Sociology and Psychology Research Journals
Volume 1 Issue 1 February 2024
ISSN: Pending…

DEMYSTIFYING QUANTILE REGRESSION: UNLOCKING ITS POTENTIAL IN PSYCHOLOGICAL RESEARCH

[1]Alejandro González García and María [2]Fernanda Rodríguez López
[1]School of Psychology, Autonomous University of Nuevo León, Mexico.
[2]Department of Basic Sciences of the School of Medicine and Health Sciences of Tecnológico de Monterrey, Mexico.

Abstract: This methodological article aims to elucidate the theoretical aspects of quantile regression, advocating for its broader recognition and utilization as a predictive tool. Despite its utility, quantile regression remains relatively obscure. To enhance comprehension and facilitate its application, we illustrate its implementation through an example in the domain of social and health psychology, specifically focusing on attitudes towards individuals living with HIV/AIDS. By providing a practical demonstration within this context, we aim to demystify quantile regression and highlight its relevance in various fields of research.
Keywords: quantile regression, predictive tool, social psychology, health psychology, attitudes towards HIV/AIDS.

INTRODUCTION

The objective of this methodological article is to present quantile regression in its theoretical aspects to promote its knowledge and use, since it is a very useful but little-known predictive tool. An example applied to the field of social and health psychology on the attitude towards people living with HIV/AIDS is used to make the presentation of the analysis technique more practical and understandable.

In psychology and health sciences, a very frequent practice for data analysis involves the dichotomization of the quantitative variables that are intended to be predicted. In the clinical setting, for example, we can see this practice when establishing cut-off points to determine the presence or absence of a target condition (HajianTilaki, 2018), thus allowing the use of binary logistic regression, which is a method that allows the introduction of continuous, ordinal or categorical variables in the predictive model, as opposed to multiple linear regression, which is an analysis technique that exclusively allows the use of quantitative variables (Stolper and Walter, 2019).

A regression technique that requires a quantitative variable as the predicted variable and that accepts any type of predictor variable is quantile regression, which is a better option than dichotomizing and estimating a binary logistic regression model (Waldmann, 2018). Indeed, when the predicted variable is quantitative, quantile regression is a better option than transforming the predicted variable into an ordinal variable (after defining k class intervals) to apply ordinal logistic regression since it makes use of all the information content of the quantitative variable (variance) and allows defining models for different quantile orders, for instance 0.25, 0.50, and 0.75 (Koenker et al., 2017; Konstantopoulos et al., 2019). Furthermore, quantile regression was developed as an alternative to ordinary least squares linear regression when the assumptions of homoscedasticity and normality of errors distribution are not fulfilled (Furno and Vistocco, 2018); consequently, about the fulfillment of assumptions, quantile regression is a very flexible non-parametric technique. Moreover, quantile regression can also be adapted to situations in which there exist correlated errors (Alhamzawi and Ali, 2018, 2020; IBM, 2021).

HISTORICAL NOTE

This regression technique was developed by Koenker and Bassett (1978) based on the works written by several authors, namely: Bošković (1757), who wrote about minimum absolute errors; Laplace (1789), whose work was related to the situation method; and Edgeworth (1887, 1888), who introduced the concept of the plural median. Initially, ordinal regression was applied in economic and business sciences; nevertheless, it was soon realized that it was an excellent option for analysis in ecology and health sciences (Cade and Noon, 2003; Koenker, 1998; Staffa et al., 2019), which are scientific fields in which it is common to find non-normal, heteroscedastic non-quantitative variables and non-linear interactions. Thanks to the development of computer statistics that, finally, this analysis technique has become popular, since it requires complex calculation procedures based on linear programming. Nowadays, statistical software packages (e.g., R, SPSS, STATA, Matlab, Eviews, and GRETL among others) can perform this regression analysis (Furno and Vistocco, 2018).

THEORETICAL BASIS AND TECHNICAL ASPECTS

If ordinary least squares regression predicts the mean values of $Y \in R$ conditional on the vector $x \in R^p$ (or vector of the scores on the predictor variables), quantile regression predicts the values of the median or other quantiles of Y conditional on the vector x. The estimation is performed by minimizing the sum of the absolute deviations. This minimization is usually solved by the simplex method, introduced by Edgeworth (1888) and developed by Barrodale and Roberts (1974). Although other computational options exist, they require large samples, demand more computational resources, and may have more convergence difficulties than the simplex method (Alhamzawi and Ali, 2018; Lustig et al., 1994; Yang et al., 2016).

Quantile regression posits the estimation of the quantile of order τ of the variable Y, $Q_Y(\tau)$, as a minimization problem (Koenker, 2005).

$$Q_Y(\tau) = F^{-1}\big(Q_Y(\tau)\big) = q \in \mathcal{R}; F\big(Q_Y(\tau)\big) = \tau \in (0,1)$$

$$Q_Y(\tau) = \underset{q\in\mathcal{R}}{\arg\min} \sum_{i=1}^{n} \rho_\tau(y_i - q) = \underset{q\in\mathcal{R}}{\arg\min} \left(\sum_{y_i \geq q} (\tau - 1)(y_i - q) + \sum_{y_i < q} \tau(y_i - q) \right)$$

$$= \underset{q\in\mathcal{R}}{\arg\min} \left(\tau \sum_{y_i < q} (y_i - q) - (1 - \tau) \sum_{y_i \geq q} (y_i - q) \right)$$

Where $Q_Y(P) = q$ = quantile function or inverse of the cumulative distribution function, $F_Y(y) = P(Y \leq y) = \tau$ = cumulative distribution function, τ = cumulative probability or quantile order, q = value of the quantile of order τ of the variable Y, and ρ_τ = loss function of the quantile of order τ of the variable Y.

$$u = y_i - q$$

$$\rho_\tau(u) = u \times (\tau - 1_{u<0})$$

$$\textit{Indicator function}: 1_{u<0} = \begin{cases} y_i - q \geq 0 & 0 \\ y_i - q < 0 & 1 \end{cases}$$

Next, the conditional quantile to a linear model based on k predictor variables is defined, and it is proposed to estimate the vector of regression weights through the minimization of the loss function of the conditional quantile (Koenker, 2005).

$$Q_{Y|X}(\tau) = \mathrm{X}\hat{\beta}$$

$$\hat{\beta} = \underset{\beta\in\mathcal{R}^k}{\arg\min} \left[\sum_{i=1}^{n} \rho_\tau(Y_i - \beta \mathrm{X}_i) \right]$$

where X = design matrix with a unit vector in the first column and the scores of the *n* participants in the *k* variables, which can be either quantitative (cofactors), ordinal or qualitative (factors). $\hat{\beta}$ = vector of estimated parameters with the intercept of the model, the regression weights of the cofactors, and the position parameters of the categories of the factors. ρ_τ = loss function of the quantile of order τ of the conditional variable Y to Xβ.
Usually, the order of the quantile is one half, that is, the median. When this quantile is chosen, which is the default option in statistical software packages (IBM, 2021; 28 Koenker, 2016), the optimization problem consists in minimizing the sum of the absolute deviations (Koenker,
2005)

$$\hat{\beta} = \arg\min_{\beta \in \mathcal{R}^k} \left[\sum_{i=1}^{n} \left| Y_i - \hat{\beta} X_i \right| \right]$$

The Statistical Package for Social Sciences (SPSS) can handle multiple cofactors (quantitative variables) and factors (nominal and ordinal variables), taking the last nominal or ordinal category of the factor as the reference category (IBM, 2021; Könker, 2016); likewise, it allows the application of two methods to estimate the parameters: the simplex method (Barrodale and Roberts, 1974; Koenker and d'Orey, 1987) and the Frisch-Newton interior-point method for nonlinear optimization (Frisch, 1956; Lustig et al., 1994). This statistical software chooses the most convenient method as a function of the computational requirements of the task; the simplex method is more suitable for small samples, whereas the Frisch-Newton method is more efficient for large sample sizes (Koenker et al., 2017). By default, the error terms are assumed to be independently and identically distributed, but this option can be changed to covariant and heteroscedastic errors. The scatter plot, where the xaxis represents the observed scores and the y-axis represents the predicted scores, can be examined to find out which assumption fits more to the data set. A funnelshaped (or an almond-shaped) point cloud indicates the presence of heteroscedastic residuals. In turn, the independence of the errors can be verified through the Wald and Wolfowitz run test (1943) and a graph of the sequence of the residuals, plotted in the order of collection. If the sequence reveals regular patterns, and a residual can be predicted by the previous one or another previous one, it is inferred that there is a serial dependence between the prediction errors.
SPSS presents the point estimates, asymptotic standard errors, significance tests with Student's t distribution with $n-p$ degrees of freedom, and 95% confidence intervals for the *p* parameters (model intercept, regression coefficients corresponding to the cofactors, and position parameters of the categories of the factors), the calculation of the Pseudo R-Squared coefficient suggested by Koenker and Machado (1999), the mean absolute error, the point estimates and 95% confidence intervals for the Y-scores and the residuals. It also computes the variance-covariance matrices and the correlations of the estimated parameters, either through the nonparametric method developed by Bofinger (1975), which is the default method, or through the parametric method proposed by Hall and Sheather (1988).
As with other regression methods, it is possible to specify nested effects and interactions between variables (IBM, 2021). Nested effects can be included in the quantile regression model when the values of one variable are only known for specific values of another variable and these two variables do not covary within their full potential range of values. The interaction between variables can be introduced in the model when there is significant and non-linear covariance between two predictor variables (Koenker et al., 2017).

EXAMPLE OF THE APPLICATION IN SOCIAL PSYCHOLOGY AND HEALTH SCIENCES

The following is an example of an application of quantile regression. It focuses exclusively on its statistical and analytical characteristics and ignores the theoretical aspects of the field of psychology; therefore, no theoretical framework, hypothesis formulation, or discussion of the data is provided. A relatively small sample size, but appropriate for the technique, was chosen to make the presentation of the analyses more manageable.

Considering the example a random sample of 40 young adult men (18 to 40 years old) drawn from a population of patients receiving medical care in a medical center located in a city in Mexico. The mean schooling of the participants is 10 years. Religiosity (X_2) is assessed through a closed-ended question. The attitudes toward gay people as well as the attitudes toward people living with HIV/AIDS (Y) are assessed through two self-report scales, namely: the 10-item Scale of Attitude toward Homosexuality (EAH-10) (Moral and Ortega, 2010; Moral and Martínez-Sulvarán, 2012) and the Scale of Attitude toward People Living with HIV/AIDS (Moral and Valle, 2020, 2021). The question regarding religiosity asks about the frequency of attendance at religious services and had five answer options: 1 = never or only in special services related to personal and cultural commitments, 2 = at least once a year motivated by religious faith or religious duty, 3 = at least once a month motivated by religious faith or religious duty, 4 = once or almost once a week, and 5 = at least once a week (Moral, 2010). Scores on the two attitude scales are percentile scores from 1 to 100; in both scales, a higher percentile score evidences a greater level of rejection toward the attitudinal object (that is, a more negative attitude).

Now, taking into account the data shown in Table 2, the objective is to estimate a model to predict an attitude of rejection toward people living with HIV/AIDS (quantitative variable measured on an interval scale) as a function of religiosity (variable of ordered categories) and the level of rejection toward gay people (quantitative variable measured on an interval scale) using quantile regression of order $\tau = 0.5$ (predicted median values).

Table 1 shows the point and interval estimates of the parameters of the predictive model as well as their asymptotic standard errors and the tests of statistical significance (Student's t-test with degrees of freedom = $n - p = 40 - 6 = 32$). Six parameters were estimated ($p =$
6): the intercept of the model b_0, the regression weight of

Table 1. Estimation and significance of the parameters of the quantile regression model of order $\tau = 0.5$ (predicted median values).

Parameter	b_i	s_{bi}	*t*	*df*	*Sig.*	*LL*	*UL*	*r*
b_0	42.4314	13.1617	3.2238	34	.0028	15.6836	69.1792	0.4839
b_1	0.5098	0.1242	4.1033	34	.0002	0.2573	0.7623	0.5755
$b2\|X2=1$	-36.3529	11.8796	-3.0601	34	.0043	-60.4951	-12.2108	0.4647
$b2\|X2=2$	-22.1569	11.9111	-1.8602	34	.0715	-46.3631	2.0493	0.3039
$b2\|X2=3$	-7	12.1154	-0.5778	34	.5672	-31.6214	17.6214	0.0986
$b2\|X2=4$	-3.2549	12.6056	-0.2582	34	.7978	-28.8726	22.3628	0.0442
$b2\|X2=5$	0							

Dependent variable = Y = attitude toward people living with HIV/AIDS. Predictor variables: X_1 = attitude toward gay people and X_2 = religiosity = {1 = very low, 2 = low, 3 = medium, 4 = high, 5 = very high}. The ordered category 5 (very high religiosity) was taken as the reference category and, as a consequence, a location or intercept parameter was not estimated. Estimated parameters (b_i): b_0 = intercept of the model, b_1 = weight of the quantitative variable (attitude toward gay people), and $b_{2|X2}$ = conditional location parameters (constants) to the value of religiosity (from 1 to 4; category 5 was used as the reference category). s_{bi} = standard deviation or error of the parameter estimates, $t = b_i/s_{bi}$ = value of the contrast statistic for the significance of the estimated

parameter, $df = n - p$ = degrees of freedom for the test of significance or difference between the size sample n and the number of estimated parameters p, Sig. = two-tailed probability in a Student's t-distribution with $n - p$ degrees of freedom, *LL* = lower limit of the interval estimate of the parameters of the quantile regression model of order 0.5 (median value) and with a confidence level at 95%, *UL* = upper limit of the aforementioned interval, $r = |t|/\sqrt{(t^2+df)}$ = effect size estimated by Cohen's d. The estimation of the parameters and their errors was carried out using the simplex method. The error terms were assumed to be independently and identically distributed.
Source: Authors

the cofactor b_1 (attitude toward gay people), and the four position parameters for religiosity $b_{2|X2=1}$, $b_{2|X2=2}$, $b_{2|X2=3}$ y $b_{2|X2=4}$ (categories ordered from 1 to 4; category 5 was used as the reference category). The statistical package chose the Barrodale-Roberts simplex method (1974) to estimate these six parameters, being this method the most suitable for the analysis of this small sample (n = 40). It was assumed that the error terms were independently distributed and had homogeneity of variance. In this model, the significant parameters were the intercept, the weight of the attitude toward gay people, and the location parameter of people with very low religiosity (first ordered category); the other three location parameters were not significant (from the second to the fourth ordered category of religiosity).

According to Ringquist (2013), for a given regression coefficient whose significance is tested using a Student's t-test with degrees of freedom df, $t = b_j/s_{bj} \sim t_{df}$, the correlation-based effect size can be estimated through the following statistic: r = |t|/√(t2+df). The effect size with this type of statistic can be interpreted using the cut-off points suggested by Cohen (1988) for the correlation coefficient: 0.1 small, 0.3 medium, 0.5 large, and 0.7 very large. Returning to the data shown in Table 1, the attitude of rejection toward gay people acts as a risk factor for rejection toward people living with HIV/AIDS, b1 = 0.51, 95% CI [0.26, 0.76], with a large effect size, 0.50 < r = |t|/√(t2+df) = 0.58 < 0.70. A very low level of religiosity, compared to a very high level of religiosity, acts as a protective factor, b2|X2=1 = −36.35, 95% CI [−60.50, −12.21] and shows a medium effect size, 0.30 < r = |t|/√(t2+df) = 0.47 < 0.50.

Table 2 shows the sample data of the 40 participants, as well as the predictions, the error of each prediction, the interval estimate of the predictions (confidence level

$$T = b_1/s_{b_1} = 0.5098/0.1242 = 4.1033 \sim t_{gl=n-p} gl = n - p = 40 - 6 = 34$$

$$Sig. = 2 \times (1 - P(t_{34} \leq t = 4.1033)) = .0002 < \alpha = .05$$

$$P\left(b_1 - {}_{1-\frac{\alpha}{2}}t_{n-p} \times s_{b_1} \leq \beta_1 \leq b_1 + {}_{1-\frac{\alpha}{2}}t_{n-p} \times s_{b_1}\right) = 1 - \alpha$$

$$P(0.5098 - {}_{0.975}t_{34} \times 0.1242 \leq \beta_1 \leq 0.5098 + {}_{0.975}t_{34} \times 0.1242) = .95$$

$$P(0.5098 - 2.0322 \times 0.1242 \leq \beta_1 \leq 0.5098 + 2.0322 \times 0.1242) = .95$$

$$P(0.2573 \leq \beta_1 \leq 0.7623) = .95$$

Table 2. Observed scores, predictions, and prediction residuals.

i	x_{i1}	x_{i2}	y_i	$\hat{y}_i$	$s_{\hat{y}_i}$	LL_i	UL_i	e_i
1	18	3	39	44.608	6.160	32.09	57.126	-5.608
2	62	1	31	37.686	4.692	28.15	47.223	-6.686
3	50	1	40	31.569	4.315	22.8	40.337	8.431

4	33	4	56	56	6.490	42.81	69.19	0
5	63	2	62	52.392	4.784	42.67	62.114	9.608
6	58	3	65	65	5.078	54.68	75.32	0
7	41	2	32	41.176	4.579	31.87	50.483	-9.176
8	0	2	26	20.275	7.373	5.29	35.259	5.725
9	15	1	22	13.725	5.681	2.18	25.271	8.275
10	63	2	54	52.392	4.784	42.67	62.114	1.608
11	40	3	71	55.824	5.055	45.55	66.097	15.176
12	71	2	47	56.471	5.201	45.9	67.041	-9.471
13	73	3	70	72.647	5.751	60.96	84.334	-2.647
14	55	2	48	48.314	4.539	39.09	57.537	-0.314
15	74	2	58	58	5.393	47.04	68.96	0
16	71	2	55	56.471	5.201	45.9	67.041	-1.471
17	49	1	9	31.059	4.300	22.32	39.798	-22.059
18	50	4	54	64.667	6.070	52.33	77.003	-10.667
19	67	1	45	40.235	4.968	30.14	50.331	4.765
20	23	2	32	32	5.447	20.93	43.07	0
21	41	3	72	56.333	5.031	46.11	66.557	15.667
22	58	5	72	72	10.506	50.65	93.35	0
23	78	4	96	78.941	6.801	65.12	92.762	17.059
24	29	1	28	20.863	4.750	11.21	30.515	7.137
25	31	3	35	51.235	5.381	40.3	62.171	-16.235
26	72	4	96	75.882	6.506	62.66	89.105	20.118
27	63	2	43	52.392	4.784	42.67	62.114	-9.392
28	49	1	38	31.059	4.300	22.32	39.798	6.941
29	47	4	65	63.137	6.095	50.75	75.525	1.863
30	14	2	28	27.412	6.127	14.96	39.864	0.588
31	41	1	23	26.980	4.330	18.18	35.781	-3.980
32	61	5	61	73.529	10.505	52.18	94.879	-12.529
33	69	1	40	41.255	5.090	30.91	51.600	-1.255
34	39	3	42	55.314	5.080	44.99	65.637	-13.314
35	34	4	42	56.510	6.451	43.4	69.620	-14.510
36	47	1	46	30.039	4.290	21.32	38.758	15.961
37	43	1	28	28	4.306	19.25	36.75	0
38	94	3	99	83.353	7.358	68.4	98.306	15.647
39	43	3	72	57.353	4.996	47.2	67.506	14.647
40	32	1	9	22.392	4.607	13.03	31.754	-13.392

i = order in data collection (from 1 to n), x_{i1} = percentile score of participant i on the attitude of rejection toward homosexuality, x_{i2} = ordered category of religiosity for participant i, y_i = percentile score of participant i on the attitude of rejection toward people living with HIV/AIDS, $\hat{y}_i$ = median score predicted for participant i by the quantile regression model (order τ = 0.5), $s_{\hat{y}_i}$ = standard deviation or error of the parameter estimates, LL = lower limit of the interval estimate of the median score for participant i and with a confidence level at 95%, UL = upper limit of the aforementioned interval, e_i = residual or sample error of prediction for participant i.
Source: Authors at 95%), and the residuals or sample prediction errors.
For instance, the first participant obtained an 18th percentile score on the scale that assessed rejection toward gay people (x_1 = 18), was classified as having a

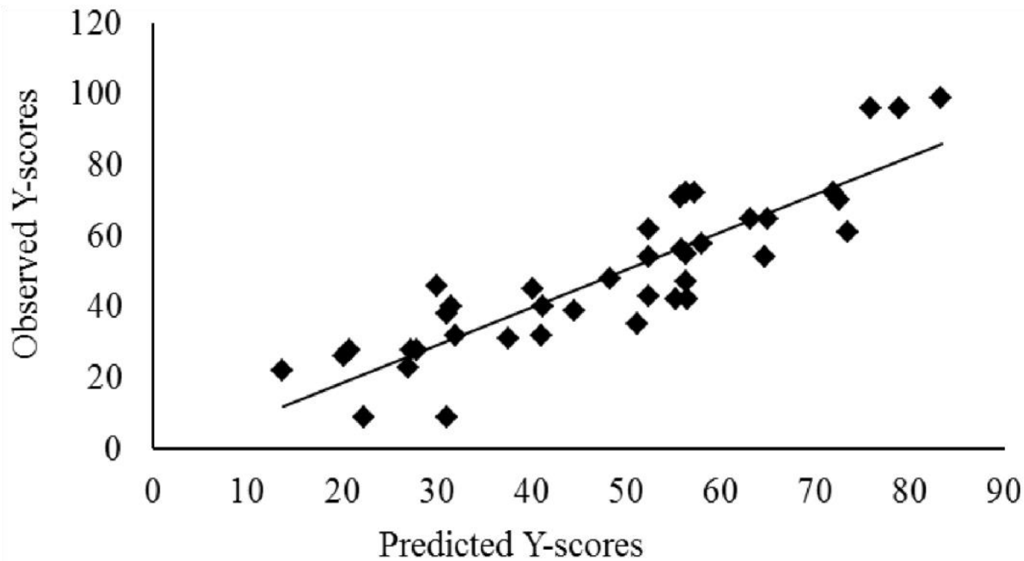

Figure 1. Scatter plot showing the relationship between predicted and observed values. Source: Authors
medium level of religiosity (x_2 = 3), reached a 39th by the quantile regression model was equal to 44.61 percentile score in the level of rejection toward people (95% *CI* [32.09, 57.61]) and the residual was −5.61. living with HIV/ AIDS (y = 39): the predicted score yielded The scatter plot between observed (X-axis) and predicted

$$\hat{y}_{i=1} = b_0 + b_1 x_{i1} + b_{2|X_2=3} = 42.4314 + 0.5098 \times 18 - 7 = 44.6078$$

$$P\left(\hat{y}_i - {}_{1-\frac{\alpha}{2}}t_{n-p} \times s_{\hat{y}_i} \leq \upsilon_i \leq \hat{y}_i + {}_{1-\frac{\alpha}{2}}t_{n-p} \times s_{\hat{y}_i}\right) = 1 - \alpha$$

$$P(44.6078 - 2.0322 \times 6.1596 \leq \upsilon_i \leq 44.6078 + 2.0322 \times 6.1596) = 1 - \alpha$$

$$P(32.0900 \leq \upsilon_i \leq 57.1257) = 0.95$$

$$e_i = y_i - \hat{y}_i = 39 - 44.6078 = -5.6078$$

(Y-axis) Y-scores shows homogeneity in the opening of the point cloud around an ascending straight line (Figure 1). On the other hand, the Wald and Wolfowitz (1943) run test allows us to maintain the null hypothesis of independence of errors. To perform this test, the residuals are arranged in the order of collection of the score vectors (i from 1 to 40); thereafter, the median of residuals is calculated, Mdn(E) = 0, and it is subsequently used

as a criterion to dichotomize them: if $e_i < Mdn$(E), $d_i = 0$; and if $e_i \geq Mdn$(E), $d_i = 1$. Afterward, the number of residuals lower than the criterion or zeros in D ($n_0 = 17$) and the number of residues higher than or equal to the criterion or ones in D ($n_1 = 23$) are counted. Additionally, the runs of zeros and ones in D are calculated ($R = 15$). Since both n_0 and n_1 are higher than 20, the exact probability is computed. The punctual probability is 0.025, the left-tailed exact probability ($R = 15 < Mdn$(R) = 20.5) equals to 0.048, and the two-tailed exact probability equals to 0.073, which is a value higher than the conventional level of significance ($\alpha = 0.05$). The null hypothesis would also hold with a two-tailed asymptotic probability and a significance level of 0.05: E(R) = 20.55, SD(R) = 3.05, $Z = (R - 0.5 - E(R))/SD(R) = -1.66$, Sig. =
$2 \times P(Z \leq -1.66) = 0.098 > \alpha = 0.05$. Likewise, the graph of the sequence of the residuals (in the order of collection of the score vectors for the predictor variables) shows a random order (Figure 2). Consequently, it is appropriate to assume that the residuals are independent and have homogeneity of variance. If these assumptions do not hold, you can change the calculation option in SPSS (IBM, 2021).
The correlation matrix between the estimated parameters, considering them as random variables, was calculated using the nonparametric method proposed by Bofingeb (1975). This matrix allows us to see that the regression coefficient of the attitude toward gay people (scale parameter) has a trivial correlation with the position parameters of religiosity (from 0.07 to 0.14) and a medium correlation with the intercept of the model (0.56). The correlations of the position parameters of

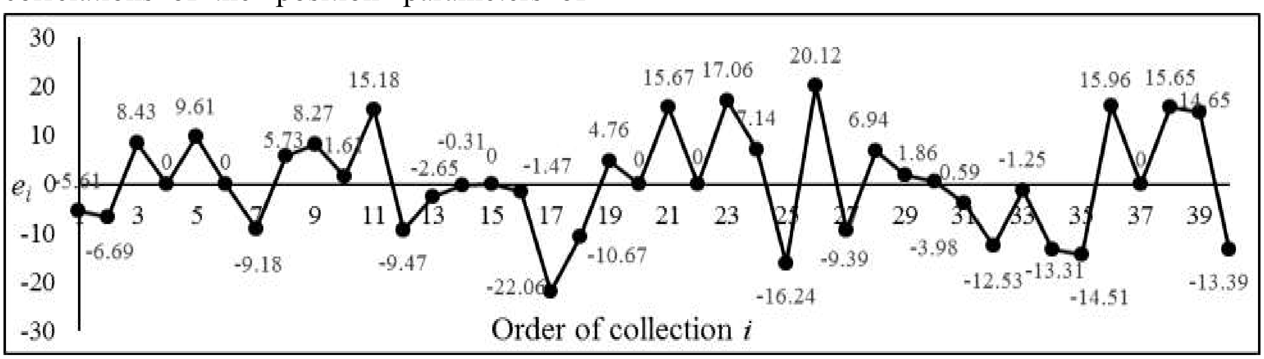

Figure 2. Diagram of the sequence of residuals *ei* in the order of collection of the score vectors for the predictor variables *i* (from 1 to 40).
Source: Authors

Table 3. Correlations of parameter estimates (quantile of order 0.5).

Parameter	b_0	b_1	$b2\|X2=1$	$b2\|X2=2$	$b2\|X2=3$	$b2\|X2=4$	$b2\|X2=4$
b_0	1	-.562	-.837	-.818	-.807	-.754	0
b_1	-.562	1	.140	.110	.112	.071	0
$b2\|X2=1$	-.837	.140	1	.854	.840	.802	0
$b2\|X2=2$	-.818	.110	.854	1	.834	.798	0
$b2\|X2=3$	-.807	.112	.840	.834	1	.784	0
$b2\|X2=4$	-.754	.071	.802	.798	.784	1	0
$b2\|X2=4$	0	0	0	0	0	0	0

Dependent variable: Y = percentile scores on attitude toward people living with HIV/AIDS. Quantile regression model of order 0.5 estimated by Barrodale-Roberts simplex method (1974), assuming that errors are independently distributed and have homogeneity of variance: b_0 (intercept) + $b_1 \times x_{i1}$ (product of the regression weight and the percentile score on the scale of attitude toward gay people) + $b_{2|X2}$ (position parameter for religiosity) = 42.43 + 0.51× x_{i1} + −36.35 (if x_2 = 1) or −22.16 (if x_2 = 2) or −7 (if x_2 = 3) or −3.25 (if x_2 = 4) or 0 (if x_2 = 5). The ordered category 5 (very high religiosity) was the reference category for the ordinal variable of religiosity. Correlations were estimated by the non-parametric method proposed by Bofingeb (1975). Source: Authors

Religiosity are very high with each other (from 0.78 to 0.85) and with the intercept of the model (from -0.84 to 0.75). The correlations between the parameters of the predictor variables (scale and position) are positive or direct, but the correlations of the predictor variables with the model intercept are negative or inverse (Table 3). This indicates low collinearity between both predictors and linearity between the ordered categories of X_2 (religiosity).

The model showed very good goodness of fit when estimated through the Pseudo R-squared coefficient proposed by Koenker and Machado (1999), which is a local measure of fit that measures the goodness of fit by comparing the sum of the weighted deviations of the final model with the sum of the intercept only model. It only takes into account the fit of the predictions to the observed data, but does not consider the number of variables in the final model or pay attention to parsimony.

$$R^2_{Q_Y(0.5)|X_1X_2} = 1 - \frac{\hat{V}_1(\tau)}{\hat{V}_0(\tau)} = 1 - \frac{\sum_{i=1}^{n} \rho_{\tau=0.5}\left(y_i - b_0 - b_2 x_{i2} - b_{1|x_{i1}}\right)}{\sum_{i=1}^{n} \rho_{\tau=0.5}\left(y_i - b_0\right)}$$

$$= 1 - \frac{\sum_{i=1}^{n} \left|y_i - b_0 - b_2 x_{i2} - b_{1|x_{i1}}\right|}{\sum_{i=1}^{n} \left|Y_i - b_0\right|} = 1 - \frac{321.9216}{687} = 0.5314$$

The mean absolute error (MAE), in this sample composed of 40 participants, was 8.05.

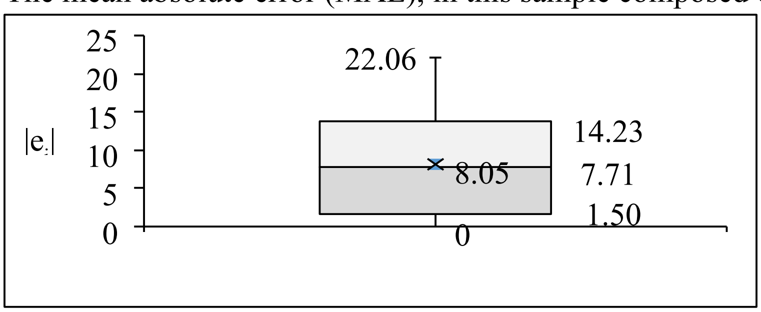

Figure 3. Box-and-whisker plot of the absolute residuals. Source: Authors

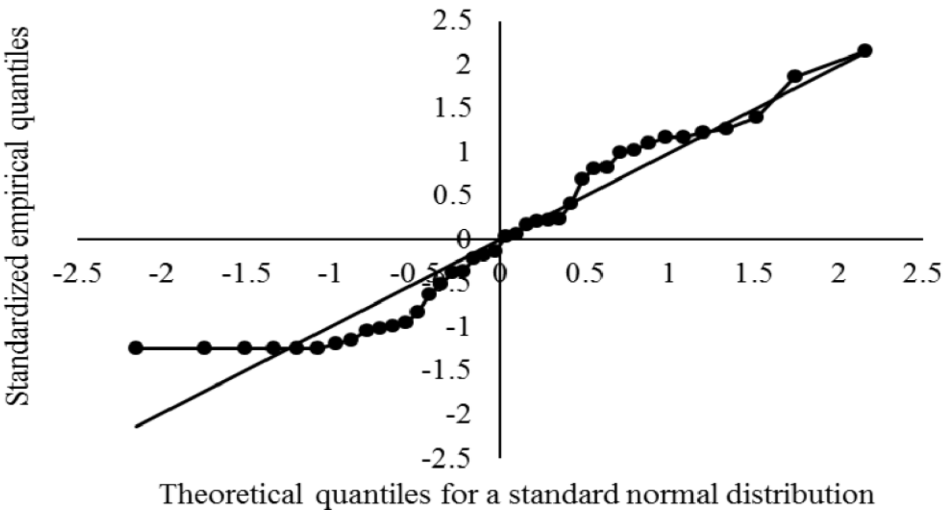

Figure 4. Normal quantile-quantile plot of the absolute residuals.
Source: Authors

$$MAE = \frac{\sum_{i=1}^{n}|E_i|}{n} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} = \frac{321.922}{40} = 8.048$$

Although the deviations from the mean converge toward a Laplace distribution, the average of the absolute deviations does not show such distributional convergence. The one-sample Anderson-Darling test can be used to reject the null hypothesis that posits that the absolute errors follow a Laplace distribution. So, at a significance level of 0.05, the null hypothesis of goodness-of-fit is rejected ($AD = 1.139$, $p = 0.025 < \alpha = 0.05$).
The distribution of absolute residuals is also far from a
normal distribution by the Anderson-Darling test
(D'Agostino, 1986): $A = 0.891$, $AD = A \times (1 + (0.75/n) + (2.25/n^2)) = 0.891 \times (1 + (0.75/40) + (2.25/402)) = 0.909$
>
$0.05AD_{40} = 0.736$, $p = 0.021 < \alpha = 0.05$) and by ShapiroWilk W test (Royston, 1992): $W = 0.926$, $p = 0.012$). As shown in Figures 3 and 4, the distribution is truncated at its left tail and has a platykurtic profile (Anscombe and Glynn, 1983) test: $b_2 = 1.933 < 3$, $Z = -2.235$, two-tailed p
$= 0.025 < \alpha = 0.05$). In the box-and-whisker plot (Figure 3), the lower whisker is cut off at zero and the boxes are wide relative to the whiskers. On the normal quantile- quantile plot centered at 0 (standardized observed and theoretical quantiles), the dotted line flattens out at the lower end in the third quadrant, reflecting a truncated sample (Figure 4). Furthermore, the curve is convex below 0 and tends to be concave above 0 (up to 1.5), which is characteristic of a leptokurtic profile (D'Agostino et al., 1990).
Since the distribution is unknown, the confidence interval for the mean absolute error can be estimated by the bias-corrected and accelerated bootstrap interval method (Efron, 1987): $PSE = 8.048$, $bias = -0.0157$, $SE = 1.037$, BCa 95% CI [5.993, 10.099]; number of bootstrap samples: 1000). The 95% confidence interval shows that it is a value significantly different from 0.

Conclusion

Binary logistic regression is a technique developed for dichotomized qualitative variables and not for dichotomized quantitative variables (Agresti, 2019). Instead, there is quantile regression, which is a good regression technique for predicting a quantitative variable without distributional requirements of normality or homogeneity of variance in the residuals (Koenker et al., 2017). This technique accepts qualitative, ordinal, and quantitative predictor variables and can even be adapted to correlated residuals (Alhamzawi and Ali, 2018, 2020; IBM, 2021). Moreover, it allows to perform analyses for different quantile orders of the predicted variable; usually, the order is 0.5 (median), but the model can also be estimated for extreme order percentiles, such as 0.25 (lower quartile), 0.75 (upper quartile), 0.10 (first decile) or 0.90 (lower decile); this fact is especially interesting when dealing with heteroscedastic data.

The quantile model for the median value would be the counterpart or equivalent to the multiple ordinary least squares linear regression model for the mean value, and the quantile models for the extreme percentiles would be the counterparts or equivalents to binary logistic regression models of the continuous variable dichotomized by the corresponding percentile; nevertheless, quantile regression would be more appropriate to the assumptions made and the measurement scales of the variables included in the model (Waldmann, 2018).

Quantile regression is a little-known technique in its theoretical foundations as well as in its aspects of calculation and interpretation in psychological research. However, as can be seen from this article, which uses an example applied to the field of social and health psychology, this technique is clear in its rationale and yields results that are easy to interpret. Therefore, its use is recommended when the data warrant it, which are common situations in research in psychology and related sciences. That is why this regression technique is becoming increasingly used in medical research
(Konstantopoulos et al., 2019; Staffa et al., 2019) and is available in statistical packages, such as SPSS (IBM, 2021) and R (Koenker, 2016).

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

REFERENCES

Agresti A (2019). An introduction to categorical data analysis (3ª Ed.). Hoboken, NJ: John Wiley & Sons.

Alhamzawi R, Ali HTM (2018). Bayesian quantile regression for ordinal longitudinal data. Journal of Applied Statistics 45(5):815-828.

Alhamzawi R, Ali HTM (2020). Brq: An R package for Bayesian quantile regression. METRON 78(3):313-328.

Anscombe FJ, Glynn WJ (1983). Distribution of the kurtosis statistics b_2 for normal samples. Biometrika 70(1):227-234.

Barrodale I, Roberts FD (1974). Solution of an overdetermined system of equations in the l_1 norm. Communications of the ACM 17(6):319320.

Bofingeb E (1975). Estimation of a density function using order statistics. Australian Journal of Statistics 17(1):1-7.

Bošković RG (1757). Elementorium universae matheseos. Reprinted in London, UK: Forgotten Books, 2019.

Cade BS, Noon BR (2003). A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment 1(8):412420.

Cohen J (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum and associates.

D’Agostino RB (1986). Tests for the normal distribution. In: R. B. D'Agostino, & M. A. Stephens (Eds.), Goodness-of-fit techniques (pp. 367-419). New York: Marcel Dekker.

D’Agostino RB, Berlanger A, D’Agostino RB Jr. (1990). A suggestion for using powerful and informative tests of normality. The American

Statistician 44(4):316−321.

Edgeworth FY (1887). On observations relating to several quantities. Hermathena 6(13):279-285.

Edgeworth FY (1888). On a new method of reducing observations relating to several quantities. Philosophical Magazine 25(154):184191.

Efron B (1987). Better bootstrap confidence intervals. Journal of the American Statistical Association 82(397):171-185.

Frisch MR (1956). La résolution des problèmes de programme linéaire par la méthode du potentiel logarithmique. Cahiers du Seminaire D'Econometrie 7-23.

Furno M, Vistocco D (2018). *Quantile regression: estimation and simulation*. Hoboken, NJ: John Wiley & Sons.

Hajian-Tilaki K. (2018). The choice of methods in determining the optimal cut-off value for quantitative diagnostic test evaluation. Statistical Methods in Medical Research 27(8):2374−2383.

Hall P, Sheather SJ (1988). On the distribution of a studentized quantile. Journal of the Royal Statistical Society, Series B (Methodological) 50(3):381-391.

Koenker R (1998). Galton, Edgeworth, Frisch, and prospects for quantile regression in econometrics. Champaign, IL: Department of

Economics, University of Illinois. http://www.econ.uiuc.edu/~roger/research/galton/Galton.pdf

Koenker R (2005). Quantile regression. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511754098

Koenker R (2016). quantreg: Quantile regression. R package version 5.21. https://CRAN.R-project.org/package=quantreg

Koenker R (2017). Quantile regression: 40 years on. Annual Review of Economics 9(1):155-176.

Koenker R, Bassett G (1978). Regression quantiles. Econometrica: journal of the Econometric Society 33-50.

Koenker R, Chernozhukov V, He H, Peng L (2017). Handbook of quantile regression. Boca Raton, FL: Chapman & Hall/CRC. https://doi.org/10.1201/9781315120256

Koenker R, d'Orey V (1987). Algorithm AS 229: Computing regression quantiles. Journal of the Royal Statistical Society, Series C (Applied Statistics) 36(3):383-393.

Koenker R, Machado JAF (1999). Goodness of fit and related inference processes for quantile regression. Journal of the American Statistical Association 94(448):1296-1310.

Konstantopoulos S, Li W, Miller S, van der Ploe A (2019). Using quantile regression to estimate intervention effects beyond the mean. Educational and Psychological Measurement 79(5):883-910.

International Business Machines Corporation (IBM) (2021). Quantile regression. In SPSS statistics. https://www.ibm.com/docs/en/spss- statistics/SaaS?topic=regression-quantile

Laplace PS (1789). On some points of the system of the world. Memoirs of the Royal Academy of Sciences of Paris.

Lustig IJ, Marsten RE, Shanno DF (1994). Interior point methods for linear programming: Computational state of the art. ORSA Journal on Computing 6(1):1-4.

Moral J (2010). Religión, significados y actitudes hacia la sexualidad: un enfoque psicosocial. Revista Colombiana de Psicología 19(1):45-59.

Moral J, Ortega ME (2010). Representación social de la sexualidad y actitudes en estudiantes universitarios mexicanos. Revista de Psicología Social 24(1):65-79.

Moral J, Martínez-Sulvarán JO (2012) Validation of the 10-items Homosexuality Attitude Scale (EAH-10). International Journal of Social Psychology 27(2):183-197.

Moral J, Valle A (2020). Propiedades psicométricas de la Escala de Actitud hacia Personas que Viven con VIH/SIDA en estudiantes de medicina mexicanos. Perspectivas Sociales 22(1):45-70.

Moral J, Valle A (2021). Factorial invariance across sexes of the Scale of Attitude toward People Living with HIV/AIDS. Journal of Behavior, Health and Social Issues 13(3):1-14.

Ringquist EJ (2013). Meta-analysis for public management and policy.

San Francisco, CA: Jossey-Bass.

Royston JP (1992). Approximating the Shapiro-Wilk W-test for nonnormality. Statistics and Computing 2(3):117-119.

Staffa SJ, Kohane DS, Zurakowski D (2019). Quantile regression and its applications: a primer for anesthesiologists. Anesthesia and Analgesia 128(4):820-830.

Stolper O, Walter A (2019). Birds of a feather: the impact of homophily on the propensity to follow financial advice. The Review of Financial Studies 32(2):524-563.

Wald A, Wolfowitz J (1943). Exact test for randomness in the nonparametric case based on serial correlation. Annals of Mathematical Statistics 14(4):378-388.

Waldmann E (2018). Quantile regression: A short story on how and why. Statistical Modelling 18(3-4):203-218.