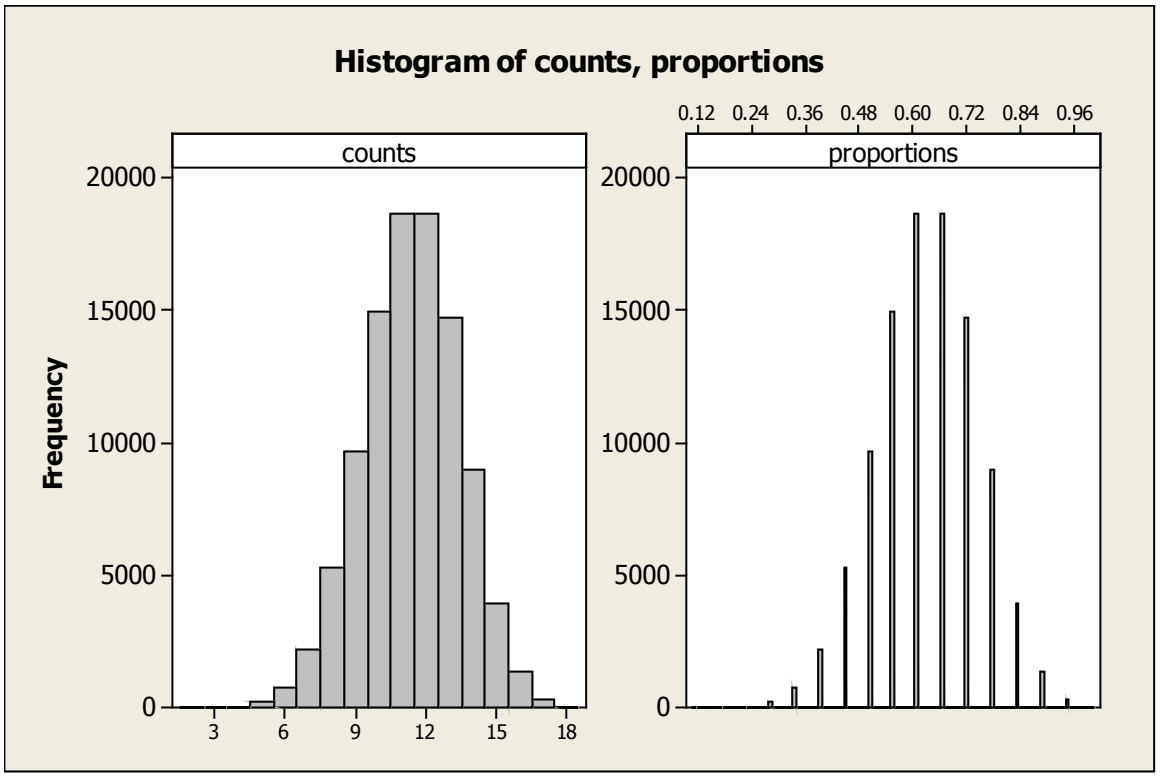
Example 1: Binomial Counts and Proportions – here n = 18 and $\pi = 0.63$
(T = 100,000 replicates for this and each of the following examples)

For the **Count** variable (Y), the expected value is $\mu_Y = n\pi = 11.34$, and $SE_Y = SQRT\{n\pi(1-\pi)\} = 2.048$; notice how closely these numbers match with the following Minitab output (Minitab incorrectly calls SE_Y 'StDev'). Furthermore, since n is 'large' (verify this!), the histogram below (left) of the sample counts is approximately normally distributed.

For the **Proportion** variable (p), the expected value is $\mu_p = \pi = 0.63$, and $SE_p = SQRT\{\pi(1-\pi)/n\} = 0.113798$; again, notice how closely these match with the following Minitab output. Also, since n is 'large', the histogram below (right) of the sample proportions is approximately normal in shape.

Descriptive Statistics: counts, proportions

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Count	100000	11.354	2.045	2.000	10.000	11.000	13.000	18.000
Proportion	100000	0.63075	0.11363	0.11111	0.55556	0.61111	0.72222	1.00000



Example 2: Discrete Uniform Counts –

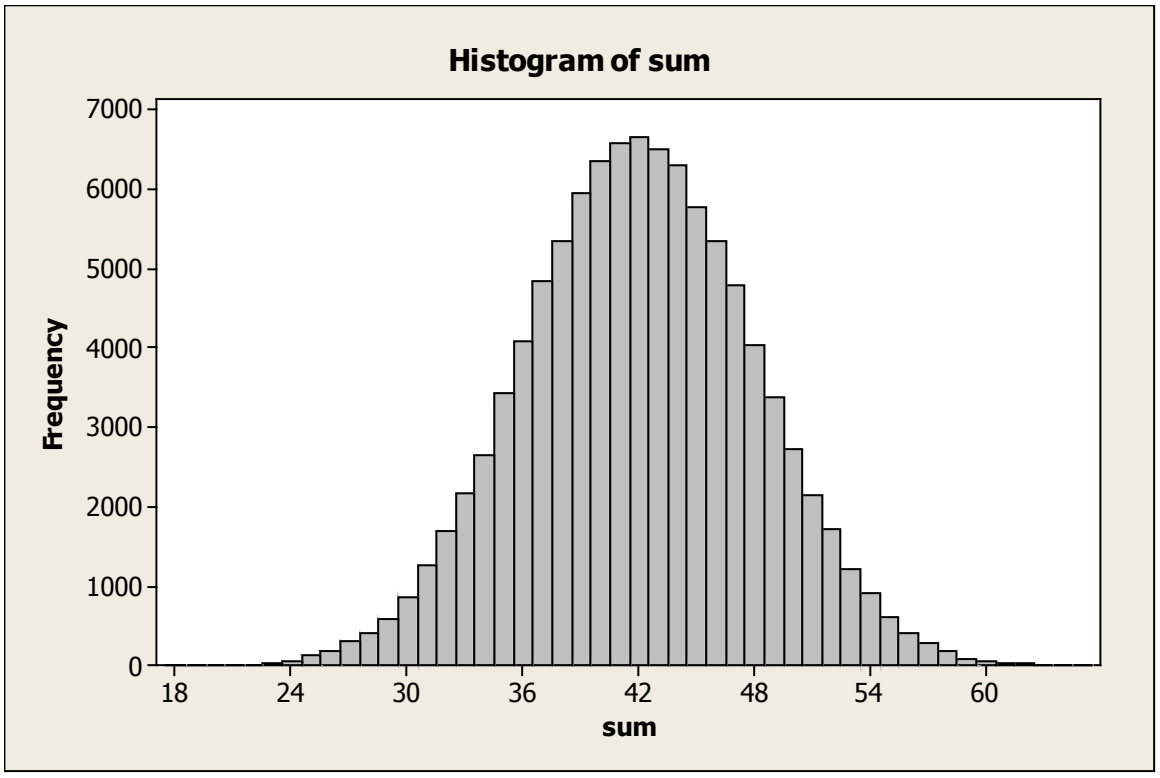
Here $Pr\{X = k\} = 1/6$ for $k = 1,2 \ldots 6$, so $\mu_X = 3.5$ and $\sigma_X^2 = 35/12$ (formulas give on pp.138-9). The underlying distribution (of the Box) looks flat or uniform – see p.138.

With a sample size of $n = 12$, the count is $Y = X_1 + X_2 + \ldots + X_{12}$, so that $\mu_Y = 12*3.5 = 42$, $\sigma_Y^2 = 12*(35/12) = 35$, and therefore $SE_Y = SQRT(35) = 5.91608$. In our simulation results summarized below, we get almost the same exact values.

Descriptive Statistics: sum

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
sum	100000	41.984	5.925	18.000	38.000	42.000	46.000	65.000

Furthermore, and quite surprising is the following result: Even though the underlying population is flat (uniform), the count variables yield a Normal-looking histogram (below). This clearly demonstrates Central Limit Theorem (or Law of Averages) – for some not-so-obvious reason, sums, counts, proportions and averages behave like Normal random variables – provided they are based on a large enough sample.



Example 3: Average of Exponential Random Variables – here we take a sample of n independent and identically distributed (denoted "**iid**") $\mathcal{E}(\theta = 1)$ RV's, and we calculate the ***sample mean***, $\bar{x}$. This underlying 'parent' population is very skewed to the right, so we expect some problems with achieving approximate normality here. Let's see how this skewness disappears as the sample size increases.

Since the mean (μ_X) and variance (σ_X^2) for an $\mathcal{E}(\theta)$ RV are θ and θ^2 respectively, here these values (and the SD) are each 1. Therefore in a sample of size n, the expected value of the sample mean (denoted $\mu_{\bar{x}}$) is also 1, the variance of the sample mean is $\sigma_X^2 / n = 1/n$, so the SE of the sample mean is $SE_{\bar{x}} = \sigma_X / \sqrt{n} = 1/\sqrt{n}$. We run our simulation here with several different sample sizes – viz, n = 4, 9, 16 and 25; the summary is given below. Notice that as the sample size increases, the expected value (denoted 'Mean' below by Minitab) gets closer and closer to $\mu_{\bar{x}} = 1$, whereas $SE_{\bar{x}} = 1/\sqrt{n}$ shrinks. More surprising is the fact that the histograms (bottom of the page) get closer and closer to Normal-looking curves as the sample size increases; for n = 4 there is a large (right) skew, same for n = 9, but for n = 25 the skew is virtually gone and the histogram is very nearly Gaussian.

Descriptive Statistics: mean4, mean9, mean16, mean25

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
mean4	100000	1.0011	0.5025	0.0258	0.6316	0.9184	1.2793	4.7168
mean9	100000	1.0009	0.3338	0.0893	0.7607	0.9641	1.2005	3.2707
mean16	100000	1.0004	0.2504	0.2794	0.8222	0.9801	1.1567	2.5570
mean25	100000	1.0003	0.2001	0.3794	0.8594	0.9868	1.1284	2.1180