Applications of Normal Quantile Plots

David Rose

June 13, 2011

The TI-84 Normal Quantile Plot

Given a ranked SRS of size n, $x_1 \leq x_2 \leq \cdots \leq x_n$ how does software associate with each x_i a standard normal z_i with equivalent quantile position? The TI-84 formula is as follows: for each $i = 1, \ldots, n$, x_i corresponds to

$$z_i = \mathsf{InvNorm}(\tfrac{2i-1}{2n}).$$

i.e., The ranked sample data is associated positionally with the midpoints of the intervals obtained by partitioning $[0, 1]$ into n equal length subintervals. In particular, x_1 is associated with $\frac{1}{2n}$, x_2 is associated with $\frac{3}{2n}$, etc., and lastly, x_n is associated with $1 - \frac{1}{2n}$.

To effect a normal quantile plot manually, place the sample data x in L2 and do an ascending sort and then place numbers $1, 2, 3, \ldots, n$ in L1. Set $Y1 = \mathsf{InvNorm}(\frac{2X-1}{2n})$ in the function register, and on the main screen of the calculator, perform $Y1(L1) \rightarrow L1$ to place the z_i's in L1. Then, the normal quantile plot for the L2 sample data is simply the scatterplot of L1 horizontally against L2 vertically. Now regression analysis is possible.

An Informal T-Test for Approximate Normality

What is wrong with the χ^2-GOF test for normality? First, the sample size should support at least five intervals with at least an expected frequency of 5 [F](some say 10 [C]) in each interval. i.e., This requires samples sizes from 25 to 50. Often when we need normality checked such as in χ^2-procedures or $F-$procedures, sample sizes are small. And, how can the null hypothesis "H_0 : The population is normal" be supported without considering the power of the test, the probability that a false H_0 will be rejected? Such tests are really tests for non-normality. We offer the following informal linear regression $t-$test for normality. Do a LinReg T-Test (on the TI-84) with the sample data X in L2 and the Z = InvNorm$(\frac{2i-1}{2n})$ in L1 for $i = 1, \ldots, n$. Then rejection of the null hypothesis "$H_0 : \rho = 0$" is affirmation of linear correlation and hence normality. (Formally, this method is considered invalid due to dependence of the ordered x's, though the sample as a whole is random. The correction for this is the excellent Shapiro-Wilk test for normality [SW], one of the most powerful of tests for normality [M].)

Informal Justification for ANOVA

To justify one way ANOVA with $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ it is required that all populations be normal with homoscedasticity, i.e., that

$$\sigma_1 = \sigma_2 = \cdots = \sigma_k.$$

But, the requirements are observable in the k independent SRS normal quantile plots. First, the plots must demonstrate linearity (normality of the populations) and secondly, having nearly parallel regression lines implies equality of the σ_i's. In fact,
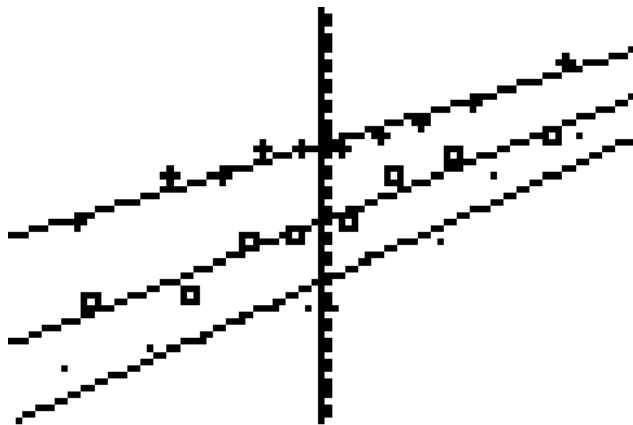
$$\widehat{x} = a + bz \Rightarrow z = \frac{\widehat{x} - a}{b} = \frac{\widehat{x} - \mu}{\sigma} \Rightarrow a = \mu \text{ and } b = \sigma.$$

$$\Rightarrow \widehat{x} = \mu + \sigma z$$

samples: $\{15, 16, 21, 22, 23, 28, 30, 32\} \subset N(24, 5) \rightarrow b = 6.18 \ (\square)$
$\{23, 28, 28, 31, 31, 31, 32, 34, 36, 40\} \subset N(32, 5) \rightarrow b = 4.62 \ (+)$
$\{8, 10, 11, 13, 14, 14, 14, 20, 21, 21, 28, 32\} \subset N(18, 5) \rightarrow b = 7.11 \ (\cdot)$

Normal Quantile Plot T-Test for Normality

sample: $\{15, 16, 21, 22, 23, 28, 30, 32\} \subset N(24, 5)$

$r^2 = 0.9565613319$

$r = 0.9780395349$

```
LinRegTTest
 y=a+bx
 β≠0 and ρ≠0
 t=11.49459795
 p=2.604258ε⁻5
 df=6
↓a=23.375
```

References

[C] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[F] R. A. Fisher, *Statistical Methods for Research Workers*, 8th edition, Oliver and Boyd, Edinburgh, 1941.

[M] Albert Madansky, *Prescriptions for Working Statisticians*, Springer-Verlag, New York, 1988.

[MMC] David S. Moore, George P. McCabe, and Bruce A. Craig, *Introduction to the Practice of Statistics*, 6th edition, W. H. Freeman and Company, New York, 2009.

[SW] S. S. Shapiro and M. B. Wilk, An analysis of variance test of normality (complete samples), *Biometrika*, **52** (1965), 591-611.