Contents

Chapter 1

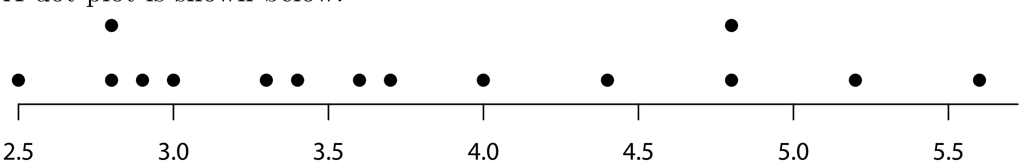
Introduction to Statistics and Data Analysis

1.1 (a) 15.

(b) $\bar{x} = \frac{1}{15}(3.4 + 2.5 + 4.8 + \cdots + 4.8) = 3.787$.

(c) Sample median is the 8th value, after the data is sorted from smallest to largest: 3.6.

(d) A dot plot is shown below.



(e) After trimming total 40% of the data (20% highest and 20% lowest), the data becomes:

2.9	3.0	3.3	3.4	3.6
3.7	4.0	4.4	4.8	

So. the trimmed mean is

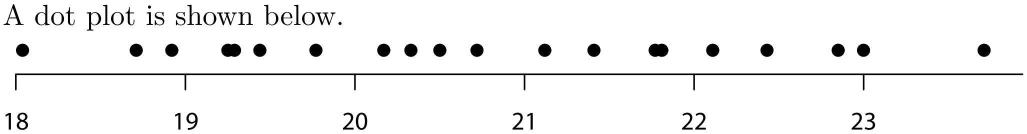
$$\bar{x}_{\text{tr20}} = \frac{1}{9}(2.9 + 3.0 + \cdots + 4.8) = 3.678.$$

(f) They are about the same.

1.2 (a) Mean=20.7675 and Median=20.610.

(b) $\bar{x}_{\text{tr10}} = 20.743$.

(c) A dot plot is shown below.



(d) No. They are all close to each other.

4.67 $E[g(X,Y)] = E(X/Y^3 + X^2Y) = E(X/Y^3) + E(X^2Y)$.
$E(X/Y^3) = \int_1^2 \int_0^1 \frac{2x(x+2y)}{7y^3}\ dx\ dy = \frac{2}{7}\int_1^2 \left(\frac{1}{3y^3} + \frac{1}{y^2}\right)\ dy = \frac{15}{84}$;
$E(X^2Y) = \int_1^2 \int_0^1 \frac{2x^2y(x+2y)}{7}\ dx\ dy = \frac{2}{7}\int_1^2 y\left(\frac{1}{4} + \frac{2y}{3}\right)\ dy = \frac{139}{252}$.
Hence, $E[g(X,Y)] = \frac{15}{84} + \frac{139}{252} = \frac{46}{63}$.

4.68 $P = I^2R$ with $R = 50$, $\mu_I = E(I) = 15$ and $\sigma_I^2 = Var(I) = 0.03$.
$E(P) = E(I^2R) = 50E(I^2) = 50[Var(I) + \mu_I^2] = 50(0.03 + 15^2) = 11251.5$. If we use the approximation formula, with $g(I) = I^2$, $g'(I) = 2I$ and $g''(I) = 2$, we obtain,

$$E(P) \approx 50\left[g(\mu_I) + 2\frac{\sigma_I^2}{2}\right] = 50(15^2 + 0.03) = 11251.5.$$

Since $Var[g(I)] \approx \left[\frac{\partial g(i)}{\partial i}\right]_{i=\mu_I}^2 \sigma_I^2$, we obtain

$$Var(P) = 50^2Var(I^2) = 50^2(2\mu_I)^2\sigma_I^2 = 50^2(30)^2(0.03) = 67500.$$

4.69 For $0 < a < 1$, since $g(a) = \sum\limits_{x=0}^{\infty} a^x = \frac{1}{1-a}$, $g'(a) = \sum\limits_{x=1}^{\infty} xa^{x-1} = \frac{1}{(1-a)^2}$ and
$g''(a) = \sum\limits_{x=2}^{\infty} x(x-1)a^{x-2} = \frac{2}{(1-a)^3}$.

(a) $E(X) = (3/4)\sum\limits_{x=1}^{\infty} x(1/4)^x = (3/4)(1/4)\sum\limits_{x=1}^{\infty} x(1/4)^{x-1} = (3/16)[1/(1-1/4)^2]$
$= 1/3$, and $E(Y) = E(X) = 1/3$.
$E(X^2) - E(X) = E[X(X-1)] = (3/4)\sum\limits_{x=2}^{\infty} x(x-1)(1/4)^x$
$= (3/4)(1/4)^2 \sum\limits_{x=2}^{\infty} x(x-1)(1/4)^{x-2} = (3/4^3)[2/(1-1/4)^3] = 2/9$.
So, $Var(X) = E(X^2) - [E(X)]^2 = [E(X^2) - E(X)] + E(X) - [E(X)]^2$
$2/9 + 1/3 - (1/3)^2 = 4/9$, and $Var(Y) = 4/9$.

(b) $E(Z) = E(X) + E(Y) = (1/3) + (1/3) = 2/3$, and
$Var(Z) = Var(X+Y) = Var(X) + Var(Y) = (4/9) + (4/9) = 8/9$, since X and Y are independent (from Exercise 3.79).

4.70 (a) $g(x) = \frac{3}{2}\int_0^1 (x^2 + y^2)\ dy = \frac{1}{2}(3x^2 + 1)$ for $0 < x < 1$ and
$h(y) = \frac{1}{2}(3y^2 + 1)$ for $0 < y < 1$.
Since $f(x,y) \neq g(x)h(y)$, X and Y are not independent.

(b) $E(X+Y) = E(X) + E(Y) = 2E(X) = \int_0^1 x(3x^2+1)\ dx = 3/4 + 1/2 = 5/4$.
$E(XY) = \frac{3}{2}\int_0^1\int_0^1 xy(x^2+y^2)\ dx\ dy = \frac{3}{2}\int_0^1 y\left(\frac{1}{4} + \frac{y^2}{2}\right)\ dy$
$= \frac{3}{2}\left[\left(\frac{1}{4}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{4}\right)\right] = \frac{3}{8}$.

(c) $Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{2}\int_0^1 x^2(3x^2+1)\ dx - \left(\frac{5}{8}\right)^2 = \frac{7}{15} - \frac{25}{64} = \frac{73}{960}$, and
$Var(Y) = \frac{73}{960}$. Also, $Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{3}{8} - \left(\frac{5}{8}\right)^2 = -\frac{1}{64}$.

(d) $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y) = 2\frac{73}{960} - 2\frac{1}{64} = \frac{29}{240}$.

4.71 (a) $E(Y) = \int_0^\infty ye^{-y/4}\ dy = 4$.

9.9 $n = 20, \bar{x} = 11.3, s = 2.45$, and $t_{0.025} = 2.093$ with 19 degrees of freedom. A 95% confidence interval for the population mean is

$$11.3 - (2.093)(2.45/\sqrt{20}) < \mu < 11.3 + (2.093)(2.45/\sqrt{20}),$$

or $10.15 < \mu < 12.45$.

9.10 $n = 12, \bar{x} = 79.3, s = 7.8$, and $t_{0.025} = 2.201$ with 11 degrees of freedom. A 95% confidence interval for the population mean is

$$79.3 - (2.201)(7.8/\sqrt{12}) < \mu < 79.3 + (2.201)(7.8/\sqrt{12}),$$

or $74.34 < \mu < 84.26$.

9.11 $n = 9, \bar{x} = 1.0056, s = 0.0245$, and $t_{0.005} = 3.355$ with 8 degrees of freedom. A 99% confidence interval for the population mean is

$$1.0056 - (3.355)(0.0245/3) < \mu < 1.0056 + (3.355)(0.0245/3),$$

or $0.978 < \mu < 1.033$.

9.12 $n = 10, \bar{x} = 230, s = 15$, and $t_{0.005} = 3.25$ with 9 degrees of freedom. A 99% confidence interval for the population mean is

$$230 - (3.25)(15/\sqrt{10}) < \mu < 230 + (3.25)(15/\sqrt{10}),$$

or $214.58 < \mu < 245.42$.

9.13 $n = 12, \bar{x} = 48.50, s = 1.5$, and $t_{0.05} = 1.796$ with 11 degrees of freedom. A 90% confidence interval for the population mean is

$$48.50 - (1.796)(1.5/\sqrt{12}) < \mu < 48.50 + (1.796)(1.5/\sqrt{12}),$$

or $47.722 < \mu < 49.278$.

9.14 $n = 15, \bar{x} = 3.7867$, $s = 0.9709$, $1 - \alpha = 95\%$, and $t_{0.025} = 2.145$ with 14 degrees of freedom. So, by calculating $3.7867 \pm (2.145)(0.9709)\sqrt{1 + 1/15}$ we obtain $(1.6358, 5.9376)$ which is a 95% prediction interval the drying times of next paint.

9.15 $n = 100, \bar{x} = 23{,}500, s = 3,900, 1-\alpha = 0.99$, and $t_{0.005} \approx 2.66$ with 60 degrees of freedom (use table from the book) or $t_{0.005} = 2.626$ if 100 degrees of freedom is used. The prediction interval of next automobile will be driven in Virginia (using 2.66) is $23{,}500 \pm (2.66)(3{,}900)\sqrt{1 + 1/100}$ which yields $13{,}075 < \mu < 33{,}925$ kilometers.

9.16 $n = 12, \bar{x} = 79.3, s = 7.8$, and $t_{0.025} = 2.201$ with 11 degrees of freedom. A 95% prediction interval for a future observation is

$$79.3 \pm (2.201)(7.8)\sqrt{1 + 1/12} = 79.3 \pm 17.87,$$

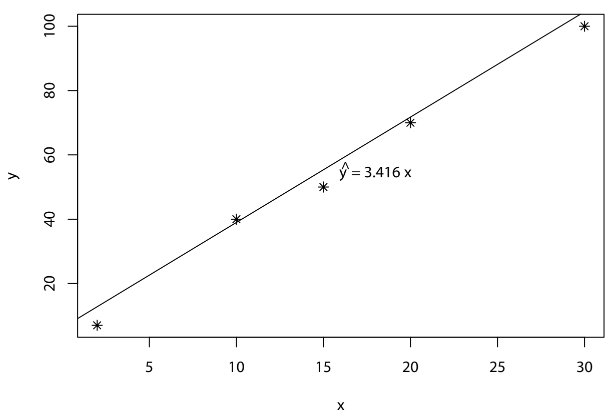
which yields $(61.43, 97.17)$.

Taking derivative of the above with respect to b_1 and setting the derivative to zero, we have $-2\sum\limits_{i=1}^{n} x_i(y_i - b_1x_i) = 0$, which implies $b_1 = \frac{\sum\limits_{i=1}^{n} x_iy_i}{\sum\limits_{i=1}^{n} x_i^2}$.

(b) $\sigma^2_{B_1} = \frac{Var\left(\sum\limits_{i=1}^{n} x_iY_i\right)}{(\sum\limits_{i=1}^{n} x_i^2)^2} = \frac{\sum\limits_{i=1}^{n} x_i^2\sigma^2_{Y_i}}{(\sum\limits_{i=1}^{n} x_i^2)^2} = \frac{\sigma^2}{\sum\limits_{i=1}^{n} x_i^2}$, since Y_i's are independent.

(c) $E(B_1) = \frac{E\left(\sum\limits_{i=1}^{n} x_iY_i\right)}{\sum\limits_{i=1}^{n} x_i^2} = \frac{\sum\limits_{i=1}^{n} x_i(\beta_1x_i)}{\sum\limits_{i=1}^{n} x_i^2} = \beta_1$.

11.29 (a) The scatter plot of the data is shown next.



(b) $\sum\limits_{i=1}^{n} x_i^2 = 1629$ and $\sum\limits_{i=1}^{n} x_iy_i = 5564$. Hence $b_1 = \frac{5564}{1629} = 3.4156$. So, $\hat{y} = 3.4156x$.

(c) See (a).

(d) Since there is only one regression coefficient, β_1, to be estimated, the degrees of freedom in estimating σ^2 is $n-1$. So,

$$\hat{\sigma}^2 = s^2 = \frac{SSE}{n-1} = \frac{\sum\limits_{i=1}^{n}(y_i - b_1x_i)^2}{n-1}.$$

(e) $Var(\hat{y}_i) = Var(B_1x_i) = x_i^2 Var(B_1) = \frac{x_i^2\sigma^2}{\sum\limits_{i=1}^{n} x_i^2}$.

(f) The plot is shown next.

14.2 The hypotheses of the three parts are,

(a) for the main effects brands,

$$H_0^{'}: \alpha_1 = \alpha_2 = \alpha_3 = 0,$$
$$H_1^{'}: \text{At least one of the } \alpha_i\text{'s is not zero;}$$

(b) for the main effects times,

$$H_0^{''}: \beta_1 = \beta_2 = \beta_3 = 0,$$
$$H_1^{''}: \text{At least one of the } \beta_i\text{'s is not zero;}$$

(c) and for the interactions,

$$H_0^{'''}: (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{33} = 0,$$
$$H_1^{'''}: \text{At least one of the } (\alpha\beta)_{ij}\text{'s is not zero.}$$

$\alpha = 0.05$.
Critical regions: (a) $f_1 > 3.35$; (b) $f_2 > 3.35$; and (c) $f_3 > 2.73$.
Computations: From the computer printout we have

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Brands	32.7517	2	16.3758	1.74
Times	227.2117	2	113.6058	12.04
Interaction	17.3217	4	4.3304	0.46
Error	254.7025	27	9.4334	
Total	531.9875	35		

Decision: (a) Do not reject $H_0^{'}$; (b) Reject $H_0^{''}$; (c) Do not reject $H_0^{'''}$.

14.3 The hypotheses of the three parts are,

(a) for the main effects environments,

$$H_0^{'}: \alpha_1 = \alpha_2 = 0, \text{ (no differences in the environment)}$$
$$H_1^{'}: \text{At least one of the } \alpha_i\text{'s is not zero;}$$

(b) for the main effects strains,

$$H_0^{''}: \beta_1 = \beta_2 = \beta_3 = 0, \text{ (no differences in the strains)}$$
$$H_1^{''}: \text{At least one of the } \beta_i\text{'s is not zero;}$$

(c) and for the interactions,

$$H_0^{'''}: (\alpha\beta)_{11} = (\alpha\beta)_{12} = \cdots = (\alpha\beta)_{23} = 0, \text{ (environments and strains do not interact)}$$
$$H_1^{'''}: \text{At least one of the } (\alpha\beta)_{ij}\text{'s is not zero.}$$

$\alpha = 0.01$.
Critical regions: (a) $f_1 > 7.29$; (b) $f_2 > 5.16$; and (c) $f_3 > 5.16$.
Computations: From the computer printout we have