Econ 205 - Cheat Sheet
Statistics for Business and Economics

Descriptive statistics:

Mean: $\bar{x}$ =`average(DATA)`, Median =`median(DATA)`, Mode =`mode(DATA)`

Variance: $\sigma^2 =$ `var.p`(DATA), $s^2 =$ `var.s`(DATA), $s^2 = \sigma^2 \left(\frac{N}{n-1}\right)$ or $\sigma^2 = s^2 \left(\frac{n-1}{N}\right)$

Standard deviation: $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} =$ `stdev.s`$(\text{DATA}) = \sigma\sqrt{\frac{N}{n-1}}$

$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}} =$ `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 $\boxed{\text{Data-Analysis-Toolpak} \gg \text{Covariance}}$, $\sigma_{XY} = s_{XY}\frac{n-1}{N}$

Correlation coefficient: $\rho_{XY} =$ `correl(data)` or $\boxed{\text{Data-Analysis-Toolpak} \gg \text{Correlation}}$

Regression model: $y = \beta_0 + \beta_1 x + \varepsilon$

Regression line estimated: $\hat{y} = b_0 + b_1 x$ in $\boxed{\text{Scatterplot} \gg \text{Add Trendline}}$ or $\boxed{\text{Data-Analysis-Toolpak} \gg}$ $\boxed{\text{Regression}}$

Probability:

Rule of complements: $P(A^c) = 1 - P(A)$

Multiplication formulat: $P(A \text{ and } B) = P(A|B) \times P(B)$

Addition formulat: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Conditional probability: $P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A|B) \times P(B)}{P(B)}$

Independence: A and B are independent if $P(A \text{ and } B) = P(A) \times P(B)$
or when $P(A|B) = P(A)$

Mean (aka expected value) of a discrete distribution: $\mu = \sum_{i=1}^{n} x_i P(x_i)$

$E[c] = c$, $Var[c] = 0$;

$E[X + c] = E[X] + c$, $Var[X + c] = Var[X]$;

$E[cX] = cE[X]$, $Var[cX] = c^2 Var[X]$

Distributions:

- Binomial distribution: $P(X = x) =$ `binom.dist`$(x, n, \pi, 0)$ and $P(X \leq x) =$ `binom.dist`$(x, n, \pi, 1)$

- Mean of binomial distribution: $\mu = E[X] = n \times \pi$ and $\sigma^2 = Var[X] = n \times \pi \times (1 - \pi)$

- Uniform distribution: $X \sim U[a, b]$, then density is $f = \frac{1}{b-a}$ and $E[X] = \frac{a+b}{2}$ and $V[X] = \frac{1}{12}(b-a)^2$

- Normal distribution:

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

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

- Student T-distribution:

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

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

Central limit theorem:
If $X \sim N(\mu, \sigma)$ then $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

If $X \sim ?(\mu, \sigma)$ 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 $n(1-\pi) \geq 5$

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 $\sigma_1^2 = \sigma_2^2$: $t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{s_P^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, 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 $\sigma_1^2 \neq \sigma_2^2$: $t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)}}$, 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 $\gg$ Regression |

SST = SSR + SSE

Coefficient of determination: $R^2 = \frac{SSR}{SST}$