STAT 231 — Lecture 12

Bartosz Antczak *Instructor: Michael Wallace* *October 4, 2017*

12.1 Checking Model Fit

Once we determine a valid probability model and estimate for $\widehat{\theta}$, we must ensure that our model is appropriate. We already covered methods that involve:

- Checking the fit of the histogram to the expected probability distribution
- Checking the empirical cumulative distribution with the cumulative distribution function for the expected continuous distribution

Today we'll look at two other methods.

12.1.1 Checking Observed Frequencies using the Assumed Model

Example 12.1.1. *Habs Goals*

Goals	0	1	2	3	4	5	6	7
Games	2	17	21	18	15	7	1	1

Let $y =$ number of goals the Habs score in a game. Let's assume that $Y \sim \text{Poisson}(\theta)$. We see that $\widehat{\theta} = 2.695$. If our model is correct, then

$$P(Y = y) = \frac{2.695e^{-2.695}}{y!}$$

We can then plot these values and compare them.

12.1.2 Examine a Gaussian qqplot

We'll only focus on Gaussian qqplots for now

If we want to check whether a particular dataset $\{y_1, y_2, \ldots, y_n\}$ are distributed normally, then the following plot $\left(Q_Z\left(\frac{i}{n+1}\right), y_{(i)}\right)$ $i \in [1, n]$ should be linear. Here, $Q_Z(p)$ is the pth quantile for the $G(0, 1)$ distribution. Other shapes the plot can take on:

- **S-shaped:** distribution is symmetric with skewness close to 0. Kurtosis should be less than 3
- **U-shaped:** non-symmetric distribution. Skewness is positive