16.1 – *p-p* plot

$F_n(x)$:= *Empirical*

$\tilde{F}(x)$:= *Estimated / Assumed / Parametric*

1. Order the observations $x_1 \leq x_2 \dots x_n$
2. The coordinates to plot are $\left(F_n(x_j), \tilde{F}(x_j)\right)$
3. If the model fits well, the data points should fit near the 45 degree line.

16.4 – Hypothesis Tests

H_0: The data came from a population with the stated model
H_1: The data did not come from such a population.

1. Kolmogrov-Smirnov

Test statistic:

$$D = \max_{t \le x \le u} \left| F_n(x) - \tilde{F}(x) \right|$$

t is the left truncation point ($t = 0$ if there is no truncation)
u is the right censored point (u is infinite is there is no censoring)

If D > a critical value at some significance value α, the reject H_0.

α	0.10	0.05	0.01
Critical Value	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

2. Anderson-Darling

Test statistic:

$$A^2 = n\int_t^u \frac{\left[F_n(x) - \tilde{F}(x)\right]^2}{\tilde{F}(x)\left[1 - \tilde{F}(x)\right]} \tilde{f}(x)\,dx$$

For individual data,

$$A^2 = -n\tilde{F}(u) + n\sum_{j=0}^{k}\left[1 - F_n(y_j)^2\right]\lg\left[1 - \tilde{F}(y_j)\right] - \lg\left[1 - \tilde{F}(y_j)\right]$$

$$+n\sum_{j=1}^{k} F_n(y_j)^2\left[\lg\tilde{F}(y_{j+1}) - \lg\tilde{F}(y_j)\right]$$

α	0.10	0.05	0.01
Critical Value	1.933	2.492	3.857

3. Chi-Square Goodness of Fit

Test statistic:

$$\chi^2 = \sum_{j=1}^{k} \frac{n\left(\hat{p}_j - p_{n,j}\right)^2}{\hat{p}_j} \equiv \sum_{j=1}^{k} \frac{\left(E_j - O_j\right)^2}{E_j}$$

Where E_j is the expected number of observations in the interval, and O_j is the observed number.

The critical value comes from the chi-square distribution with degrees of freedom = k - #estimated parameters – 1

4. Likelihood Ratio Test

H_0: The data came from model A
H_1: The data came from model B
where A is a special case of B, e.g. B is Poisson, A is Gamma.

$$teststat, T = 2\lg\left[\frac{Likelihood(\Theta_A)}{Likelihood(\Theta_B)}\right]$$

Tested against Chi-sq, where d.f. = no. of free parameters in the model H_1 – no. of free parameters in the model H_0

16.5 – Swarchz-Bayesian Criterion

The lower the better.

$$SBC = Loglikelihood - \frac{\#obs}{2}\lg n$$