Chapter 12.

Box 12.1 Summarizing Observations

Given: A sample $\{x_1, x_2, \ldots, x_n\}$ of n observations.

1. Sample arithmetic mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. Sample geometric mean: $\dot{x} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$
3. Sample harmonic mean: $\ddot{x} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$
4. Sample median: $\begin{cases} x_{((n-1)/2)} & \text{if } n \text{ is odd} \\ 0.5(x_{(n/2)} + x_{((1+n)/2)}) & \text{otherwise} \end{cases}$

   Here $x_{(i)}$ is the ith observation in the sorted set.
5. Sample mode = observation with the highest frequency (for categorical data).
6. Sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
7. Sample standard deviation: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
8. Coefficient of variation $= s/\bar{x}$
9. Coefficient of skewness $= \frac{1}{ns^3}\sum_{i=1}^{n}(x_i - \bar{x})^3$
10. Range: Specify the minimum and maximum.
11. Percentiles: $100p$-percentile $x_p = x_{(\lfloor 1+(n-1)p \rfloor)}$.
12. Semi-interquartile range SIQR $= \frac{Q_3 - Q_1}{2} = \frac{x_{0.75} - x_{0.25}}{2}$
13. Mean absolute deviation $= \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$

Chapter 13.

Box 13.1 Confidence Intervals

1. Given: A sample $\{x_1, x_2, \ldots, x_n\}$ of n observations.
   $\bar{x}$ = sample mean; s = sample standard deviation

   (a) Standard error of the sample mean: $\sigma_{\bar{x}} = \dfrac{s}{\sqrt{n}}$

   (b) $100(1-\alpha)\%$ two-sided confidence interval for the mean:
   $\bar{x} \mp z_{1-\alpha/2} s/\sqrt{n}$
   If $n \leq 30$†: $\bar{x} \mp t_{[1-\alpha/2;n-1]} s/\sqrt{n}$

   (c) $100(1-\alpha)\%$ one-sided confidence interval for the mean:
   $(\bar{x}, \bar{x} + z_{1-\alpha} s/\sqrt{n})$ or $(\bar{x} - z_{1-\alpha} s/\sqrt{n}, \bar{x})$
   If $n \leq 30$†: $(\bar{x}, \bar{x} + t_{[1-\alpha;n-1]} s/\sqrt{n})$ or $(\bar{x} - t_{[1-\alpha;n-1]} s/\sqrt{n}, \bar{x})$

2. To compare two systems using unpaired observations:

   (a) The standard error of the mean difference: $s = \sqrt{\dfrac{s_a^2}{n_a} + \dfrac{s_b^2}{n_b}}$

   (b) The effective number of degrees of freedom:

$$\nu = \frac{(s_a^2/n_a + s_b^2/n_b)^2}{\dfrac{1}{n_a+1}\left(\dfrac{s_a^2}{n_a}\right)^2 + \dfrac{1}{n_b+1}\left(\dfrac{s_b^2}{n_b}\right)^2} - 2$$

   (c) The confidence interval for the mean difference:
   $(\bar{x}_a - \bar{x}_b) \mp t_{[1-\alpha/2;\nu]} s$

3. If n_1 of the n observations belong to a certain class, the following statistics can be reported for the class:

   (a) Proportion of the observations in the class: $p = \dfrac{n_1}{n}$

   (b) $100(1-\alpha)\%$ two-sided confidence interval for the proportion‡:
   $p \mp z_{1-\alpha/2}\sqrt{\dfrac{p(1-p)}{n}}$

   (c) $100(1-\alpha)\%$ one-sided confidence interval for the proportion‡:

$$\left(p, p + z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}}\right) \quad \text{or} \quad \left(p - z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}}, p\right)$$

† Only for samples from normal populations.
‡ Provided $np \geq 10$.

Chapter 13 (continued)

13.9 DETERMINING SAMPLE SIZE

The confidence level of conclusions drawn from a set of measured data depends upon the size of the data set. The larger the sample, the higher is the associated confidence. However, larger samples also require more effort and resources. Thus, the analyst's goal is to find the smallest sample size that will provide the desired confidence. In this section, we present formulas for determining the sample sizes required to achieve a given level of accuracy and confidence. Three different cases: single-system measurement, proportion determination, and two-system comparison are considered. In each case, a small set of preliminary measurements are done to estimate the variance, which is then used to determine the sample size required for the given accuracy.

13.9.1 Sample Size for Determining Mean

Suppose we want to estimate the mean performance of a system with an accuracy of $\pm r\%$ and a confidence level of $100(1-\alpha)\%$. The number of observations n required to achieve this goal can be determined as follows.

We know that for a sample of size n, the $100(1-\alpha)\%$ confidence interval of the population mean is

$$\bar{x} \mp z\frac{s}{\sqrt{n}}$$

13.9.2 Sample Size for Determining Proportions

This technique can be extended to determination of proportions. The confidence interval for a proportion was shown in Section 13.8 to be

$$\text{Confidence interval for proportion} = p \mp z\sqrt{\frac{p(1-p)}{n}}$$

To get a half-width (accuracy of) r,

$$p \mp r = p \mp z\sqrt{\frac{p(1-p)}{n}}$$

$$r = z\sqrt{\frac{p(1-p)}{n}}$$

$$n = z^2\frac{p(1-p)}{r^2}$$

The desired accuracy of r percent implies that the confidence interval should be $(\bar{x}(1-r/100), \bar{x}(1+r/100))$. Equating the desired interval with that obtained with n observations, we can determine n:

$$\bar{x} \mp z\frac{s}{\sqrt{n}} = \bar{x}\left(1 \mp \frac{r}{100}\right)$$

$$z\frac{s}{\sqrt{n}} = \bar{x}\frac{r}{100}$$

$$n = \left(\frac{100zs}{r\bar{x}}\right)^2$$

Here, z is the normal variate of the desired confidence level.

chapter 14

Box 14.1 Simple Linear Regression

1. Model: $y_i = b_0 + b_1 x_i + e_i$
2. Parameter estimation: $b_1 = \dfrac{\Sigma xy - n\bar{x}\,\bar{y}}{\Sigma x^2 - n(\bar{x})^2}$

$$b_0 = \bar{y} - b_1 \bar{x}$$

3. Allocation of variation: $\text{SSY} = \sum_{i=1}^{n} y_i^2$

$$\text{SS0} = n\bar{y}^2$$
$$\text{SST} = \text{SSY} - \text{SS0}$$
$$\text{SSE} = \Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy$$
$$\text{SSR} = \text{SST} - \text{SSE}$$

4. Coefficient of determination $R^2 = \dfrac{\text{SSR}}{\text{SST}} = \dfrac{\text{SST} - \text{SSE}}{\text{SST}}$
5. Standard deviation of errors $s_e = \sqrt{\dfrac{\text{SSE}}{n-2}}$
6. Degrees of freedoms: $\text{SST} = \text{SSY} - \text{SS0} = \text{SSR} + \text{SSE}$

$$n - 1 = n - 1 = 1 + (n-2)$$

(Continued)

Box 14.1 Continued

7. Standard deviation of parameters: $s_{b_0} = s_e \left[\dfrac{1}{n} + \dfrac{\bar{x}^2}{\Sigma x^2 - n\bar{x}^2}\right]^{1/2}$

$$s_{b_1} = \frac{s_e}{\left[\Sigma x^2 - n\bar{x}^2\right]^{1/2}}$$

8. Prediction: Mean of future m observations:

$$\hat{y}_p = b_0 + b_1 x_p$$

$$s_{\hat{y}_p} = s_e \left[\frac{1}{m} + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\Sigma x^2 - n\bar{x}^2}\right]^{1/2}$$

9. All confidence intervals are computed using $t_{[1-\alpha/2;n-2]}$.
10. Model assumptions:
    (a) Errors are independent and identically distributed normal variates with zero mean.
    (b) Errors have the same variance for all values of x
    (c) Errors are additive.
    (d) x and y are linearly related.
    (e) x is nonstochastic and is measured without error.
11. Visual tests:
    (a) Scatter plot of y versus x should be linear.
    (b) Scatter plot of errors versus predicted responses should not have any trends.
    (c) The normal quant le-quantile plot of errors should be linear.

If any test fails or if the io y_{max}/y_{min} is large, curvilinear regressions and transformations sho: ɔe investigated.

Chapter 17

17.2 COMPUTATION OF EFFECTS

In general, any 2^2 design can be analyzed using the method of Example 17.1. In the general case, suppose y_1, y_2, y_3, and y_4 represent the four observed responses. The correspondence between the factor levels and the responses is shown in Table 17.2. The model for a 2^2 design is

$$y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B$$

Substituting the four observations in the model, we get

$$y_1 = q_0 - q_A - q_B + q_{AB}$$

$$y_2 = q_0 + q_A - q_B - q_{AB}$$

$$y_3 = q_0 - q_A + q_B - q_{AB}$$

$$y_4 = q_0 + q_A + q_B + q_{AB}$$

17.4 ALLOCATION OF VARIATION

The importance of a factor is measured by the proportion of the total variation in the response that is explained by the factor. Thus, if two factors explain 90 and 5% of the variation of the response, the second factor may be considered unimportant in many practical situations.

The sample variance of y can be computed as follows:

$$\text{Sample variance of } y = s_y^2 = \frac{\sum_{i=1}^{2^2}(y_i - \bar{y})^2}{2^2 - 1}$$

Here, $\bar{y}$ denotes the mean of responses from all four experiments. The numerator on the right-hand side of the above equation is called the **total variation of y or Sum of Squares Total (SST):**

$$\text{Total variation of } y = \text{SST} = \sum_{i=1}^{2^2}(y_i - \bar{y})^2$$

For a 2^2 design, the variation can be divided into three parts:

$$\text{SST} = 2^2 q_A^2 + 2^2 q_B^2 + 2^2 q_{AB}^2 \qquad (17.1)$$

Before presenting a derivation of this equation, it is helpful to understand its meaning. The three parts on the right-hand side represent the portion of the total variation explained by the effect of A, B, and interaction AB, respectively. Thus, $2^2 q_A^2$ is the portion of SST that is explained by the factor A. It is called the sum of squares due to A and is denoted as SSA. Similarly, SSB is $2^2 q_B^2$ and SSAB (due to interaction AB) is $2^2 q_{AB}^2$. Thus,

$$\text{SST} = \text{SSA} + \text{SSB} + \text{SSAB}$$

These parts can be expressed as a fraction; for example,

$$\text{Fraction of variation explained by } A = \frac{\text{SSA}}{\text{SST}}$$

When expressed as a percentage, the fraction provides an easy way to gauge the importance of the factor A. Those factors which explain a high percentage of variation are considered important

chapter 17 (continued).

17.5 GENERAL 2^k FACTORIAL DESIGNS

A 2^k experimental design is used to determine the effect of k factors, each of which have two alternatives or levels. We have already discussed the special case of two factors ($k = 2$) in the last two sections. Now we generalize the analysis to more than two factors.

The analysis techniques developed so far for 2^2 designs can be easily extended to a 2^k design. Given k factors at two levels each, a total of 2^k experiments are required. The analysis produces 2^k effects. These include k main effects, $\binom{k}{2}$ two-factor interactions, $\binom{k}{3}$ three-factor interactions, and so on. The sign table method of analyzing the results and allocating the variation is also valid. We illustrate this with an example.

The SST can be computed from the effects as follows:

$$\text{SST} = 2^3(q_A^2 + q_B^2 + q_C^2 + q_{AB}^2 + q_{AC}^2 + q_{BC}^2 + q_{ABC}^2$$