# Formula sheet 2

## Sample statistics

Sample mean:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Sample variance:

$$s_x^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right)$$

## Linear transformations

Mean:

$$E(aX+b) = aE(X) + b$$

Variance:

$$V(aX+b) = a^2 V(X)$$

## Linear combinations

Mean:

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

Variance:

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{Cov}(X_i, X_j)$$

Central Limit Theorem: $n > 30$

$$\bar{X} \overset{\text{approx.}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

## Binomial:

$$p(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x},\ x \in \{0, 1, \ldots, n\}$$

with $E(X) = np, V(X) = np(1-p)$

## One sample inference

Two sided C.I. (mean):

$$\bar{x} \mp t_{(1-\alpha/2, n-1)} \frac{s}{\sqrt{n}}$$

Sample size:

$$n \geq \left\lceil \left(2z_{1-\alpha/2}\frac{\sigma}{\text{width}}\right)^2 \right\rceil$$

Test statistic:

$$T.S. = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \overset{\mathrm{H}_0}{\sim} t_{n-1}$$

Two sided C.I. (proportion):

$$\tilde{p} \mp z_{1-\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}},$$

where $\tilde{n} := n + 4$, and $\tilde{p} := (x+2)/\tilde{n}$

Test statistic:

$$T.S. = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{\mathrm{H}_0}{\sim} N(0, 1)$$

Two sided C.I. (variance):

$$\left(\frac{(n-1)s^2}{\chi^2_{(1-\alpha/2, n-1)}}, \frac{(n-1)s^2}{\chi^2_{(\alpha/2, n-1)}}\right)$$

Test statistic:

$$T.S. = \frac{(n-1)s^2}{\sigma_0^2} \overset{\mathrm{H}_0}{\sim} \chi^2_{n-1}$$

Sign test statistic:
$B =$ number of observations strictly greater than the hypothesized null $p$ percentile value

$$B \overset{\mathrm{H}_0}{\sim} \mathrm{Bin}(n, 1-p)$$

Wilcoxon signed-rank test:

1. Calculate the differences $d_i = x_i - \mu_0$ for each observation, i.e. center data according to $\mathrm{H}_0$.
2. Discard any $d_i = 0$ (as long as they are not more that 10% of the data, otherwise research "adjusted" methods).
3. Rank $|d_i|$ (i.e. rank ignoring the sign).
4. Calculate the test statistic $S_+ =$ sum of the ranks corresponding to positive $d_i$'s.

## Two sample inference

Two sided C.I. (means):

$$\bar{x} - \bar{y} \mp t_{1-\alpha/2,\nu}\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$

where

$$\nu = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{(s_X^2/n_X)^2}{n_X-1} + \frac{(s_Y^2/n_Y)^2}{n_Y-1}}$$

Test statistic:

$$T.S. = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}} \overset{\mathrm{H}_0}{\sim} t_\nu$$

If $\sigma_X^2 = \sigma_Y^2$, then replace $s_X^2$ and $s_Y^2$ with

$$s_p^2 = \left(\frac{(n_X - 1)}{n_X + n_Y - 2}\right) s_X^2 + \left(\frac{(n_Y - 1)}{n_X + n_Y - 2}\right) s_Y^2$$

and $\nu = n_X + n_Y - 2$.

Two sided C.I. (proportions):

$$\tilde{p}_X - \tilde{p}_Y \mp z_{1-\alpha/2}\sqrt{\frac{\tilde{p}_X(1-\tilde{p}_X)}{\tilde{n}_X} + \frac{\tilde{p}_Y(1-\tilde{p}_Y)}{\tilde{n}_Y}}$$

where $\tilde{n}_X = n_X + 2$ and $\tilde{p}_X = (x+1)/\tilde{n}_X$.

Test statistic:

$$T.S. = \frac{\hat{p}_X - \hat{p}_Y - \Delta_0}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}} \overset{\mathrm{H}_0}{\sim} N(0,1)$$

Two sided C.I. (ratio of variances):

$$\left(\frac{s_X^2}{s_Y^2}\frac{1}{F_{(1-\alpha/2;n_X-1,n_Y-1)}}, \frac{s_X^2}{s_Y^2}\frac{1}{F_{(\alpha/2;n_X-1,n_Y-1)}}\right)$$

Test statistic:

$$T.S. = \frac{\frac{s_X^2}{s_Y^2}}{\Delta_0} \overset{\mathrm{H}_0}{\sim} F_{n_X-1,n_Y-1}$$

Wilcoxon rank-sum test:

1. rank all the data irrespective of sample
2. calculate the sum of the ranks associated with the smallest sample

Wilcoxon signed-rank test:

1. calculating the differences $d_i = (x_i - y_i) - \Delta_0$,
2. proceed from step 2 in the Wilcoxon signed-rank one population setting.

Levene's test:

$$T.S. \overset{\mathrm{H}_0}{\sim} F_{t-1,N-t}$$

## Contigency tables

Expected value:

$$E_{ij} = \frac{n_{i+}n_{+j}}{n}$$

Test statistic:

$$T.S. = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij} - E_{ij})^2}{E_{ij}} \overset{\mathrm{H}_0}{\sim} \chi^2_{(r-1)(c-1)}$$

Standardized residuals:

$$r_{ij}^\star = \frac{n_{ij} - E_{ij}}{\sqrt{E_{ij}(1-p_{i+})(1-p_{+j})}}$$