4. Analysis of Variance (ANOVA)

4.1. Introduction: Comparing Means

Consider the problem of testing

$$H_0\colon \mu_1 = \mu_2 \quad \text{against} \quad H_1\colon \mu_1 \neq \mu_2$$

in two independent samples of two different populations of possibly unequal size n_1 and n_2 containing normally distributed observations with both unknown means μ_i and variances σ_i^2, $i = 1, 2$, which we assume to be equally large. Following STAT1030, the appropriate test statistic is

$$t = \frac{\bar{x_1} - \bar{x_2}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2) \text{ under } H_0.$$

$$\text{with } s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \text{ where}$$

$$s_1^2 = \frac{\sum_{i=1}^{n_1}(x_{1i} - \bar{x_1})^2}{n_1 - 1} \text{ and } s_2^2 = \frac{\sum_{i=1}^{n_2}(x_{2i} - \bar{x_2})^2}{n_2 - 1}.$$

Example (ignore the ANOVA output for now)

Two groups consisting of different people receive two different sleeping pills (med 1/2) and the impact upon sleep prolongation is measured:

Patient number:	Sleep prolongation med 1	 med 2
1	0.70	1.90
2	-1.60	0.80
3	-0.20	1.10
4	-1.20	0.10
5	-0.10	-0.10
6	3.40	4.40
7	3.70	5.50
8	0.80	1.60
9	0.00	4.60
10	2.00	3.40
Average:	0.750	2.330
Stddev.:	1.789	2.002

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	10	7.5	0.75	3.20056
Column 2	10	23.3	2.33	4.009

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	12.48	1	12.482	3.46263	0.079187	4.4139
Within Groups	64.89	18	3.60478			
Total	77.37	19				

Pill 2 is more effective than Pill 1 in sample, but we need to consult the independent sample t-test to check that this holds also out of sample:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{-1.58}{\sqrt{\frac{1.79^2 + 2.00^2}{10}}} = -1.861,$$

where we have used that if $n_1 = n_2$, then $s^2 = (s_1^2 + s_2^2)/2$. The p-value of the test is T.DIST.2T(-1.861;18)=0.079, insufficient to reject the null hypothesis that the medication has no effect at a 5% significance level.

When assessing whether two populations have the same mean using the ANOVA F-test, we compare the variation between the means of both groups with the variation within the group. This is why this method is called analysis of variance. If the within-group variation is small, the same variation between the groups produces a larger statistic and a more significant result. A large F-statistic is thus evidence against the null $H_0 : \mu_1 = \mu_2$ in favour of the alternative $H_1 : \mu_1 \neq \mu_2$.

The advantage of the ANOVA F-test is that it may be generalized to test k means:

$$H_0 \colon \ \mu_1 = \mu_2 = \cdots = \mu_k.$$

It's application requires that

- the groups are independent of each other,
- all groups have the same variance,
- the observations in all groups are normally distributed.

Example (continued.)

We split up the deviation of each persons sleep prolongation x_{ij}, $i = 1, \dots, 10$, $j = 1, 2$ from the overall mean $\bar{x}$ into a component due to the different treatments $\bar{x}_j - \bar{x}$, where $\bar{x}_j$, $j = 1, 2$ denote the mean of the sleep prolongation for the respective treatment (between group deviation), and a component unexplained by the different treatments, $x_{ij} - \bar{x}_j$ (within group deviation):

$$x_{ij} - \bar{x} = (\bar{x}_j - \bar{x}) + (x_{ij} - \bar{x}_j).$$

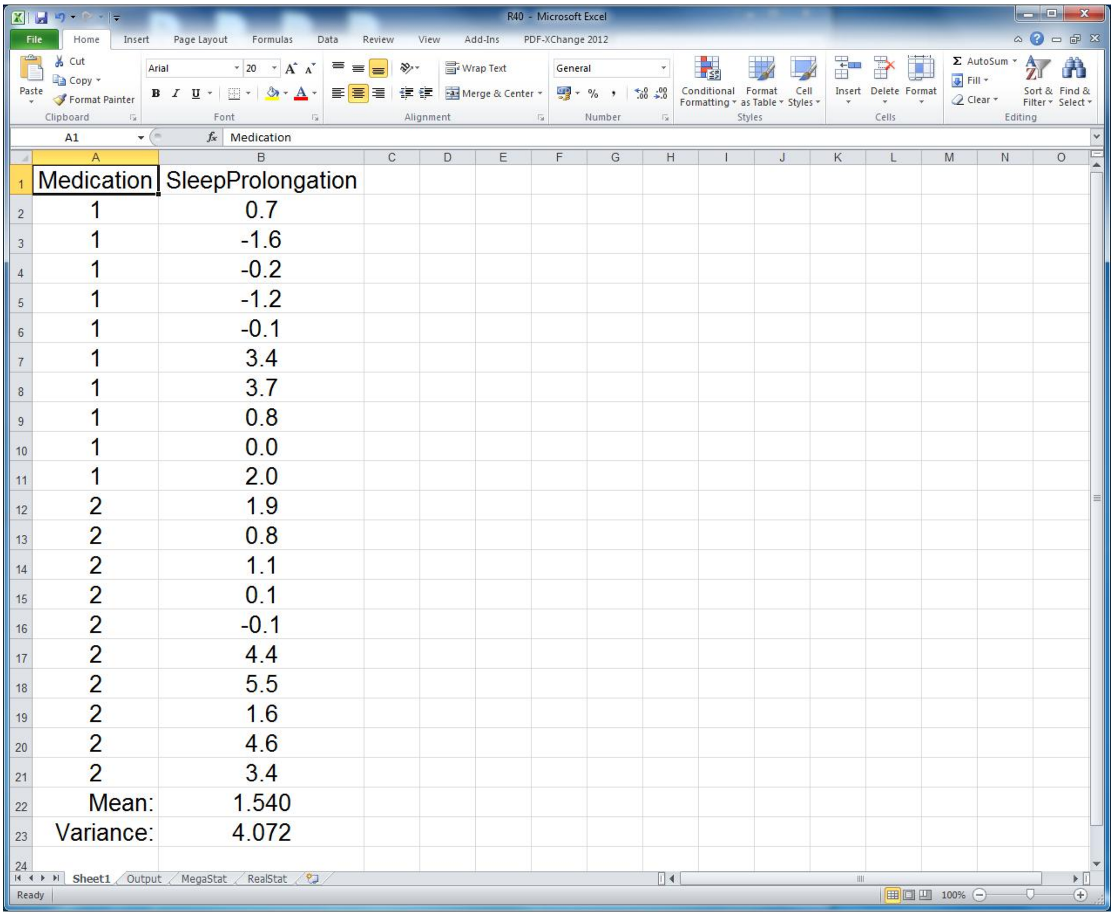
Squaring this expression and summing up yields after some algebra:

$$\underbrace{\sum_{i=1}^{n_1}(x_{i1} - \bar{x})^2 + \sum_{i=1}^{n_2}(x_{i2} - \bar{x})^2}_{\text{SST}=(n-1)s^2}$$

$$= \underbrace{\sum_{i=1}^{n_1}(\bar{x_1} - \bar{x})^2 + \sum_{i=1}^{n_2}(\bar{x_2} - \bar{x})^2}_{\text{SSM or SSTR or SSB}} + \underbrace{\sum_{i=1}^{n_1}(x_{i1} - \bar{x_1})^2 + \sum_{i=1}^{n_2}(x_{i2} - \bar{x_2})^2}_{\text{SSE or SSW}}$$

$$= n_1(\bar{x_1} - \bar{x})^2 + n_2(\bar{x_2} - \bar{x})^2 + (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2,$$

where TR stands for treatment, B stands for between groups, and W stands for within groups (T,M,E see regression ANOVA).



A1 Medication

	A	B
1	Medication	SleepProlongation
2	1	0.7
3	1	-1.6
4	1	-0.2
5	1	-1.2
6	1	-0.1
7	1	3.4
8	1	3.7
9	1	0.8
10	1	0.0
11	1	2.0
12	2	1.9
13	2	0.8
14	2	1.1
15	2	0.1
16	2	-0.1
17	2	4.4
18	2	5.5
19	2	1.6
20	2	4.6
21	2	3.4
22	Mean:	1.540
23	Variance:	4.072
24		

ANOVA: Single Factor

DESCRIPTION Alpha 0.05

Groups	*Count*	*Sum*	*Mean*	*Variance*	*SS*	*Std Err*
1	10	7.5	0.75	3.2006	28.805	0.600398
2	10	23.3	2.33	4.009	36.081	0.600398

ANOVA

Sources	*SS*	*df*	*MS*	*F*	*P value*	*F crit*
Between Groups	12.482	1	12.482	3.4626	0.079187	4.413873
Within Groups	64.886	18	3.60478			
Total	77.368	19	4.072			

The associated degrees of freedom are:
DFT=19 ($n = 20$ observations minus 1 for the calculation of the grand mean),
DFB=1 (2 sample means minus 1 for the calculation of the grand mean),
DFW=18 ($n = 20$ observations minus 2 for the calculation of the sample means).

The mean squares are obtained as:

$$\text{MSB} = \frac{\text{SSB}}{\text{DFB}} = \frac{12.48}{1} = 12.48,$$
$$\text{MSW} = \frac{\text{SSW}}{\text{DFW}} = \frac{64.89}{18} = 3.605.$$

Since we want to reject when the between group variation is large compared to the within group variation we calculate the F-statistic as

$$F = \frac{MSB}{MSW} = \frac{12.48}{3.605} = 3.463,$$

which is just the square of -1.861, the t-statistic we obtained earlier, and the p-value is F.DIST.RT(3.463;1;18)=0.079, just the same as before using the t-statistic. We explain in the following why this is the case.

Why is $F = t^2$?

This is just a matter of algebraic manipulations. Squaring the t-test statistic yields

$$t^2 = \frac{\frac{n_1 \cdot n_2}{n_1+n_2}(\bar{x}_1 - \bar{x}_2)^2}{s^2}.$$

$$\text{with} \quad s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}.$$

Now

$$\text{SSW} = (n_1-1)s_1^2 + (n_2-1)s_2^2 = (n_1+n_2-2)s^2$$

such that

$$\text{MSW} = \text{SSW}/(n_1+n_2-2) = s^2.$$

Similiarly it can be shown by inserting

$$\bar{x} = \frac{n_1\bar{x_1} + n_2\bar{x_2}}{n_1+n_2} \quad \text{into}$$

$$\text{MSB} = \text{SSB} = n_1(\bar{x_1} - \bar{x})^2 + n_2(\bar{x_2} - \bar{x})^2,$$

$$\text{that} \quad \text{MSB} = \frac{n_1 \cdot n_2}{n_1+n_2}(\bar{x}_1 - \bar{x}_2)^2.$$

Therefore: $F = \dfrac{MSB}{MSW} = t^2$.

It is also possible to show that squaring a $t(\nu)$-distributed random variable will always yield a $F(1,\nu)$-distributed random variable, which explains the identical p-values for the t- and for the F-test.

Appropriateness of the F-test

Recall from STAT1030 that the fraction of two independent χ^2-distributed variables divided by their degrees of freedom follow the F-distribution as follows:

$$X_i \sim \chi^2(\nu_i)\ (i = 1, 2):\quad \frac{X_1/\nu_1}{X_2/\nu_2} \sim F(\nu_1, \nu_2).$$

If the ANOVA assumptions hold and additionally $\mu_1 = \mu_2 = \cdots = \mu_k$ as stated in H_0, then it turns out that MSB=SSB/DFB is an unbiased estimator of σ^2 and additionally

$$\text{SSB}/\sigma^2 \sim \chi^2(\text{DFB}).$$

Furthermore it turns out that under ANOVA assumptions MSW=SSW/DFW is also an unbiased estimator of σ^2 and additionally

$$\text{SSW}/\sigma^2 \sim \chi^2(\text{DFW})$$

independent of MSB and no matter whether H_0 holds true or not, implying that under H_0:

$$F = \text{MSB}/\text{MSW} \sim F(\text{DFB,DFW}).$$

4.2. One-Way ANOVA for $k \geq 2$ groups

In ANOVA we have k independent random samples, each one corresponding to a population subject to a different treatment. We have:

- $N = n_1 + n_2 + \ldots + n_k$ total observations.
- k sample means: $\bar{x_1}, \bar{x_2}, \ldots, \bar{x_k}$.
  These may be used to calculate an estimator of the population variance (MSB), if the population means are equal. In that case we expect the variance among the sample means to be small.
- k sample variances: $s_1^2, s_2^2, \ldots, s_k^2$.
  These may be used to find a second estimator of the population variance (MSW), no matter whether the population means are equal or not.

(Note the alternative notations MSE for MSW, and also MSTR and MSM for MSB.)

ANOVA is a statistical method to determine whether several population means are equal. That is, we test

H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$ against
H_1: not all $\mu_i,\ i = 1, \ldots, k$ are equal.

ANOVA is a joint test in the sense that it tests the equality of all population means simultaneously rather than pairwise. This is achieved by comparing the two estimators of the population variance (hence the name).

The required assumptions of ANOVA are:

- We assume *independent random sampling* from each of the k populations.

- We assume that the k populations:
  - are normally distributed,
  - with means μ_j not necessarily equal,
  - but with *equal variances* $\sigma_i^2 =: \sigma^2$.

Checking the assumptions

1) Independence is taken care of in the sampling or experimental design. Similiar to regressions, the whole approach breaks down if the independence assumption is not satisfied.

2) Use the Shapiro-Wilk test from the Real Statistics 'Descriptive Statistics and Normality' window in order to check for Normality in each of the groups. If you find any p-value below 0.05, it means that the normality assumption is not met.

Deviations from normality are acceptable as long as

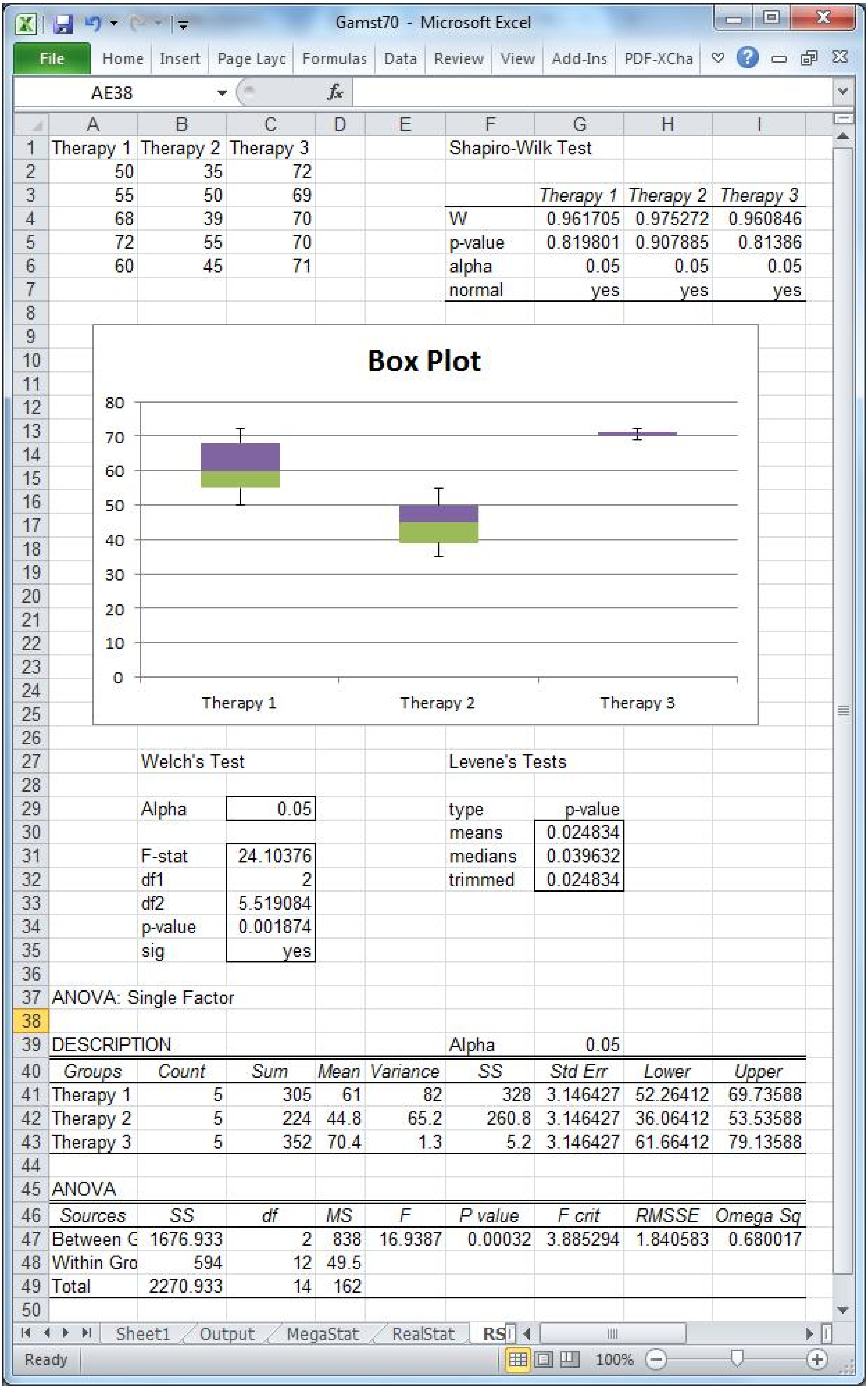
- The populations are symmetrical and unimodal
- There are at least 10 observations in each group

Even symmetry can be dispensed with if all sample sizes are equally large (so called balanced design), all populations have the same shape and there are many observations.

3) To test for Homogeneity of variance Real Statistics provides Levene's test, available as an option within Single factor ANOVA. The output contains 3 p-values. The first p-value (means) is the original and most powerful version of the test and requires that all groups have normally distributed observations. The second version (medians) is less powerful, but robust against skewness. The last version (trimmed) is robust against excess kurtosis.

The null hypothesis is that all groups have the same variance, so a p-value below 0.05 implies that the homogeneity of variance assumption is not met.

ANOVA is not very sensitive to violations of the homogeneity of variance assumption, as long as the sizes of the different samples are about the same. As a rule of thumb, ANOVA is applicable as long as the highest variance is no more than twice the lowest variance.



Therapy 1	Therapy 2	Therapy 3
50	35	72
55	50	69
68	39	70
72	55	70
60	45	71

Shapiro-Wilk Test

	Therapy 1	*Therapy 2*	*Therapy 3*
W	0.961705	0.975272	0.960846
p-value	0.819801	0.907885	0.81386
alpha	0.05	0.05	0.05
normal	yes	yes	yes

Welch's Test

Alpha	0.05
F-stat	24.10376
df1	2
df2	5.519084
p-value	0.001874
sig	yes

Levene's Tests

type	p-value
means	0.024834
medians	0.039632
trimmed	0.024834

ANOVA: Single Factor

DESCRIPTION

Alpha 0.05

Groups	*Count*	*Sum*	*Mean*	*Variance*	*SS*	*Std Err*	*Lower*	*Upper*
Therapy 1	5	305	61	82	328	3.146427	52.26412	69.73588
Therapy 2	5	224	44.8	65.2	260.8	3.146427	36.06412	53.53588
Therapy 3	5	352	70.4	1.3	5.2	3.146427	61.66412	79.13588

ANOVA

Sources	*SS*	*df*	*MS*	*F*	*P value*	*F crit*	*RMSSE*	*Omega Sq*
Between G	1676.933	2	838	16.9387	0.00032	3.885294	1.840583	0.680017
Within Gro	594	12	49.5					
Total	2270.933	14	162					

When the ANOVA assumptions do not hold

Often violations of the assumption are just due to some outliers, which may be easily detected with explorative tools such as histograms and boxplots. Checking ‘Outliers and Missing Data’ within the ‘Descriptive Statistics and Normality’ window of the Real Statistics toolbox will mark all observations which are more than 2.5 standard deviations away from the mean.

If the assumption of equal variances is not satisfied but all other assumptions are in order, you may replace standard ANOVA with Welch’s test. It is an extension of Student’s t-test for populations with unequal variances to more than 2 groups and is available as one of the options within Single Factor ANOVA.

1) When the normality tests reject. . .

Skewness and excess kurtosis in the data may be corrected by replacing the original observations x_i with some suitably transformed observations $f(x_i)$:

- Skewness to the right may be corrected by square-root and logarithmic transformations: $f(x) = \sqrt{x}$ or $f(x) = \log(x)$.

- Skewness to the left may be corrected by $f(x) = \arcsin\sqrt{x}$ or $f(x) = \log\left(\frac{x}{1-x}\right)$, which also corrects thin-tailedness.

- Fat-tailedness (= excess kurtosis) may be corrected by $f(x) = \log\left(\sqrt{\frac{1+x}{1-x}}\right)$.

2) When the variances are not homogeneous

If the variance is a function of the mean alone,

$$\sigma^2 = \phi(\mu),$$

then it is always possible to find a suitable transformation function $f(x)$ such that homogeneity of variance holds for the transformed data.

In Poisson type distributions, where

$$\sigma^2 = \mu = \lambda$$

(λ =parameter of the Poisson distribution) variance is stabilized by the square-root transformation $f(x) = \sqrt{x}$.

Recall that Poisson type distributions describe counting processes of rare events, such as the number of working accidents in different companies.

If the coefficient of variance is a constant,

$$V = \sigma/\mu = \text{const.},$$

like for example in the exponential distribution, then variance is stabilized by the logarithmic transformation $f(x) = \log(x)$. Constant coefficients of variance occur in situations where effects are proportional rather than additive, that is, the variance of different treatments differ consistently by certain percentages (rather than a certain number of units).

The variance of proportions (that is, numbers in the range [0,1]) may be stabilized by the transformation $f(x) = \arcsin\sqrt{x}$.

In general, if the variance is a function of the mean alone, $\sigma^2 = \phi(\mu)$, then an approximate variance stabilizing transformation is given by

$$f(x) = \int \frac{dx}{\sqrt{\phi(x)}}.$$

3) When transformations don't help either

Don't forget that ANOVA is rather robust to skewness and also to differences in variance as long as the sample sizes are equally large. Recall also that you may apply Welch's variance weighted ANOVA when homogeneity of variance is the only problem.

A last resort are non-parametric tests:

- The Mann-Whitney test in the case of two classes, which is just a special case of the
- Kruskall-Wallis test for arbitarily many samples.

When the test statistics are distorted by a large spread due to extreme outliers, which one doesn't want to eliminate, one may still try the

- median test.

Setting up an ANOVA table

Consider k independent samples:

$$x_{11}, x_{12}, \cdots, x_{1n_1},$$
$$\vdots$$
$$x_{k1}, x_{k2}, \cdots, x_{kn_k},$$

where $n_1 + \cdots + n_k = N$.

We wish to test whether all observations come from the same distribution or not.

1) The Sum of Squares Principle

We calculate the grand mean from all observations as

$$\bar{\bar{x}} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} = \frac{1}{N} \sum_{i=1}^{k} n_i \bar{x}_i.$$

The total deviations are the differences between the data points x_{ij} and the grand mean $\bar{\bar{x}}$:

$$\mathrm{Tot}_{ij} = x_{ij} - \bar{\bar{x}}.$$

The total deviations $x_{ij} - \overline{\overline{x}}$ are split up into group effects or treatment deviations $\bar{x_i} - \overline{\overline{x}}$ and the residuals $e_{ij} = x_{ij} - \bar{x_i}$:

$$x_{ij} - \overline{\overline{x}} = (\bar{x_i} - \overline{\overline{x}}) + (x_{ij} - \bar{x_i}).$$

Squaring this expression and summing up yields the SST = Total Sum of Squares which may again be split into two components:

the SSB = Between Groups Sum of Squares

and

the SSW = Within Groups Sum of Squares:

$$\begin{aligned} \text{SST} &= \sum_i \sum_j (x_{ij} - \overline{\overline{x}})^2 = (N-1)s^2 \\ &= \sum_i \overbrace{\sum_j (\overline{x_i} - \overline{\overline{x}})^2}^{n_i(\overline{x_i} - \overline{\overline{x}})^2} + \sum_i \overbrace{\sum_j (x_{ij} - \overline{x_i})^2}^{(n_i - 1)s_i^2} \\ &= \text{SSB} + \text{SSW}. \end{aligned}$$

Aczel uses the terms sum of squares for treatment (SSTR) instead of SSB and sum of squares for error (SSE) instead of SSW.

2) Additivity of degrees of freedom

The degrees of freedom of SST are $\underline{(N-1)}$:

N total observations in all k samples, less one degree of freedom lost with the calculation of the grand mean.

The degrees of freedom of SSB are $\underline{(k-1)}$:

k sample means, less one degree of freedom lost with the calculation of the grand mean.

The degrees of freedom of SSW are $\underline{(N-k)}$:

N total observations in all samples, less one degree of freedom lost with each calculation of any of the k sample means.

The degrees of freedom are additive in the same way as are the sums of squares:

$$\text{DFT} = \text{DFB} + \text{DFW}.$$

3) The Mean Squares and the F-test

Dividing the components of the total sum of squares by their respective degrees of freedom yields the mean squares

$$\text{MSB} = \frac{\text{SSB}}{k-1} \quad \text{and} \quad \text{MSW} = \frac{\text{SSW}}{N-k}.$$

If all ANOVA assumptions,

- independent samples,
- normally distributed observations,
- homogeneity of variance

and the null hypothesis

- equal means in all groups

hold true, then the test statistic

$$F = \frac{MSB}{MSW}$$

follows the F distribution with degrees of freedom $k-1$ and $N-k$, $(F \sim F(k-1, N-k))$.

Large values of the F statistic lead to rejection of the null hypothesis of equal means in all groups.

What if the F-test rejects?

Mathematically, the model may be written

$$X_{ij}=\mu_i+\epsilon_{ij}=\mu+\overbrace{(\mu_i-\mu)}^{\alpha_i}+\epsilon_{ij},\ \ \epsilon_{ij}\sim N(0,\sigma^2),$$

where X_{ij} denotes the j'th observation of population (treatment) i, μ is the overall mean response to the treatments, and $\alpha_i = \mu_i - \mu$ measures the effect of the specific treatment in class i. This implies that testing

$$H_0\colon \mu_1 = \mu_2 = \cdots = \mu_k$$

is equivalent to testing

$$H_0\colon \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0.$$

In the fixed effects model we choose intentionally k specific populations due to some common characteristic we want to study or due to some specific treatments we want to investigate. In this setup the α_i are regarded as fixed but unknown quantities to be estimated and the alternative hypothesis is just

$$H_1\colon \text{not all } \alpha_i \text{ are zero.}$$

The task is then to estimate these α_i's in case the null hypothesis is rejected.

4) Estimation of the effects

In the fixed effects model, the effects are estimated from the difference of the groups means from the grand mean $\bar{\bar{x}}$, which is the arithmetic mean of all onservations from all groups:

$$a_i = \bar{x_i} - \bar{\bar{x}}, \quad i = 1, 2, \ldots k.$$

<u>Example:</u> (sleeping pills continued).

The p-value of the F-test was 7.9% $>$ 5%, hence we accept $\mu_1 = \mu_2 = 1.54$, which is equivalent to $\alpha_1 = \alpha_2 = 0$ (implying no effect of medication change).

If the F-test had rejected, we would have estimated μ_1 by $\bar{x_1} = 0.75$ and μ_2 by $\bar{x_2} = 2.33$, such that the <u>sample effects</u> become

$$\begin{aligned} a_1 &= 0.75 - 1.54 = -0.79 \\ a_2 &= 2.33 - 1.54 = +0.79, \end{aligned}$$

which we would have used as our estimates for α_1 and α_2, imlying that medication 2 is better than medication 1.

4.3. The random effects model

In the fixed effects model discussed above the treatments are selected specifically by the experimenter because they are of particular interest. The purpose of the experiment is to make inferences about the means of the particular populations from which the samples are drawn. If, however, we want to make a broad generalization concerning a larger set of populations and not just the k populations from which we sample, then the appropriate model is called a *random-effects model*.

In the random-effects model the k sampled populations are considered to be a random sample of populations drawn from a larger set. The hypothesis of interest is not that $\mu_1 = \mu_2 = \cdots = \mu_k$. Rather we want to determine whether there is variablility among the population means of the larger set.

The following *model assumptions* are made:

1. The k samples represent independent random samples from k populations randomly selected from a larger set of populations.
2. Each of the populations in the larger set is normally distributed, and therefore each of the sampled populations is also normal.
3. Each of the populations in the larger set has the same variance σ^2, and thus each of the k sampled populations also has variance σ^2.
4. The effects $\alpha_1, \alpha_2, \ldots, \alpha_k$ are independent normally distributed random variables, each with mean 0 and common variance σ_A^2.

Mathematicaly the model may be written as:

$$X_{ij} = \mu + \overbrace{(\bar{X}_i - \mu)}^{A_i} + \overbrace{(X_{ij} - \bar{X}_i)}^{\epsilon_{ij}} = \mu + A_i + \epsilon_{ij},$$

where $\quad A_i \sim N(0, \sigma_A^2), \quad \epsilon_{ij} \sim N(0, \sigma^2)$.

The null hypothesis is now stated as:

$$H_0\colon\ \sigma_A^2 = 0,$$

which we test against

$$H_1\colon\ \sigma_A^2 \neq 0,$$

If the null hypothesis holds, then as before:

$$X_{ij} = \mu + \epsilon_{ij} \sim N(\mu, \sigma^2),$$

which implies that the ANOVA table looks just the same as in the fixed effects model and H_0 is also tested *in exactly the same way* as in the fixed effects model.

The difference between the models becomes first apparent in different expected mean squares, when the null hypothesis is rejected.

The expected mean squares for the treatments is in the fixed effects model:

$$E(\mathrm{MSB}) = \sigma^2 + \frac{1}{k-1}\sum_{i=1}^{k} n_i(\mu_i - \mu)^2,$$

which in the special case of equal sample sizes $n_i = n$ in each group simplifies to:

$$E(\mathrm{MSB}) = \sigma^2 + n\frac{\sum(\mu_i - \mu)^2}{k-1} = \sigma^2 + n{\cdot}s^2(\mu),$$

where we have written $s^2(\mu)$ for the sample variance of the population means μ_i.

Similiarly the expected mean squares in the random effects model is

$$E(\mathrm{MSB}) = \sigma^2 + n_0{\cdot}\sigma_A^2,$$

where

$$n_0 = \frac{N^2 - \sum_{i=1}^{k} n_i^2}{N(k-1)} \quad \text{and} \quad N = \sum_{i=1}^{k} n_i.$$

Note that $n_0 = n$ in the special case of equal sampel sizes $n_i = n$ in all groups.

Estimating $E(\text{MSB})$ by MSB, and σ^2 by MSW yields the following estimator of σ_A^2:

$$s_A^2 = \frac{\text{MSB} - \text{MSW}}{n_0}$$

Example. (Milton/Arnold: Example 13.7.1)

A utility company has a large stock of voltmeters that are used interchangeably by many employees. Below are differences between 4 applied voltages and the corresponding readings for 6 randomly selected voltmeters:

Voltmeter					
1	***2***	***3***	***4***	***5***	***6***
0.18	-0.15	-0.25	1.95	-0.90	1.10
-1.31	1.85	0.77	1.03	-0.50	1.21
0.15	0.63	1.65	0.65	0.25	0.68
-0.81	0.45	1.24	1.25	-0.88	0.92

We wish to test whether the average readings of all voltmeters in stock are identical.

The null and alternative hypotheses are:

$$H_0\colon \sigma_A = 0 \qquad \text{versus} \qquad H_1\colon \sigma_A \neq 0.$$

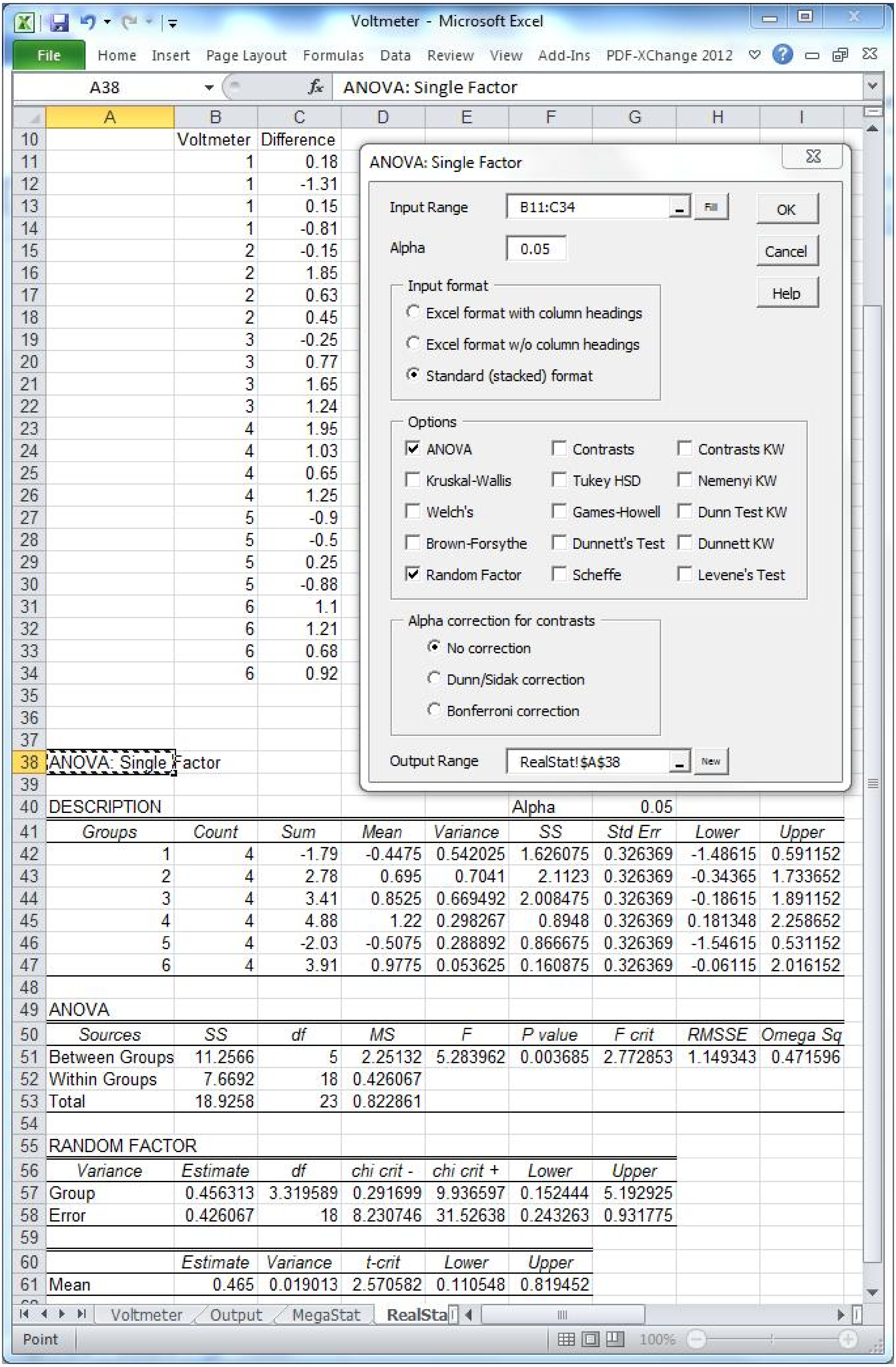
The ANOVA table provides evidence of differences in average reading errors among the voltmeters in stock.

We can estimate how much of the variability is due to differences in meters and how much is due to random sampling error. To do this, we estimate the variance components σ^2 and σ_A^2 as

$$s^2 = \text{MSW} = 0.426 \qquad \text{and}$$

$$s_A^2 = \frac{\text{MSB} - \text{MSW}}{n_0} = \frac{2.251 - 0.426}{4} = 0.456.$$

The Random Factor Option within the Real Statistics Single Factor ANOVA tool provides also 95% confidence intervals for the variance components and the overall mean.



Voltmeter	Difference
1	0.18
1	-1.31
1	0.15
1	-0.81
2	-0.15
2	1.85
2	0.63
2	0.45
3	-0.25
3	0.77
3	1.65
3	1.24
4	1.95
4	1.03
4	0.65
4	1.25
5	-0.9
5	-0.5
5	0.25
5	-0.88
6	1.1
6	1.21
6	0.68
6	0.92

ANOVA: Single Factor

Input Range B11:C34

Alpha 0.05

Input format

○ Excel format with column headings

○ Excel format w/o column headings

● Standard (stacked) format

Options

☑ ANOVA	☐ Contrasts	☐ Contrasts KW
☐ Kruskal-Wallis	☐ Tukey HSD	☐ Nemenyi KW
☐ Welch's	☐ Games-Howell	☐ Dunn Test KW
☐ Brown-Forsythe	☐ Dunnett's Test	☐ Dunnett KW
☑ Random Factor	☐ Scheffe	☐ Levene's Test

Alpha correction for contrasts

● No correction

○ Dunn/Sidak correction

○ Bonferroni correction

Output Range RealStat!A38

ANOVA: Single Factor

DESCRIPTION Alpha 0.05

Groups	Count	Sum	Mean	Variance	SS	Std Err	Lower	Upper
1	4	-1.79	-0.4475	0.542025	1.626075	0.326369	-1.48615	0.591152
2	4	2.78	0.695	0.7041	2.1123	0.326369	-0.34365	1.733652
3	4	3.41	0.8525	0.669492	2.008475	0.326369	-0.18615	1.891152
4	4	4.88	1.22	0.298267	0.8948	0.326369	0.181348	2.258652
5	4	-2.03	-0.5075	0.288892	0.866675	0.326369	-1.54615	0.531152
6	4	3.91	0.9775	0.053625	0.160875	0.326369	-0.06115	2.016152

ANOVA

Sources	SS	df	MS	F	P value	F crit	RMSSE	Omega Sq
Between Groups	11.2566	5	2.25132	5.283962	0.003685	2.772853	1.149343	0.471596
Within Groups	7.6692	18	0.426067					
Total	18.9258	23	0.822861					

RANDOM FACTOR

Variance	Estimate	df	chi crit -	chi crit +	Lower	Upper
Group	0.456313	3.319589	0.291699	9.936597	0.152444	5.192925
Error	0.426067	18	8.230746	31.52638	0.243263	0.931775

	Estimate	Variance	t-crit	Lower	Upper
Mean	0.465	0.019013	2.570582	0.110548	0.819452