IPDA1005 Introduction to Probability and Data Analysis

Test 2 Practice Questions with Solution

1. The heights of boys in a primary school is known to be normally distributed. The average height is 130cm and the standard deviation is 5cm.

   (a) Find the probability that a randomly chosen boy has height

   i. **[3 marks]** more than 135cm

   Solution:

   $$\begin{aligned} P(X > 135) &= P\left(Z > \frac{135 - 130}{5}\right) \\ &= P(Z > 1) \\ &= 0.1587 \end{aligned}$$

   ii. **[3 marks]** between 125cm and 140cm

   Solution:

   $$P(125 < X < 140) = P\left(-1 < Z < 2\right) = 0.8186$$

   (b) **[4 marks]** Below what value are the heights of 95% of boys?

   Solution: $0.95 = P(X < x) = P\left(Z < \frac{x-130}{5}\right)$, so $\frac{x-130}{5} = 1.645$, and hence $x = 138.225$.

2. Assume that X has normal distribution with mean 2 and variance 4. Find

   (a) **[3 marks]** $P(X \leq 4)$

   Solution: $P(X \leq 4) = P\left(Z \leq \frac{4-2}{2}\right) = P(Z \leq 1) = .8413$

   (b) **[3 marks]** $P(X < -1)$

Solution: $P(X < -1) = P\left(Z \le \frac{-1-2}{2}\right) = P(Z \le -1.5) = .0668$

(c) [**4 marks**] $P(0 < X < 3)$

Solution: $P(0 < X < 3) = P\left(\frac{0-2}{2} \le Z \le \frac{3-2}{2}\right) = P(-1 \le Z \le 0.5) = 0.6915 - 0.1587) = 0.5328$

(d) [**4 marks**] an interval around the mean where X will lie with probability 0.9.

Solution: $P(\mu - a\sigma < X < \mu + a\sigma) = .9 \Rightarrow P(-a < Z < a) = .9 \Rightarrow P(0 < Z < a) = .45 \Rightarrow a = 1.645$, so $P(2 - (1.645)(2) < X < 2 + (1.645)(2)) = .9$. Thus the desired interval is $(2 - (1.645)(2), 2 + (1.645)(2)) = (-1.29, 5.29)$

3. [**4 marks**] If the height of men in a particular country is normally distributed with unknown mean and standard deviation, what is the probability that the height of a randomly chosen man falls within 1.5 standard deviations of the mean?

Solution:

$P(\mu - 1.5\sigma < X < \mu + 1.5\sigma) = P(-1.5 < Z < 1.5) = 2P(0 < Z < 1.5) = 2(.4332) = .8664$

4. The net weight of a 100 gram can of tuna fish is normally distributed with mean 105 grams and standard deviation 3 grams.

(a) [**3 marks**] What proportion of all 100 gram cans will actually have net weight 100 grams or more?

Solution: $P(X \ge 100) = P\left(Z \ge \frac{100-105}{3}\right) = P\left(Z \ge \frac{-5}{3}\right) = P\left(Z \le \frac{5}{3}\right) = .9525$

(b) [**4 marks**] What should the mean be changed to if the manufacturers wanted the proportion of cans that weigh 100 grams or more to be 0.95?

Solution: $P(X \ge 100) = .95 \Rightarrow P(Z \ge \frac{100-\mu}{\sigma}) = .95 \Rightarrow P(Z \le \frac{100-\mu}{3}) = .05 \Rightarrow \frac{100-\mu}{\sigma} = -1.645 \Rightarrow \mu = 104.935$.

(c) [**3 marks**] If 5 cans are selected at random, what is the probability that at least one of them will have net weight less than 100 grams?

> **Solution:** $1 - .9525^5 = .2160$

5. A city with a large population contains 30% immigrant workers. A sample of size 100 was taken from the city's population, and let X be the number of immigrant workers in the sample. Using a suitable approximation, compute the following probabilities.

   (a) [**4 marks**] $P(X \geq 40)$

   > **Solution:** Here the condition that np and $n(1-p)$ are both at least 10 is satisfied, so the appropriate approximation is normal distribution with mean $\mu = np = 30$ and variance $\sigma^2 = np(1-p) = 21$. $P(X \geq 40) = P(X \geq 39.5) \approx P\left(Z \geq \frac{39.5-30}{\sqrt{21}}\right) = P(Z \geq 2.07) = 0.0192$

   (b) [**4 marks**] $P(25 \leq X \leq 32)$

   > **Solution:** $P(25 \leq X \leq 32) \approx P(24.5 \leq X \leq 32.5) \approx P\left(\frac{24.5-30}{\sqrt{21}} \leq Z \leq \frac{32.5-30}{\sqrt{21}}\right) = P(-1.2 \leq Z \leq .546) = 0.5924$

6. A student attempts 48 true-false questions in a test. For each question, independently, he has $\frac{3}{4}$ probability of answering correctly. Using an appropriate approximation, find the probability that he answers

   (a) [**4 marks**] at least 40 questions correctly

   > **Solution:**
   > Here the condition that np and $n(1-p)$ are both at least 10 is satisfied, so the appropriate approximation is normal distribution with mean $\mu = np = 36$ and variance $np(1-p) = 9$. $P(X \geq 40) = P(X \geq 39.5) \approx P\left(Z \geq \frac{39.5-36}{3}\right) = P(Z \geq 1.17) = 1 - .8790 = .1210$

   (b) [**4 marks**] exactly 37 questions correctly.

   > **Solution:** $P(X = 37) = P(36.5 \leq X \leq 37.5) \approx P\left(\frac{36.5-36}{3} \leq Z \leq \frac{37.5-36}{3}\right) = P(.167 \leq Z \leq .5) = .1251$

7. From a bag containing 3 black balls, 2 blue balls and 3 green balls, a random sample of 4 balls is selected. Let X be the number of black balls and Y be the number of blue balls.

   (a) Find the joint probability distribution of X and Y.

   (b) Find the marginal distributions of X and Y.

(c) Find the conditional distributions of Y given $X = x$ for each value of x. Calculate the corresponding conditional expectations.

(d) Decide whether X and Y are independent.

(e) Compute the covariance and correlation between X and Y.

Solution:

(a) The number of different ways of selecting 4 balls from the 8 balls is $\binom{8}{4} = 70$. So

$$P(X = x, Y = y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{4-x-y}}{70}$$

So the joint probability distribution is given by

	0	1	2	3	Total
0	0	$\frac{3}{70}$	$\frac{9}{70}$	$\frac{3}{70}$	$\frac{15}{70}$
1	$\frac{2}{70}$	$\frac{18}{70}$	$\frac{18}{70}$	$\frac{2}{70}$	$\frac{40}{70}$
2	$\frac{3}{70}$	$\frac{9}{70}$	$\frac{3}{70}$	0	$\frac{15}{70}$
Total	$\frac{5}{70}$	$\frac{30}{70}$	$\frac{30}{70}$	$\frac{5}{70}$	1

(b) From the Total row and Total column, we get the marginal distributions of X and Y to be

x	0	1	2	3
$p_X(x)$	$\frac{5}{70}$	$\frac{30}{70}$	$\frac{30}{70}$	$\frac{5}{70}$

and the marginal distribution of Y is given by

y	0	1	2
$p_Y(y)$	$\frac{15}{70}$	$\frac{40}{70}$	$\frac{15}{70}$

(c) The conditional distribution of Y given $X = 0$ is given by

y	1	2
$p_{Y\|X}(y\|0)$	$\frac{2}{5}$	$\frac{3}{5}$

$E(Y|X = 0) = \frac{8}{5} = 1.6$.

The conditional distribution of Y given $X = 1$ is given by

y	0	1	2
$p_{Y\|X}(y\|1)$	$\frac{1}{10}$	$\frac{3}{5}$	$\frac{3}{10}$

$E(Y|X=1) = \frac{6}{5} = 1.2.$

The conditional distribution of Y given $X = 2$ is given by

y	0	1	2
$p_{Y\|X}(y\|2)$	$\frac{3}{10}$	$\frac{3}{5}$	$\frac{1}{10}$

$E(Y|X=2) = \frac{4}{5} = 0.8.$

The conditional distribution of Y given $X = 3$ is given by

y	0	1
$p_{Y\|X}(y\|3)$	$\frac{3}{5}$	$\frac{2}{5}$

$E(Y|X=3) = \frac{2}{5} = 0.4.$

(d) As the conditional distributions of Y are different from the unconditional distribution, X and Y are not independent.

(e)

$$E(XY) = 0\left(\frac{20}{70}\right) + 1\left(\frac{18}{70}\right) + 2\left(\frac{27}{70}\right) + 3\left(\frac{2}{70}\right) + 4\left(\frac{3}{70}\right) = \frac{9}{7}$$

$$E(X) = 0\left(\frac{5}{70}\right) + 1\left(\frac{30}{70}\right) + 2\left(\frac{30}{70}\right) + 3\left(\frac{5}{70}\right) = \frac{3}{2}$$

$$E(X^2) = 0\left(\frac{5}{70}\right) + 1\left(\frac{30}{70}\right) + 4\left(\frac{30}{70}\right) + 9\left(\frac{5}{70}\right) = \frac{39}{14}$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{39}{14} - \left[\frac{3}{2}\right]^2 = \frac{15}{28}$$

$$E(Y) = 0\left(\frac{15}{70}\right) + 1\left(\frac{40}{70}\right) + 2\left(\frac{15}{70}\right) = 1$$

$$E(Y^2) = 0\left(\frac{15}{70}\right) + 1\left(\frac{40}{70}\right) + 4\left(\frac{15}{70}\right) = \frac{10}{7}$$

$$\text{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{10}{7} - 1 = \frac{3}{7}$$

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{9}{7} - \frac{3}{2} = \frac{-3}{14}$$

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\text{SD}(X)\text{SD}(Y)} = \frac{\frac{-3}{14}}{\sqrt{\left(\frac{15}{28}\right)\left(\frac{3}{7}\right)}} = \frac{-1}{\sqrt{5}}$$

8. Joint distribution of two random variables X and Y is given by

		x		
		1	2	3
	1	0.05	0.1	0.1
y	2	0.05	0.1	0.1
	3	0.1	0.2	0.2

(a) Find the marginal distribution of Y.

(b) Find the conditional distribution of X given $Y = 2$.

(c) Decide whether X and Y are independent, justifying your answer.

(d) What is the covariance between X and Y?

Solution:

(a) The marginal distribution of Y is given by

y	1	2	3
$p_Y(y)$	0.25	0.25	0.5

(b) The conditional distribution of X given $Y = 2$ is given by

x	1	2	3
$p_{X\|Y}(x\|2)$	0.2	0.4	0.4

(c) As $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all values of x and y, X and Y are independent.

(d) As X and Y are independent, $Cov(X, Y) = 0$.

9. Let X and Y have joint density

$$f(x, y) = \frac{6 - x - y}{8}, \quad 0 \le x \le 2, 2 \le y \le 4.$$

(a) Find $f_X(x)$.

Solution: For $0 \le x \le 2$,

$$\begin{aligned} f_X(x) &= \frac{1}{8}\int_2^4 (6 - x - y)dy \\ &= \frac{1}{8}\left[(6 - x)y - \frac{y^2}{2}\right]_2^4 \\ &= \frac{1}{8}[2(6 - x) - 6] \\ &= \frac{3 - x}{4}. \end{aligned}$$

(b) Find $P(1 \le Y \le 3|X = 2)$.

Solution: $f_{Y|X}(y|x) = \frac{6-x-y}{2(3-x)}$ for $0 \le x \le 2, 2 \le y \le 4$. Hence $f_{Y|X}(y|2) = \frac{4-y}{2}$ for $2 \le y \le 4$. Thus

$$P(1 \le Y \le 3|X = 2) = \int_2^3 \frac{4-y}{2} dy = 2 - \frac{5}{4} = \frac{3}{4}.$$

10. The random variables X and Y represent the proportion of marks students obtained in Test 1 and Test 2, respectively, and they have joint pdf

$$f(x, y) = \frac{12}{7} x(x + y), 0 \le x \le 1, 0 \le y \le 1.$$

(a) Find the marginal densities of X and Y.

Solution:

$$f_X(x) = \int_0^1 \frac{12}{7} x(x + y) dy = \frac{6}{7} x(2x + 1), \quad 0 \le x \le 1$$

$$f_Y(y) = \int_0^1 \frac{12}{7} x(x + y) dx = \frac{2}{7}(3y + 2), \quad 0 \le y \le 1$$

(b) Find the expectations and variances of X and Y.

Solution:

$$E(X) = \int_0^1 \frac{6}{7} x^2(2x + 1) = \frac{5}{7},$$

$$E(Y) = \int_0^1 \frac{2}{7}(3y^2 + 2y) dy = \frac{4}{7},$$

$$E(X^2) = \int_0^1 \frac{6}{7} x^3(2x + 1) = \frac{39}{70}$$

and

$$E(Y^2) = \int_0^1 \frac{2}{7}(3y^3 + 2y^2) dy = \frac{17}{42}.$$

So the variances are given by

$$\text{Var}(X) = \frac{39}{70} - \frac{25}{49} = \frac{23}{490},$$

$$\text{Var}(Y) = \frac{17}{42} - \frac{16}{49} = \frac{23}{294}.$$

(c) Find $E(XY)$.

Solution:
$$E(XY) = \int_0^1 \int_0^1 \frac{12}{7} x^2 y(x+y) dx dy = \frac{17}{42}$$

(d) Find the covariance between X and Y.

Solution:

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{17}{42} - \left(\frac{5}{7}\right)\left(\frac{4}{7}\right) = -\frac{1}{294}$$

(e) Find the correlation between X and Y.

Solution:
$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{SD(X)SD(Y)} = 0.05613$$

11. Let X and Y be two random variables with $V(X) = 1$, $V(Y) = 4$, $Cov(X,Y) = -1$. Find

(a) $Cov(2X+4, 4Y+2)$

Solution: $Cov(2X+4, 4Y+2) = 2(4)Cov(X,Y) = -8$

(b) $Cov(2X+1, 4-2X)$

Solution: $Cov(2X+1, 4-2X) = 2(-2)Cov(X,X) = (-4)V(X) = -4$

(c) $V(2X-3Y)$

Solution: $V(2X-3Y) = 4V(X) + 9V(Y) - 12Cov(X,Y) = 4 + 36 + 12 = 52$

(d) $Corr(X,Y)$

Solution: $Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)} = \frac{-1}{(1)(2)} = \frac{-1}{2}$

(e) $Corr(3X+1, -2Y+2)$

Solution: $Corr(3X+1, -2Y+2) = \frac{1}{2}$

(f) Are X and Y independent?

Solution: As the covariance between them is not zero, X and Y cannot be independent.

12. Consider the following sample data set.

$$5.5,\ 14.4,\ 10.7,\ 7.7,\ 6.3,\ 6,\ 7.6,\ 5.8,\ 7.8,\ 7.3,\ 2.0,\ 12.5,\ 9.1$$

(a) Compute the mean and variance of the sample.

Solution:

$$\overline{x} = \frac{1}{n}\sum x_i = \frac{102.7}{13} = 7.9$$

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\overline{x}^2\right) = \frac{1}{12}\left(935.67 - 13(7.9^2)\right) = 10.362$$

(b) Calculate the median and the interquartile range.

Solution: Data in ascending order are

$$2.0,\ 5.5,\ 5.8,\ 6,\ 6.3,\ 7.3,\ 7.6,\ 7.7,\ 7.8,\ 9.1,\ 10.7,\ 12.5,\ 14.4$$

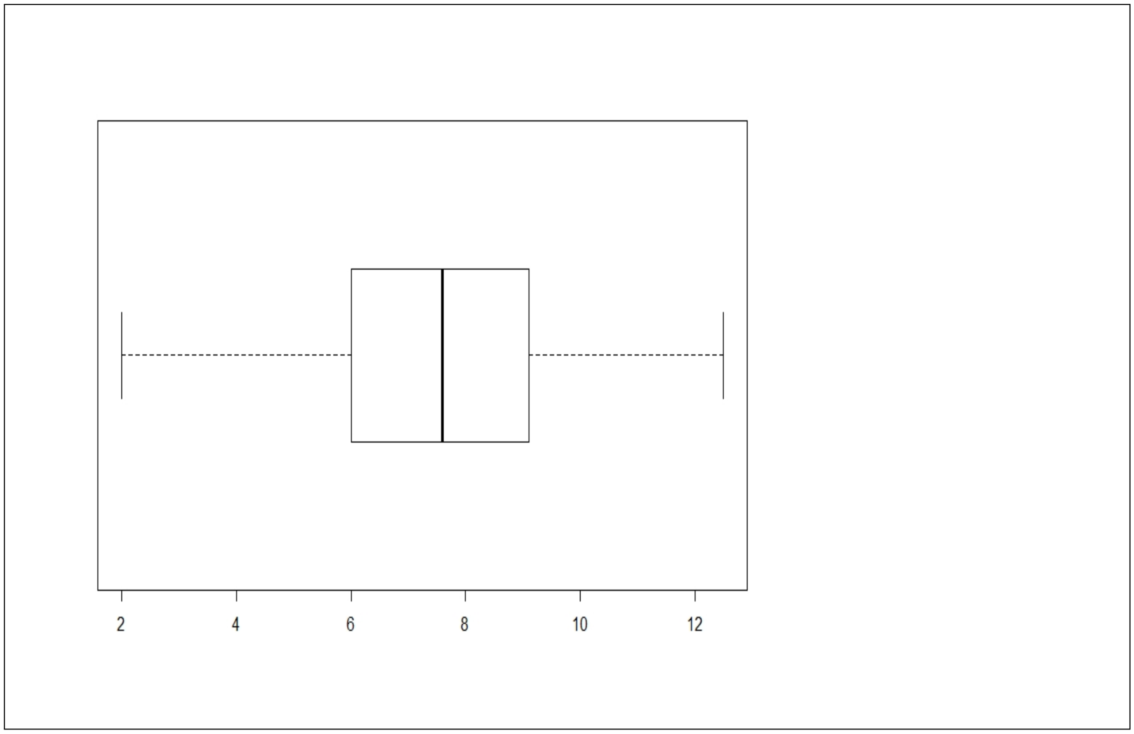
Median $= 7.6$. $Q_1 = \frac{5.8+6}{2} = 5.9$ and $Q_3 = \frac{9.1+10.7}{2} = 9.9$, so $IQR = 9.9 - 5.9 = 4$.

(c) Draw a left-to-right boxplot for the data and identify any outliers.

Solution:

$$(Q_1 - 1.5IQR, Q_3 + 1.5IQR) = (-0.1, 15.9)$$

All observations are within this interval, so there are no outliers.



13. In a large US university, it is known that the proportion of females among the students is **0.47**, and the mean GPA (Grade Point Average) of all students is **3.4**. A sample of 150 students was taken at random, and it was found that they had an average GPA of **3.6** and a standard deviation of **0.8**. It was also found that **51%** of the selected students were females.

    (a) For each number in boldface, state what it represents and decide whether it is a parameter or a statistic.

    > **Solution:** **0.47** and **3.4** are parameters, while **3.6**, **0.8** and **51%** are statistics. **0.47** is the population proportion of females, **3.4** is the population mean of GPA, **3.6** and **0.8** are the sample mean and sample standard deviation of GPA, and **51%** is the sample proportion of females.

    (b) The quantities that you identified as statistics, decide what they estimate, if any.

    > **Solution:** **3.6**, **0.8** and **51%** estimate the population mean of GPA, population standard deviation of GPA and population proportion of females respectively.

14. The following is the hours of sleep 6 patients had before and after taking a sleep inducing drug. We want to test the drug is effective.

Patient	1	2	3	4	5	6
Before	5	4	6	5.5	6	6.5
After	6.5	4.5	5	6	7.5	4

(a) Set up appropriate hypotheses and test at 1% level. State the p-value.

(b) Construct a 90% confidence interval for the mean difference.

(c) What is the sample size needed for the 90% confidence interval to have a margin of error no more than 0.5?

Solution:

(a) The samples are not independent and a paired t-test is appropriate here. The null and alternative hypotheses are given by $H_0 : \mu_d = 0$ vs. $H_0 : \mu_d > 0$.

The values for the difference d are given by 1.5, 0.5, -1, 0.5, 1.5 and -2.5. The mean and the standard deviation of this data set are $\overline{d} = 0.0833$ and $s = 1.56$.

$$t = \frac{\overline{d} - 0}{\frac{1.56}{\sqrt{6}}} = 0.13$$

with a high p-value, so have no reason to reject the null hypothesis. The drug has no effect.

The p-value is given by `1-pt(0.13,5)` $= 0.4508$

(b) As $t_{5,.05} = 2.015$, the 90% confidence interval for the mean difference is given by

$$0.0833 \pm 2.015 \left(\frac{1.56}{\sqrt{6}} \right) = 0.0833 \pm 1.2833.$$

(c) The required sample size is

$$n = \left(\frac{1.645(1.56)}{.5} \right)^2 = 26.34.$$

Rounding up, we need 27 observations.

15. A telephone survey was conducted to estimate p, the proportion of households with no children. A random sample of size 500 yielded 220 households with no children. Test the hypothesis that $p = 0.48$ against a two-sided alternative at the 10% and 5% levels of significance using

(a) the critical value method

(b) the p-value method.

For $\alpha = 0.05$, state your conclusion.

Solution: $H_0 : p = 0.48$ vs. $H_1 : p \neq 0.48$.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.44 - .48}{\sqrt{\frac{.48(.52)}{500}}} = -1.79.$$

(a) We reject H_0 if $|Z| \geq z_{\frac{\alpha}{2}}$. As $z_{.05} = 1.645$ and $z_{.025} = 1.96$, we do not reject H_0 at 5% level, but we do at 10% level.

(b) The p-value is $2P(Z > 1.79) = 2(0.0367) = .0734$. We do not reject H_0 at 5% level, but we do at 10% level.

At the 5% level, we do not have sufficient evidence to conclude that the proportion of households with no children is not 0.48.

16. A random sample of size 25 was taken from a normal distribution with mean 10 and variance 16. Let $\overline{X}$ and S denote the sample mean and the sample standard deviation respectively.

(a) State the sampling distribution of S^2.

(b) Using `R`, compute $P\left(\overline{X} \leq 10.8\right)$.

(c) Using `R`, compute $P\left(S \geq 5\right)$.

Solution:

(a) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \Rightarrow \frac{24S^2}{16} \sim \chi^2_{24} \Rightarrow S^2 \sim \left(\frac{2}{3}\right)\chi^2_{24}$

(b) $P\left(\overline{X} \leq 10.8\right)$ is given by `pnorm(10.8,10,4/5)` $= .8413$.

(c) $P\left(S \geq 5\right) = P\left(\frac{24S^2}{16} \geq 37.5\right) = 0.0390$ from `1-pchisq(37.5,24)`.

17. A food inspector examined 12 jars of a certain brand of peanut butter and determined the percentages of impurities. The result is given below:

$$2.3, 1.9, 2.1, 2.8, 2.3, 3.6, 1.4, 1.8, 2.1, 3.2, 2.0, 1.9$$

Construct a 90% confidence interval for the standard deviation.

Solution: From the data, we get $s = 0.625$. From the table, the 11 d.f. χ^2 values for 0.05 and 0.95 are 19.675 and 4.575. So the confidence interval for the variance is given by

$$\left(\frac{11 \times 0.625^2}{19.675}, \frac{11 \times 0.625^2}{4.575}\right) = (0.2184, 0.939)\,.$$

Thus the confidence interval for the standard deviation is $(.47, .97)$.

18. From past experience, it is assumed that standard deviation of measurements on sheet metal stampings is 0.41. A new set of 30 stampings are used to test the accuracy of this assumption, and a sample standard deviation of 0.49 was obtained. Test the hypothesis that $\sigma = 0.41$ against the alternative hypothesis that $\sigma > 0.41$. Use the 0.05 level of significance. State the p-value.

Solution: $H_0 : \sigma = 0.41$ vs. $H_1 : \sigma > 0.41$. $n = 30, s = 0.49, \alpha = 0.05$. We reject the null hypothesis if $\chi^2 \geq \chi^2_{29,.05} = 42.557$. $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{29(.49)^2}{.41^2} = 41.42 < 42.557$, so do not reject H_0.

The p-value is obtained by `1-pchisq(41.42,29)` to be 0.0633.

19. The amount of soup in the 200ml cans is normally distributed with mean 205ml and standard deviation 10ml.

(a) For a randomly selected can of soup, find the probability that the amount of soup in it falls below the claimed 200ml.

Solution: Let X be the amount of soup. $P(X \leq 200) = P\left(Z \leq \frac{200-205}{10}\right) = P(Z \leq -.5) = .3085$

(b) For a randomly selected can of soup, find the probability that the amount of soup in it falls between 190ml and 210ml.

Solution:

$$\begin{aligned} P(190 < X < 210) &= P\left(\frac{190-205}{10} \leq Z \leq \frac{210-205}{10}\right) \\ &= P(-1.5 \leq Z \leq 0.5) = 0.6915 - 0.0668 \\ &= 0.6247 \end{aligned}$$

(c) A random sample of size 50 was drawn from this population.

i. What is the mean and standard deviation of the sample mean $\overline{X}$?

> **Solution:** $E(\overline{X}) = \mu = 205$ and $SD(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{50}} = \sqrt{2} = 1.414$.

ii. State the distribution of $\overline{X}$.

> **Solution:** Normal with mean 205 and variance $\frac{100}{50} = 2$.

iii. Compute $P\left(\overline{X} < 203\right)$

> **Solution:** $P\left(\overline{X} < 203\right) = P\left(Z < \frac{203-205}{\frac{10}{\sqrt{50}}}\right) = P\left(Z < -1.414\right) = .0793$.

iv. What happens to the answer to (ii) and (iii) if the normality assumption is removed?

> **Solution:** If the normality assumption is removed, we still have the distribution to be approximately normal with mean 205 and variance 2. This is due to the central limit theorem and the large sample size (at least 20).

20. A random sample of 12 cans of string beans is taken from a canning plant and the sample mean and the sample standard deviation for the net weight (in ounces) of beans was found to be 15.97 and 0.15 respectively. Construct a 90% confidence interval for the population mean. How would your answer change if the population standard deviation had been known and is given to be 0.15?

> **Solution:** $n = 12$, $\overline{x} = 15.97$, $s = 0.15$ and $\alpha = 0.1$. $t_{n-1,\frac{\alpha}{2}} = t_{11,.05} = 1.796$. So the 90% confidence interval is given by
>
> $$15.97 \pm 1.796\left(\frac{0.15}{\sqrt{12}}\right) = 15.97 \pm 0.078 = (15.892, 16.048)\,.$$
>
> If the population standard deviation had been known to be 0.15, the 90% confidence

interval would be

$$15.97 \pm 1.645 \left(\frac{0.15}{\sqrt{12}} \right) = 15.97 \pm 0.071 = (15.899, 16.041).$$

21. A study was conducted on adults who are coffee drinkers. A random sample of size 240 contained 80 coffee drinkers.

    (a) Construct a 95% confidence interval for the proportion of coffee drinkers among adults.

    Solution:

    $$\begin{aligned} \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= \frac{1}{3} \pm z_{.025} \sqrt{\frac{\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)}{240}} \\ &= \frac{1}{3} \pm 1.96(.03) \\ &= .33 \pm .06 \\ &= (0.27, 0.39). \end{aligned}$$

    (b) How many people must be questioned if we are to obtain an estimate of the proportion of coffee drinkers which we can be 95% sure is within 0.02 of the true value?

    Solution: We can use $\hat{p} = \frac{1}{3}$ as p^*.

    $$n = \left(\frac{z_{\frac{\alpha}{2}}}{m} \right)^2 p^*(1-p^*) = \left(\frac{1.96}{.02} \right)^2 \left(\frac{1}{3} \right) \left(\frac{2}{3} \right) = 2134.2.$$

    Round up to get $n = 2135$.

22. The proportion of voters who favour the Conservative Party is 0.3. If $\hat{p}$ the sample proportion from a sample of size 100, state the mean and variance of $\hat{p}$, and its approximate distribution.

Solution: $E(\hat{p}) = p = 0.3$ and $\text{Var}(\hat{p}) = \frac{p(1-p)}{n} = \frac{(.3)(.7)}{100} = .0021$.
As $np = 30$ and $n(1-p) = 70$ are both at least 10, by the central limit theorem, $\hat{p} \approx N(.3, .0021)$

23. What is the sample size required if we need to estimate a population proportion p such that a 90% confidence interval has margin of error no more than 0.15?

Solution: $n = \left(\frac{z_{\frac{\alpha}{2}}}{2m}\right)^2 = \left(\frac{z_{.05}}{2(.15)}\right)^2 = \left(\frac{1.645}{.3}\right)^2 = 30.06$. Rounding up, we take $n = 31$.

24. A state Department of Juvenile Corrections believes that less than 20% of offenders admitted to its training schools have been convicted of car theft. Scrutiny of a random sample of 125 admission records reveals that 19 of them were for car theft. Construct a 95% confidence interval for the proportion of car thieves among the offenders admitted to the training schools.

Solution: $\hat{p} = \frac{19}{125} = 0.152$.
$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .152 \pm z_{.025}\sqrt{\frac{.152(.848)}{125}} = .152 \pm 1.96(.032) = .152 \pm .063$

25. We are interested in the proportion p of parts being produced in a manufacturing plant that are defective. For a random sample of 100 parts, 10 are found to be defective. Construct 90%, 95% and 99% confidence intervals for p.

Solution: $\hat{p} = \frac{10}{100} = 0.1$
As $z_{.05} = 1.645$, $z_{.025} = 1.96$ and $z_{.005} = 2.576$, , the confidence intervals are respectively given by

$$.1 \pm 1.645\sqrt{\frac{(.1)(.9)}{100}} = .1 \pm 1.645(.03) = .1 \pm .049,$$

$$.1 \pm 1.96(.03) = .1 \pm .059$$

and

$$.1 \pm 2.576(.03) = .1 \pm .077$$

26. A random sample of size 36 was drawn from a normal population with mean 2 and variance 25.

(a) What is the mean of the sample mean $\overline{X}$?

(b) What is the variance of $\overline{X}$?

(c) State the distribution of $\overline{X}$.

(d) What happens to the answer to (c) if the normality assumption is removed?

Solution:

(a) $E(\overline{X}) = \mu = 2$

(b) $\text{Var}(\overline{X}) = \frac{\sigma^2}{n} = \frac{25}{36}$.

(c) Normal with mean 2 and variance $\frac{25}{36}$.

(d) If the normality assumption is removed, we still have the distribution to be approximately normal with mean 2 and standard deviation $\frac{5}{6}$. This is due to the central limit theorem and the large sample size (at least 20).

27. A random sample of size 9 was drawn from a normal population with mean -1 and variance 4.

(a) What is the mean and standard deviation of the sample mean $\overline{X}$?

(b) State the distribution of $\overline{X}$.

(c) Compute $P\left(\overline{X} > 0\right)$

(d) What happens to the answer to (b) and (c) if the normality assumption is removed?

Solution:

(a) $E(\overline{X}) = \mu = -1$ and $SD(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{2}{3}$.

(b) Normal with mean -1 and variance $\frac{4}{9}$.

(c) $P\left(\overline{X} > 0\right) = P\left(Z > \frac{0-(-1)}{\frac{2}{3}}\right) = P\left(Z > 1.5\right) = 1 - .9332 = .0668$.

(d) If the normality assumption is removed, we would not know the distribution of $\overline{X}$. n is not large enough (it should be at least 20) for central limit theorem to apply. Consequently, the answer to (c) may also be wrong.

28. A sample of size 16 was drawn from a normal population with mean μ and standard deviation $\sigma = 2.5$. If the sample mean is 3.6, construct a 99% confidence interval for μ.

Solution: $\overline{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 3.6 \pm z_{.005} \frac{2.5}{\sqrt{16}} = 3.6 \pm 2.576(.625) = 3.6 \pm 1.61 = (1.99, 5.21)$

29. The diameter of the ball used for a particular sport should be 5 cm. A sample of size 20 was taken and diameters are measured. The sample mean was found to be 4.9 cm and the population standard deviation was known to be 0.25. Assuming the population is normal, construct a 95% confidence interval for the population mean.

Solution: $\overline{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 4.9 \pm z_{.025} \frac{.25}{\sqrt{20}} = 4.9 \pm 1.96(.0559) = 4.9 \pm 0.11 = (4.79, 5.01)$

30. As a pollution control measure, a coal fired power station incorporates a scrubber in its smoke stack. An environmental protection officer took discharge specimens from the smoke stack at various times on 15 randomly chosen days. The 15 specimens were subjected to careful laboratory analysis to determine their SO_2 content, and it was found that the sample average was 69mg/Mj with a standard deviation of 33gm/Mj. Construct a 95% confidence interval for the mean sulphur dioxide emission.

Solution: $n = 15$, $\overline{x} = 69$, $s = 33$ and $\alpha = 0.05$. $t_{n-1,\frac{\alpha}{2}} = t_{14,.025} = 2.145$. So the 95% confidence interval is given by

$$69 \pm 2.145 \left(\frac{33}{\sqrt{15}} \right) = 69 \pm 18.277$$

31. Results published in The New England Journal of Medicine in 1988 give the haemoglobin level for a sample of seven heart-lung transplant patients as:

$$10.0, 10.1, 10.2, 11.8, 13.4, 10.2, 11.3$$

(a) Give a 99% confidence interval for the mean haemoglobin level for all such patients.

(b) What assumption are you making about the distribution of haemoglobin levels in order for the above confidence level to be valid?

Solution: The sample mean and the sample standard deviation are given by $\overline{x} = 11$ and $s = 1.264$.

(a) The 99% confidence interval is given by

$$\overline{x} \pm t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 11 \pm t_{6,.005} \frac{1.264}{\sqrt{7}} = 11 \pm 3.707(.478) = 11 \pm 1.77 = (9.23, 12.77)$$

(b) We assume that the distribution of haemoglobin levels is normal.

32. Two independent samples of size 40 and 50 from a normal population yielded sample means 13.5 and 9.3. The corresponding sample variances were 14.4 and 112.5.

(a) Find an approximate 99% confidence interval for the difference between the means.

> **Solution:**
>
> $$13.5 - 9.3 \pm t_{39,.005}\sqrt{\frac{14.4}{40} + \frac{112.5}{50}} = 4.2 \pm 2.704(1.6155) = 4.2 \pm 4.368$$

(b) Test, at 0.01 level of significance, the hypothesis that the means of these populations are same against the alternative hypothesis that the first population has a larger mean.

> **Solution:**
> The null and the alternative hypotheses are $H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 > 0$. We reject the null hypothesis if $|t| \geq t_{39,0.01} \approx 2.423$.
>
> $$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{13.5 - 9.3}{\sqrt{\frac{14.4}{40} + \frac{112.5}{50}}} = \frac{4.2}{1.616} = 2.599.$$
>
> As the computed t-value is larger than the critical value $t_{39,.01} = 2.423$, we reject the null hypothesis and conclude that the first population has a larger mean.

33. A random sample of size 20 from a normal population resulted in a sample variance of 23.04.

(a) Construct a 95% confidence interval for σ^2 and σ.

(b) Using a .05 level of significance, test the hypothesis $H_0 : \sigma = 6.7$ against the alternative $H_0 : \sigma < 6.7$.

(c) Using a .05 level of significance, test the hypothesis $H_0 : \sigma = 5$ against the alternative $H_0 : \sigma \neq 5$.

> **Solution:**
>
> (a) The chi-square critical values are obtained from the tables as
>
> $$\chi_{19,.95} = 10.12 \quad \chi_{19,.025} = 32.85 \quad \chi_{19,.975} = 8.91$$
>
> The 95% confidence interval for σ^2 is given by
>
> $$\left(\frac{(19)\, s^2}{\chi^2_{19,.025}}, \frac{(19)\, s^2}{\chi^2_{19,.975}}\right) = \left(\frac{19(23.04)}{32.85}, \frac{19(23.04)}{8.91}\right)$$
> $$= (13.326, 49.131)$$

Taking square roots, the 95% confidence interval for σ is given by $(3.65, 7.01)$

(b) The test statistic is given by

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{19(23.04)}{6.7^2} = 9.752$$

Comparing this with $\chi^2_{19,.95} = 10.12$, we decide that we have sufficient evidence to conclude that $\sigma < 6.7$.

(c) As $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{19(23.04)}{5^2} = 17.51$ falls between $\chi_{19,.025} = 32.85$ and $\chi_{19,.975} = 8.91$, we do not have sufficient evidence to reject H_0. We cannot conclude that the population standard deviation is not 5.

34. A random sample of size 64 from a Poisson(θ) population resulted in $\overline{x} = 3.8$.

(a) Construct a 95% confidence interval for θ.

(b) At the 0.01 level of significance, test the hypothesis that $\theta = 4.5$ against the alternative that $\theta < 4.5$.

Solution:

(a) The 95% confidence interval is given by

$$\overline{x} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{x}}{n}} = 3.8 \pm 1.96(\sqrt{\frac{3.8}{64}}) = 3.8 \pm 0.4776 = (3.3224, 4.2776)$$

(b) We are testing $H_0 : \theta = 4.5$ against $H_1 : \theta < 4.5$. The test statistic is given by

$$z = \frac{\overline{x} - \theta_0}{\sqrt{\frac{\theta_0}{n}}} = \frac{3.8 - 4.5}{\sqrt{\frac{4.5}{64}}} = -2.64$$

p-value is given by $P(Z \leq -2.64) = 0.004$. As the p-value is smaller than 0.01, we reject the null hypothesis. We have sufficient evidence to conclude that θ is less than 4.5.

35. Sample proportions obtained from two independent samples of size 25 and 36 were 0.32 and 0.25 respectively.

(a) Construct a 95% confidence interval for the difference of proportions.

Solution: The confidence interval for is given by

$$.32 - .25 \pm z_{.025}\sqrt{\frac{.32(.68)}{25} + \frac{.25(.75)}{36}} = .07 \pm 1.96(.118) = .07 \pm .2312.$$

(b) Test equality of the population proportions at the 10% significance level.

Solution: Here we are testing $H_0 : p_1 = p_2$ against $H_1 : p_1 \neq p_2$. $\hat{p} = \frac{x_1+x_2}{n_1+n_2} = \frac{25\times.32+36\times.25}{36+25} = \frac{17}{61} = 0.279$.

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\left(1-\hat{p}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{.07}{\sqrt{.279\left(.721\right)\left(\frac{1}{25} + \frac{1}{36}\right)}} = 0.599$$

which is too small for us to reject the null hypothesis.