FORK 1002 Statistics
Exercise set 1: Basic statistics

1. Exercises on descriptive statistics.

   Consider the following sample of observations from a study of the effect of Viagra on a variable called "libido":

Person i	Libido (y_i)	Dose (x_i)
1	3	1
2	2	1
3	1	1
4	1	1
5	4	1
6	5	2
7	2	2
8	4	2
9	2	2
10	3	2
11	7	3
12	4	3
13	5	3
14	3	3
15	6	3

Source: Page 350 in Andy Field (2009), *Discovering Statistics using SPSS*, 3rd. Edition, SAGE

   $x_i = 1$: Placebo. $x_i = 2$: Low dose of Viagra. $x_i = 3$: High dose of Viagra.

   (a) Let the sum of the libido values be equal to 52, that is, $\sum y_i = 52$. What is the sample mean of libido?

   (b) What is the sample mean of libido for those who received placebo?

   (c) What is the sample mean of libido for those who received viagra (that is, either a low or high dose)?

   (d) Let $\sum_{i=1}^{15}(y_i - \overline{y})^2 = 43.73$, where $\overline{y}$ is the sample mean of y_i. Compute the sample variance and sample standard deviation

   (e) What is the median value of libido?

   (f) What is the mode of libido?

   (g) What are the maximum and minimum values of libido, and what is the sample range?

2. Exercises on weighted sums.

   The table below lists the possible returns associated with an investment, and the probability associated with each possible return. A negative return means a loss.

Return in %	Probability
-20	0.10
-10	0.15
10	0.45
25	0.25
30	0.05

   (a) Compute the weighted return using the probabilities as weights

   (b) Compute the weighted variance using the probabilities as weights

   (c) Compute the weighted standard deviation

3. More exercises on sums.

   Let x_i and y_i be defined as in exercise 1 (the Viagra sample), and compute the following sums:

   (a) $\sum_{i=1}^{4} y_i$

   (b) $\sum_{i=2}^{5} y_i^{i-1}$

   (c) $\sum_{i=3}^{6} ay_i$, where $a = 0,5$

   (d) $\sum_{i=5}^{8} a_i y_i$, where $a_5 = 0,2, a_6 = 0,1, a_7 = 0,4, a_8 = 0,3$

   (e) $\sum_{i=1}^{2} (2y_i + 3x_i)$

   (f) $\sum_{i=3}^{5} (i + 3)$

   (g) $\sum_{i=10}^{12} x_i y_i$

   (h) $\sum_{i=1}^{3} \sum_{j=1}^{2} x_i y_j$

4. Exercises on the normal distribution.

   Let z denote the value of a test expression that is normally distributed:

   (a) What is the probability that z is less than: i) 0? ii) 1? iii) 1.9? iv) -1.2?

   (b) What is the probability that z is greater than: i) 0.5? ii) 1.3? iii) 1.6? iv) -1.9?

   (c) What is the probability that z lies between: i) -0.3 and 0.2? ii) -1.7 and -0.1? iii) -0.4 and 0.9?

5. Exercises on the t-distribution.

   Let $t(df)$ denote the value of a test expression that is t-distributed with df degrees of freedom:

   (a) What is the probability that: i) $t(9)$ is greater than 1.1? ii) $t(15)$ is greater than 1.341? iii) $t(40)$ is greater than 1.684? iv) $t(80)$ is greater than 1.99?

   (b) What is the probability that: i) $t(30)$ is less than 1.055? ii) $t(50)$ is less than 2.403? iii) $t(60)$ is less than -1.045? iv) $t(20)$ is less than 0?

   (c) What is the probability that: i) $t(13)$ lies between -0.870 and 1.35? ii) $t(27)$ lies between -1.703 and 1.703? iii) $t(40)$ lies between -0.851 and 0.681?

6. Exercises on confidence intervals.

   A $100 \cdot (1-\alpha)\%$ confidence interval for a population mean is usually computed as

$$\begin{aligned} U &= \overline{y} + t_{\alpha/2}(df) \cdot s/\sqrt{n} \\ L &= \overline{y} - t_{\alpha/2}(df) \cdot s/\sqrt{n}, \end{aligned}$$

   where U and L are the upper and lower bounds of the interval, respectively, $df = n-1$ is the degrees of freedom, s is the sample standard deviation and n is the number of observations

   (a) Compute a 90% interval for the population mean of libido

   (b) Compute a 95% interval for the population mean of libido

   (c) Compute a 99% interval for the population mean of libido

7. Exercise on testing the difference between two means.

   The hypothesis that Viagra has an effect on libido can be tested with the test expression

$$\frac{\overline{y}_1 - \overline{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

   where $\overline{y}_1$ is the mean libido for those who received placebo, $\overline{y}_2$ is the mean libido for those who received viagra, n_1 and n_2 are the sample sizes, and $s_1^2 = 1.7$ and $s_2^2 = 2.77$ are the sample variances. Suppose the test expression is t-distributed with $df = n_{min} - 1$, where n_{min} is the smallest of n_1 and n_2. Perform the test using a 1% significance level:

   (a) Define the null and alternative hypotheses

   (b) Obtain the critical values and the rejection area

(c) Compute the value of the test expression

(d) Conclude

8. Exercises on p-values.

   Suppose your null hypothesis is $H_0 : \mu = 0$:

   (a) If the alternative hypothesis is $H_1 : \mu > 0$, and if the value of a normally distributed test expression is 0.64, then what is the p-value of the test?

   (b) If the alternative hypothesis is $H_1 : \mu < 0$, and if the value of a normally distributed test expression is 0.15, then what is the p-value of the test?

   (c) If the alternative hypothesis is $H_1 : \mu \neq 0$, and if the value of a normally distributed test expression is -0.31, then what is the p-value of the test?

   (d) If the alternative hypothesis is $H_1 : \mu > 0$, and if the value of a $t(19)$-distributed test expression is 1.729, then what is the p-value of the test?

   (e) If the alternative hypothesis is $H_1 : \mu < 0$, and if the value of a $t(19)$-distributed test expression is -1.729, then what is the p-value of the test?

   (f) If the alternative hypothesis is $H_1 : \mu < 0$, and if the value of a $t(35)$-distributed test expression is -0.682, then what is the p-value of the test?

   (g) If the alternative hypothesis is $H_1 : \mu \neq 0$, and if the value of a $t(21)$-distributed test expression is 2.518, then what is the p-value of the test?

9. Computer exercises:

   (a) Load the dataset *viagra.xls*. [Hint for SPSS: File → Open → Data]

   (b) Compute the sample mean, sample standard deviation and sample range of the variable *libido*. [Hint for SPSS: Analyze → Descriptive Statistics → Descriptives..., put *libido* into the "Variable(s)" box, make sure "Mean", "Std. Deviation", "Range", "Maximum" and "Minimum" are ticked via "Options", then press "Continue", press "OK"]

   (c) Make a new variable equal to $dose - 1$ and call it *dosenew* [Hint for SPSS: Transform → Compute Variable..., write "dosenew" in the "Target Variable" box, write "dose-1" in the "Numeric Expression" box, press "OK"]

   (d) Compute a 90% confidence interval for the mean of *libido*. [Hint for SPSS: Analyze → Compare Means → One Sample T Test, select the *libido* variable, click on "Options", write "90" in the "Confidence Interval Percentage" box, press "Continue", press "OK"]

   (e) Make a bar diagram of *libido* [Hint for SPSS: Graphs → Chart Builder..., choose "Bar" in the "Choose From" box, drag the chosen bar type into the "Chart preview..." box, drag *libido* into the "x-axis" area, press "OK"]