



University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): JL Hutton and E Stanghellini

Article Title: Modelling health scores with the skew-normal distribution

Year of publication: 2009

Link to published article:

<http://www2.warwick.ac.uk/fac/sci/statistics/crism/research/2009/paper09-39>

Publisher statement: None

Modelling health scores with the skew-normal distribution

J. L. Hutton,

Department of Statistics, University of Warwick,

E. Stanghellini,

Dipartimento di Economia Finanza e Statistica, Università di Perugia

September 22, 2009

Abstract

Health care interventions which use quality of life or health scores often provide data which are skewed and bounded. The scores are typically formed by adding up responses to a number of questions. Different questions might have different weights, but the scores will be bounded, and are often scaled to the range 0 to 100. If improvement in health over time is measured, scores will tend to cluster near the 'healthy' or 'good' boundary as time progresses, leading to a skew distribution. Further, some patients will drop out as time progresses, so the scores reflect a selected population.

We fit models based on the skew-normal distribution to data from a randomised controlled trial of treatments for sprained ankles, in which scores were recorded at baseline and 1, 3 and 9 months. We consider the extent to which skewness in the data can be explained by the clustering at the boundary via a comparison between a censored normal and a censored skew-normal model.

As this analysis is based on the complete data only, a formula for the distortion of the treatment effects due to informative drop-out is given. This allows us to assess under which conditions the conclusions drawn on the complete data may be either reinforced or reversed, when the informative drop-out process is taken into account.

1 INTRODUCTION

Some outcomes of interest in medical research can be measured directly: death, and blood pressure are simple examples. Other outcomes, such as mental or physical health, have to be assessed indirectly. The measures used for this range from binary, when questions such as 'are you in good health?' are answered 'yes' or 'no', to (almost) continuous, when answers to a series of questions about quality of life are summed to give a weighted average. The latter measures have finite range. Multiple valued and continuous scores might be scaled so that the 'population norm' is at the centre of the range, but many scores assessing mental health or physical activity have good health at one end of the range.

Symmetrically distributed random variables are less suited as models for such scores. Although one might consider transforming the scale so that standard models with Gaussian errors can be used, this is not always possible. Non-symmetric distributions, such as the skew-normal or gamma are then useful. An advantage of the skew-normal model for longitudinal data is

the elegant form of the multivariate distribution and the fact that it contains the normal distribution as a particular case [1]. Furthermore, skewness might be explained by the effect of informative drop-out [2]. The approach proposed allows one (a) to check whether the asymmetry in the data may be removed after accounting for the censoring mechanism and (b) to assess the distortion in estimated treatment effect induced by a possible informative drop-out mechanism.

2 MOTIVATING EXAMPLE

A randomised controlled trial of four treatments (1. tubigrip, i.e. elastic bandage, 2. below-knee plaster-cast, 3. splints or 4. boots) for acute, severe ankle sprain had quality of ankle function as the primary outcome [3]. Demographic data including age, sex, and employment were collected on the 564 patients randomised. The protocol specified an intention to treat analysis, which we adopt. Ankle function was measured by the Foot and Ankle Score (FAOS), a questionnaire on pain, symptoms, activities of daily living, ability to participate in sports and quality of life. Patients were asked to complete the questionnaire when they were randomised (baseline) and at 1, 3 and 9 months after injury. The outcomes of FAOS are weighted averages which are scaled to the range 0 to 100; extreme pain or limitation scores 0, and no symptoms score 100. Two of the five domains of FAOS, sport and activities of daily living had 39% responses missing, while the remaining three were 75% complete, so this study reports analysis of the score based on pain, symptoms, and quality of life.

The primary question was how plaster-cast, splints and boots compared with tubigrip, and size of any difference in FAOS. Physical healing is known to depend on age, so it is sensible to include age in the models. As expected, the mean and skewness of the score distribution, marginal w.r. to treatment and age, increased with time, see Figure 1, and was concentrated at 100 in the later two times. The standard deviation is roughly constant, see Table 1. Most improvement in FAOS occurred in the first month, as illustrated in Figure 2.

The impact of missingness on the treatment estimates should also be considered. Postal questionnaires were received from 83%, 82% and 76% of participants at months 1, 3, and 9 respectively. The drop-out was not entirely monotone. The means for those who have missing data are generally higher than the complete cases, indicating a possible tendency for patients who have recovered not to return their questionnaires. Age also seems to play a role, as on average, the age of complete cases is higher than the others. Further, a qualitative study in which non-respondents were interviewed reported that nearly half the non-respondents thought they had made a full recovery by 3 months, the second follow-up time [4].

3 SKEW-NORMAL MODELS

An obvious approach to a score on the range $[0,100]$ is to transform using a scaled logit to the range $(-\infty, \infty)$. However, QQ plots from fitting models on the transformed score give a poor fit, as the transformation does not reduce the concentration at the maximum, and there is the disadvantage of not working on the original scale familiar to the health care professionals. The best Box-Cox transformation for the scores at 9 months achieves an almost uniform distribution. Other naturally skew distributions, such as the Gamma and Weibull, and the Beta distribution, which has a finite range, do not fit well, and are not as simply generalized to the multivariate case.

Let $\mathbf{Y} = (Y_0, Y_1, Y_2, Y_3)^T$ be the vector of the scores at the four occasions. We estimated:

$$\mathbf{Y} = \mathbf{XB} + \varepsilon \quad (1)$$

with $\varepsilon \sim SN_4(0, \Omega, \alpha)$ and \mathbf{X} the design matrix with covariates: 'Age' and the indicators of the treatment group. Figure 3 shows the matrix of the scatter plot of the residuals of the model, with the contour level of the corresponding bivariate distribution obtained from the fitted joint distribution after marginalization. For occasions 1 and 2 the model does not exhibit substantial departure from the normal assumptions, while for the two subsequent occasions the skewness increases substantially. The covariance matrix of the residuals is left unconstrained, and the estimated correlations *increase* as the intervals between measurements *increase*; this is an effect of the boundary.

These considerations are also supported by the comparison of the estimates with those obtained with a multivariate regression models. For the first two occasions, the OLS estimates are not different from the ones obtained with model (1), while they diverge substantially for the following occasions. However, it is difficult to disentangle the true asymmetry in the distribution from the effect due to the clustering of the residuals towards the value 0, which is induced by the boundary at score 100.

The conclusions are also confirmed by a series of univariate regression models of score at times 0, 1, and 2 against age, indicators for treatment groups and score at the preceding time with skew-normal residuals. Results show a reasonably normally distribution of the residuals of these models but not for score at months 9, which is an important measure of long-term healing. The histogram of the residuals of such a model, see Figure 4, exhibits skewness as well as high frequency of values around 0. We therefore focus on the score at month 9, and consider a regression with skew-normal errors, normal errors with censoring and skew-normal errors with censoring.

4 CENSORED MODELS

We focus here on the final outcome, i.e. the score measured after 9 months. From the previous analysis it seems natural to build a univariate regression model for the final score that takes into account the censoring mechanism. It would also be interesting to check whether the skewness in the data is induced by the threshold at score 100 or is genuinely in the phenomenon under study. Although there are many choices of skew distributions, we believe that the skew-normal is again a suitable tool to perform this task, as it contains the normal one as a particular case. In this section we perform a complete data analysis. We defer the discussion on the distortion induced by the possibly informative drop-out mechanism to the next section.

Let Y be the observed score at 9 months and \mathbf{x} be the vector of covariates, which include 'Age' and three dummy variables D_2, D_3, D_4 for each of the corresponding treatment group. We therefore specify the following model:

$$Y^* = \beta^T \mathbf{x} + \eta \quad (2)$$

with η distributed as $SN(0, \sigma, \alpha)$ and β a vector of unknown regression coefficients. The observed Y is defined as:

$$\begin{cases} Y = Y^* & \text{if } Y^* < 100 \\ Y = 100 & \text{otherwise.} \end{cases} \quad (3)$$

The likelihood for a generic i -th patient then becomes:

$$\left[\frac{2}{\sigma} \phi \left(\frac{\phi - \mu_i}{\sigma} \right) \Phi(\alpha \sigma^{-1}(y_i - \mu_i)) \right]^{1-a_i} \times [2\Phi_2(0, \sigma^{-1}(100 - \mu_i); \rho)]^{a_i}$$

in which $a_i = I[y_i^* \geq 100]$, $\mu_i = \beta^T \mathbf{x}_i$, $\phi(\cdot)$ and $\Phi(\cdot)$ denotes the standard normal density and distribution function, while $\Phi_2(\cdot)$ denote the distribution function of a standard bivariate normal with ρ as a correlation coefficient, $\rho = \alpha/\sqrt{1 + \alpha^2}$. As working parameters, instead of using the direct parametrization (DP) above, we here use the so-called centred parametrisation (CP) see [1], since the shape of the log-likelihood of the non-censored model improves greatly by such a reparametrisation of the model. The optimization is performed with numerical methods.

In Table 2, panels (a) and (b), the estimates of the uncensored and censored skew-normal models are presented. Note that in the CP, while the regression coefficients retain their interpretation, the skewness parameter is the usual univariate index of skewness γ_1 [5, p. 8]. The estimated skewness under the censored model is about half that of the uncensored model, and the standard deviation is slightly greater, as is expected. The estimated effect of the boot, adjusted for age and previous scores becomes significant, a reduction in score of 4.4. The score at the previous occasion is strongly predictive of the final score. Although the estimated coefficients for the first two scores are marginally significant at 5%, these scores cannot both be omitted. Including an earlier score allows the final score to depend on the rate of improvement from an earlier score, in addition to the level of the previous score.

The analysis of the residuals, evaluated by truncating at 100 the predicted values above 100, shows some evidence of asymmetry in the data. In Figure 5 the Healy's plot is presented when either the normal distribution (left panel) or the skew-normal distribution (right panel) is fitted to the residuals of the skew-normal censored model, showing a substantial improvement of the second model. However, these naive residuals are not independent, and several alternatives have been studied, including Cox-Snell, martingale, deviance and Schoenfeld residuals. In this applied context, we think this definition is appropriate for assessing whether the fitted model is adequate.

A natural comparison is with the censored normal model. In Table 2 panel (c) the details are presented. The censored normal model has a similar dispersion to the censored skew-normal and a smaller intercept to accommodate the symmetry. There are no major changes in the regression coefficients and in their significance levels. Twice the difference in the log-likelihood is 2.12, which suggests the assumption that the underlying model has normal errors is adequate. However, the ratio of the skewness parameter, 0.271, to its Wald standard error, 0.118, is 2.3. The likely explanation of the apparent contradiction is that even with the CP, the profile likelihood for skewness is not near a quadratic. The estimated number of scores above 100 under the two models is 43 for the censored normal and 28 for the censored skew-normal. Both models underestimate the true number, that is 50, but it is yet another indication that the censored model is a better model for the data at hand. The naive residuals, observed minus fitted, from the censored normal distribution have a rather symmetric distribution.

5 ASSESSING DISTORTION INDUCED BY THE DROP-OUT MECHANISM

Let Z be the binary random variable that takes value 1 if an observation does not drop-out. A standard model for the drop-out mechanism is the following, [6, 7]:

$$\begin{aligned} Z^* &= \delta^T \mathbf{x} + \eta_1 \\ Y &= \beta^T \mathbf{x} + \eta_2, \end{aligned} \quad (4)$$

with \mathbf{x} and $Z = I[Z^* \geq 0]$ always observed and Y observed only if $Z^* \geq 0$. We assume $Cov(\eta_1, \eta_2) = \mathbf{V}$, with $\mathbf{V} = \{v_{ij}\}$ an unrestricted covariance matrix.

The interest is in estimating the regression coefficient β of Y on \mathbf{x} in the overall population. Let $\gamma = v_{12}v_{11}^{-1}$. Given that Y is observed only on those patients who did not drop out, the regression coefficient β evaluated in this selected population is distorted. The amount of the distortion depends on the correlation between the residuals η_1 and η_2 . Let $v_{11} = \sigma_1^2$. Under the assumption that η_1 is Gaussian, see [6, 7]:

$$E[Y \mid \mathbf{x}, Z^* \geq 0] = \beta \mathbf{x} + \gamma \sigma_1 \lambda \left(\frac{\delta \mathbf{x}}{\sigma_1} \right), \quad (5)$$

where $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$ is the well-known inverse Mill's ratio. This equation gives rise to the well-known two-stage estimating procedure, see [6]. Procedures to estimate the parameters based on maximum likelihood are also available, see [8] p. 566. Due to colinearity in the explanatory variables, the estimates are reliable mainly when some zero restrictions are imposed on the coefficients β and δ . However, several studies have shown that results may depend heavily on which components are restricted to be zero, a choice that is usually rather arbitrary; see [7] for the details.

A possible alternative that we pursue here is to give bounds on the distortion induced on the estimates of β , when these are estimated from the complete cases only. Let $\tilde{\beta}(\mathbf{x})$ and $\tilde{\delta}(\mathbf{x})$ be the coefficients of the linear regression function of Y and Z^* , respectively, in the selected population, i.e. in the population with $Z^* > 0$. From the derivations in [9] we know that:

$$\beta - \tilde{\beta}(\mathbf{x}) = \gamma(\delta - \tilde{\delta}(\mathbf{x})). \quad (6)$$

Note that, as expected, when $\gamma = v_{12} = 0$, then $\tilde{\beta}(\mathbf{x}) = \beta$. Since $|\tilde{\delta}(\mathbf{x})| \leq |\delta|$, see [11], we notice that the distortion is bounded in modulo by $\gamma\delta$, so there is no distortion in the j -th element of β , i.e. β_j , if the corresponding element of δ is zero. Note that the derivations in [9] are based on the assumption that Y and \mathbf{X} are jointly Gaussian, but may be easily modified for η_1 being Gaussian and \mathbf{x} given. A sketch of the proof is in the Appendix.

As there are repeated measurements, we let $\mathbf{W}_t = (Y_t, Z_t^*)^T$ and we assume that $\mathbf{W}_t \sim N(\mu_t, \mathbf{V}_t)$, with $\mu_t = (\beta_t, \delta_t)^T \mathbf{x}_t$, where \mathbf{x}_t is the vector of all the information available at time t ; $\mathbf{V}_t = \{v_{ijt}\}$. Setting $v_{12t} = 0$, for some t , corresponds to the assumption that, for occasion t , the drop-out mechanism is random in the terminology of Diggle and Kenward [10]. We here allow \mathbf{V}_t to be an unrestricted covariance matrix.

For each occasion, we assess which elements of δ are significantly different from zero. In doing so, we force the the pattern of missingness to be monotone. This justifies the choice of covariates \mathbf{x}_t as $\mathbf{x}_t^T = [\mathbf{x}_0, y_{t-1}]$, with \mathbf{x}_0 the relevant covariates at $t = 0$, first 'Age', D_2, D_3, D_4 , and all previous scores, denoted by Y_{t-1} . Table 3 reports the estimated coefficients in the Probit models. The only covariate which has a significant effect on the probability to drop-out is 'Age',

for occasions after 1 and 9 months. This implies that (a) the treatment contrasts are robust and (b) the effect of age on the first and third occasion, if estimated from the complete cases only, may be distorted.

If we now focus on the probability of drop-out at 9 months, the estimated coefficient of 'Age' is -0.021 (s.e. 0.009). By taking into account that the σ_1 is not identified in the Probit model, we see that this corresponds to $\delta_{\text{Age}}/\sigma_1$. This effect is of the same magnitude as the estimated age effect in the censored normal model, $\beta_{\text{Age}} = -0.014$. Ignoring for now the random variation induced by the sampling, the maximum possible distortion on β_{Age} is $-0.021\gamma\sigma_1 = -0.021v_{12}$. It then follows that, if $v_{12} < 0$, the negative age effect previously found is larger than the true one, i.e. if more younger people had responded, the estimated negative effect of age would decrease in modulo. If instead $v_{12} > 0$, then the true effect of age is larger in modulo than the effect previously found.

6 DISCUSSION

The skew-normal distribution proved to be a useful model for scores taken during the process of healing. When sufficient time has elapsed for many patients to have achieved full recovery, models which use censoring to allow for the finite range are generally more appropriate. These models explain a substantial part of the skewness of the distributions, and provide a good estimate of the proportion of people achieving full health (11.5% as estimate of 13.3% in this example). Probit regression of the missingness process indicates that the complete case estimates are robust for the treatment contrast. Future work will jointly model the missingness and score processes with censored skew-normal models.

APPENDIX

For simplicity let $\mathbf{x} = x$ be a single r.v. Let

$$\tilde{\beta}(x) = \frac{dE[Y \mid x, Z^* \geq 0]}{dx}$$

and

$$\tilde{\delta}(x) = \frac{dE[Z^* \mid \mathbf{x}, Z^* \geq 0]}{dx}$$

be the derivatives w.r. x of the nonlinear regression function expressing $E[Y \mid x, Z^* \geq 0]$ and $E[Z^* \mid x, Z^* \geq 0]$. From eq. (5):

$$\tilde{\beta}(x) = \frac{dE[Y \mid x, Z^* \geq 0]}{dx} = \beta + \gamma\sigma_1 \frac{d\lambda(u)}{du} \frac{du}{dx}$$

with $u = \frac{\delta x}{\sigma_1}$. Let $U \sim N(0, 1)$. We know that $\text{Var}(U \mid U \geq -c) = 1 - \lambda(c)[c + \lambda(c)]$, which shows that $\lambda(c)[c + \lambda(c)]$ is bounded between 0 and 1. Therefore, as $\frac{d\lambda(u)}{du} = -\lambda(u)[u + \lambda(u)]$, then

$$\begin{aligned} \frac{dE[Y \mid x, Z^* \geq 0]}{dx} &= \beta + \gamma\sigma_1 \{-\lambda(u)[u + \lambda(u)]\} \frac{\delta}{\sigma_1} \\ &= \beta + \gamma(\tilde{\delta}(x) - \delta). \end{aligned}$$

The last result follows from:

$$\tilde{\delta}(x) = \delta - \delta\{\lambda(u)[u + \lambda(u)]\}$$

see e.g. [8], p. 522. It then follows that $\tilde{\delta}(x)$ is always smaller in modulo than δ for all x , so the distortion is null if $\delta = 0$. Generalization to \mathbf{x} being a vector of r.v.'s follows analogously. See also [9] Prop. 5(ii).

Acknowledgement Thanks to Prof Sallie Lamb, Prof Matthew Cooke and the CAST team for permission to use these data. The CAST trial was funded by a grant from the UK Department of Health through its Health Technology Assessment Programme, project no. 01/14/10. Prof Stanghellini's visits to University of Warwick were supported by CRiSM, which is part-funded by EPSRC, and by the MIUR research grant PRIN 2007XECZ7L_003 .

References

- [1] Azzalini A, Capitanio A. Statistical application of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society, Series B*, 1999; **61**, 3, 579-602.
- [2] Skinner, CJ. Discussion on "Informative drop-out in longitudinal data analysis" by PJ Diggle and MG Kenward. *Journal of the Royal Statistical Society Series C*, 1994; **43**, 76-77.
- [3] Lamb SE, Marsh JL, Hutton JL, Nakash RA, Cooke MW. Mechanical supports for acute, severe ankle sprains: a pragmatic, multi-centre, randomised controlled trial. *The Lancet*, 2009; **373**, 575-581.
- [4] Nakash RA, Hutton JL, Lamb SE, Gates S, Fisher J. Response and non-response to postal questionnaire follow-up in a clinical trial - a qualitative study of the patient's perspective. *BMC Med Res Meth.* 2008; **14**, 226-235.
- [5] Wetherill GB. *Intermediate statistical methods*. London: Chapman and Hall, 1981.
- [6] Heckman JJ. Sample bias as a specification error. *Econometrica*, 1979; **47**, 153-61.
- [7] Copas, JB, Li HG. Inference for Non-random Samples (with discussion). *Journal of the Royal Statistical Society, Series B*, 1997; **59**, 55-95.
- [8] Wooldridge JM. *Econometrics Analysis of Cross Section and Panel data*. Cambridge, MA: The MIT Press, 2003.
- [9] Marchetti GM, Stanghellini E. A note on distortions induced by truncation, with application to linear regression systems. *Statistics & Probability Letters*, 2008; **78**, 824-829.
- [10] Diggle, PJ, Kenward MG. Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society, Series C*, 1994; **43**, 49-93.
- [11] Goldberger AS. Linear regression after selection. *Journal of Econometrics*, 1981; **15**, 357-66.

Table 1: Summary statistics for FAOS

| Group | | Month of measurement | | | |
|-----------------|---------|----------------------|-------|-------|-------|
| | | 0 | 1 | 3 | 9 |
| 1. Tubigrip | Mean | 37.7 | 58.7 | 68.6 | 78.3 |
| | St.dev. | 13.0 | 19.1 | 20.4 | 20.1 |
| | Skew | -0.15 | -0.11 | -0.36 | -0.88 |
| 2. Plaster-cast | Mean | 37.2 | 64.9 | 75.3 | 82.2 |
| | St.dev. | 13.8 | 18.4 | 17.9 | 17.6 |
| | Skew | 0.32 | -0.07 | -0.49 | -1.34 |
| 3. Splints | Mean | 39.1 | 60.1 | 75.6 | 80.4 |
| | St.dev. | 19.1 | 19.4 | 18.9 | 18.3 |
| | Skew | 1.46 | 0.13 | -0.35 | -1.02 |
| 4. Boot | Mean | 41.1 | 58.8 | 72.6 | 78.9 |
| | St.dev. | 18.9 | 20.4 | 19.4 | 20.7 |
| | Skew | 0.91 | -0.09 | -0.71 | -1.14 |

Table 2: Maximum likelihood estimates: ordinary and censored skew-normal models, and censored normal model

| | (a) | | (b) | | (c) | |
|-----------|---------------|-------|-------------|-------|-----------------|-------|
| | Uncensored SN | | Censored SN | | Censored Normal | |
| | Estimate | S.E. | Estimate | S.E. | Estimate | S.E. |
| Intercept | 35.248 | 4.553 | 31.315 | 4.729 | 22.727 | 4.636 |
| D2 | -0.579 | 1.804 | -2.177 | 2.166 | -2.019 | 2.181 |
| D3 | -0.822 | 1.874 | -1.234 | 2.258 | -0.696 | 2.275 |
| D4 | -2.670 | 1.802 | -4.433 | 2.151 | -4.324 | 2.156 |
| Age | -0.015 | 0.060 | -0.072 | 0.071 | -0.014 | 0.072 |
| y_0 | 0.115 | 0.054 | 0.103 | 0.064 | 0.124 | 0.065 |
| y_1 | 0.079 | 0.049 | 0.085 | 0.060 | 0.119 | 0.061 |
| y_2 | 0.519 | 0.056 | 0.633 | 0.059 | 0.688 | 0.056 |
| s.d. | 13.120 | 0.519 | 14.419 | 0.601 | 14.431 | 0.106 |
| skewness | -0.520 | 0.107 | -0.271 | 0.118 | 0 | - |
| Loglik | -1488.99 | | -1368.387 | | -1369.447 | |

S.E.: estimated standard error

Table 3: Bias in regression using complete case analyses: Probit fit for missing observations at 1, 3 and 9 months

| | Missing at 1 month | | | Missing at 3 months | | | Missing at 9 months | | |
|-------|--------------------|-------|---------|---------------------|-------|---------|---------------------|-------|---------|
| | Estimate | S.E. | p-value | Estimate | S.E. | p-value | Estimate | S.E. | p-value |
| D2 | -0.219 | 0.181 | 0.226 | -0.061 | 0.246 | 0.804 | -0.242 | 0.236 | 0.305 |
| D3 | -0.036 | 0.174 | 0.824 | 0.335 | 0.235 | 0.154 | 0.076 | 0.225 | 0.736 |
| D4 | -0.197 | 0.178 | 0.267 | 0.001 | 0.236 | 0.997 | -0.154 | 0.228 | 0.499 |
| age | -0.013 | 0.006 | 0.032 | -0.005 | 0.008 | 0.550 | -0.021 | 0.009 | 0.016 |
| y_0 | -0.008 | 0.005 | 0.119 | 0.002 | 0.007 | 0.743 | -0.004 | 0.007 | 0.606 |
| y_1 | | | | 0.002 | 0.008 | 0.728 | 0.005 | 0.006 | 0.425 |
| y_2 | | | | | | | -0.004 | 0.006 | 0.500 |

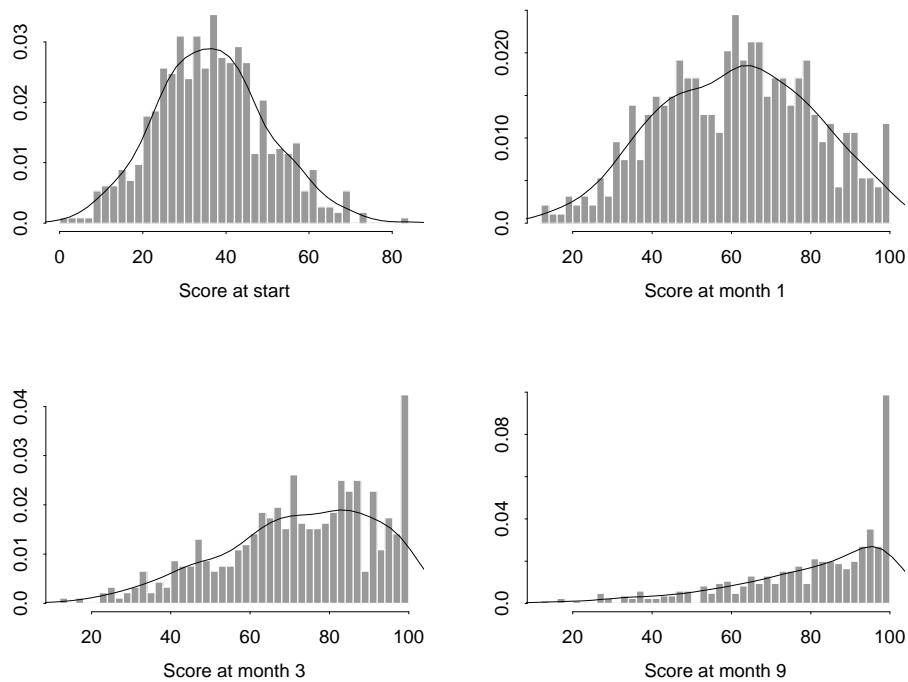


Figure 1: Histograms of FAOS scores

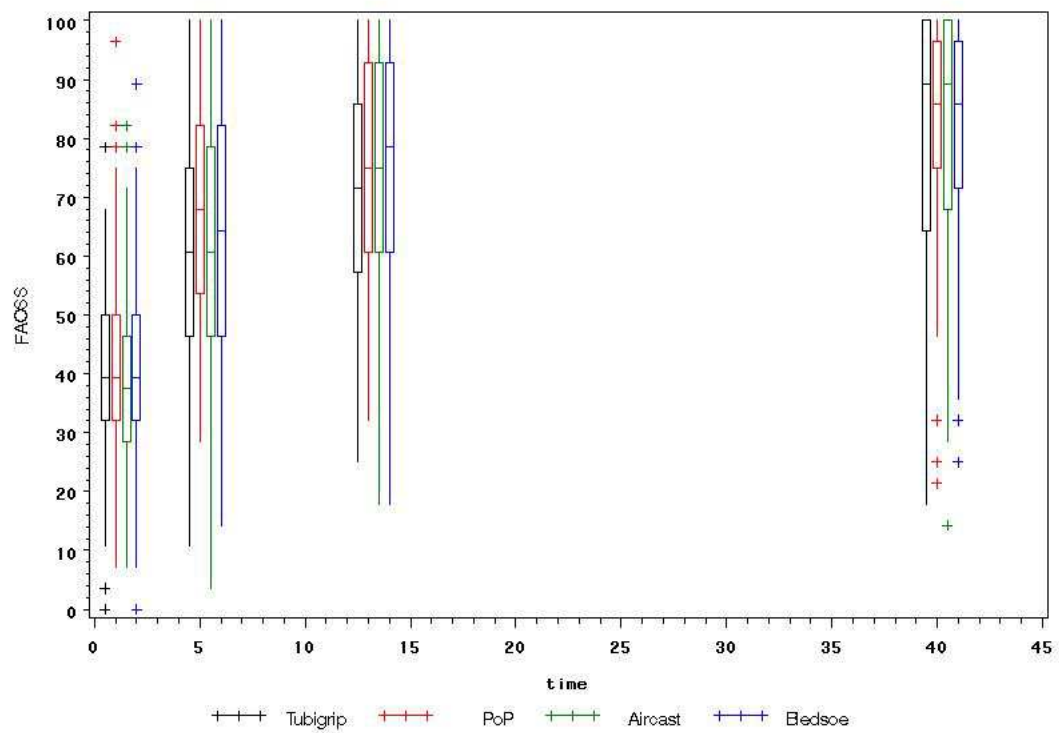
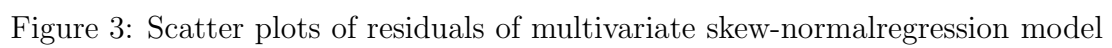


Figure 2: Boxplots of FAOS scores by treatment and time



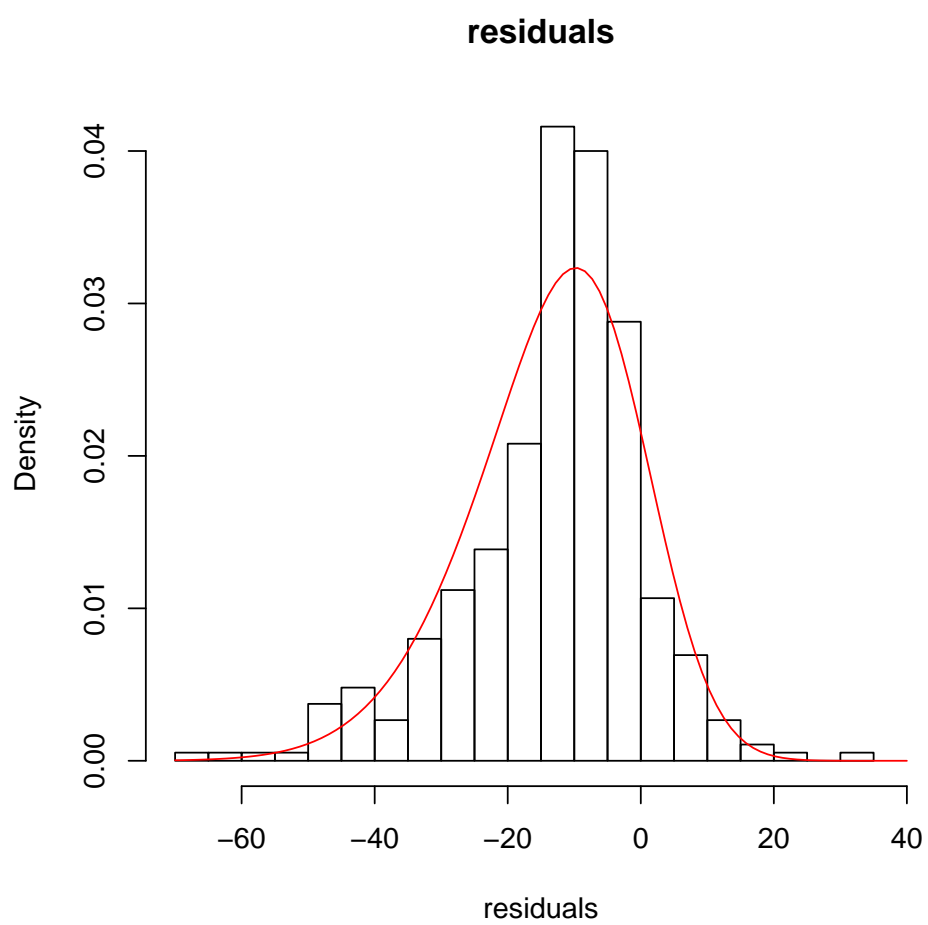


Figure 4: Histogram: residuals from univariate skew-normal regression model for score at 9 months

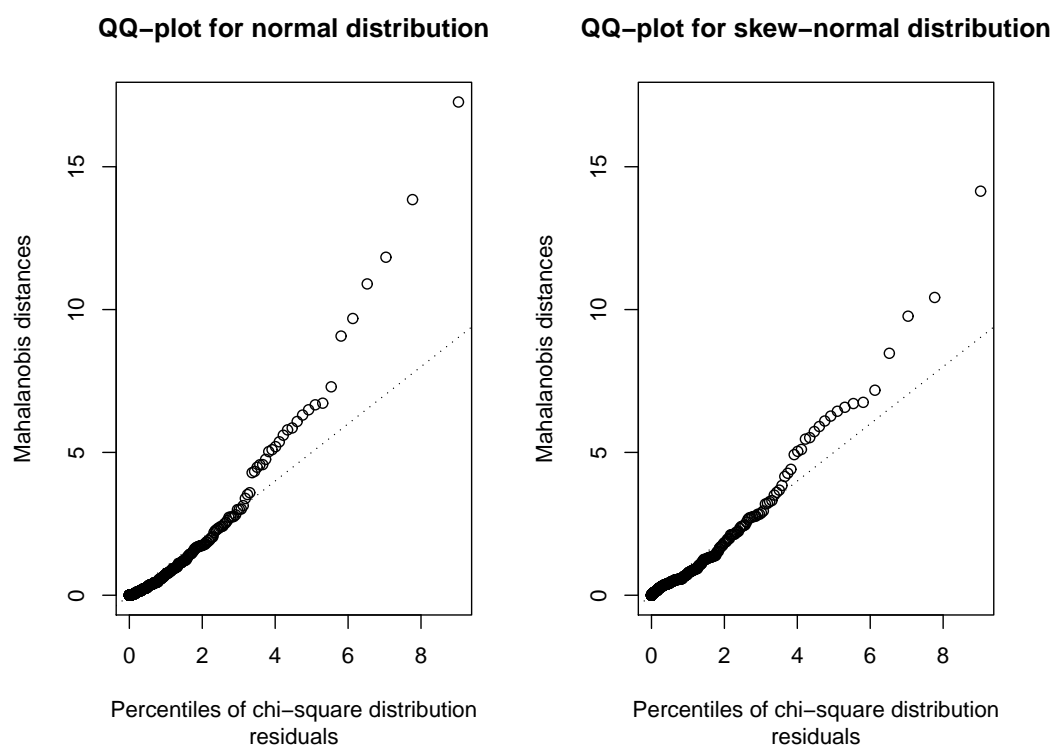


Figure 5: Healy's plot of residuals of the censored skew-normal regression model for score at 9 months