

6.8: Moments

Introduction

Moments are used to describe the shape of a distribution. For those of you who remember your calculus, moments were discussed as a method to find the center of mass, or balancing point (Herman and Strang 2018). For distributions, the center and shape moments follow from the **expected value** of the probability function.

Note:

Expected value of a statistic is calculated by multiplying the likelihood of each possible outcome in a sample space by that outcome, then adding up all of those product values. From probability theory it is the weighted average of the outcomes of a random variable. A simpler way to think of the expected value is that if one were to guess the height of a person, the expected value is the average height of the population from which the person would be selected.

Four moments apply for describing the shape of a distribution. The 1st moment describes the middle, the 2nd describes the spread from the middle, the 3rd describes symmetry about the middle, and the 4th describes the shape, whether peaked and sharp, or **leptokurtic**, or broad and flattened, or **platykurtic**.

Equations for the moments

Over the years, several equations have been proposed to estimate skewness and kurtosis. The above formulas are just one example from the list (Joanes and Gill 1998).

Pearson's standardized moments:

$$\frac{\mu_n}{\sigma^n} \equiv \frac{E[X - \mu]^n}{\sigma^2}$$

where E is the expected value of the random variable. The expected value concept follows from rules of probability — basically, the average of a large number, n , of X .

Four moments can be used to describe the shape of a distribution.

1st moment, μ (mean): population mean, [3.2 – Measures of Central Tendency](#)

2nd moment, σ^2 (variance): population variance, [3.3 – Measures of dispersion](#)

3rd moment, $\bar{\mu}_3$ (skew):

$$\frac{\mu_3}{\sigma^3} = \frac{\sqrt{n(n-1)}}{n-2} \left[\frac{\frac{1}{n} \sum (X - \mu)^3}{\left(\frac{1}{n} \sum (X - \mu)^2 \right)^{\frac{3}{2}}} \right]$$

4th moment, $\bar{\mu}_4$ (kurtosis):

$$\frac{\mu_4}{\sigma^4} = 3\sigma^4 = \frac{\mu_4}{\sigma_2^2} - 3$$

Estimating moments in R and R Commander

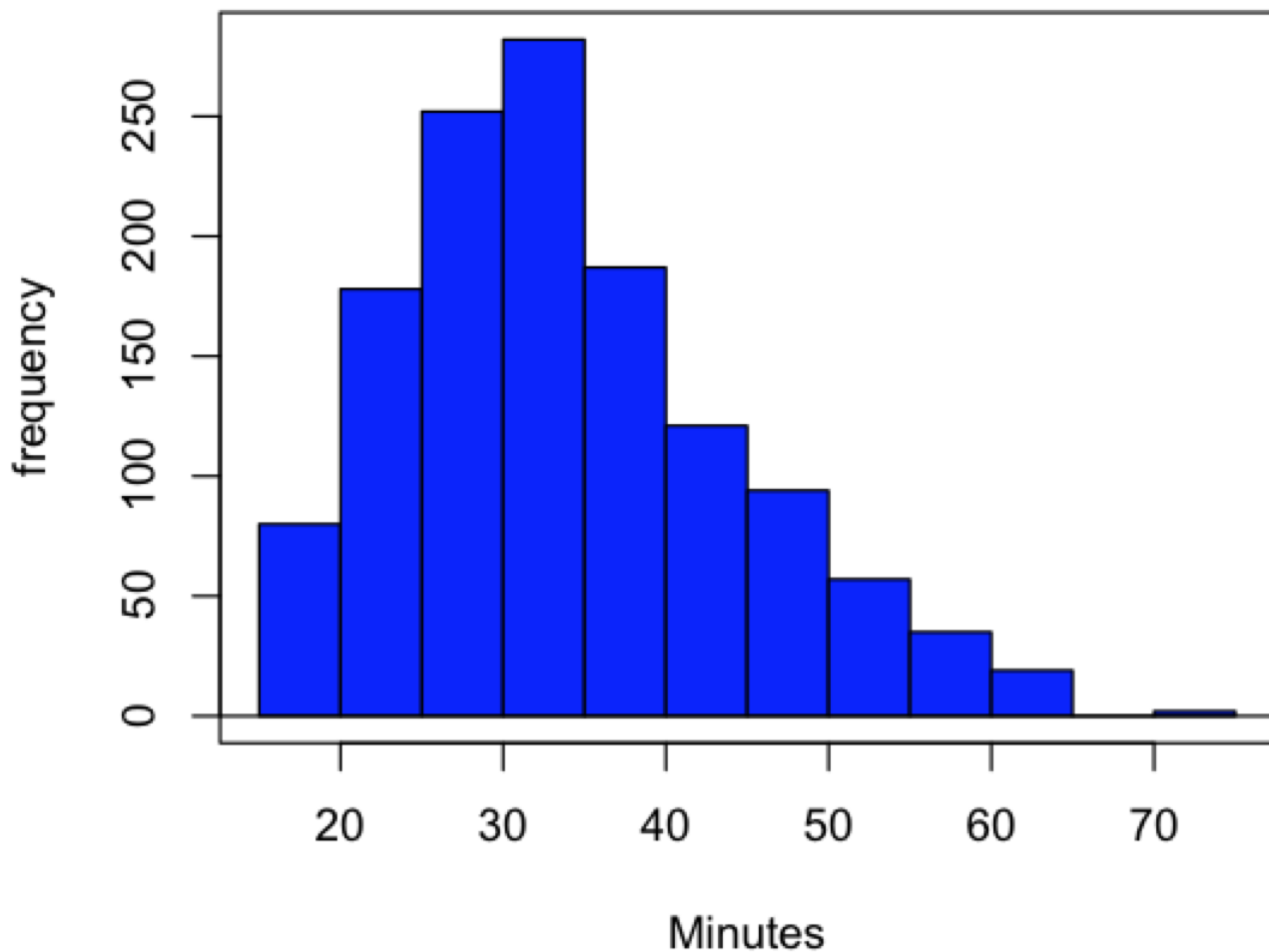


Figure 6.8.1: Histogram of finishing times in minutes for 1307 runners at the 2016 Banana 5K race.

In R Commander, we select **Statistics** → **Summaries** → **Numerical summaries...**, which brings up a popup menu. First, select the variable, in this case Minutes, from the Data tab (not shown). Next, click on Statistics tab to choose options (Fig. 6.8.2).

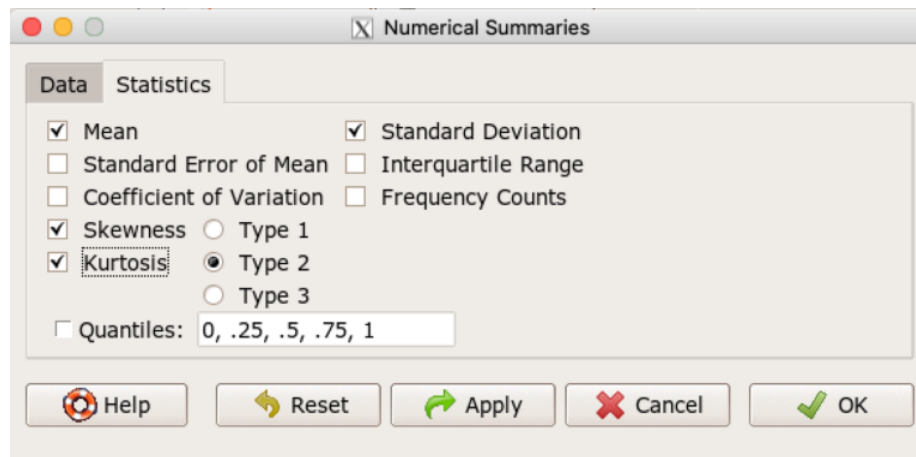


Figure 6.8.2: Numerical Summaries menu in Rcmdr Statistics.

For estimates of the moments, check Mean, Standard Deviation, Skewness, and Kurtosis. Note that Rcmdr gives you the choice among three different Types of skewness and kurtosis. Type 1 include the equations provided on this page, corresponding to definitions dating back to the 1940s. Type 2 is the default and corresponds to equations used by other professional statistics

package (SAS, SPSS). For large sample size, the different types will tend to agree. Caution applies to smaller data sets — the different types may disagree (Joanes and Gill 1998).

Large sample size, $n = 1307$

Type 1

mean	sd	skewness	kurtosis	n
34.42999	10.31437	0.6159258	-0.01593882	1307

Type 2

mean	sd	skewness	kurtosis	n
34.42999	10.31437	0.6166337	-0.01139521	1307

Type 3

mean	sd	skewness	kurtosis	n
34.42999	10.31437	0.615219	-0.02050335	1307

Small sample size

To test the claim about sample size and the moment statistics, draw a random sample of 30 from the larger data set. Sample without replacement:

```
sample.banana <- data.frame(sample(banana5K$Minutes, 30, replace = FALSE))
```

I forgot to specify a new variable name, so R used the whole command as the variable name. I could go back and fix my function call, or simply rename the variable as follows:

```
names(sample.banana)[c(1)] <- c("Minutes")
```

The random sample yielded a distribution (Fig. 6.8.3).

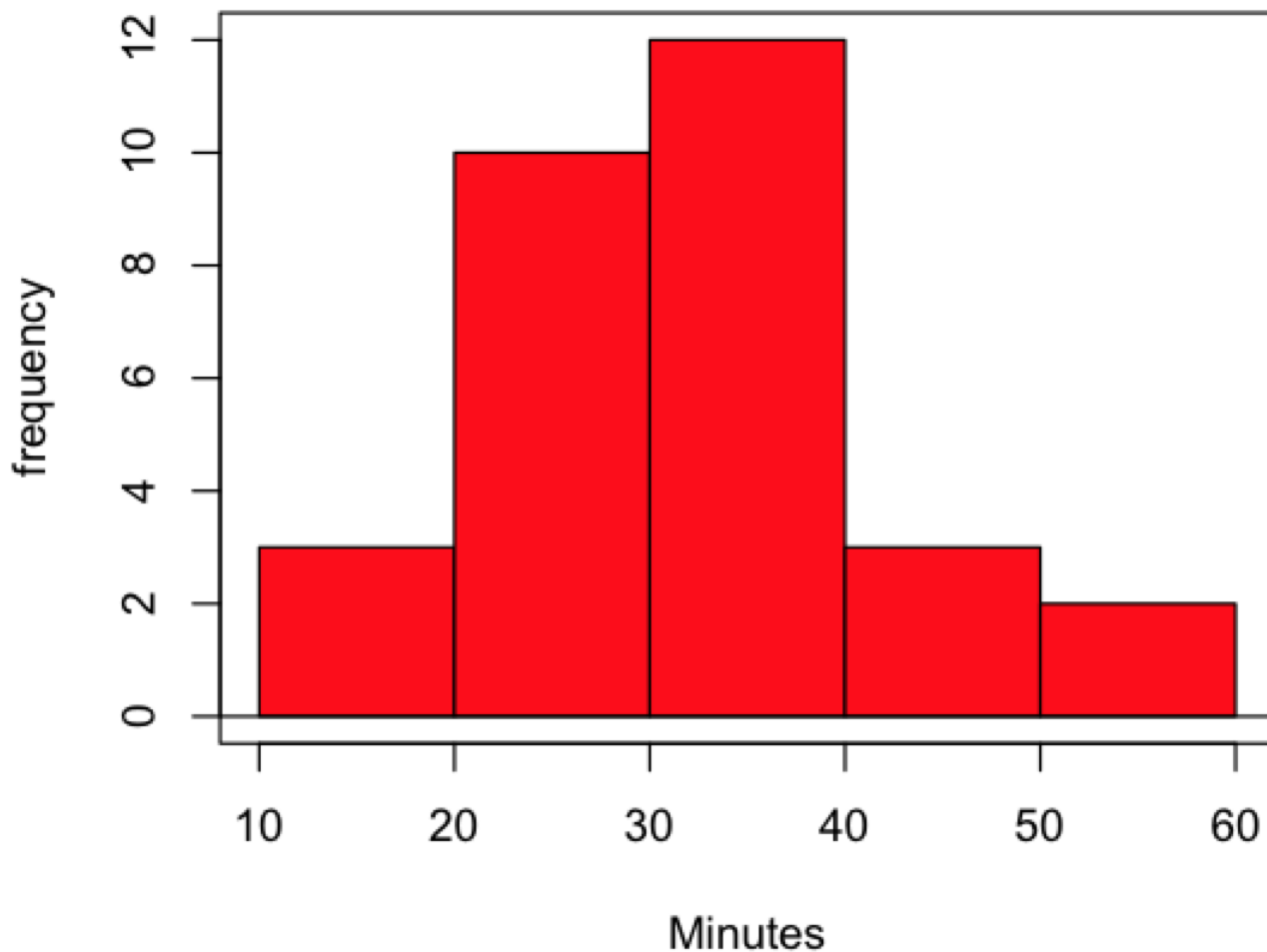


Figure 6.8.3: Histogram of finishing times in minutes for random sample of 30 drawn from 1307 runners at 2016 Banana 5K race.

Repeat Numerical summaries on small data set, $n = 30$

Type 1:

mean	sd	skewness	kurtosis	n
33.16667	10.00373	0.5538637	0.5024438	30

Type 2:

mean	sd	skewness	kurtosis	n
33.16667	10.00373	0.5834511	0.8276415	30

Type 3:

mean	sd	skewness	kurtosis	n
33.16667	10.00373	0.5264025	0.2728392	30

Conclusion: We can compare consistency of the estimators by calculating coefficient of variation. The three types of skewness estimators differed by only 1% and 5% for large and small sample size, respectively. In contrast, the three types of kurtosis estimators differed by 29% and 52% for large and small sample size, respectively.

Questions

1. Explore the consistency of skewness and kurtosis estimates by calculating and comparing coefficient of variation estimates. R Commander provide a nice way to draw randomly from various defined distributions. Draw two data sets of 15 (small) and 1000 (large), from the chi-square distribution (1 degree of freedom) and a minimum of one other continuous distribution.

Example, draw random sample of 1000 from chi-square distribution. Rcmdr: Distributions → Continuous distributions → Chi-squared distribution → Sample from chi-squared distribution...

Enter name for the variable, enter degrees of freedom (e.g., 1), number of samples (e.g., 1000), and number of observations (variables, columns). Leave Sample means checked under data sets.

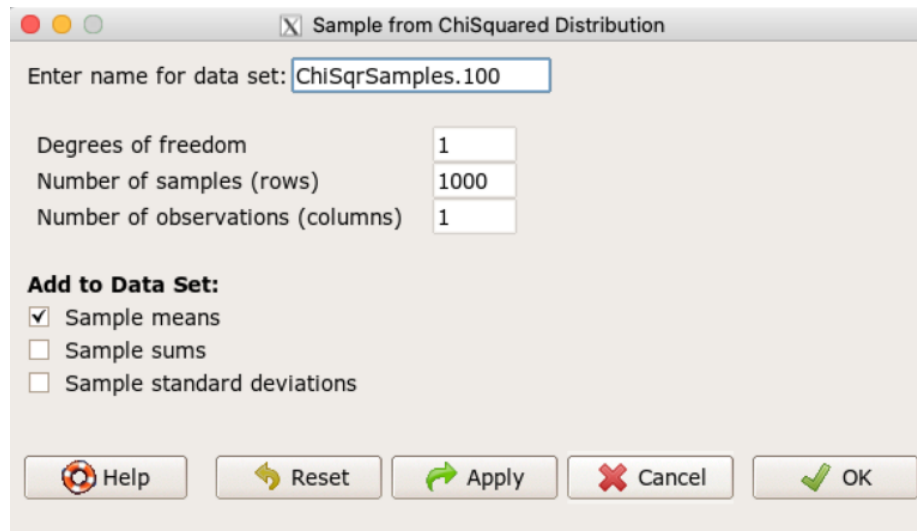


Figure 6.8.4: Sample from Chi Squared Distribution menu in Rcmdr.

This results in a new data set. Get “moments” from Numerical summaries and calculate coefficient of variations. Which moments have the most consistency regardless of the kind of distribution.

2. Make histograms for each of your created data sets. Describe what you see about the shape of the plotted distributions.

This page titled [6.8: Moments](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.