Econ 205 - Cheat Sheet
Statistics for Business and Economics

Descriptive statistics:
Mean: $\bar{x}$ `=average(DATA)`, Median `=median(DATA)` , Mode `=mode(DATA)`
Variance: $\sigma^2 = \texttt{var.p}\left(\text{DATA}\right)$, $s^2 = \texttt{var.s}\left(\text{DATA}\right)$, $s^2 = \sigma^2\left(\frac{N}{n-1}\right)$ or $\sigma^2 = s^2\left(\frac{n-1}{N}\right)$
Standard deviation: $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \texttt{stdev.s}(\text{DATA}) = \sigma\sqrt{\frac{N}{n-1}}$
$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \texttt{stdev.p}(\text{DATA}) = s\sqrt{\frac{n-1}{N}}$
Coefficient of variation: $CV = \frac{s}{\bar{x}}$ or $CV = \frac{\sigma}{\mu}$
P^{th}-percentile `=percentile.exc(data,P/100)` or location formula $L_P = (n+1)\frac{P}{100}$
Covariance: σ_{XY} `covariance.s(data)` or Data-Analysis-Toolpak > Covariance, $\sigma_{XY} = s_{XY}\frac{n-1}{N}$
Correlation coefficient: $\rho_{XY} = \texttt{correl}(\texttt{data})$ or Data-Analysis-Toolpak > Correlation
Regression model: $y = \beta_0 + \beta_1 x + \varepsilon$
Regression line estimated: $\hat{y} = b_0 + b_1 x$ in Scatterplot > Add Trendline or Data-Analysis-Toolpak > Regression

Probability:
Rule of complements: $P\left(A^c\right) = 1 - P\left(A\right)$
Multiplication formulat: $P\left(A \text{ } and \text{ } B\right) = P\left(A|B\right) \times P\left(B\right)$
Addition formulat: $P\left(A \text{ } or \text{ } B\right) = P\left(A\right) + P\left(B\right) - P\left(A \text{ } and \text{ } B\right)$
Conditional probability: $P\left(A|B\right) = \frac{P(A \text{ } and \text{ } B)}{P(B)} = \frac{P(A|B)\times P(B)}{P(B)}$
Independence: A and B are independent if $P(A \text{ } and \text{ } B) = P(A) \times P(B)$
or when $P\left(A|B\right) = P\left(A\right)$
Mean (aka expected value) of a discrete distribution: $\mu = \sum_{i=1}^{n} x_i P(x_i)$
$E\left[c\right] = c$, $Var\left[c\right] = 0$;
$E\left[X+c\right] = E\left[X\right] + c$, $Var\left[X+c\right] = Var\left[X\right]$;
$E\left[cX\right] = cE\left[X\right]$, $Var\left[cX\right] = c^2 Var\left[X\right]$

Distributions:

- Binomial distribution: $P\left(X = x\right) = \texttt{binom.dist}\left(x, n, \pi, 0\right)$ and $P\left(X \leq x\right) = \texttt{binom.dist}\left(x, n, \pi, 1\right)$
- Mean of binomial distribution: $\mu = E\left[X\right] = n \times \pi$ and $\sigma^2 = Var\left[X\right] = n \times \pi \times (1-\pi)$
- Uniform distribution: $X \sim U\left[a, b\right]$, then density is $f = \frac{1}{b-a}$ and $E\left[X\right] = \frac{a+b}{2}$ and $V\left[X\right] = \frac{1}{12}\left(b-a\right)^2$
- Normal distribution:

$X \sim N\left(\mu, \sigma\right) : P\left(X < x\right) = P\left(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right)$ where $Z = \frac{X-\mu}{\sigma} \sim N\left(0, 1\right)$ then:

$$\begin{aligned} \text{to-the-left} &: P\left(Z < z\right) = \texttt{norm.s.dist}\left(z\right), \text{ or} \\ \text{to-the-right} &: P\left(Z > z\right) = 1 - \texttt{norm.s.dist}\left(z\right), \text{ and} \\ \text{the } \pi^{th} \text{ percentile } P_\pi &= \texttt{norm.s.inv}\left(\pi\right) \end{aligned}$$

- Student T-distribution:

-if σ unknown then $T = \frac{X-\mu}{s} \sim T\left(n-1\right)$:

$$\begin{aligned} \text{to-the-left} &: & P\left(T < t\right) = \texttt{t.dist}\left(t, n-1, 1\right), \text{ or} \\ \text{to-the-right} &: & P\left(T > t\right) = 1 - \texttt{t.dist}\left(t, n-1, 1\right), \text{ and} \\ \text{the } \pi^{th} \text{ percentile } P_\pi &= & \texttt{t.inv}\left(\pi, n-1\right) \end{aligned}$$

Central limit theorem:
If $X \sim N\left(\mu, \sigma\right)$ then $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
If $X \sim ?\left(\mu, \sigma\right)$ then $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ if $n \geq 30$
If proportion $\pi \sim ?$, then sample proportion $\hat{p} \sim N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right)$ if $n\pi \geq 5$ and $n\left(1-\pi\right) \geq 5$

Confidence intervals:
CI_α (z): $\left[\bar{x} \pm \overbrace{z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}}^{W}\right]$ and conf. int. (t): $\left[\bar{x} \pm \overbrace{t_{\alpha/2} \times \frac{s}{\sqrt{n}}}^{W}\right]$
Sample size formula: $n = \left(z \times \frac{\sigma}{W}\right)^2$

Hypothesis testing:
z_{obs} formula: $z_{obs} = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ or $z_{obs} = \frac{\hat{p}-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$
t_{obs} formula: $t_{obs} = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$.

Two sample tests:
Case 1: equal variances $\sigma_1^2 = \sigma_2^2$: $t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{s_P^2\left(\frac{1}{n_1}+\frac{1}{n2}\right)}}$, where $s_P^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ and $v = n_1 + n_2 - 2$
Case 2: unequal variances $\sigma_1^2 \neq \sigma_2^2$: $t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n2}\right)}}$, where $v = \frac{\left(s_1^2/n_1+s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1}+\frac{\left(s_2^2/n_2\right)^2}{n_2-1}}$

Regression analysis:
Data-Analysis-Toolpak > Regression
SST = SSR + SSE
Coefficient of determination: $R^2 = \frac{SSR}{SST}$