ST260 Formula Sheet
Compatible with Weiers: Intro. to Business Statistics, Duxbury, Inc.

Sample mean:
\[ \bar{x} = \frac{\sum x_i}{n} \]

Sample variance:
\[ s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum (\text{deviations})^2}{n-1} \]

Computing formula:
\[ s^2 = \frac{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2}{n-1} = \frac{\sum x_i^2 - nx^2}{n-1} = \frac{2}{n-1} \]

Sample Standard Deviation:
\[ s = +\sqrt{s^2} \]

Standardizing any random variable:
\[ Z = \frac{X - \mu}{s} = \left\{ \begin{array}{l}
\text{number of standard deviations} \\
\text{that } X \text{ is from its mean}, \mu.
\end{array} \right. \]

If \( X \) is “normal”, \( Z \) is “standard normal”:
If \( X \sim N(\mu, \sigma) \), then \( Z = \frac{X - \mu}{s} \sim N(0,1) \).

Linear transformation of \( X \):
If \( X \sim N(\mu, \sigma) \), and \( Y = a + bX \), then \( Y \sim N(a + b\mu, |b|\sigma) \).

Regression Analysis

Definition forms:
Equation 1:
\[ \sum (x_i - \bar{x})(y_i - \bar{y}) \]
Equation 2:
\[ \sum (x_i - \bar{x})^2 \]
Equation 3:
\[ \sum (y_i - \bar{y})^2 \]

The following “computational” forms may be easier to use:
\[ \sum x_i y_i \left( \frac{\sum x_i}{n} \right) \left( \frac{\sum y_i}{n} \right) \\
= \sum x_i y_i - n \bar{x} \bar{y} \]
\[ \sum x_i^2 \left( \frac{\sum x_i}{n} \right)^2 \\
= \sum x_i^2 - n \bar{x}^2 \]
\[ \sum y_i^2 \left( \frac{\sum y_i}{n} \right)^2 \\
= \sum y_i^2 - n \bar{y}^2 \]

Estimated “slope”:
\[ b_1 = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2} = \frac{1}{2} \]

Estimated “intercept”:
\[ b_0 = \bar{y} - b_1 \bar{x} \]

The Prediction Equation is: Page 612
\[ \hat{y} = b_0 + b_1 x \]

ANOVA summary values:
\[ SSR = b_1^2 \cdot 2 \]
\[ SSE = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 \]
\[ SST = \sum (y_i - \bar{y})^2 \]

Each residual is: \( e_i = y_i - \hat{y}_i \)

\[ t \text{-statistic} = \text{number of standard errors the estimated coefficient is from zero.} \]

\[ P \text{-value} = \text{the probability of observing a future } t \text{-statistic value as extreme or more extreme from zero, than this one.} \]

Correlation coefficient:
\[ r = \frac{1}{\sqrt{2}} \sqrt{3} \]

Relation between slope and correlation: same “signs” and
\[ b_1 = r \left( \frac{\text{Std.Dev. } y}{\text{Std.Dev. } x} \right) \]

How good is the regression?
\[ \begin{align*}
\text{Coefficient of determination, } r^2: \\
&= \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad \text{Page 628}
\end{align*} \]

Standard Error of Estimate:
\[ s_{y.x} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} \]

Logarithmic transformation:
If the scatterplot of \( Y \) vs. \( X \) shows exponential growth or decay, use “log \( y \)” in place of “\( y \)”:

Regression result: \( \log \hat{y} = b_0 + b_1 x \)

Estimation result: \( \hat{y} = 10^{b_0 + b_1 x} \)
Rules for Probability:  Page 167

“and” = “both” = “joint.”  (\(\cap\))

“or” = “either or both.”  (\(\cup\))

Complement = “will not occur”:  \(P(A^c) = P(\text{Not } A)\)  Page 160

1 - \(P(A)\)

Addition Law:  Page 169

\[ P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \]

Special Case: If A and B are mutually exclusive, then

\[ P(A \text{ or } B) = P(A) + P(B) - 0 \]

Binomial distribution:  (discrete)  Page 207

\[ X \sim \text{Bino}(n, \pi) \]

\[ \pi = \text{probability of a success on any one trial;} \]

\[ 1 - \pi = \text{probability of a failure on any one trial;} \]

\[ P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}, \quad x = 0, 1, 2, \ldots, n \]

where \(\binom{n}{x} = \frac{n!}{x!(n-x)!}\) is the number of combinations & \(n! = n(n-1)(n-2) \ldots (2)(1)\)

For any binomial, \(\mu_{\text{Bino}} = np, \quad s_{\text{Bino}} = \sqrt{np(1-p)}\)  Page 207, 259

Distribution of “sample means”  Page 210

\[ \mu_X = \mu \quad \text{(The mean of the population of all possible X-bars is the same mean as the original population;)} \]

\[ s_X = \frac{s}{\sqrt{n}} \quad \text{its standard deviation is smaller than that of the original population.)} \]

Let \(\bar{X}\) be the ran. var. of the original population.

Let \(\bar{X}\) be the ran. var. of population of all possible \(\bar{X}\) values.  When is the population of all possible \(\bar{X}\) variables “normally” distributed?

\[ \begin{align*}
\text{\fbox{Anytime the original population is normal.}} & \quad \text{\fbox{Page 286}} \\
\text{\fbox{Anytime the original population is NOT normal.}} & \quad \text{\fbox{Page 286}} \\
\text{\fbox{The } \bar{X}\text{-population is still approximately normal, if } n \text{ is LARGE (} n \geq 30), \text{ by the C.L.T.}} & \quad \text{\fbox{Page 288-90}}
\end{align*} \]

If \(\bar{X} \sim N\left(\mu, \frac{s}{\sqrt{n}}\right)\), then \(Z = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim N(0, 1)\)

If \(X_1 \sim N(\mu_1, s_1) \text{ a nd } X_2 \sim N(\mu_2, s_2)\), then

\[
(\bar{X}_1 \pm \bar{X}_2) \sim N\left(\mu_1 \pm \mu_2, \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right).
\]

Conditional Probability:

\[ P(A | B) = \frac{P(A \text{ and } B)}{P(B)} \]

General Multiplication Law:

\[ P(A \text{ and } B) = P(B) \cdot P(A | B) = P(A) \cdot P(B | A) \]

Special Case:

If A and B are independent, then

\[ P(A | B) = P(A) \quad \text{and} \quad P(B | A) = P(B) \]

\[ \therefore P(A \text{ and } B) = P(A) \cdot P(B) \]

Expected Value and Variance for a any discrete random variable X:

\[ X = x_1, x_2, \ldots, x_k \]

\[ P(X = x_i) = p_i, \quad i = 1, 2, \ldots, k \]

The Expected Value of X is:

\[ E(X) = \mu_X = \sum x_i p_i \]

The Variance of X is:

\[ \sigma_X^2 = \sum (x_i - \mu_X)^2 p(x_i) \]

The Standard Deviation of X (Risk) is:

\[ \sigma_X = \sqrt{\text{Var}(X)} \]

Binomial distribution:  (discrete)  Page 207

\[ X \sim \text{Bino}(n, \pi) \]

\[ \pi = \text{probability of a success on any one trial;} \]

\[ 1 - \pi = \text{probability of a failure on any one trial;} \]

\[ P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}, \quad x = 0, 1, 2, \ldots, n \]

where \(\binom{n}{x} = \frac{n!}{x!(n-x)!}\) is the number of combinations & \(n! = n(n-1)(n-2) \ldots (2)(1)\)

For any binomial, \(\mu_{\text{Bino}} = np, \quad s_{\text{Bino}} = \sqrt{np(1-p)}\)  Page 207, 259

Population Parameter  Point Estimator  Margin of Error

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimator</th>
<th>m.o.e. at (1-(\alpha))100% confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean, (\mu)  if (\sigma) is known:  (by the C.L.T.)</td>
<td>(\bar{X})</td>
<td>(Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}})</td>
</tr>
<tr>
<td>Mean, (\mu)  if (\sigma) is unknown:</td>
<td>(\bar{X})</td>
<td>(t_{(\alpha/2, n-1)} \cdot \frac{s}{\sqrt{n}})</td>
</tr>
<tr>
<td>Proportion, (\pi):  Page 327  (\hat{p} = X/n)</td>
<td>(\hat{p})</td>
<td>(Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}})</td>
</tr>
<tr>
<td>Diff. of two means, (\mu_1 - \mu_2):  Page 425  (\bar{X}_1 - \bar{X}_2)</td>
<td>(\bar{X}_1 - \bar{X}_2)</td>
<td>(Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})</td>
</tr>
<tr>
<td>Diff. of two proportions, (\pi_1 - \pi_2):  Page 435  (\hat{p}_1 - \hat{p}_2)</td>
<td>(\hat{p}_1 - \hat{p}_2)</td>
<td>(Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}})</td>
</tr>
<tr>
<td>Slope of regression line, (\beta):  Page 633  (b_1)</td>
<td>(b_1)</td>
<td>(t_{(\alpha/2, n-2)} \cdot \frac{s_{y</td>
</tr>
<tr>
<td>Mean from a regression when (X = \bar{x}):  Page 622</td>
<td>(\hat{y} = a + bx)</td>
<td>(t_{(\alpha/2, n-2)} \cdot \frac{s_{y</td>
</tr>
</tbody>
</table>

Revised Aug '03

Eld Mansfield