DATA5100 - PROJECT CHECKPOINT 8 – ANDREX IBIZA – 2024-02-27

Sampling Distribution of Sample Mean Weight

I made a sampling distribution of sample means for samples of size n = 500 for the weights of the penguins in the PalmerPenguins dataset. See Figure 1 for the graph of the distribution (red) with a superimposed normal curve (blue). The sampling distribution of sample means is approximately normal. This is confimed by the normal QQ plot. See Figure 2.

The population mean weight is μ = 4201.754 with standard deviation of σ = 801.9545. The expected mean of the sample means is $\mu_{\bar{x}}$ = μ = 4201.754 with standard error $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ = 146.4162. The actual mean of the sample means is $\mu_{\bar{x}}$ = 4202.233 with standard error $\sigma_{\bar{x}}$ = 143.2586.

These results support the conclusions of the Central Limit Theorem.

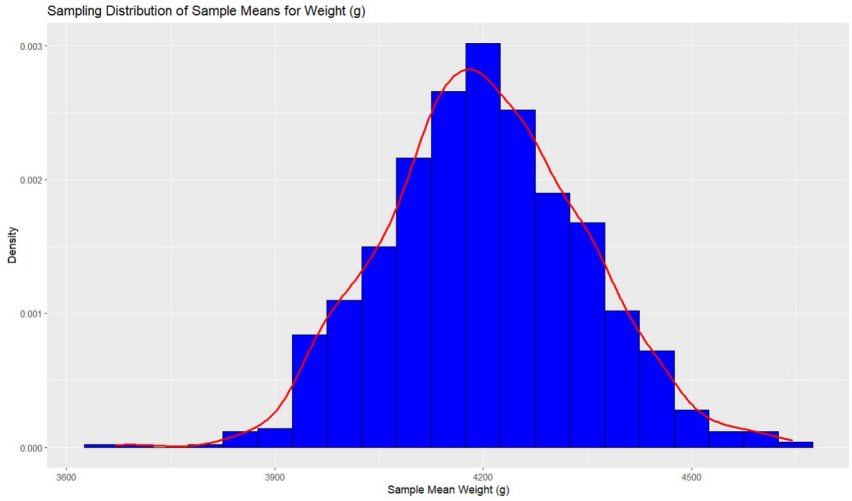

Figure 1

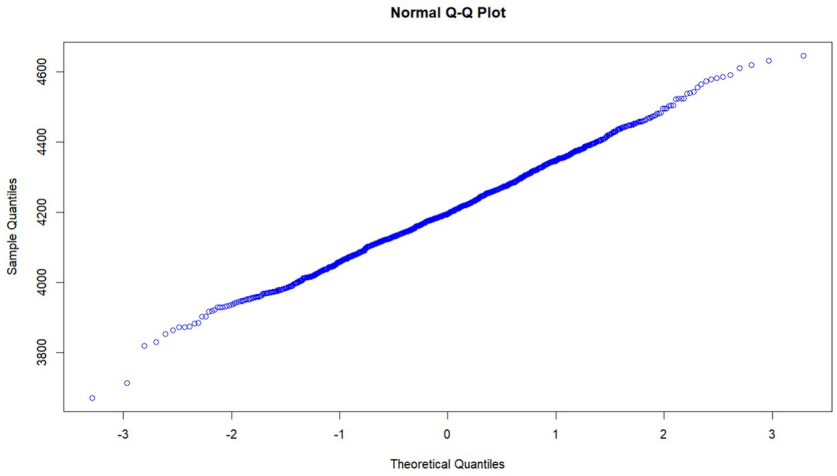

Figure 2

DATA5100 - PROJECT CHECKPOINT 8 – ANDREX IBIZA – 2024-02-27

Sampling Distribution of Sample Proportions

I made a sampling distribution of sample proportions for samples of size n = 500 for the proportion of the players in the PalmerPenguins dataset that are species Adelie. See Figure 3 for the graph of the distribution (red) with a superimposed normal curve (blue). The sampling distribution of sample proportions is approximately normal. This is confimed by the normal QQ plot. See Figure 4.

The population proportion of Adelie penguins is p = 0.4418605. The expected mean of the sample proportions is $\mu_{\hat{p}} = \text{p} = 0.4418605$ with standard error $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = 0.022209$. The actual mean of the sample proportion is $\mu_{\hat{p}} = 0.4417294$ with standard error $\sigma_{\hat{p}} = 0.02191825$.

My results are as expected.

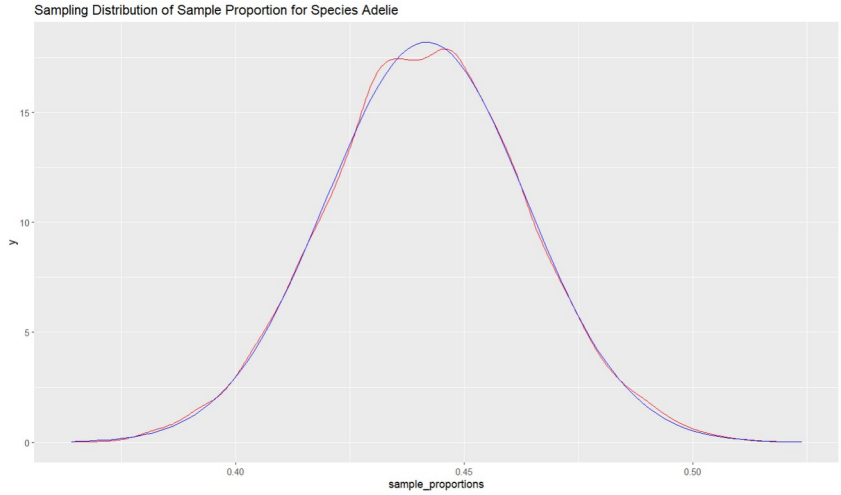

Figure 3

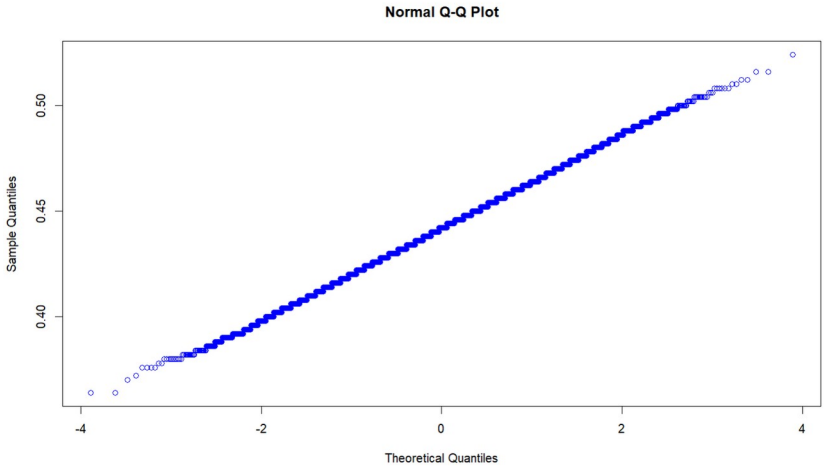

Figure 4

DATA5100 - PROJECT CHECKPOINT 8 – ANDREX IBIZA – 2024-02-27

This week, I changed the name of the “body_mass_g” column to “weight” for ease of use.