

Análisis estadístico

Se trata de evaluar el efecto de los cuestionarios autoevaluados en la calificación final de la asignatura de Biología.

Los datos se recogen en dos asignaturas: Biología Médica y Biología en Enfermería. En el primer caso, se pasaron 23 cuestionarios autoevaluados y en el segundo 9. Relacionadas con las calificaciones en estos cuestionarios se analizan tres variables: porcentaje de cuestionarios realizados, calificación media por cuestionario realizado y calificación media total, estas dos últimas sobre 10.

La calificación en la asignatura se resume con las siguientes características: calificación final parte teórica, calificación final parte práctica y calificación final parte teórica + práctica. Todas las calificaciones se consideran en la escala de 0 a 10. Además, se evalúa el número de convocatorias consumidas (1 o 2) y si la calificación final se corresponde con la convocatoria ordinaria o extraordinaria.

No se consideran los alumnos no presentados en ninguna de las dos convocatorias.

Todo el análisis se hace utilizando el software R versión 4.0.4 [R Core Team, 2021]. Se establece el nivel de significación en 0.05, marcando en negrita y cursiva los p-valores que resultan significativos y sólo en cursiva aquellos que se quedan al borde de la significación ($p < 0.1$).

Análisis descriptivo. Atendiendo al tipo de variable estadística se utilizan los siguientes estadísticos descriptivos,

- *Variables cuantitativas.* Se utilizan el tamaño muestral válido (n), la media, desviación típica (DT), un intervalo de confianza (IC) del 95% para la media, mediana (Med.), percentiles 25 y 75 (P_i), rango intercuartílico (IQR), mínimo (Min.) y máximo (Max.). Además, se utiliza el contraste de Kolmogorov-Smirnov (KS) para comprobar la hipótesis de normalidad.
- *Variables cualitativas.* Los estadísticos descriptivos son los porcentajes de cada nivel junto con un IC del 95% para ese porcentaje.

Relación entre pares de variables. Se calculan, para cada grado por separado, los coeficientes de correlación entre cada par de variables cuantitativas. Se utiliza el coeficiente de correlación de Spearman junto con un intervalo de confianza del 95% obtenido a partir de 5000 muestras bootstrap. Se representa la matriz de correlaciones a partir de un correlograma. Para evaluar la relación entre una variable cuantitativa y otra cualitativa se comparan las puntuaciones de la variable cuantitativa, contrastando la hipótesis de igualdad de medias para dos muestras independientes, utilizando el contraste t-Student. Cuando no pueda asumirse homogeneidad de varianzas, que se evalúa utilizando el test de Levene, se utiliza el contraste de Welch. En el caso en el que no sea posible asumir normalidad, se utilizará el contraste no paramétrico U de Mann-Whitney.

Efecto de los cuestionarios en los resultados finales. Se analiza el efecto de los cuestionarios en los resultados finales. Se considera como variable confusora la convocatoria,

si el resultado final se ha obtenido en la convocatoria ordinaria o en la extraordinaria. Para ello se ajustan modelos de regresión lineal.

Un modelo de regresión lineal tendrá el aspecto,

$$Y = b_0 + b_1X_1 + \dots + b_pX_p$$

donde Y es la variable dependiente o principal y X_1, \dots, X_p las variables independientes o explicativas. Los coeficientes b_i , se estiman para minimizar los residuos y su valor nos da información de la relación entre Y y la correspondiente X_i cuando el resto de las variables incluidas en el modelo permanecen constantes. En nuestros modelos Y será una de las 3 calificaciones finales (teórica, práctica o teórica+práctica), y como variables independientes se incluirán, en principio, las tres variables relacionadas con la realización de los cuestionarios, junto a la variable confusora convocatoria.

Como medidas de la bondad del ajuste del modelo utilizaremos el coeficiente de determinación o coeficiente de correlación múltiple al cuadrado (R^2) que se define como la proporción de la variabilidad explicada por el modelo, junto con el contraste de regresión, para evaluar si el modelo es significativo o no, y por tanto si existe o no una relación significativa entre las variables analizadas.

Por cada variable dependiente, se ajusta un modelo para el grado de medicina y otro diferente para el de enfermería.

Asunciones del modelo.

Para que el modelo de regresión sea válido se deben verificar los siguientes supuestos,

- Linealidad. Se asume que la variable respuesta depende linealmente de las variables explicativas. Si la respuesta no aparenta ser lineal habrá que introducir en el modelo transformaciones no-lineales o términos de interacción.
- Normalidad y homogeneidad de los residuos. Los residuos de un modelo se calculan como las diferencias entre los valores predichos por el modelo y los observados en la variable principal. Para tener un buen modelo de regresión, además de que los residuos deben ser pequeños, deben distribuirse según una normal y con varianza homogénea.
- Independencia. Los residuos son independientes entre sí.
- Colinealidad. Si dos variables independientes están muy relacionadas y ambas se incluyen en el modelo, el modelo ajustado muy posiblemente no será válido. En este caso, hay dos variables consideradas independientes que están muy relacionadas entre sí: el porcentaje de cuestionarios realizados y la nota total en ellos. Nos quedamos con la primera de ellas para nuestros modelos de regresión.
- Valores atípicos y datos influyentes. Pueden existir observaciones que afecten mucho al modelo, de forma que eliminando una única observación pudiera cambiar todo el ajuste del modelo. Conviene identificar estos puntos y evaluar si eliminarlos o no para el ajuste del modelo.

Vamos a verificar estos requisitos utilizando el **análisis de los residuos**. Se describe la distribución de estos residuos y se verifica la hipótesis de normalidad utilizando el contraste

de Kolmogorov-Smirnov y el QQ-plot. Se construye además el plot de residuos vs valores predichos, que sirve para determinar si hay algún patrón sistemático, lo que sugeriría falta de linealidad, o si existen crecimientos o decrecimientos de la variabilidad, en contra de la hipótesis de homogeneidad de varianza. Además, se utiliza el contraste de Breusch-Pagan para comprobar la hipótesis de homocedastidad en los residuos.

La hipótesis de independencia se comprueba utilizando el contraste de Durbin-Watson.

Para comprobar si existe multicolinealidad utilizamos el Factor de inflación de la Varianza (vif). Se considera que valores por encima de 5 indican problemas de colinealidad.

Por último, se trata de identificar posibles outliers. Aquellos puntos cuyo residuo, en valor absoluto, sea mayor que 3, podrían ser outliers. Se espera que aproximadamente el 99% de los residuos estén en la banda ± 3 .

Referencias

[R Core Team, 2021] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

URL <https://www.R-project.org/>. (15/02/2021)

EFFECTO DE LOS CUESTIONARIOS EN LOS RESULTADOS FINALES

GRADO: MEDICINA

Y: CALIFICACIÓN TEÓRICA

Modelo,

	\hat{b}_i	EE	IC 95% para \hat{b}_i		$H_0: \hat{b}_i = 0$		VIF
			Inf.	Sup.	t-valor	p-valor	
(Intercept)	-0.961	1.357	-3.637	1.716	-0.708	0.4798	
% test	0.008	0.006	-0.004	0.021	1.2993	0.1954	1.2377
Nota test	0.286	0.159	-0.028	0.599	1.7995	0.0735	1.2493
Conv=Ordinaria	3.251	0.441	2.381	4.121	7.3705	<0.0001	1.1546

EE = Error Estándar; VIF = Factor de inflación de varianza

Interpretación de los coeficientes:

- Con la misma puntuación media por cuestionario y en la misma convocatoria, por cada unidad que aumenta el porcentaje de cuestionarios realizados, la calificación media en la parte teórica aumenta en 0.008 puntos. Este efecto no es estadísticamente distinto de 0.
- Con el mismo porcentaje de cuestionarios realizados y en la misma convocatoria, por cada punto que aumenta la nota por cuestionario realizado, la calificación media en la parte teórica aumenta en 0.286 puntos. Este efecto no es estadísticamente distinto de 0, aunque se queda al borde de la significación.
- Con el mismo porcentaje de cuestionarios realizados y con la misma puntuación media por cuestionario, en convocatoria ordinaria la puntuación en la parte teórica es 3.25 puntos mayor de media. Este efecto es estadísticamente distinto de 0.

No se detectan problemas de multicolinealidad.

Bondad de ajuste

R ²	Contraste de regresión	
	F-valor	p-valor
0.3199	29.3241	<0.0001

El modelo es significativo y explica el 32% de la variabilidad total.

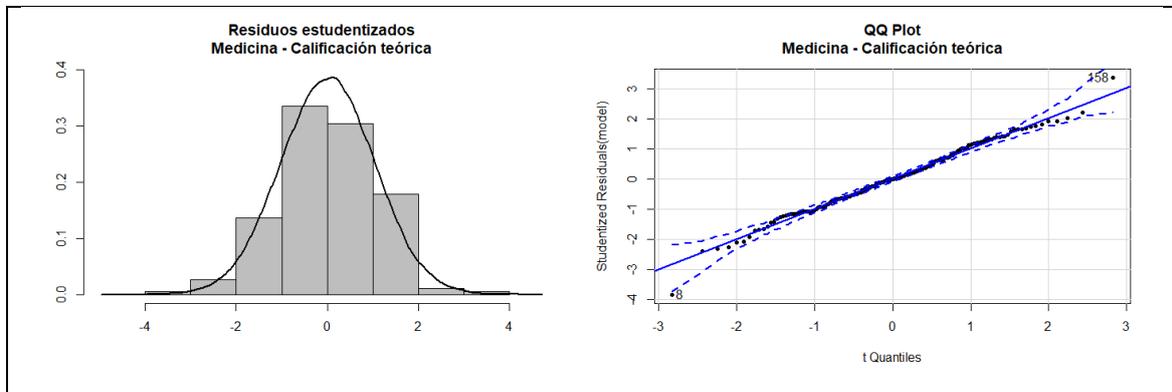
Análisis de los residuos

Distribución,

N	Media	DT	IC 95% para la media		Med.	Mín.	Max.	H ₀ : normalidad	
			Inf.	Sup.				KS	p-valor
191	0.01	1.03	-0.14	0.16	0	-3.84	3.36	0.035	0.8072

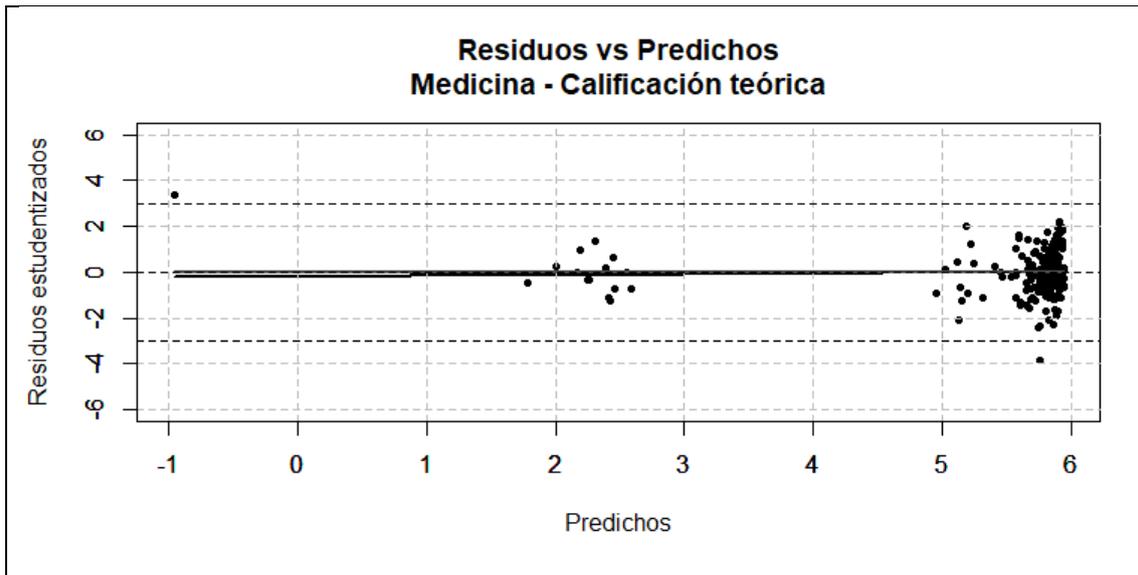
DT=Desviación Típica; Med. = Mediana; Mín.=Mínimo; Máx. = Máximo, KS = Kolmogorov-Smirnov

Gráficamente



No hay evidencias que nos hagan pensar que la distribución de los residuos es no-normal.

Gráfico de residuos vs predichos,



El estadístico de Breusch-Pagan 2.36 con p-valor 0.501 no rechaza la hipótesis de que la varianza de los residuos sea constante. El contraste de Durbin-Watson no rechaza la hipótesis de independiente (DW= 1.81; p-valor=0.2).

Hay dos puntos fuera de la banda ± 3 , pero es compatible con que aproximadamente el 99% esté dentro de esa banda. No hay outliers.

Y: CALIFICACIÓN PRÁCTICA

Modelo,

	\hat{b}_i	EE	IC 95% para \hat{b}_i		H ₀ : $\hat{b}_i = 0$		VIF
			Inf.	Sup.	t-valor	p-valor	
(Intercept)	7.149	1.386	4.414	9.884	5.1564	<0.0001	
% test	-0.006	0.006	-0.018	0.007	-0.8964	0.3712	1.2377
Nota test	0.113	0.162	-0.208	0.433	0.6935	0.4889	1.2493
Conv=Ordinaria	1.134	0.451	0.245	2.023	2.5167	0.0127	1.1546

EE = Error Estándar; VIF = Factor de inflación de varianza

Interpretación de los coeficientes:

- Con la misma puntuación media por cuestionario y en la misma convocatoria, por cada unidad que aumenta el porcentaje de cuestionarios realizados, la calificación media en la parte práctica disminuye en -0.006 puntos. Este efecto no es estadísticamente distinto de 0.
- Con el mismo porcentaje de cuestionarios realizados y en la misma convocatoria, por cada punto que aumenta la nota por cuestionario realizado, la calificación media en la parte práctica aumenta en 0.113 puntos. Este efecto no es estadísticamente distinto de 0.
- Con el mismo porcentaje de cuestionarios realizados y con la misma puntuación media por cuestionario, en convocatoria ordinaria la puntuación en la parte práctica es 1.134 puntos mayor de media. Este efecto es estadísticamente distinto de 0.

No se detectan problemas de multicolinealidad.

Bondad de ajuste

R ²	Contraste de regresión	
	F-valor	p-valor
0.041	2.664	0.0493

El modelo es significativo, aunque se queda muy al borde de la significación y explica sólo el 5% de la variabilidad total.

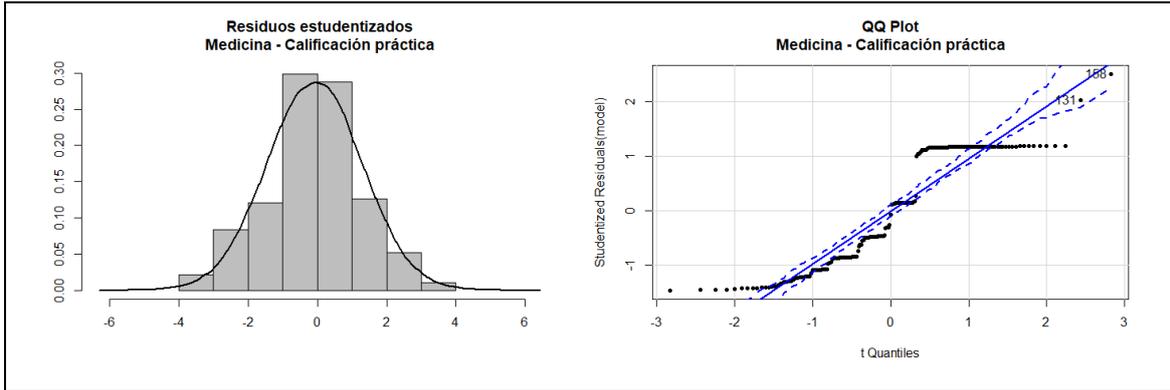
Análisis de los residuos

Distribución,

N	Media	DT	IC 95% para la media		Med.	Mín.	Máx.	H ₀ : normalidad	
			Inf.	Sup.				KS	p-valor
191	-0.07	1.37	-0.27	0.12	-0.06	-3.58	3.66	0.051	0.2575

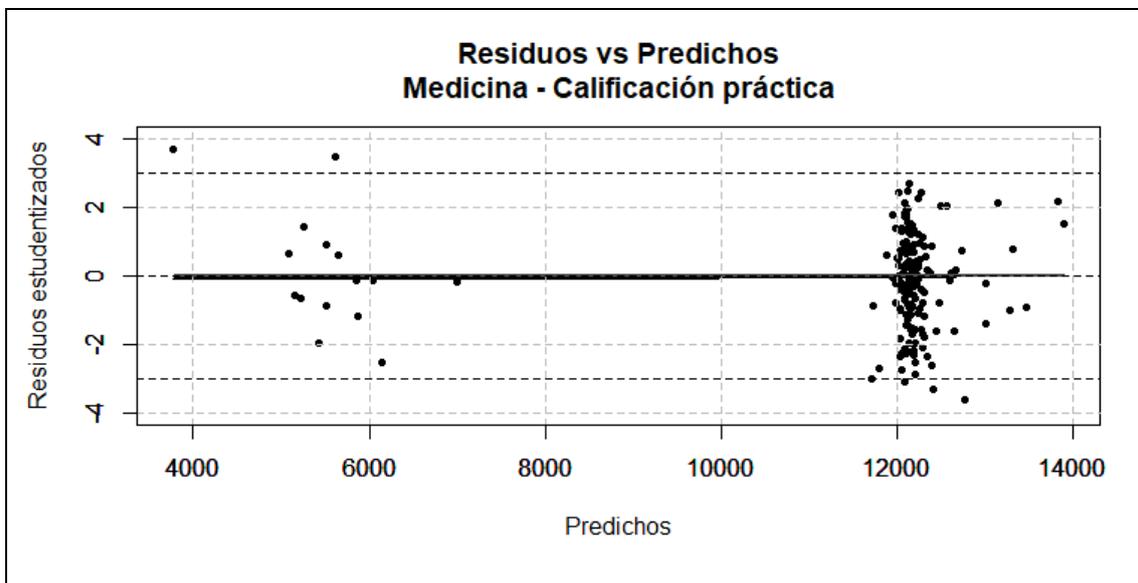
DT=Desviación Típica; Med. = Mediana; Mín.=Mínimo; Máx. = Máximo, KS = Kolmogorov-Smirnov

Gráficamente



No hay evidencias que nos hagan pensar que la distribución de los residuos es no-normal.

Gráfico de residuos vs predichos,



El estadístico de Breusch-Pagan 5.23 con p-valor 0.1559 no rechaza la hipótesis de que la varianza de los residuos sea constante. El contraste de Durbin-Watson no rechaza la hipótesis de independiente (DW= 2.15; p-valor=0.278).

Hay algunos puntos con residuos fuera de la banda ± 3 , pero muy cerca de este límite. No hay outliers.

Y: CALIFICACIÓN TEÓRICA+PRÁCTICA

Modelo,

	\hat{b}_i	EE	IC 95% para \hat{b}_i		$H_0: \hat{b}_i = 0$		VIF
			Inf.	Sup.	t-valor	p-valor	
(Intercept)	2.515	1.148	0.25	4.78	2.1903	0.0297	
% test	0.002	0.005	-0.008	0.013	0.4134	0.6798	1.2377
Nota test	0.212	0.134	-0.054	0.477	1.5739	0.1172	1.2493
Conv=Ordinaria	2.344	0.373	1.607	3.08	6.279	<0.0001	1.1546

EE = Error Estándar; VIF = Factor de inflación de varianza

Interpretación de los coeficientes:

- Con la misma puntuación media por cuestionario y en la misma convocatoria, por cada unidad que aumenta el porcentaje de cuestionarios realizados, la calificación media en la parte teórica+práctica aumenta en 0.002 puntos. Este efecto no es estadísticamente distinto de 0.
- Con el mismo porcentaje de cuestionarios realizados y en la misma convocatoria, por cada punto que aumenta la nota por cuestionario realizado, la calificación media en la parte teórica+práctica aumenta en 0.212 puntos. Este efecto no es estadísticamente distinto de 0.
- Con el mismo porcentaje de cuestionarios realizados y con la misma puntuación media por cuestionario, en convocatoria ordinaria la puntuación en la parte teórica+práctica es 2.34 puntos mayor de media. Este efecto es estadísticamente distinto de 0.

No se detectan problemas de multicolinealidad.

Bondad de ajuste

R ²	Contraste de regresión	
	F-valor	p-valor
0.2391	19.5823	<0.0001

El modelo es significativo y explica el 24% de la variabilidad total.

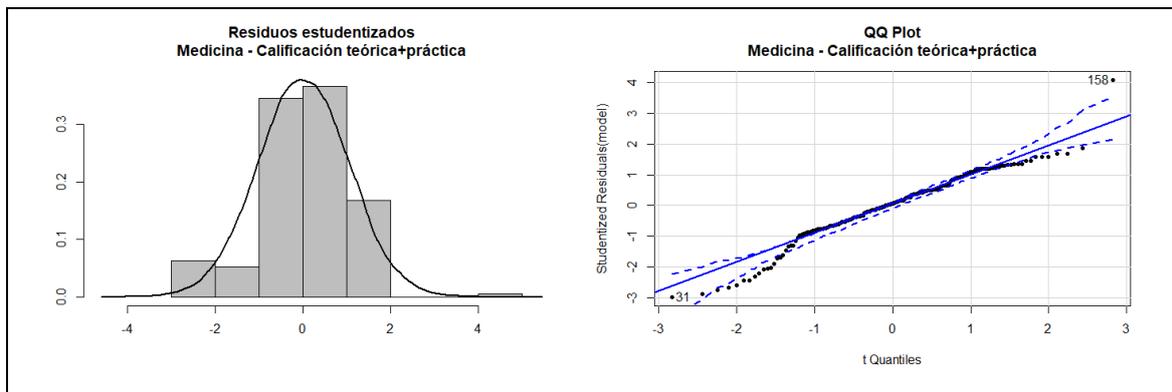
Análisis de los residuos

Distribución,

N	Media	DT	IC 95% para la media		Med.	Mín.	Max.	H ₀ : normalidad	
			Inf.	Sup.				KS	p-valor
191	0.01	1.04	-0.14	0.16	0.07	-2.97	4.08	0.064	0.0537

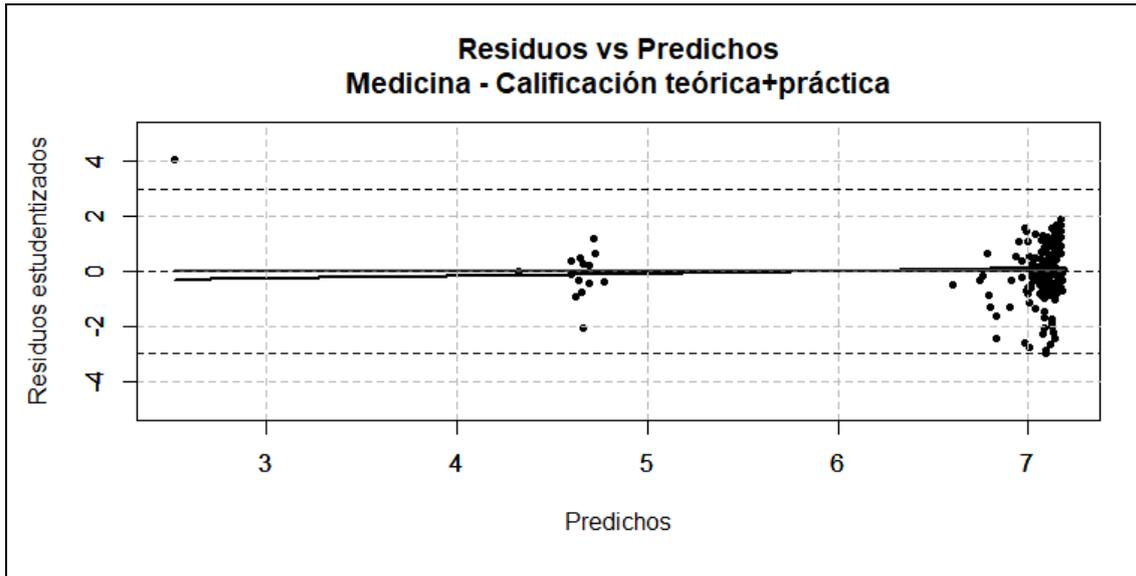
DT=Desviación Típica; Med. = Mediana; Mín.=Mínimo; Máx. = Máximo, KS = Kolmogorov-Smirnov

Gráficamente



No hay evidencias que nos hagan pensar que la distribución de los residuos es no-normal.

Gráfico de residuos vs predichos,



El estadístico de Breusch-Pagan 4.06 con p-valor 0.2549 no rechaza la hipótesis de que la varianza de los residuos sea constante. El contraste de Durbin-Watson no rechaza la hipótesis de independiente (DW= 1.97; p-valor=0.85).

Hay un punto fuera de la banda ± 3 , pero es compatible con que aproximadamente el 99% esté dentro de esa banda. No hay outliers.

GRADO: ENFERMERÍA

Y: CALIFICACIÓN TEÓRICA

Modelo,

	\hat{b}_i	EE	IC 95% para \hat{b}_i		$H_0: \hat{b}_i = 0$		VIF
			Inf.	Sup.	t-valor	p-valor	
(Intercept)	-5.296	1.836	-8.923	-1.67	-2.8854	0.0045	
% test	0.006	0.011	-0.015	0.026	0.5365	0.5924	1.0539
Nota test	0.814	0.19	0.438	1.19	4.2793	<0.0001	1.0923
Conv=Ordinaria	2.367	0.332	1.711	3.022	7.1307	<0.0001	1.079

EE = Error Estándar; VIF = Factor de inflación de varianza

Interpretación de los coeficientes:

- Con la misma puntuación media por cuestionario y en la misma convocatoria, por cada unidad que aumenta el porcentaje de cuestionarios realizados, la calificación media en la parte teórica aumenta en 0.006 puntos. Este efecto no es estadísticamente distinto de 0.
- Con el mismo porcentaje de cuestionarios realizados y en la misma convocatoria, por cada punto que aumenta la nota por cuestionario realizado, la calificación media en la parte teórica aumenta en 0.814 puntos. Este efecto es estadísticamente distinto de 0.
- Con el mismo porcentaje de cuestionarios realizados y con la misma puntuación media por cuestionario, en convocatoria ordinaria la puntuación en la parte teórica es 2.367 puntos mayor de media. Este efecto es estadísticamente distinto de 0.

No se detectan problemas de multicolinealidad.

Bondad de ajuste

R ²	Contraste de regresión	
	F-valor	p-valor
0.3813	31.2192	<0.0001

El modelo es significativo y explica el 31% de la variabilidad total.

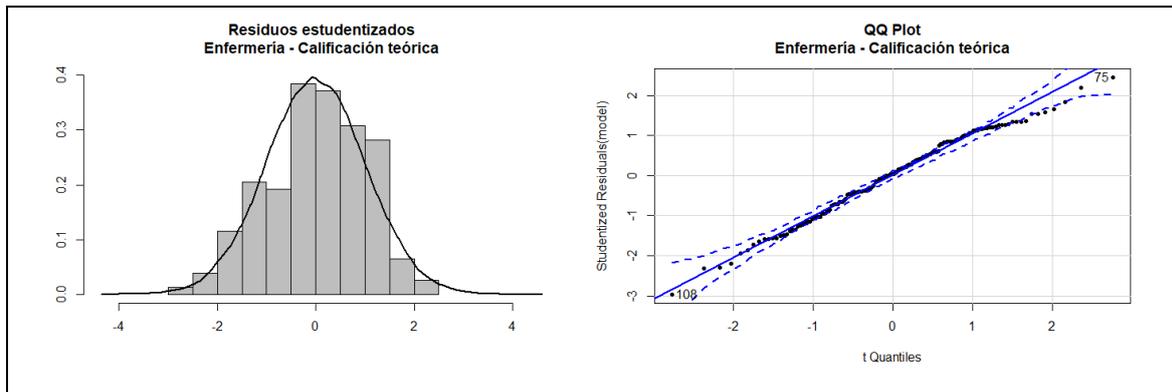
Análisis de los residuos

Distribución,

N	Media	DT	IC 95% para la media		Med.	Mín.	Max.	H ₀ : normalidad	
			Inf.	Sup.				KS	p-valor
156	0	1.01	-0.16	0.16	0.03	-2.96	2.45	0.059	0.2094

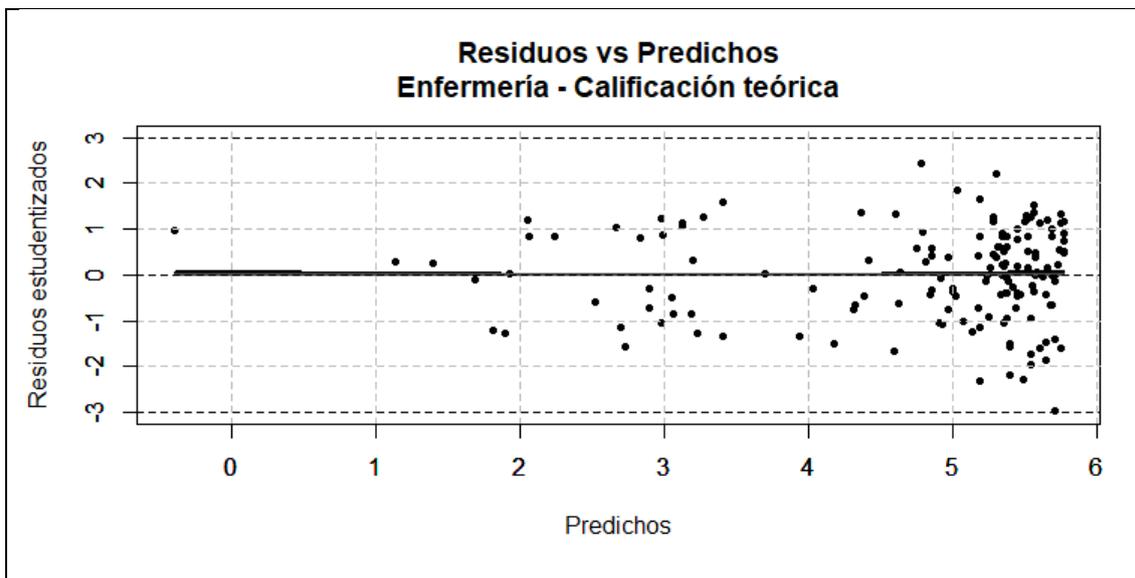
DT=Desviación Típica; Med. = Mediana; Mín.=Mínimo; Máx. = Máximo, KS = Kolmogorov-Smirnov

Gráficamente



No hay evidencias que nos hagan pensar que la distribución de los residuos es no-normal.

Gráfico de residuos vs predichos,



El estadístico de Breusch-Pagan 2.19 con p-valor 0.533 no rechaza la hipótesis de que la varianza de los residuos sea constante. El contraste de Durbin-Watson no rechaza la hipótesis de independiente (DW= 2.03; p-valor=0.902).

No hay puntos fuera de la banda ± 3 .

Y: CALIFICACIÓN PRÁCTICA

Modelo,

	\hat{b}_i	EE	IC 95% para \hat{b}_i		H ₀ : $\hat{b}_i = 0$		VIF
			Inf.	Sup.	t-valor	p-valor	
(Intercept)	-2.209	2.521	-7.189	2.772	-0.8762	0.3823	
% test	-0.005	0.014	-0.033	0.024	-0.3371	0.7365	1.0539
Nota test	0.886	0.261	0.37	1.401	3.3907	0.0009	1.0923
Conv=Ordinaria	2.067	0.456	1.166	2.967	4.5349	<0.0001	1.079

EE = Error Estándar; VIF = Factor de inflación de varianza

Interpretación de los coeficientes:

- Con la misma puntuación media por cuestionario y en la misma convocatoria, por cada unidad que aumenta el porcentaje de cuestionarios realizados, la calificación media en la parte práctica disminuye en -0.005 puntos. Este efecto no es estadísticamente distinto de 0.
- Con el mismo porcentaje de cuestionarios realizados y en la misma convocatoria, por cada punto que aumenta la nota por cuestionario realizado, la calificación media en la parte práctica aumenta en 0.886 puntos. Este efecto es estadísticamente distinto de 0.
- Con el mismo porcentaje de cuestionarios realizados y con la misma puntuación media por cuestionario, en convocatoria ordinaria la puntuación en la parte práctica es 2.067 puntos mayor de media. Este efecto es estadísticamente distinto de 0.

No se detectan problemas de multicolinealidad.

Bondad de ajuste

R ²	Contraste de regresión	
	F-valor	p-valor
0.2167	14.0146	<0.0001

El modelo es significativo y explica el 22% de la variabilidad total.

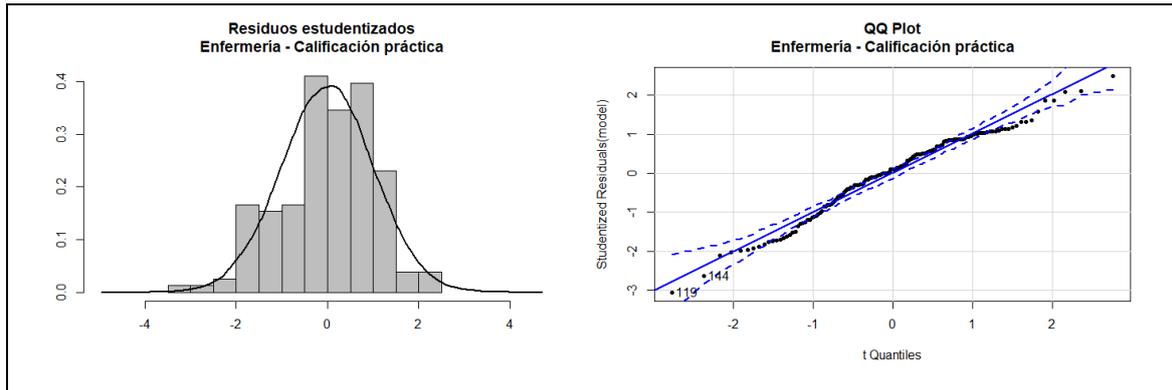
Análisis de los residuos

Distribución,

N	Media	DT	IC 95% para la media		Med.	Mín.	Max.	H ₀ : normalidad	
			Inf.	Sup.				KS	p-valor
156	0	1.02	-0.16	0.16	0.1	-3.05	2.49	0.069	0.0703

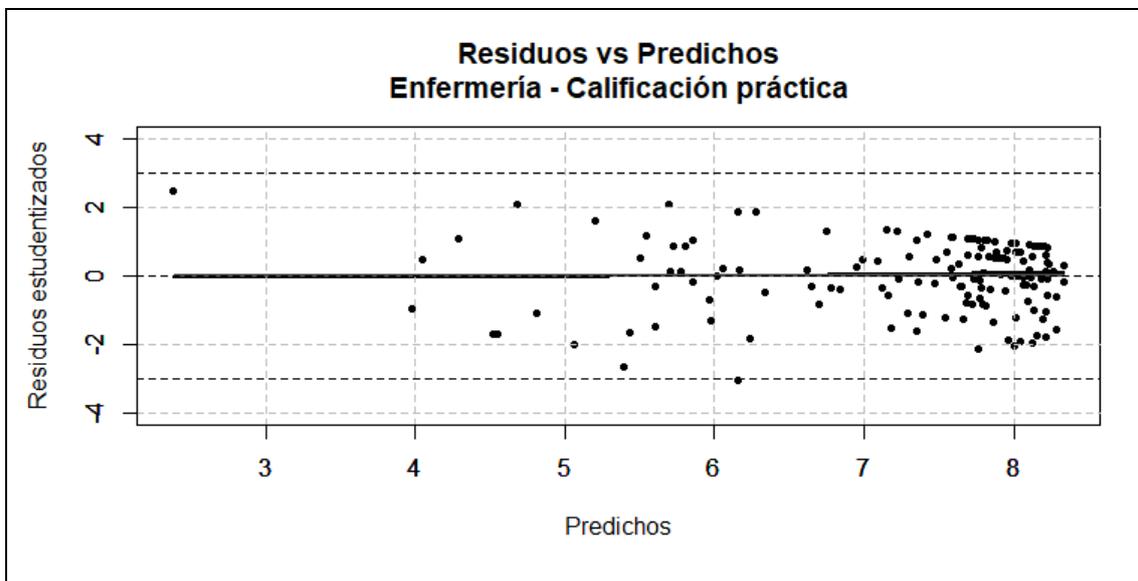
DT=Desviación Típica; Med. = Mediana; Mín.=Mínimo; Máx. = Máximo, KS = Kolmogorov-Smirnov

Gráficamente



No hay evidencias que nos hagan pensar que la distribución de los residuos es no-normal.

Gráfico de residuos vs predichos,



El estadístico de Breusch-Pagan 5.23 con p-valor 0.1559 no rechaza la hipótesis de que la varianza de los residuos sea constante. El contraste de Durbin-Watson no rechaza la hipótesis de independiente (DW= 2.14; p-valor=0.366).

No hay puntos con residuos fuera de la banda ± 3 .

Y: CALIFICACIÓN TEÓRICA+PRÁCTICA

Modelo,

	\hat{b}_i	EE	IC 95% para \hat{b}_i		$H_0: \hat{b}_i = 0$		VIF
			Inf.	Sup.	t-valor	p-valor	
(Intercept)	-3.973	1.729	-7.389	-0.557	-2.298	0.0229	
% test	0.001	0.01	-0.018	0.021	0.1148	0.9087	1.0539
Nota test	0.845	0.179	0.491	1.198	4.7149	<0.0001	1.0923
Conv=Ordinaria	2.238	0.313	1.621	2.856	7.1598	<0.0001	1.079

EE = Error Estándar; VIF = Factor de inflación de varianza

Interpretación de los coeficientes:

- Con la misma puntuación media por cuestionario y en la misma convocatoria, por cada unidad que aumenta el porcentaje de cuestionarios realizados, la calificación media en la parte teórica+práctica aumenta en 0.001 puntos. Este efecto no es estadísticamente distinto de 0.
- Con el mismo porcentaje de cuestionarios realizados y en la misma convocatoria, por cada punto que aumenta la nota por cuestionario realizado, la calificación media en la parte teórica+práctica aumenta en 0.845 puntos. Este efecto es estadísticamente distinto de 0.
- Con el mismo porcentaje de cuestionarios realizados y con la misma puntuación media por cuestionario, en convocatoria ordinaria la puntuación en la parte teórica+práctica es 2.238 puntos mayor de media. Este efecto es estadísticamente distinto de 0.

No se detectan problemas de multicolinealidad.

Bondad de ajuste

R ²	Contraste de regresión	
	F-valor	p-valor
0.3922	32.6873	<0.0001

El modelo es significativo y explica el 39% de la variabilidad total.

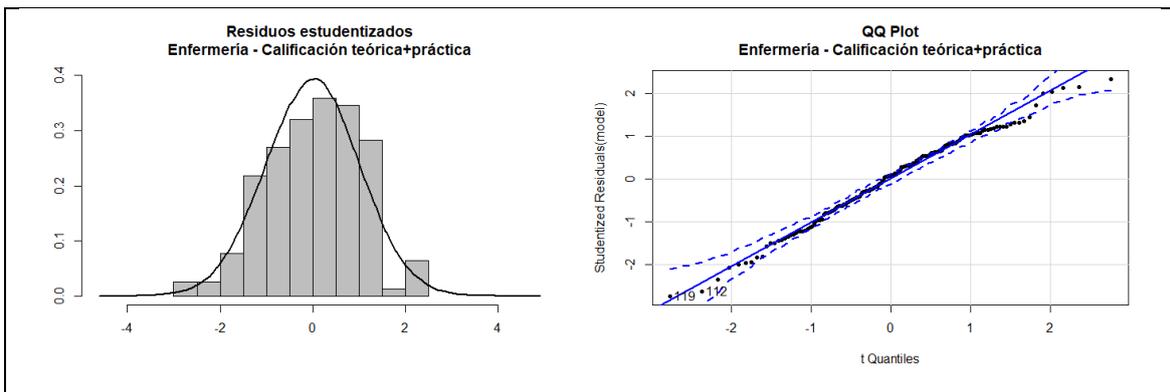
Análisis de los residuos

Distribución,

N	Media	DT	IC 95% para la media		Med.	Mín.	Máx.	H ₀ : normalidad	
			Inf.	Sup.				KS	p-valor
156	0	1.01	-0.16	0.16	0.08	-2.73	2.34	0.057	0.2417

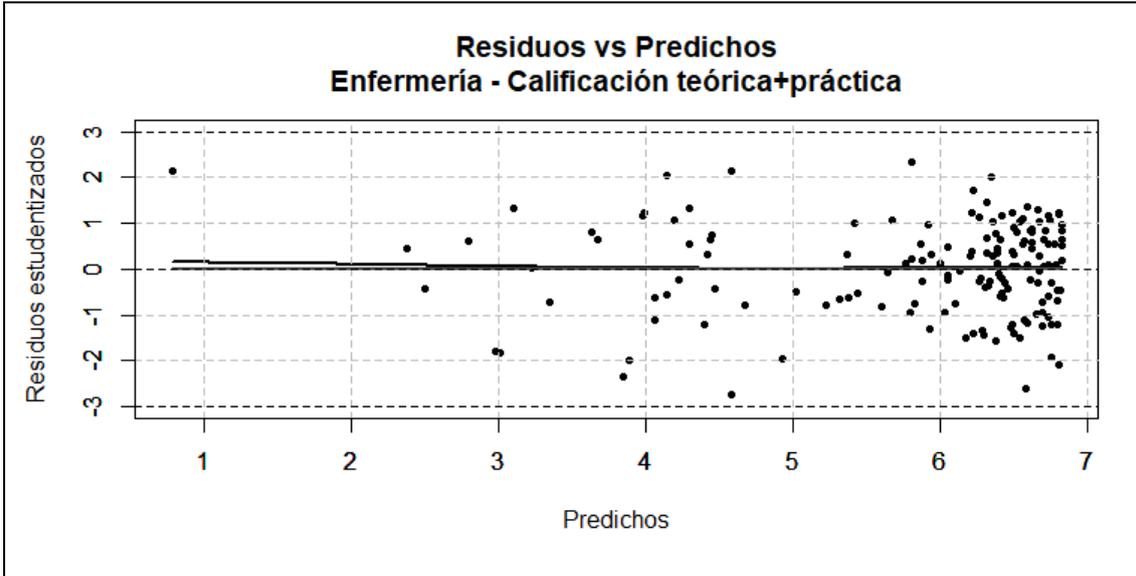
DT=Desviación Típica; Med. = Mediana; Mín.=Mínimo; Máx. = Máximo, KS = Kolmogorov-Smirnov

Gráficamente



No hay evidencias que nos hagan pensar que la distribución de los residuos es no-normal.

Gráfico de residuos vs predichos,



El estadístico de Breusch-Pagan 4.06 con p-valor 0.2549 no rechaza la hipótesis de que la varianza de los residuos sea constante. El contraste de Durbin-Watson no rechaza la hipótesis de independiente (DW= 2.1; p-valor=0.518).

No hay puntos fuera de la banda ± 3 .