Probability and Statistics I

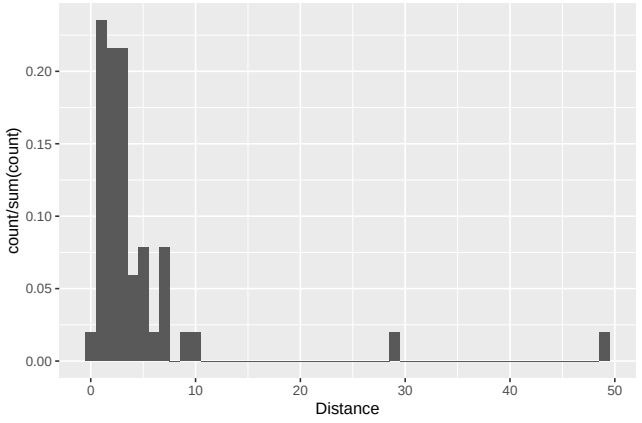
18. Chapter 4 Summary Exercise

How far do Western students live from campus?

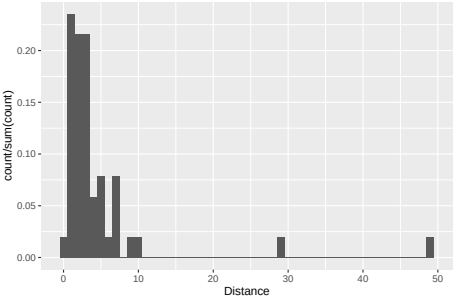
Use Google Maps to compute the distance from your home to campus. Enter the result in the spreadsheet at:

`https://uwoca-my.sharepoint.com/:x:/g/personal/sbonner6_uwo_ca/Ee9EAgVPP2pJkSgS7SPQCZoBhubeOhoSTgMd-2QrGf9KIg?e=o27rxM`

How far do Western students live from campus?



How far do Western students live from campus?



$$n = 51, \quad \hat{\mu} = 4.634, \quad \widehat{\sigma^2} = 58.011$$

How far do Western students live from campus?

Use the data summaries on the previous slide to estimate the parameters assuming:

- the distribution is normal.
- the distribution is gamma.

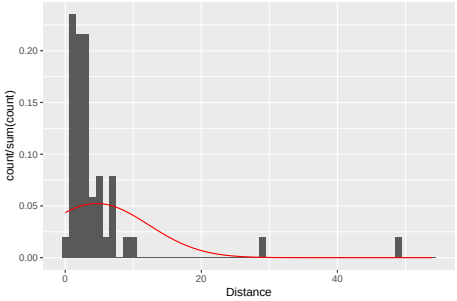
For each distribution find:

1. The value d_1 so that 97.5% of students live less than d_1 km from campus.
2. The value d_2 so that 97.5% of students live greater than d_2 km from campus.

Which distribution do you believe will fit the data better?

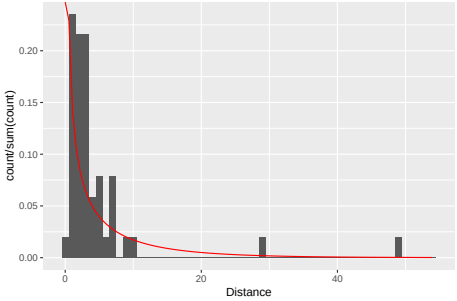
Normal Distribution

$$\hat{\mu} = 4.634, \quad \widehat{\sigma^2} = 58.011$$



Gamma Distribution

$$\hat{\alpha} = 0.37, \quad \hat{\beta} = 12.519$$



Percentile Plots

Suppose that the data come from a distribution with percentiles η_p. Then we would expect the k-th largest data point to be close to $\eta_{(k-.5)/n}$[1].

With 51 observations, we would expect

- the smallest observation to be close to $\eta_{\frac{.5}{51}}$
- the next smallest observation to be close to $\eta_{\frac{1.5}{51}}$
- $\cdots$
- the biggest observation to be close to $\eta_{\frac{50.5}{51}}$

Generally, if $x_{(1)}, \ldots, x_{(n)}$ represent the *ordered* data then

$$x_{(k)} \approx \eta_{\frac{k}{n}}.$$

Percentile Plots

Suppose that the data come from a distribution with percentiles η_p. Then we would expect the k-th largest data point to be close to $\eta_{(k-.5)/n}$[2].

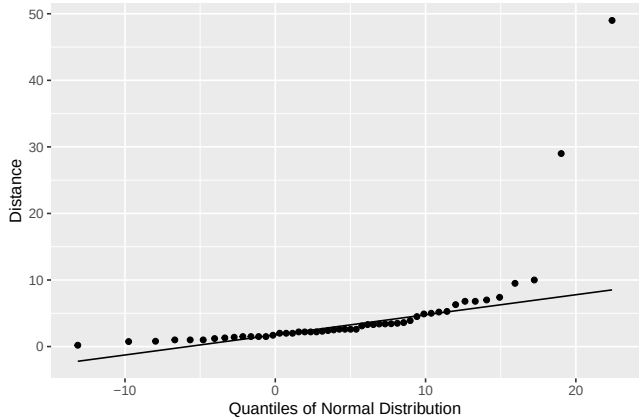
Generally, if $x_{(1)}, \ldots, x_{(n)}$ represent the *ordered* data then

$$x_{(k)} \approx \eta_{\frac{k}{n}}.$$

A percentile plot plots $x_{(k)}$ vs $\eta_{\frac{k}{n}}$ for $k = 1, \ldots, n$. If the data were generated from that distribution then the points will lie close to a straight line.

[2] The extra $-.5$ keeps us below 100%.

Percentile Plots – Normal



Percentile Plots – Gamma