Ivan Smirnov

Statistics Study Sheet

:: Key Terms ::

- statistic – numerical descriptive of sample
- parameter – numerical descriptive of population
- expected value – what we expect the mean to be
- confidence interval – the # in $\mu \pm$ # that a certain percentage of data is certain to fall within this range
- margin of error – general formula (invNorm * (0.5 + {CONFIDENCE INTERVAL / 2}) x ($\mu / \sqrt{n}$)

:: Symbols ::

n – number of trials
p – probability of success
q – probability of failure (used in binomial probabilities)
r – # of successes (used in binomial probabilities)
$\bar{x}$ – mean (statistic)
μ – mean (parameter)
s^2 – variance (statistic)
σ^2 – variance (parameter)
s – standard deviation (statistic)
σ – standard deviation (parameter)
$\hat{p}$ – proportion (statistic)
p – proportion (parameter)

:: Methods ::

A. one list of data – use List → 1-var stats. You can get the mean, median, mode, quarters, and std-dev
B. two lists of related data (x,y) – use 1-var stats, but pass L_1, L_2 as the argument
C. Binomial Distribution – use binomcdf(n,p,r)
    $\sqrt{(npq)} = \sigma$ and $\mu = np$
D. Normal distribution $\mu=0$, and $\sigma=1$
    use z-values to manipulate data
    invNorm(z-score, μ, σ)= x-value
E. Distribution of means. If not normal or unknown, n > 30 must be true
    new σ = (regular σ)/($\sqrt{n}$)
F. Matrix (a chart) use X^2 test (chi-squared test)
G. Two lists – use ANOVA test

:: Problem Solving Checklist ::

1. Which formula should you use? (Consult study sheets)
2. are you finding t-value or z-score?
3. Is the test one tailed or two tailed?

:: Syntax Help ::

normalpdf(#)-gives the probability of that EXACT number occurring
normalcdf(lower limit, upper limit, mean, std-dev)
invNorm(are on the left, mean, sigma)
binompdf(n,p,r) – probability of r occurring
binomcdf(n,p,r) – sum of probabilities from 0 to r
tcdf(lower, upper, degrees of freedom)
invT(area, degrees of freedom)

*NOTE – Don't blindly trust the calculator, always check to make sure the answer makes sense!

:: Extra ::

Chebyshev formula gives us % of data within k std-devs IFF k>1
% data = $1 - (1/k^2)$
Note – a number greater than 2.5std-devs away from the mean is a rare event, an outlier

Normal distribution
68% within 1σ
95% within 2σ
99.7 within 3σ

Confidence Interval Formulas

All confidence intervals are of the form

Point Estimate $\pm$ E where E = Maximal Margin of Error

Parameter	Point Estimate	E	Assumptions
μ	$\bar{x}$	$z_c \sigma_{\bar{x}}$	σ known $\bar{x}$'s normal
μ	$\bar{x}$	$t_c \cdot \frac{s}{\sqrt{n}}$ $df = n - 1$	σ unknown $\bar{x}$'s normal
p	$\hat{p} = \frac{r}{n}$	$z_c \sqrt{\frac{\hat{p}\hat{q}}{n}}$	$n\hat{p} > 5$ $n\hat{q} > 5$
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	σ_1, σ_2 known $\bar{x}_1 - \bar{x}_2$'s normal IRS
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$t_c \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $df = \mathrm{MIN}(n_1 - 1, n_2 - 1)$	σ_1, σ_2 unknown and assumed unequal $\bar{x}_1 - \bar{x}_2$'s normal IRS
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$t_c s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ $s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$ $df = n_1 + n_2 - 2$	σ_1, σ_2 unknown and assumed equal $\bar{x}_1 - \bar{x}_2$'s normal IRS
μ_d	$\bar{d}$	$t_c \cdot \frac{s_d}{\sqrt{n}}$ $df = n - 1$	d's normal DRS
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$z_c \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$	$n_1\hat{p}_1 > 5$ $n_1\hat{q}_1 > 5$ $n_2\hat{p}_2 > 5$ $n_2\hat{q}_2 > 5$

Hypothesis Tests

All test statistics are of the form $\dfrac{\text{Point Estimate - Parameter}}{\text{Standard Error}}$

Parameter	Point Estimate	Test Statistic	Assumptions
μ	$\bar{x}$	$z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}}$	σ known $\bar{x}$'s normal
μ	$\bar{x}$	$t = \dfrac{\bar{x} - \mu}{s / \sqrt{n}}$ $df = n - 1$	σ unknown $\bar{x}$'s normal
p	$\hat{p} = \dfrac{r}{n}$	$z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$	$np > 5$ $nq > 5$
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	σ_1, σ_2 known $\bar{x}_1 - \bar{x}_2$'s normal IRS
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ $df = \text{MIN}(n_1 - 1, n_2 - 1)$	σ_1, σ_2 unknown and assumed unequal $\bar{x}_1 - \bar{x}_2$'s normal IRS
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ $s = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$ $df = n_1 + n_2 - 2$	σ_1, σ_2 unknown and assumed equal $\bar{x}_1 - \bar{x}_2$'s normal IRS
μ_d	$\bar{d}$	$t = \dfrac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$ $df = n - 1$	d's normal DRS
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\bar{p}\bar{q}}{n_1} + \dfrac{\bar{p}\bar{q}}{n_2}}}$ $\bar{p} = \dfrac{r_1 + r_2}{n_1 + n_2}$	$n_1\bar{p} > 5$ $n_1\bar{q} > 5$ $n_2\bar{p} > 5$ $n_2\bar{q} > 5$