

## Formulas and tables



Can be used at the exam but Students should use their own copy and no other material!

## Rules for rounding

Suggestion: use 1 or 2 decimals for averages and probabilities, 2 or 3 decimals for std

If the decimal number after the position we want to round to is:

<5 → truncate to floor

>5 → round to the next number (floor + 1)

=5 → look one number further until a number ≠5 is found and apply the same rules

Examples:

| Original value | →1 decimal | →2 decimals |
|----------------|------------|-------------|
| 12.422         | 12.4       | 12.42       |
| 11.237         | 11.2       | 11.24       |
| 10.251         | 10.2       | 10.25       |
| 10.257         | 10.3       | 10.26       |
| 14             | 14.0       | 14.00       |

## Descriptive Statistics (i)

Arithmetic mean – for frequency tables - weighted

$$\bar{x} = \frac{\sum x_i}{n} \quad \bar{x} = \frac{\sum_{i=1}^k x_i n_i}{n} \quad \bar{x}_p = \frac{\sum_{j=1}^K x_j p_j}{\sum_{j=1}^k p_j}$$

$$\text{density} = \frac{\text{frequency}}{\text{width}}$$

$$\Leftrightarrow \text{frequency} = \text{width} \times \text{density}$$

$$n \text{ odds} \rightarrow \text{rank}(\text{Median}) = \frac{n+1}{2}$$

$$n \text{ even} \rightarrow \text{rank}(\text{Median}) = \frac{n}{2}, \frac{n}{2} + 1$$

$$n \text{ odds} \rightarrow \text{rank}(Q1) = \frac{n+1}{4} \quad \text{rank}(Q3) = 3 \frac{n+1}{4}$$

$$n \text{ even} \rightarrow \text{rank}(Q1) = \frac{n}{4}, \frac{n}{4} + 1 \quad \text{rank}(Q3) = 3 \frac{n}{4}, 3 \frac{n}{4} + 1$$

## Descriptive Statistics (ii)

Variance

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \left( \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 \right) \cdot \frac{n}{n-1}$$

st.dev. =  $\sqrt{\text{variance}}$

$$\text{variation coeff} = \frac{\text{st.dev.}}{\bar{x}} \cdot (100)$$

Variance for frequency tables

$$\frac{\sum_{i=1}^k (x_i - \bar{x})^2 n_i}{n-1}$$

Range=Max-Min

$$\text{IQR} = Q3 - Q1 = 31 - 17 = 14$$

Properties of Normal  $N(\mu, \sigma)$

$$\Pr(\mu - \sigma < X < \mu + \sigma) = 68\%$$

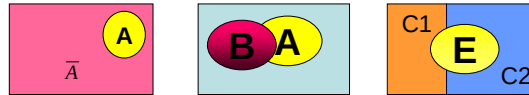
$$\Pr(\mu - 2\sigma < X < \mu + 2\sigma) = 95\%$$

$$\Pr(\mu - 3\sigma < X < \mu + 3\sigma) = 99.7\%$$

$$Q_1 = \mu - 0.67 \cdot \sigma$$

$$Q_3 = \mu + 0.67 \cdot \sigma$$

## Basic rules for computing probabilities



$$p(\bar{A}) = 1 - p(A)$$

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

$$p(A \cap B) = p(A | B) \cdot p(B) \\ = p(B | A) \cdot p(A)$$

$$p(A | B) = \frac{p(A \cap B)}{p(B)}$$

$$p(A \cap B) \stackrel{\text{independent}}{=} p(B) \cdot p(A)$$

Bayes Formula

$$p(C1 | E) = \frac{p(E | C1) \cdot p(C1)}{p(E | C1) \cdot p(C1) + p(E | C2) \cdot p(C2)} \quad \text{Being } C1 \cup C2 = \Omega$$

## Computing probabilities using the Binomial and Poisson distributions

prob(success)= $\pi$ , X = number of successes in N trials

$$p(X = x) = \binom{N}{x} \pi^x (1 - \pi)^{N-x}$$

The expected (mean) number of successes is  $N \cdot \pi$

prob(success) $\approx 0$  i.e. rate of success =  $\lambda$ , X = number of successes in  $N \gg 0$  trials or along a continuous interval of follow-up, or area

$$p(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

The expected (mean) number of successes is  $\lambda = N \cdot \pi$

With large N it is also possible to approximate the Binomial with a Normal distribution with mean= $N \cdot \pi$  and variance= $N \cdot \pi(1 - \pi)$

$$\binom{N}{k} = \frac{N!}{(N-k)! \cdot k!} \\ k! = k \cdot (k-1) \cdot (k-2) \cdot \dots \cdot 2 \cdot 1 \\ 5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \\ 0! = 1 \\ e^{-\lambda} = \frac{1}{e^{\lambda}} \\ k^0 = 1$$

## Computing probabilities using the Normal distribution

$$\Pr(X \leq x) = \text{Area}_{N(\mu, \sigma^2)}(X \leq x) = \text{Area}_{N(0,1)}\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

|                                                                                                                      |                                                  |                                                   |
|----------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|---------------------------------------------------|
| $X_i \xrightarrow{\text{standardization}} Z_i$<br>Values on the original scale, with parameters $\mu$ and $\sigma^2$ | Standardization:<br>$Z = \frac{X - \mu}{\sigma}$ | Inverse transform :<br>$X = \mu + \sigma \cdot Z$ |
|----------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|---------------------------------------------------|

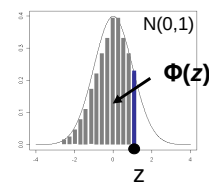
For  $z > 0$  we read  $\Phi(z)$  on the table here given (watch out, other textbooks might return areas with different shapes)

Areas with shape different than this one are obtained using the symmetry property of the curve, and the fact that total area = 1.

$$\Phi(0) = 0.5 \quad \Phi(+\infty) = 1$$

$$\text{Area}(Z < -z) = \text{Area}(Z > z)$$

$$\Phi(-z) = 1 - \Phi(z)$$



## (1-α) % level CI for the mean μ

Hypotheses: in the population the phenomenon is distributed according a Normal with mean  $\mu$  (unknown) and variance  $\sigma^2$  known

The formula is valid also if the population is not Normal or the variance of the population  $\sigma^2$  is not known\* IF the sample is large ( $n \geq 30$ )

\*(replace  $\sigma^2$  with the variance estimated from the sample)

$$\text{Formula: } \left( \bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

Quantile:  $N(0,1)$

| $1 - \alpha$ | $z_{\frac{\alpha}{2}}$ |
|--------------|------------------------|
| 90%          | 1.64                   |
| 95%          | 1.96                   |
| 99%          | 2.58                   |

Notice: when the variance  $\sigma^2$  is not known and the sample size is small the quantiles are taken from the Student T distribution with  $(n-1)$  degrees of freedom

## Z- or T-test on one mean

$$H_0: \mu = \mu_0$$

Hypotheses: in the population the phenomenon is distributed according a Normal with mean  $\mu$  (unknown) and variance  $\sigma^2$  known

The formula is valid also if the population is not Normal or the variance of the population  $\sigma^2$  is not known\* **IF** the sample is large ( $n \geq 30$ )

\*(replace  $\sigma^2$  with the variance estimated from the sample)

Test statistic:

the sample mean, standardized:

$$t = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

Compute the p-value on the table of  $N(0,1)$

Or

Fix  $\alpha$  and compare t with the threshold(s) of the rejection region delimited by the quantile(s):

$$\begin{array}{ll} H_1 & z \\ \mu > \mu_0 (\mu < \mu_0) & z_\alpha (-z_\alpha) \\ \mu \neq \mu_0 & \pm z_{\frac{\alpha}{2}} \end{array}$$

Notice: when the variance  $\sigma^2$  is not known and the sample is small the quantiles are taken from the Student T distribution with  $(n-1)$  degrees of freedom

## Z- or T-test to compare 2 means (indep. samples)

$$H_0: \mu_1 = \mu_2 \text{ i.e. } H_0: \delta = \mu_1 - \mu_2 = 0$$

Hypotheses: the two groups come from independent populations with mean respectively  $\mu_1$  e  $\mu_2$  (unknown) and the same unknown variance  $\sigma^2$ ; both samples are large ( $n_1, n_2 > 30$ )

Compute:  $s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$

Test statistic : the difference of the two sample means, standardized:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Compute the p-value on the table of  $N(0,1)$

Or:

Fix  $\alpha$  and compare t with the threshold(s) of the rejection region:

$$\begin{array}{ll} H_1 & z \\ \delta > 0 (\delta < 0) & z_\alpha (-z_\alpha) \\ \delta \neq 0 & \pm z_{\frac{\alpha}{2}} \end{array}$$

Notice: when the samples are small the quantiles are taken from the Student T distribution with  $(n_1 + n_2 - 1)$  degrees of freedom

## (1-α) % level CI and test for the prob. π

Hypothesis: the sample is large (n ≥ 30)

Formula (1-α)% CI:

$$\left( \hat{\pi} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}, \hat{\pi} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \right)$$

Quantile: N(0,1)

|     |                        |
|-----|------------------------|
| 1-α | $z_{\frac{\alpha}{2}}$ |
| 90% | 1.64                   |
| 95% | 1.96                   |
| 99% | 2.58                   |

Test statistic for  $H_0: \pi = \pi_0$ : the sample proportion, standardized:

$$t = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}}$$

Compute the p-value on the table of N(0,1)  
Or:

Fix α and compare t with the threshold of the rejection region delimited by the quantile(s) from N(0,1)

$$\begin{array}{l} H_1 \\ \pi > \pi_0 \quad (\pi < \pi_0) \quad z_{\alpha} \quad (-z_{\alpha}) \\ \pi \neq \pi_0 \quad \pm z_{\frac{\alpha}{2}} \end{array}$$

## Chi-Squared test for the association between two qualitative variables (cross-tab. rxc)

$H_0$ : X (r modalities) and Y (c modalities) independent vs  $H_1$ : X, Y associated

Including: comparison of two probabilities, by constructing the 2x2 cross-table

The test is valid if in the table all expected frequencies are >5

Compute the expected frequencies under the hypothesis of independence  $H_0$ :

$$\text{Expected} = \frac{\text{tot row} \cdot \text{tot column}}{\text{general tot}}$$

Test statistic :

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

Fix α and compare the test statistic with the threshold from a Chi-Squared table for (r-1)(c-1) degrees of freedom (df)

| df | α=0.1  | α=0.05 | α=0.01 |
|----|--------|--------|--------|
| 1  | 2.706  | 3.841  | 6.635  |
| 2  | 4.605  | 5.991  | 9.21   |
| 3  | 6.251  | 7.815  | 11.345 |
| 4  | 7.779  | 9.488  | 13.277 |
| 5  | 9.236  | 11.07  | 15.086 |
| 6  | 10.645 | 12.592 | 16.812 |

## Measures of association of quantitative variables

Covariance:

$$\text{cov}_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum_{i=1}^n x_i y_i}{n} - \bar{x}\bar{y}$$

Linear Correlation Coefficient :

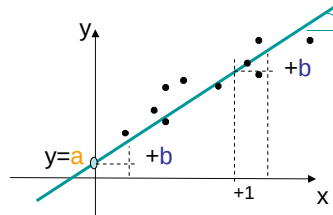
$$r = \frac{\text{cov}_{xy}}{\text{std}_x \cdot \text{std}_y}$$

Regression line:

$$y = a + b \cdot x$$

$$b = \frac{\text{cov}_{xy}}{\text{var}_x} = r_{xy} \frac{\text{std}_y}{\text{std}_x}$$

$$a = \bar{y} - b \cdot \bar{x}$$



## Test on the slope of a regression line

$H_0: \beta = 0$  i.e. no association between X and Y

Compute the slope b and its standard error SE(b):

$$s_{RES}^2 = \frac{\sum (y - Y)^2}{n - 2} \quad SE(b) = \frac{s_{RES}}{\sqrt{\sum (x_i - \bar{x})^2}}$$

Test statistic:

the sample slope, standardized:

$$t = \frac{b}{SE(b)}$$

Compute the p-value on the table of N(0,1) [only n large]

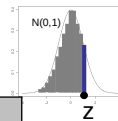
Or

Fix  $\alpha$  and compare t with the threshold(s) of the rejection region delimited by the quantile(s):

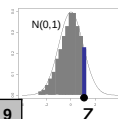
$$\begin{array}{ll} H_1 & z \\ \beta > 0 (\beta < 0) & z_\alpha (-z_\alpha) \\ \beta \neq 0 & \pm z_{\frac{\alpha}{2}} \end{array}$$

Notice: when the sample is small the quantiles are taken from the Student T distribution with (n-2) degrees of freedom

| Table of Normal (0,1), z from 0 to 1.99 |       |       |       |       |       |       |       |       |       |       |
|-----------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| z                                       | 0     | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |
| 0.0                                     | 0.500 | 0.504 | 0.508 | 0.512 | 0.516 | 0.520 | 0.524 | 0.528 | 0.532 | 0.536 |
| 0.1                                     | 0.540 | 0.544 | 0.548 | 0.552 | 0.556 | 0.560 | 0.564 | 0.567 | 0.571 | 0.575 |
| 0.2                                     | 0.579 | 0.583 | 0.587 | 0.591 | 0.595 | 0.599 | 0.603 | 0.606 | 0.610 | 0.614 |
| 0.3                                     | 0.618 | 0.622 | 0.626 | 0.629 | 0.633 | 0.637 | 0.641 | 0.644 | 0.648 | 0.652 |
| 0.4                                     | 0.655 | 0.659 | 0.663 | 0.666 | 0.670 | 0.674 | 0.677 | 0.681 | 0.684 | 0.688 |
| 0.5                                     | 0.691 | 0.695 | 0.698 | 0.702 | 0.705 | 0.709 | 0.712 | 0.716 | 0.719 | 0.722 |
| 0.6                                     | 0.726 | 0.729 | 0.732 | 0.736 | 0.739 | 0.742 | 0.745 | 0.749 | 0.752 | 0.755 |
| 0.7                                     | 0.758 | 0.761 | 0.764 | 0.767 | 0.770 | 0.773 | 0.776 | 0.779 | 0.782 | 0.785 |
| 0.8                                     | 0.788 | 0.791 | 0.794 | 0.797 | 0.800 | 0.802 | 0.805 | 0.808 | 0.811 | 0.813 |
| 0.9                                     | 0.816 | 0.819 | 0.821 | 0.824 | 0.826 | 0.829 | 0.831 | 0.834 | 0.836 | 0.839 |
| 1.0                                     | 0.841 | 0.844 | 0.846 | 0.848 | 0.851 | 0.853 | 0.855 | 0.858 | 0.860 | 0.862 |
| 1.1                                     | 0.864 | 0.867 | 0.869 | 0.871 | 0.873 | 0.875 | 0.877 | 0.879 | 0.881 | 0.883 |
| 1.2                                     | 0.885 | 0.887 | 0.889 | 0.891 | 0.893 | 0.894 | 0.896 | 0.898 | 0.900 | 0.901 |
| 1.3                                     | 0.903 | 0.905 | 0.907 | 0.908 | 0.910 | 0.911 | 0.913 | 0.915 | 0.916 | 0.918 |
| 1.4                                     | 0.919 | 0.921 | 0.922 | 0.924 | 0.925 | 0.926 | 0.928 | 0.929 | 0.931 | 0.932 |
| 1.5                                     | 0.933 | 0.934 | 0.936 | 0.937 | 0.938 | 0.939 | 0.941 | 0.942 | 0.943 | 0.944 |
| 1.6                                     | 0.945 | 0.946 | 0.947 | 0.948 | 0.949 | 0.951 | 0.952 | 0.953 | 0.954 | 0.954 |
| 1.7                                     | 0.955 | 0.956 | 0.957 | 0.958 | 0.959 | 0.960 | 0.961 | 0.962 | 0.962 | 0.963 |
| 1.8                                     | 0.964 | 0.965 | 0.966 | 0.966 | 0.967 | 0.968 | 0.969 | 0.969 | 0.970 | 0.971 |
| 1.9                                     | 0.971 | 0.972 | 0.973 | 0.973 | 0.974 | 0.974 | 0.975 | 0.976 | 0.976 | 0.977 |



| Table of Normal (0,1), z from 2.00 to 3.99 |       |       |       |       |       |       |       |       |       |       |
|--------------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| z                                          | 0     | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |
| 2.0                                        | 0.977 | 0.978 | 0.978 | 0.979 | 0.979 | 0.980 | 0.980 | 0.981 | 0.981 | 0.982 |
| 2.1                                        | 0.982 | 0.983 | 0.983 | 0.983 | 0.984 | 0.984 | 0.985 | 0.985 | 0.985 | 0.986 |
| 2.2                                        | 0.986 | 0.986 | 0.987 | 0.987 | 0.987 | 0.988 | 0.988 | 0.988 | 0.989 | 0.989 |
| 2.3                                        | 0.989 | 0.990 | 0.990 | 0.990 | 0.990 | 0.991 | 0.991 | 0.991 | 0.991 | 0.992 |
| 2.4                                        | 0.992 | 0.992 | 0.992 | 0.992 | 0.993 | 0.993 | 0.993 | 0.993 | 0.993 | 0.994 |
| 2.5                                        | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 | 0.995 | 0.995 | 0.995 | 0.995 | 0.995 |
| 2.6                                        | 0.995 | 0.995 | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 | 0.996 |
| 2.7                                        | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 |
| 2.8                                        | 0.997 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 |
| 2.9                                        | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.999 | 0.999 | 0.999 |
| 3.0                                        | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 |
| 3.1                                        | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 |
| 3.2                                        | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 |
| 3.3                                        | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 3.4                                        | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 3.5                                        | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 3.6                                        | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 3.7                                        | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 3.8                                        | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 3.9                                        | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |



## Student T table of quantiles; quantiles of $N(0,1)$

| $\alpha$ | 0.1           | 0.05          | 0.01          | $\alpha$ | 0.1           | 0.05          | 0.01          |
|----------|---------------|---------------|---------------|----------|---------------|---------------|---------------|
| df       | $t(\alpha/2)$ | $t(\alpha/2)$ | $t(\alpha/2)$ | df       | $t(\alpha/2)$ | $t(\alpha/2)$ | $t(\alpha/2)$ |
| 1        | 6.314         | 12.706        | 63.657        | 16       | 1.746         | 2.120         | 2.921         |
| 2        | 2.920         | 4.303         | 9.925         | 17       | 1.740         | 2.110         | 2.898         |
| 3        | 2.353         | 3.182         | 5.841         | 18       | 1.734         | 2.101         | 2.878         |
| 4        | 2.132         | 2.776         | 4.604         | 19       | 1.729         | 2.093         | 2.861         |
| 5        | 2.015         | 2.571         | 4.032         | 20       | 1.725         | 2.086         | 2.845         |
| 6        | 1.943         | 2.447         | 3.707         | 21       | 1.721         | 2.080         | 2.831         |
| 7        | 1.895         | 2.365         | 3.499         | 22       | 1.717         | 2.074         | 2.819         |
| 8        | 1.860         | 2.306         | 3.355         | 23       | 1.714         | 2.069         | 2.807         |
| 9        | 1.833         | 2.262         | 3.250         | 24       | 1.711         | 2.064         | 2.797         |
| 10       | 1.812         | 2.228         | 3.169         | 25       | 1.708         | 2.060         | 2.787         |
| 11       | 1.796         | 2.201         | 3.106         | 26       | 1.706         | 2.056         | 2.779         |
| 12       | 1.782         | 2.179         | 3.055         | 27       | 1.703         | 2.052         | 2.771         |
| 13       | 1.771         | 2.160         | 3.012         | 28       | 1.701         | 2.048         | 2.763         |
| 14       | 1.761         | 2.145         | 2.977         | 29       | 1.699         | 2.045         | 2.756         |
| 15       | 1.753         | 2.131         | 2.947         | 30       | 1.697         | 2.042         | 2.750         |

### $\alpha$ for 2-sided test

For a 1-sided test  
choose the column  
corresponding to  $2 \cdot \alpha$

e.g. for a 1-sided test at  
level 5% use the first  
column (0.1)

For  $df > 30$  use the  
 $N(0,1)$ :

| $\alpha$            | 0.1  | 0.05 | 0.01 |
|---------------------|------|------|------|
| $\alpha$ in 2 tails | 1.64 | 1.96 | 2.58 |
| $\alpha$ in 1 tail  | 1.28 | 1.64 | 2.33 |