1) See Chapter 4.4 and 4.5 in Box, Hunter & Hunter (course book).

2) a) Checking the Normal assumption of residuals.

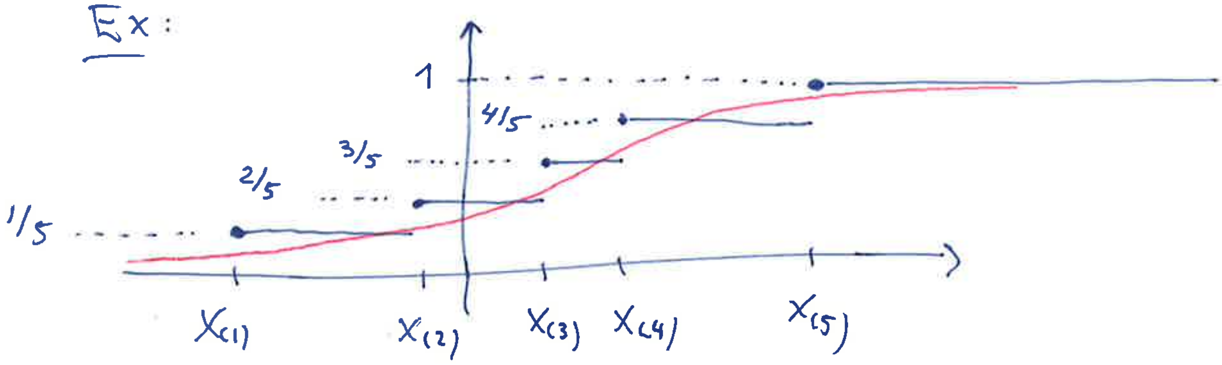
b) Assume, for example, $n = 5$.

Sorted data: $x_{(1)} < x_{(2)} < x_{(3)} < x_{(4)} < x_{(5)}$

Empirical distribution function:

$$F_n(x_{(i)}) = \frac{i}{n}$$

Ex:



Construction of qq-plot:

* Choose $p = \frac{i}{n}$ for $i = 1, \dots, n$

* Calculate: empirical quantile $= F_n^{-1}(p) = x_{(i)}$

  theoretical quantile $= F^{-1}(p)$

* If a good fit $\underbrace{F_n^{-1}(p)}_{= x_{(i)}} \simeq F^{-1}(p)$

$\Rightarrow$ The points $(x, y) = (x_{(i)}, F^{-1}(i/n))$ should fit a line with slope $\sim 45°$

3) a) Short description (see Box, Hunter & Hunter for more details!)

$H_0: \mu_A = \mu_B$

$H_1: \mu_A \neq \mu_B$

* Consider all possible treatment assignments. Nbr of possible assignments:

$$\binom{10}{5} = \frac{10!}{5!\,5!} = 252$$

* For each treatment assignment compute $d_i = \bar{y}_A - \bar{y}_B \; ; \; i = 1, \dots, 252$.

* $\{d_1, d_2, \dots, d_{252}\}$ enumerates all pre-randomization outcomes <u>assuming</u> <u>no treatment effect</u>.

* Since each treatment assignment is equally likely under H_0, a probability distribution of the experimental results, if H_0 is true, can be stated as:

$$\hat{F}(y) = \frac{\#(d_i \leq y)}{252} = \frac{\sum_{i=1}^{252} I(d_i \leq y)}{252}$$

* Given the <u>observed</u> difference d_i and $\hat{F}(y)$, calculate a p-value. (small p-value implies evidence against H_0).

b) If the number of treatment assignments is very large this method is of little practical interest.

4)

a) $H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

$$S_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (y_i - \bar{y})^2 = 85,82 \qquad ; n_1 = 10$$

$$S_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2 = 87,73 \qquad ; n_2 = 10$$

$$F_0 = \frac{S_1^2}{S_2^2} = 0,98$$

$$F_{0,025, 9, 9} = 4,03 .$$

$$F_{0,0975, 9, 9} = \frac{1}{F_{0,025, 9, 9}} = \frac{1}{4,03} = 0,248$$

Since $F_{0,0975, 9, 9} < F_0 < F_{0,025, 9, 9}$

we cannot reject H_0.

b) Since we cannot reject $H_0: \sigma_1^2 = \sigma_2^2$

calculate the pooled variance.

$$S_p^2 = \frac{(n_1-1) S_1^2 + (n_2-1) S_2^2}{n_1 - 1 + n_2 - 1} = 86,775 \implies S_p = 9,32$$

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{70,4 - 70,2}{9,32 \sqrt{\frac{1}{10} + \frac{1}{10}}} = 0,048$$

Since $t_{0,025, 18} = 2,101$ we cannot reject.

c) Recall that: $\sum_{i=1}^{n} X_i^2 \in \chi^2(n)$ when $X_i \in N(0,1)$

This gives $T^2 = \frac{X^2}{Y/f}$ where $X^2 \in \chi^2(1)$.

Hence $T^2 = \frac{X^2/1}{Y/f} \in F(1, f)$.

5)

a) Confounding relations: $D = ABC$. $\Rightarrow I = ABCD$.

$AB = CD$

$AC = BD$

$AD = BC$

$A = BCD$

$B = ACD$

$C = ABD$

Resolution of the design is IV.

b) All the main effects are confounded with a three order effect which are assumed to negligable. Therefore you can estimate the main effects alone.

c) Calculate the mean effects:

$\bar{E}_A = 5{,}67$, $\bar{E}_B = -0{,}155$, $\bar{E}_C = -2{,}905$, $\bar{E}_D = -7{,}19$

$\bar{E}_{AB} = \bar{E}_{CD} = 0{,}645$, $\bar{E}_{AC} = \bar{E}_{BD} = 0{,}065$, $\bar{E}_{AD} = \bar{E}_{BC} = -0{,}14$.

$$s_p^2 = \text{Pooled variance} = \frac{\sum_{i=1}^{2^3} s_i^2}{2^3} = 0{,}4474$$

$$\sigma_{\text{mean effect}} = \frac{2 s_p}{\sqrt{2^k n}} = \frac{2 \cdot \sqrt{0{,}4474}}{\sqrt{8 \cdot 2}} = 0{,}33$$

$\Rightarrow$ Confidence interval: $(-3\sigma_{\text{mean effect}}, +3\sigma_{\text{mean effect}}) = (-0{,}99, +0{,}99)$

$\boxed{\text{Hence, A, C and D are active at a 3-sigma-level.}}$

6) a) A randomized block design.
Groups are the blocks.

b)

	SS	df	MS	F-ratio	p-value
Diet	19,73	2	9,87	0,8555	0,46
Group	1507,73	4	376,93	32,6821	$5,3 \cdot 10^{-5}$
Residuals	92,27	8	11,53		
Total	1619,73	14			

c) The diets do not seem to differ (large p-value). However, the groups seem to have significant impact. It seems that group 1 has lost least weight, and groups 4 an 5 the most.
Errors are assumed to be independent and $N(0, \sigma^2)$ distributed.

d) One could look at the relative loss of weight instead.
Obviously, the persons in group 5 should loose most weight (in pounds) since they are the heaviest.