MATH 58 / 58B — A VERY BRIEF SUMMARY

Parameter	**Statistic**	**Standard Error**	**Distribution**	**Assumptions**
p	$\hat{p}$	$\sqrt{p_0(1-p_0)/n}$ (HT) $\sqrt{\hat{p}(1-\hat{p})/n}$ (CI)	Z Z	successes & failures ≥ 10 p_0 is the number in H_0
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$ (2 independent samples)	$\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}$ (HT) $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ (CI)	Z Z	successes & failures ≥ 10 in each group
$\ln(RR) = \ln(p_1/p_2)$ take inverse ln (e) for CI	$\ln(\hat{p}_1/\hat{p}_2)$	$\sqrt{\frac{1}{a} - \frac{1}{a+c} + \frac{1}{b} - \frac{1}{b+d}}$	Z	large samples (Math 58B only)
$\ln(OR)$ take inverse ln (e) for CI	$\ln(\hat{OR})$	$\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$	Z	large samples (Math 58B only)
μ	$\overline{x}$	$\sigma/\sqrt{n}$ (σ known) $s/\sqrt{n}$ (σ unknown)	Z $t, \mathrm{df} = n-1$	$n \geq 30$ or normal
$\mu_1 - \mu_2$	$\overline{x}_1 - \overline{x}_2$ (2 independent samples)	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ ($\sigma_1 \neq \sigma_2$) $\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}$ ($\sigma_1 = \sigma_2$)	t, df $\approx \min\{n_1, n_2\} - 1$ t, df $= n_1 + n_2 - 2$	$n_1, n_2 \geq 30$ or normal
μ_d	$\overline{x}_d$ (matched pairs)	$s_d/\sqrt{n}$ ($n = \#$pairs)	t, df $= n-1$	$n \geq 30$ or normal
X_{n+1}	$\overline{x}$	$\sqrt{s^2 + \frac{s^2}{n}}$	t, df $= n-1$	normal

To find the p-value of a HT, look up the score $\dfrac{\text{statistic} - \text{hypothesized value}}{\text{SE}}$ on the specified distribution in the direction of H_A.

A $(100-\alpha)\%$ CI is of the form statistic $\pm$ multiplier $\times$ SE. The multiplier is the $(100-\alpha/2)\%$ point from the specified distribution.

Pooled proportion (under the assumption $p_1 = p_2$): $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$.

Pooled standard error (under the assumption $\sigma_1 = \sigma_2$): $s_p = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$

Inference for two variables (explanatory variable not necessarily binary)

Two categorical variables (χ^2 test). χ^2-statistic $= \displaystyle\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(\text{observed}_{ij}-\text{expected}_{ij})^2}{\text{expected}_{ij}}$, where $\text{expected}_{ij} = \dfrac{(\text{total of row } i)\times(\text{total of column } j)}{\text{grand total}}$.
Under the null hypothesis assumption assumption, the statistic follows a χ^2 distribution with $(r-1)(c-1)$ degrees of freedom.

Categorical explanatory, quantitative response variables (ANOVA). F-statistic$=\dfrac{\text{MST}}{\text{MSE}}$, where $\text{MST} = \dfrac{n_1(\bar{x}_1-\bar{x})^2+\cdots+n_I(\bar{x}_I-\bar{x})^2}{I-1}$
and $\text{MSE} = \dfrac{(n_1-1)s_1^2+\cdots+(n_I-1)s_I^2}{N-I}$. Under the null hypothesis assumption, the statistic follows an F distribution with $(I-1, N-I)$ degrees of freedom.

Two quantitative variables (Simple Linear Regression).

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i-\bar{x})(y_i-\bar{y})}{s_x s_y}, \qquad b_1 = r\frac{s_y}{s_x}, \qquad b_0 = \bar{y}-b_1\bar{x}, \qquad \hat{y} = b_0+b_1x, \qquad SE(b_1) = \frac{s}{s_x}\frac{1}{\sqrt{n-2}}.$$

Under the null hypothesis $H_0 : \beta_1 = 0$, the standardized slope statistic $\dfrac{b_1-0}{SE(b_1)}$ follows a t distribution with $n-2$ degrees of freedom.