Summary

Sample Mean	Sample Proportion	Sample Variance	Sample Standard Deviation	Z-Score
$\bar{X} = \frac{1}{n}\sum x_i$	$\hat{p} = \frac{k}{n}$	$S^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$	$S = \sqrt{S^2}$	$Z = \frac{X - Mean}{SD}$

Probability

General Addition Rule $P(E \cup F) = P(E) + P(F) - P(E \cap F)$	General Multiplication Rule $P(E \cap F) = P(E\|F)P(F)$
Two Independent Events $P(E \cap F) = P(E)P(F)$	Law of Total Prob. $P(E) = P(E\|B_1)P(B_1) + ... + P(E\|B_k)P(B_k)$
Bayes' Rule $P(B_i\|E) = \frac{P(E\|B_i)P(B_i)}{P(E\|B_1)P(B_1)+P(E\|B_2)P(B_2)+...+P(E\|B_k)P(B_k)}$	Complement Rule P(E') = 1 - P(E)
Permutation $P_k^n = \frac{n!}{(n-k)!}$	Combination $C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}$

Random Variables and Probability Distribution

Expected Value (Discrete) $E[X] = \mu = \sum xp(x)$	Variance (Discrete) $\sigma^2 = \sum(x-\mu)^2 p(x)$	Z-score $Z = \frac{X-\mu}{\sigma}$
Expected value (Binomial) $\mu = np$	Variance (Binomial) $\sigma^2 = np(1-p)$	Probability (Binomial) $p(x) = \binom{n}{k}p^k(1-p)^{n-k}$
Expected value (Geometric) $\mu = \frac{1}{p}$	Variance (Geometric) $\sigma^2 = \frac{1-p}{p^2}$	Probability (Geometric) $p(x) = (1-p)^{x-1}p, x = 1, 2, 3, ...$
E[y] = E$[a_1x_1 + a_2x_2 + ... + a_nx_n]$ $\mu_y = a_1\mu_1 + a_2\mu_2 + ... + a_n\mu_n$	Var(y) = Var$(a_1x_1 + a_2x_2 + ... + a_nx_n)$ $\sigma_y^2 = b_1^2\sigma_1^2 + b_2^2\sigma_2^2 + ... + b_n^2\sigma_n^2$; $\sigma_y = \sqrt{Var(y)}$	Constant, a_o E$[a_o] = a_o$; Var$(a_o) = 0$

Probability in density curves

Rule	SD within the mean	Area under the curve
Empirical	1	68%
Empirical	2	95%
Empirical	3	99.7%
Chebyshev	k	$\geq (1 - \frac{1}{k^2})100\%$

Normal Approximation to a Binomial Distribution

X is Binomial	X is Normal
P(a < X < b)	P(a + 0.5 < X < b - 0.5)
P(a $\leq$ X $\leq$ b)	P(a - 0.5 < X < b + 0.5)

Sampling Distribution

Statistic	Mean	Standard Error	Distribution
$\bar{X}$	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	Normal
$\bar{X}$	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{S}{\sqrt{n}}$	T with df = n - 1
$\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	Normal

Parameter Estimation

Parameter	Point estimator	Confidence Interval	Additional
p	$\hat{p}$	$\hat{p} \pm Z_{crit}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$n = p(1-p)(\frac{1.96}{B})^2$ at 95%
μ	$\bar{X}$	$\bar{X} \pm Z_{crit}\frac{\sigma}{\sqrt{n}}$ or $\bar{X} \pm t_{crit}\frac{S}{\sqrt{n}}$	$n = (\frac{1.96\sigma}{B})^2 at 95\%$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$(\hat{p}_1 - \hat{p}_2) \pm Z_{crit}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	.
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$(\bar{x}_1 - \bar{x}_2) \pm t_{crit}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$	df $= \frac{(V_1+V_2)^2}{\frac{V_1^2}{n_1-1}+\frac{V_2^2}{n_2-1}}$; $V_1 = \frac{S_1^2}{n_1}$ and $V_2 = \frac{S_2^2}{n_2}$

Hypothesis Testing

Testing	Test Statistic	σ	Degrees of Freedom	Additional
μ	$Z = \frac{\bar{X}-\mu_o}{\sigma/\sqrt{n}}$	known	.	.
μ	$t = \frac{\bar{X}-\mu_o}{S/\sqrt{n}}$	unknown	n - 1	.
p	$Z = \frac{\hat{p}-p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$	.	.	.
$\mu_1 - \mu_2$	$t = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_{1o}-\mu_{2o})}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$	unknown	$\frac{(V_1+V_2)^2}{\frac{V_1^2}{n_1-1}+\frac{V_2^2}{n_2-1}}$	$V_1 = \frac{S_1^2}{n_1}$ and $V_2 = \frac{S_2^2}{n_2}$
$p_1 - p_2$	$Z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1}+\frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}}$	.	.	$\hat{p}_c = \frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$
Goodness of Fit	$X^2 = \sum \frac{(O-E)^2}{E}$	.	k - 1	$E = np_{io}$
Homogeneity or Independence	$X^2 = \sum_i \sum_j \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$	.	(I - 1)(J - 1)	$E_{ij} = \frac{n_{i.}n_{.j}}{n}$
.	$X^2 = \sum_i \sum_j \frac{n_{ij}^2}{E_{ij}} - n$	.	(I - 1)(J - 1)	$E_{ij} = \frac{n_{i.}n_{.j}}{n}$

Simple Linear Regression: Inferential Methods

S_e	S_b	S_{xx}	Conf. Interval of b	Test Statistic for b
$\sqrt{\frac{SSE}{n-2}}$	$\frac{S_e}{\sqrt{S_{xx}}}$	$\sum(x-\bar{x})^2$	$b \pm t_{crit}S_b$	$\frac{b-\beta_o}{S_b}$

Summarizing Bivariate Data

Correlation	y-intercept	slope	Sum Sq Error	Sum Sq Total
$r = \frac{\sum Z_x Z_y}{n-1}$	$a = \bar{Y} - b\bar{X}$	$b = \frac{\sum(x-\bar{X})(y-\bar{Y})}{\sum(x-\bar{X})^2}$	$SSE = \sum(y-\hat{y})^2$	$SSTo = \sum(y-\bar{Y})^2$
Coef. Determination	**Standard Error**			
$r^2 = 1 - \frac{SSE}{SSTo}$	$S_e = \sqrt{\frac{SSE}{n-2}}$			

ANOVA

Source	df	SS	MS	F
Treatment	k - 1	$\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$	$\frac{SST}{k-1}$	$\frac{MST}{MSE}$
Error	n - k	$\sum_{i=1}^{k}(n_i - 1)S_i^2$	$\frac{SSE}{n-k}$	.
Total	n - 1	SST + SSE	.	.