Hypothesis testing – a cheat sheet

There are two main groups of hypothesis tests:

1. tests about mean
2. tests about variance (standard deviation)

1. Hypothesis tests about mean

How many mean values are present according to the text of the task?

a. one
b. two
c. more than two

1a. Hypothesis test about mean (one mean value)

- the test is called hypothesis test about a population mean
- we're interested if the population mean is equal to a specific value which is known (a constant)
- notation (H0): $\mu = \mu_0$
    - if the population parameters are known (μ, σ^2, σ) we use the formula (1) to calculate the test statistic and the critical value is calculated using a function NORMSINV ($1-\alpha/2$).

$$u = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \qquad (1)$$

    - if the population parameters are not known, we have to use the sample statistics instead ($\bar{x}, s_1^2, s_1$). Then we have to decide on the sample size. If the sample size > 30, we use the formula (2) to calculate the test statistic and the critical value is calculated using a function NORMSINV ($1-\alpha/2$). If the sample size <= 30, we use the formula (3) to calculate the test statistic and the critical value is calculated using a function TINV ($\alpha; n-1$).

$$u = \frac{\bar{x} - \mu_0}{\frac{s_1}{\sqrt{n}}}$$ (2), we use a normal distribution

$$t = \frac{\bar{x} - \mu_0}{\frac{s_1}{\sqrt{n}}}$$ (3), we use a Student t distribution

1b. Hypothesis test about mean (two mean values)

The samples could be:

i independent

ii matched (dependent)

1bi. Hypothesis test about the difference between means of two populations (independent samples)

- we're interested if the means are equal (with no difference) or not
- notation (H0): $\mu_1 = \mu_2$
    - if the population parameters are known (μ, σ^2, σ) we use the formula (4) to calculate the test statistic and the critical value is calculated using a function NORMSINV ($1-\alpha/2$).

$$u = \frac{\mu_1 - \mu_2}{\sqrt{\frac{n_2\sigma_1^2 + n_1\sigma_2^2}{n_1.n_2}}} \qquad (4)$$

    - if the population parameters are not known, we have to use the sample statistics instead ($\bar{x}, s_1^2, s_1$). Then we have to decide on the sample size. If the sample size of BOTH samples is > 30 (i.e. n1 > 30 AND n2 > 30), we use the formula (5) to calculate the test statistic and the critical value is calculated using a function NORMSINV ($1-\alpha/2$). If the sample size of at least one sample is <= 30 (i.e. either n1 <= 30 OR n2 <= 30), we use the formula (6) to calculate the test statistic and the critical value is calculated using a function TINV ($\alpha;(n_1+n_2-2)$).

$$u = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{n_2 s_{11}^2 + n_1 s_{12}^2}{n_1.n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_2 s_{11}^2 + n_1 s_{12}^2}{n_1.n_2}}}$$ (5), we use a normal distribution

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_{11}^2 + (n_2-1)s_{12}^2}{n_1+n_2-2}}} \cdot \sqrt{\frac{n_1.n_2}{n_1+n_2}}$$ (6), we use a Student t distribution

1bii. Hypothesis test about the difference between means of two populations (matched samples)

- we're interested if the means are equal (with no difference) or not
- notation (H0): $\mu_d = 0$
- this test is almost always performed on small samples (n <= 30) so we'll not know the population parameters at all, we'll use the sample statistics instead
    - to calculate the test statistic, we use the formula (7) and the critical value is calculated using a function TINV ($\alpha; n-1$).
    - the test can also be performed using Tools/Data Analysis/t-test: paired two sample for means

$$t = \frac{\bar{d}}{\sqrt{\frac{\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}{n \cdot (n-1)}}}$$ (7), we use a Student t distribution

1c. Hypothesis test about mean (more than two mean values)

- the test is called Analysis of variance (ANOVA)

2. Hypothesis tests about variance (standard deviation)

How many values of variance are present according to the text of the task?

a. one
b. two
c. more than two

Note: We do not need to check for the population parameters and sample size (as we need to while conducting hypothesis testing about mean). That’s why hypothesis tests about variance are much easier ☺.

2a. Hypothesis test about variance (one value of variance)

- the test is called hypothesis test about a population variance
- we’re interested if the population variance is equal to a specific value which is known (a constant)
- notation (H0): $\sigma^2 = \sigma_0^2$
    - to calculate the test statistic, we use the formula (8) and the critical values (there will be two critical values) are calculated using a function CHIINV $(1-\alpha/2; n-1)$ and $(\alpha/2; n-1)$.

$$\chi^2 = \frac{(n-1)\cdot s_1^2}{\sigma_0^2} \quad (8),$$ we use a χ^2 (Chi-square; read as „kai“) distribution

2b. Hypothesis test about variance (two values of variance)

- the test is called hypothesis test about variances of two populations
- we’re interested if variances of two populations are equal or not
- notation (H0): $\sigma_1^2 = \sigma_2^2$
    - to calculate the test statistic, we use the formula (9) and the critical value is calculated using a function FINV $(\alpha; n_1 - 1; n_2 - 1)$. This test is always specified as one-tailed test and so the numerator of the ratio (s_{11}^2) should be greater than the denominator of the ratio (s_{12}^2), i.e. **denote the population providing largest sample variance as population 1**

$$F = \frac{s_{11}^2}{s_{12}^2} \quad (9),$$ we use a Fisher (F) distribution

- **Note:** if the data set is present, i.e. you do not have only the sample statistics, you can compute the *F-test: Two Sample for variances using Tools/Data Analysis* (variable 1 and 2 range should be the same as population 1 and 2 according to the values of sample variances, see above).

- we can continue in hypothesis testing using a t-test (hypothesis test about the difference between means of two populations) as follows:
  - if **variances of two populations are equal**, than we compute the *t-test: two-sample assuming equal variances* (the test can be performed using Tools/Data Analysis/t-test: two-sample assuming equal variances)
  - if **variances of two populations are not equal**, than we compute the *t-test: Two-sample assuming unequal variances* (Behrens-Fisher test; (the test can be performed using Tools/Data Analysis/t-test: two-sample assuming unequal variances))

2c. Hypothesis test about variance (more than two values of variance)

- the test is called Barttlet or Cochran test
- since Excel is not capable of performing such a test, we'll not discuss it in detail