

Formulas and Terms in Math 102 / Core 143

In a histogram,

of scores (or total probability of outcomes) in an interval
= area under the histogram above that interval

For a list of n scores (x -values):

Average $\bar{x} = \mu = \frac{\sum x}{n}$, Standard deviation $\sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$

N -th percentile = x -value for which $N\%$ of scores are $\leq x$

Median = 50-th percentile, $IQR = 75$ -th – 25-th percentile

Converting a score x to standard units (z -value or t -value):

$$z \text{ or } t = \frac{x - \bar{x}}{\sigma} \quad \text{or} \quad \frac{x - EV}{SE} \quad (\text{see below})$$

Correlation: $r = \frac{\sum(z_x z_y)}{n}$ (where z_q means q in standard units)

Regression line of y on x (for predicting y from x , or for estimating average y within a vertical strip at x):

$$y - \bar{y} = r \left(\frac{\sigma_y}{\sigma_x} \right) (x - \bar{x})$$

Root Mean Square (RMS) error for regression of y on x (= approximate standard deviation of data in any vertical strip, if scatter diagram is homoscedastic): $\sigma_y \sqrt{1 - r^2}$

Probability:

Multiplication rule: $\text{Prob}(A \text{ and } B) = \text{Prob}(A) \cdot \text{Prob}(B \text{ given } A)$

Independent events: $\text{Prob}(B \text{ given } A) = \text{Prob}(B)$

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

Mutually exclusive events: $\text{Prob}(A \text{ and } B) = 0$

Binomial probabilities: If an event has probability p on each trial, the probability of its occurring exactly k times in n independent trials:

$$C(n, k) \cdot p^k \cdot (1 - p)^{n-k}, \quad \text{where} \quad C(n, k) = \frac{n!}{k!(n - k)!}$$

For a sample of size n from a population's scores with average μ and standard deviation σ :

EV of sum of sample = $n\mu$

SE of sum = $\sigma \cdot \sqrt{n}$

EV of average of sample = μ

SE of average = σ/\sqrt{n}

Special case: Population is 0's and 1's (for yeses and nos, or men and women, or ...), fraction of 1's is p , for a sample of size n :

EV of count = np

SE of count = $\sqrt{p(1 - p)} \cdot \sqrt{n}$

EV of % (or proportion) = p

SE of % (or proportion) = $\sqrt{p(1 - p)}/\sqrt{n}$

$k\%$ confidence interval for the average of a population:

$$\bar{x} \pm (z \text{ or } t) \cdot (SE \text{ for average})$$

where z or t comes from table with $k\%$ of area between $-z$ and z or between $-t$ and t . (Similar for “proportion” in place of “average” — but it fails if sample proportion is so near 0% or 100% that interval includes these values.)

For z -tests of significance (“Sample average is different from what theory predicts; is it different enough to believe that population average is different, i.e., that theory is wrong?”), value for μ usually comes from null hypothesis; approximate (bootstrap) population σ with sample standard deviation. Or, if $n < 30$, find t -value, approximating σ with $s = SD^+$ found by multiplying the usual formula by $\sqrt{n/(n-1)}$, and use t -table with $n-1$ degrees of freedom.

For significance test of whether μ 's in two populations are equal: Null hypothesis is they are, so EV of difference of averages of samples of each is 0;

$$SE \text{ for difference of sample averages} = \sqrt{(SE \text{ of first})^2 + (SE \text{ of second})^2}$$

(For more than 2 populations, use one-way ANOVA.)

[Special case, not in Friedman et al.: For significance test of difference of proportions in two populations:

$$\text{Pooled estimate for common population proportion is } \hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2},$$

$$\text{and } SE \text{ for difference} = \sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}.]$$

For frequency distributions, deciding whether differences in proportions of various categories (represented in columns), either between the rows of a table or between a given row of frequencies and some expected row: Null hypothesis is that there is no difference in proportions between rows [so that table of “expected” frequencies is given by (row sum)·(column sum)/(grand total)] or from the expected row; and

$$\chi^2 = \sum [(\text{observed} - \text{expected})^2 / \text{expected}]$$

degrees of freedom: in table distributions, $(\# \text{ of rows} - 1) \cdot (\# \text{ of columns} - 1)$

in list distributions, $\# \text{ in list} - 1$;