ylab= "Densidad", xlab= "t")
abline(v=3.94, col="blue")
abline(v=-3.309, col="red")
abline(v=-3.309, col="red")
```



Las líneas rojas indican el valor crítico de t a un nivel de confianza de 99.9%. Las líneas azules indican el valor t para $\hat{\beta}_{educ}$.

e. ¿Cuál es el supuesto clave para poder estimar el estadístico t ?

Que la estimación de β sigue una distribución normal.

f. Estime el intervalo de confianza al 99%

```
confint(est_0, level=0.99)[2,]
```

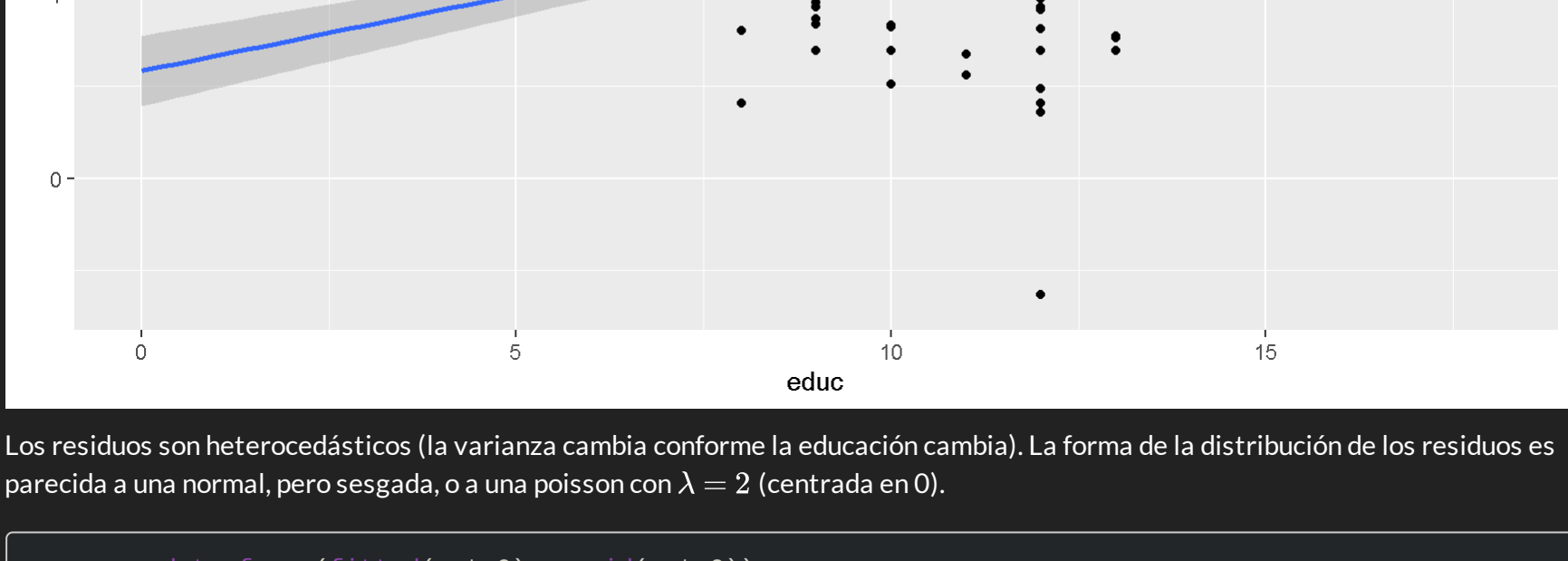
	0.5% <dbl>	99.5% <dbl>
educ	0.06318261	0.1023061
1 row		

```
#Paso por paso:
Low_interval <- est_0$coefficients[["educ"]]+qt(0.005, df=526-1-1)*est_0sse[["educ"]]
High_interval <- est_0$coefficients[["educ"]]-qt(0.005, df=526-1-1)*est_0sse[["educ"]]
list(c("educ", "0.5%"=round(Low_interval,3), "99.5%"=round(High_interval,3)))
```

```
## [[1]]
## 0.5% 99.5%
## "educ" "0.063" "0.102"
```

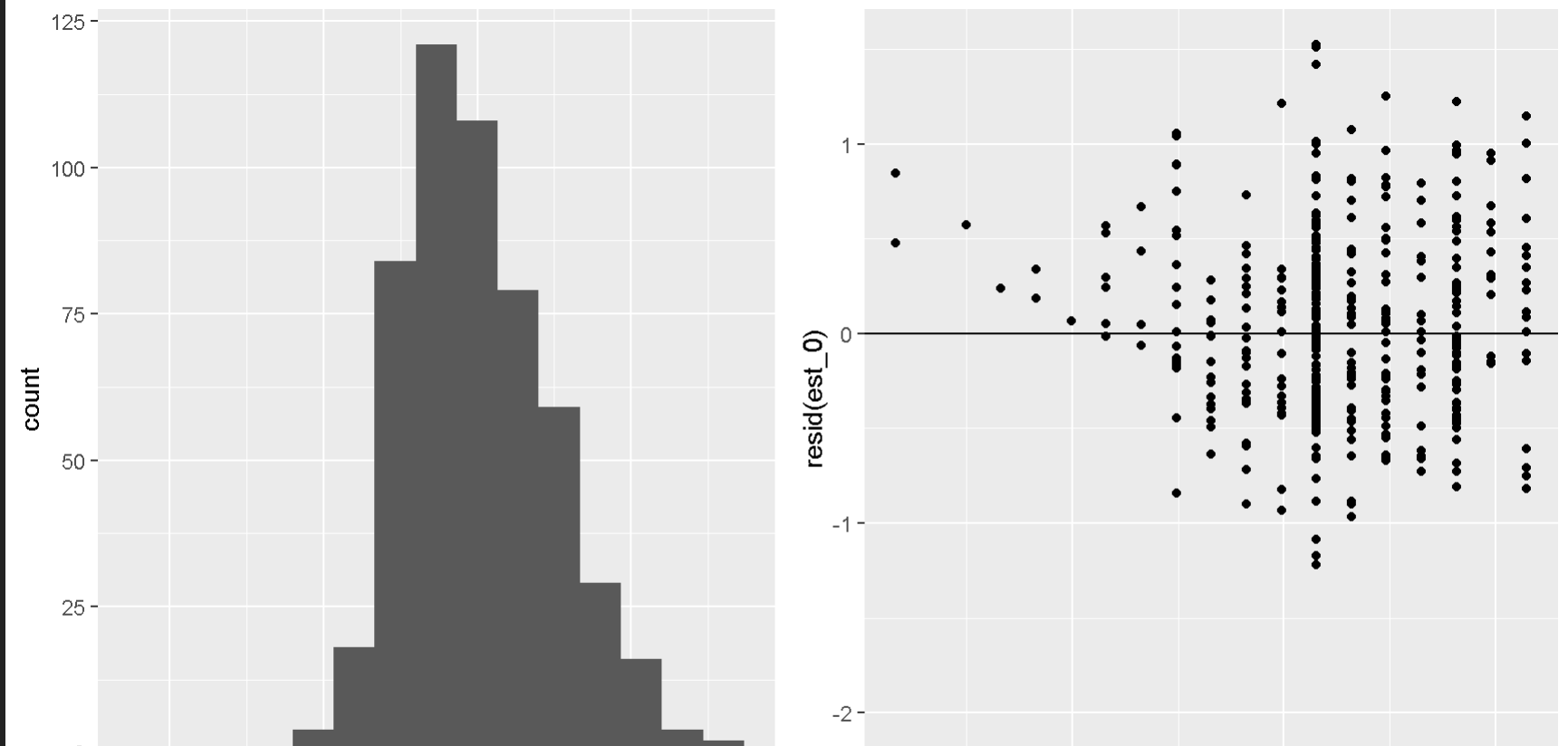
g. Draw (with R) a scatter plot of Y on X and add the regression line. Are the residuals homoscedastic? Will they distribute normally? Draw (manually) how do you think the distribution of residuals will look?

```
ggplot(data=wage1, aes(educ, lwage)) + geom_point() + geom_smooth(method=lm) + theme_grey()
```



Los residuos son heterocedásticos (la varianza cambia conforme la educación cambia). La forma de la distribución de los residuos es parecida a una normal, pero sesgada, o a una poisson con $\lambda = 2$ (centrada en 0).

```
one_g <- data.frame(fitted(est_0), resid(est_0))
hist_g <- ggplot(one_g, aes(resid(est_0))) + geom_histogram(bins=15)
variance_g <- ggplot(one_g, aes(fitted(est_0), resid(est_0))) + geom_point() + geom_hline(yintercept= 0)
grid.arrange(hist_g, variance_g, ncol=2)
```



2. Wage1

```
exper2 <- wage1$exper^2
wageLM <- lm(lwage ~ educ + exper + exper2 + female, data=wage1)
```

a. Interprete el efecto del primer año de experiencia.

```
FirstExper <- summary(WageLM)$coefficients[3,1] +
2*summary(WageLM)$coefficients[4,1]*0
```

El primer año de experiencia provoca un cambio 3.89% en el salario por hora.

b. Interprete el efecto de la experiencia al pasar de 7 a 8 años de experiencia y de 17 a 18 años de experiencia, respectivamente.

```
SevenExper <- summary(WageLM)$coefficients[3,1] +
2*summary(WageLM)$coefficients[4,1]*7
SeventeenExper <- summary(WageLM)$coefficients[3,1] +
2*summary(WageLM)$coefficients[4,1]*17
```

El octavo año de experiencia provoca un cambio 2.93% en el salario por hora.

El decimo séptimo año de experiencia provoca un cambio 1.56% en el salario por hora.

c. ¿Con cuántos años de experiencia se alcanza el máximo sueldo, ceteris paribus?

$$\beta_{exper} + 2\beta_{exper^2} \cdot \text{exper} = 0 \rightarrow \frac{-\beta_{exper}}{2\beta_{exper^2}} = \text{exper}$$

```
ExperMax <- summary(WageLM)$coefficients[3,1] / summary(WageLM)$coefficients[4,1] * -0.5
```

El máximo sueldo se alcanza con 28.36 años de experiencia.

d. ¿Es estadísticamente significativa la experiencia en la ecuación de sueldos estimada? (vea sección 4.4 en Wooldridge ed.7) y explore el comando 'glht' en R (instalado con el paquete multcomp)

```
K <- diag(length(coef(WageLM)))[-1,]
rownames(K) <- names(coef(WageLM))[-1]
confint(glht(WageLM, linfct=K))
```

```
##
## Simultaneous Confidence Intervals
## Fit: lm(formula = lwage ~ educ + exper + exper2 + female, data = wage1)
##
## Quantile = 2.4297
## 95% family-wise confidence level
##
## Linear Hypotheses:
## Estimate lwr upr
## educ == 0 0.0641361 0.0672329 0.1610399
## exper == 0 0.0389100 0.0271001 0.0506299
## exper2 == 0 -0.0006868 -0.0009469 -0.0004251
## female == 0 -0.3371868 -0.4254378 -0.2489357
```

Si, tiene al de menos 95% de confianza.

e. ¿Cuál es el valor predicho del sueldo en dólares por hora para un hombre con 10 años de experiencia y 12 años de educación? ¿Cuál es el valor predicho para una mujer con las mismas características?

```
CoeffM <- matrix(as.numeric(summary(WageLM)$coefficients[,1]), nrow=length(summary(WageLM)$coefficients[,1])
ValueM <- matrix(c(1, 12, 10, 100, 0), ncol=5)
dataF <- matrix(data.frame(1, 12, 10, 100, 1), ncol=5)
PredictWageM <- sum(as.numeric(ValueM)*as.numeric(CoeffM))
PredictWageF <- sum(as.numeric(ValueF)*as.numeric(CoeffM))
```

Si es hombre, 5.588, si es mujer 3.989 dólares la hora.

f. Interprete el coeficiente female. ¿Qué representa la constante del modelo en este caso?

Las mujeres ganan 33.72% menos que los hombres.

3. Retornos a la educación y sexo

a. Estime una regresión con el logaritmo del sueldo como dependiente solamente para la muestra de mujeres y otra solamente para la muestra de hombres. Escriba la ecuación estimada y pegue los resultados incluyendo errores estándar robustos 'hetero'.

```
est_2 <- feols(lwage ~ educ + exper + expersq,
data = subset(wage1, female=="f"), se="hetero")
est_3 <- feols(lwage ~ educ + exper + expersq,
data = subset(wage1, female=="m"), se="hetero")
etable(est_2, est_3)
```

	est_2 <chr>	est_3 <chr>
Dependent Var:	lwage	lwage
(Intercept)	0.2661 (0.1684)	0.1573 (0.1388)
educ	0.0792*** (0.0132)	0.0903*** (0.0097)
exper	0.0224*** (0.0063)	0.0540*** (0.0065)
expersq	-0.0004** (0.0001)	-0.0009*** (0.0001)
S.E. type	Heteroskedas.rob.	Heteroskedast.rob.
Observations	252	274
R2	0.22509	0.39359
1-10 of 11 rows		Previous 2 Next

b. ¿De qué tamaño es la diferencia entre los retornos a la educación (por año adicional) del hombre y los de la mujer?

De 1.12% por año adicional de educación (en favor de los hombres).

c. Construya el intervalo de confianza al 95% del coeficiente de educación para ambas submuestras.

```
#Para mujeres:
Low_interval <- est_2$coefficients[["educ"]]+qt(0.025, df=526-1-1)*est_2sse[["educ"]]
High_interval <- est_2$coefficients[["educ"]]-qt(0.025, df=526-1-1)*est_2sse[["educ"]]
list(c("educ", "0.5%"=round(Low_interval,3), "99.5%"=round(High_interval,3)))
```

```
## [[1]]
## 0.5% 99.5%
## "educ" "0.053" "0.105"
```

```
#Para hombres:
Low_interval <- est_3$coefficients[["educ"]]+qt(0.025, df=526-1-1)*est_3sse[["educ"]]
High_interval <- est_3$coefficients[["educ"]]-qt(0.025, df=526-1-1)*est_3sse[["educ"]]
list(c("educ", "0.5%"=round(Low_interval,3), "99.5%"=round(High_interval,3)))
```

```
## [[1]]
## 0.5% 99.5%
## "educ" "0.071" "0.109"
```

d. ¿Con el 95% de significancia estadística son significativamente distintos los coeficientes de educación para la muestra de hombres y de mujeres?

Los intervalos de confianza se superponen, por lo que no son estadísticamente distintos.

e. ¿Es estadísticamente distinto el retorno al primer año de experiencia para hombres y mujeres?

```
wage1$exper1 <- as.integer(wage1$exper==1)
est_4 <- feols(lwage ~ educ + exper1 + exper + expersq,
data = subset(wage1, female=="f"), se="hetero")
est_5 <- feols(lwage ~ educ + exper1 + exper + expersq,
data = subset(wage1, female=="m"), se="hetero")
#Para mujeres:
confint(est_4, level=0.95)[3,]
```

	2.5% <dbl>	97.5% <dbl>
exper1	-0.2946913	0.1520556
1 row		

```
#Para hombres:
confint(est_5, level=0.95)[3,]
```

	2.5% <dbl>	97.5% <dbl>
exper1	-0.4010029	-0.05295639
1 row		

No son estadísticamente distintos.

4. Interacción

```
est_6 <- feols(lwage ~ educ + exper + exper*female + educ*female +
female, data=wage1)
```

a. Cree una tabla incluyendo estos resultados y los de 3.a.

```
etable(est_6, est_2, est_3, se="hetero")
```

	est_6 <chr>	est_2 <chr>	est_3 <chr>
Dependent Var:	lwage	lwage	lwage
(Intercept)	0.2498 (0.1508)	0.2661 (0.1684)	0.1573 (0.1388)
educ	0.0118*** (0.0099)	0.0792*** (0.0132)	0.0903*** (0.0097)
exper	0.0149*** (0.0024)	0.0224*** (0.0063)	0.0540*** (0.0065)
female	0.0850 (0.2348)		
exper x female	-0.0108*** (0.0029)		
educ x female	-0.0195 (0.0171)		
expersq		-0.0004** (0.0001)	-0.0009*** (0.0001)
1-10 of 14 rows			Previous 2 Next

b. Observando el resultado obtenido en esta regresión ¿Cuál es el retorno a un año extra de educación para los hombres? ¿Para mujeres? Explique cómo lo obtuvo.

Para los hombres, es de 10.18% en el salario por cada año extra de educación. Para las mujeres, es de 8.23%.

Cálculo para los hombres:

$$\hat{\beta}_{educ}$$

Para las mujeres:

$$\hat{\beta}_{educ} + \hat{\beta}_{educ:female}$$

c. ¿De qué tamaño es la diferencia en los retornos a la educación entre hombres y mujeres? ¿Es significativamente distinta la diferencia en el retorno a la educación entre hombres y mujeres? Explique.

De 1.95% por año adicional de educación (en favor de los hombres).

Es significativamente distinta si:

$$\text{Retorno de hombres} \neq \text{Retorno de mujeres}$$

$$\rightarrow \hat{\beta}_{educ} \neq \hat{\beta}_{educ} + \hat{\beta}_{educ:female}$$

$$\rightarrow \hat{\beta}_{educ:female} \neq 0$$

```
confint(est_6, level=0.95)[["educ:female",]]
```

	2.5% <dbl>	97.5% <dbl>
educ:female	-0.04782198	0.008824567
1 row		

El intervalo de confianza al 95% de la $\hat{\beta}_{educ:female}$, incluye al 0. Por lo tanto, no podemos decir que es significativamente distinta la diferencia en el retorno a la educación entre hombres y mujeres.

d. ¿De qué tamaño es la diferencia en el retorno a un año extra de experiencia entre hombres y mujeres? ¿Es significativamente distinta esta diferencia?

De 1.08% dólares la hora por año extra de experiencia (en favor de los hombres).

Es significativamente distinta si:

$$\hat{\beta}_{exper:female} \neq 0$$

```
confint(est_6, level=0.95)[["exper:female",]]
```

	2.5% <dbl>	97.5% <dbl>
exper:female	-0.01642645	-0.0051581
1 row		

El intervalo de confianza al 95% de $\hat{\beta}_{exper:female}$ no incluye al 0. Por lo tanto, la diferencia al retorno es significativa.

e. En esta muestra y ceteris paribus ¿es más alto el logaritmo del sueldo promedio de las mujeres o el de los hombres? ¿En qué medida se encuentran?

Reescribiendo est_6 si es hombre o si es mujer, respectivamente:

$$\hat{lwage}_{male} = 0.2498 + 0.1018(educ) + 0.0149(exper)$$

$$\hat{lwage}_{female} = 0.3348 + 0.0823(educ) + 0.0041(exper)$$

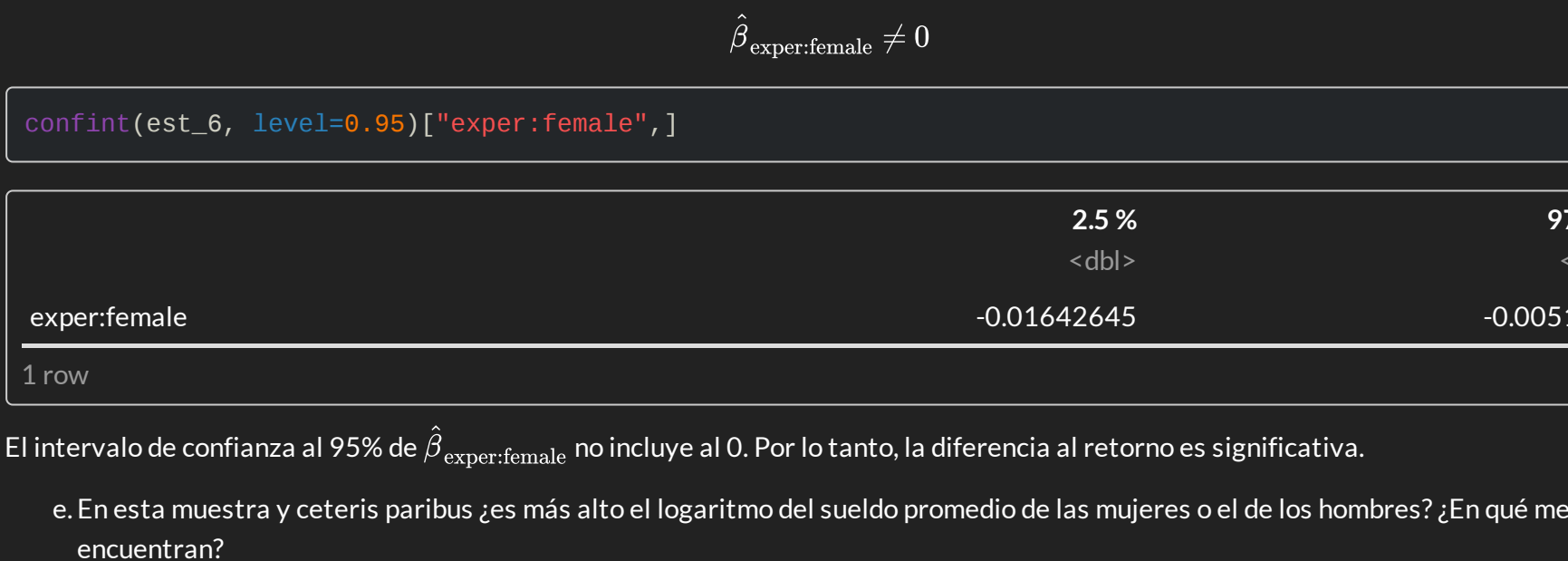
Revisando cuando \hat{lwage}_{female} , es mayor que \hat{lwage}_{male} y viceversa (*ceteris paribus*: considerando experiencia y educación iguales en ambos estimados):

$$\hat{lwage}_{male} < \hat{lwage}_{female} \rightarrow \text{exper} < 7.8703 - 1.8055(educ)$$

$$\hat{lwage}_{male} > \hat{lwage}_{female} \rightarrow \text{exper} > 7.8703 - 1.8055(educ)$$

Y serán iguales cuando:

$$\hat{lwage}_{male} = \hat{lwage}_{female} \rightarrow \text{exper} = 7.8703 - 1.8055(educ)$$



Para las distintas combinaciones de educación y experiencia (*ceteris paribus* entre hombres y mujeres), la zona gris oscura es cuando el estimado del logaritmo del sueldo es mayor para las mujeres. Para la zona gris clara, el estimado del logaritmo del sueldo es mayor para los hombres. Para la línea negra, cuando son iguales.

5. F-test

```
est_7 <- feols(lwage ~ educ + exper + exper*female + educ*female +
female + trade + services + profserv, data=wage1)
```

a. Escriba la hipótesis nula a probarse y estime (a mano) un test-F para verificar si en conjunto estas tres variables agregadas son significativas o no ¿qué concluye?

$$H_0 : \beta_{trade} = \beta_{services} = \beta_{profserv} = 0$$

```
SSR_r <- sum(resid(est_6)^2)
SSR_nr <- sum(resid(est_7)^2)
q <- (526-5-1) - (526-8-1)
F_test <- ((SSR_r-SSR_nr)/q)/(SSR_nr/(526-8-1))
VC.F <- qt(p=0.01, df=526, df2=517)
F_test
```

```
## [1] 18.68277
```

El valor crítico de la distribución F es 0.815. La F estimada está en la zona de rechazo (es 18.083), por lo que las tres variables agregadas al modelo son significativas en conjunto.

b. Estime la siguiente regresión e interprete el coeficiente servicios, trade y femtrade:

```
wage1$femtrade <- wage1$female*wage1$trade #adding a column with interactions
est_8 <- feols(wage ~ educ + female + exper + services + trade + femtrade, data = wage1)
etable(est_8, se = "hetero") # Robust Standard Errors (HCL)
```

	est_8 <chr>
Dependent Var:	wage
(Intercept)	-0.5338 (0.8406)
educ	0.5665*** (0.0629)
female	-2.268*** (0.3109)
exper	0.0581*** (0.0097)
services	-1.539*** (0.3637)
trade	-1.692*** (0.4601)
femtrade	0.4111 (0.5213)
1-10 of 14 rows	

$\beta_{services}$: Una persona que trabaja en servicios se estima que gane 1.539 dólares menos que una persona que no trabaja en servicios (sin tener clara cuál sería la base a comparar, pues varias industrias de la bases de datos no se incluyeron en la regresión, pero si se tomó en cuenta su educación y experiencia).

β_{trade} : Una persona que trabaja en trading se estima que gane 1.692 dólares menos que una persona que no trabaja en trading.

$\beta_{femtrade}$: Una persona que sea mujer y trabaje en trading gana 0.411 dólares más que una persona que sea mujer pero no trabaje en trading.