The University of Sydney

MATH1005 Statistics

maths.usyd.edu.au/MATH1005/

Semester 2 **Solutions MULTIPLE CHOICE QUESTIONS** 2015

Numerical and Graphical Summaries

1. The following table gives the number of ice creams sold in a coffee shop on each day in January 2002 in a Canadian city:

2	0	0	1	1	0	2	1
3	3	6	7	0	4	1	0
1	1	3	2	1	0	8	0
0	4	5	1	0	2	3	

What is the median?

(a) 0.0 (b) 1.0 (c) 2.0 (d) 3.0 (e) none of the above.

Solution

There are 31 observations, so when arranged in ascending order, the median is the unique middle value: $x_{(16)}$. As the 9 smallest values are all 0 and the 10th-smallest to the 17th-smallest are all 1, $\tilde{x} = x_{(16)} = 1$. (b).

2. The mean number of ice creams sold by the shop in Q1 is:

(a) 4.1 (b) 0.9 (c) 5.0 (d) 2.0 (e) 0.0.

Solution

$$\bar{x} = \frac{9 \times 0 + 8 \times 1 + 4 \times 2 + 4 \times 3 + 8 + 5 + 6 + 7 + 8}{31} = \frac{62}{31} = 2. \text{ (d).}$$

3. From a set of 8 values $x_1, \ldots, x_8$ we have $\sum_{i=1}^{8} x_i = 427$ and $\sum_{i=1}^{8} x_i^2 = 22805$. The mean and variance of these 8 values is (2dp):

(a) 53.38 and 1.19 (b) 53.38 and 1.41 (c) 53.40 and 1.41 (d) 53.38 and 1.98 (e) none of the above.

Solution

The mean is $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8} \times 427 = 53.375$.

The variance is $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}(\sum_{i=1}^{n} x_i^2 - \frac{(\sum x_i)^2}{n}) = \frac{1}{8-1}\left[22805 - \frac{(427)^2}{8}\right] =$ 1.98 (2dp) (d).

4. The following stem-leaf gives examination results as integers out of 100.

```
3 | 8
4 | 0468
5 | 235568
6 | 0145588
7 | 11445
8 | 3
```

The mean of this sample is (2dp):

(a) 60.17 (b) 43.24 (c) 58.85 (d) 97.24 (e) cannot calculate.

Solution

The sum of the values $\{x_i\}$ can be computed

- directly: by writing out the data in full. $\sum_{i=1}^{24} x_i = 38 + 40 + 44 + \ldots + 57 + 83 = 1444$
- exploiting the stem-and-leaf-display:

$$\begin{aligned}\sum_{i=1}^{24} x_i &= 38 \\ &\quad +(0+4+6+8)+4\times 40 \\ &\quad +(2+3+5+5+6+8)+6\times 50 \\ &\quad +\ldots \\ &= 1444\end{aligned}$$

Hence the mean is $\bar{x} = 1444/24 = 60.17$ (2dp) (a).

5. Given the following bivariate data

i	1	2	3	4	5
x_i	1	6	8	3	5
y_i	3	1	7	4	2

the value of S_{xy} is:

(a) 87 (b) 78.2 (c) 304 (d) 5 (e) 8.8.

Solution

$$S_{xy} = \sum_{i=1}^{5} x_i y_i - \frac{1}{5}\left(\sum_{i=1}^{5} x_i\right)\left(\sum_{i=1}^{5} y_i\right) = 87 - \frac{1}{5}(23 \times 17) = 8.8 \text{ (e)}.$$

6. For a set of 12 pairs of observations (x, y), the following summaries were obtained:

$\sum_{i}^{12} x_i = 25, \sum_{i}^{12} y_i = 432, \sum_{i}^{12} x_i^2 = 59, \sum_{i}^{12} y_i^2 = 15648, \sum_{i}^{12} x_i y_i = 880.5.$

The values of S_{xx}, S_{yy} and S_{xy} are (respectively):

(a) 6.92, 15648, and 880.5 (b) 6.92, 96, and 880.5 (c) 6.92, 96, and -19.5

(d) -19.5, 96, and 880.5 (e) none of the above.

Solution

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} = 880.5 - \frac{(25)(432)}{12} = -19.5.$$

From here we see the only possible correct answer is (c).

For the benefit of later questions we compute the other two quantities as well:

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum x_i)^2}{n} = 59 - \frac{(25)^2}{12} = 6.91667.$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{(\sum y_i)^2}{n} = 15648 - \frac{(432)^2}{12} = 96.$$

7. The sample correlation coefficient between x and y in the previous question is (2dp):

(a) 7.57 (b) -7.57 (c) -0.08 (d) -0.76 (e) 1.00.

Solution

The correlation coefficient is $r = \dfrac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \dfrac{-19.5}{\sqrt{(6.917)(96)}} = -0.757$ (d).

8. The predicted value of y at $x = 5$ from the least squares regression line in the previous question is (2dp):

(a) 27.77 (b) 47.77 (c) 41.87 (d) 55.97 (e) can not calculate.

Solution

The slope of the least-squares line is: $b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{-19.5}{6.91667} = -2.82$ (2dp).
The intercept is: $a = \bar{y} - b\bar{x} = 432/12 - (-2.82) \times 25/12 = 41.8735$.
Hence the LSR line is: $\hat{y} = 41.8735 - 2.82x$.
Substituting $x = 5$, we get $\hat{y} = 41.8735 - 2.82(5) \approx 27.77$.

9. The first quartile Q_1 of the 10 observations $\{5, 3, 4, 6, 0, 2, 2, 2, 1, 7\}$ is closest to

(a) 4.5 (b) 2.0 (c) 1.5 (d) 2.5 (e) none of the above.

Solution

Sorted data:$\{x_{(i)}\} = \{0, 1, 2, 2, 2, 3, 4, 5, 6, 7\}$.

$$Q_1 = \frac{x_{(\lceil \frac{10}{4} \rceil)} + x_{(\lfloor \frac{10}{4}+1 \rfloor)}}{2} = \frac{x_{(3)} + x_{(3)}}{2} = x_{(3)} = 2 \text{ (b)}.$$

10. For the 10 observations in the previous question, the mean is closest to

(a) 2.5 (b) 3.0 (c) 4.0 (d) 2.0 (e) 3.2

Solution

$$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{32}{10} = 3.2 \text{ (e)}.$$

11. Consider the following stem and leaf diagram, where 0 |0 represents 0.0:

```
0 | 00246
1 |1258
2 |12379
3|
4|
5 |0
```

The interquartile range (IQR) is

(a) 1.8 (b) 1.9 (c) 2.0 (d) 2.1 (e) 2.2

Solution

$$Q_1 = \frac{x_{(\lceil \frac{15}{4} \rceil)} + x_{(\lfloor \frac{15}{4}+1 \rfloor)}}{2} = \frac{x_{(4)} + x_{(4)}}{2} = x_{(4)} = 0.4.$$
$$Q_3 = \frac{x_{(\lceil \frac{3\times 15}{4}) \rceil)} + x_{(\lfloor \frac{3\times 15}{4}+1 \rfloor)}}{2} = \frac{x_{(12)} + x_{(12)}}{2} = x_{(12)} = 2.3$$
$IQR = 2.3 - 0.4 = 1.9$ (b).

12. For the 15 observations in the previous question, the number of outliers is

(a) 0 (b) 1 (c) 2 (d) 3 (e) 4

Solution

Lower threshold: $LT = Q_1 - 1.5IQR = 0.4 - 1.5 \times 1.9 = -2.45$.
Upper threshold: $UT = Q_3 + 1.5IQR = 2.3 + 1.5 \times 1.9 = 5.15$.
Hence there are no lower outliers and 0 upper outliers, giving (a).

13. A correlation coefficient of $r = -0.99$ is reported for a sample of pairs (x_i, y_i) Without any further information this implies:

(a) the points (x_i, y_i) are scattered about a straight line of slope 0.99.

(b) with a probability of 99% the relationship between x and y is best described with a straight line.

(c) the points (x_i, y_i) lie on a straight line with slope -0.99.

(d) the relationship between x and y is more likely to be non linear.

(e) the straight line of best least-squares fit has negative slope.

Solution

(e)

14. For a sample $x_1, x_2, \ldots, x_{18}$ we have $\sum_{i=1}^{18} x_i = 91$ and $\sum_{i=1}^{18} x_i^2 = 799$. To 1dp, the sample standard s_x is

(a) 20.0 (b) 19.9 (c) 4.5 (d) 2.2 (e) 4

Solution

$$s = \sqrt{\frac{1}{n-1}(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i=1}^{n} x_i^2)} = \sqrt{\frac{1}{17}(799 - \frac{1}{18}(91)^2)} \approx 4.5 \text{ (c).}$$

15. The following table shows the age (x, in years) and height (y, in cm) of 10 childen:

Age x	9	11	12	11	9	11	10	13	10	11
Height y	113	135	146	129	113	132	127	155	128	131

with

$$\sum_{i=1}^{10} x_i = 107, \sum_{i=1}^{10} y_i = 1309, \sum_{i=1}^{10} x_i y_i = 14148, S_{xx} = 14.1, S_{yy} = 1494.9.$$

Performing a linear regression of y on x, the intercept a and the slope b (to 1dp) are respectively:

(a) 10.0; 23.4 (b) 23.4; 12.1 (c) 23.4; 10.0 (d) 10.0; 25.4 (e) none of the above

Solution

$S_{xy} = 14148 - \frac{1}{10}107 \times 1309 = 141.7$.

$b = \frac{S_{xy}}{S_{xx}} = \frac{141.7}{14.1} \approx 10.04965$.

$a = \bar{y} - b\bar{x} = \frac{1309}{10} - (10.04965)\frac{107}{10} \approx 23.36875$.

So $(a, b) =$ answer (c).

Probability and Distributions

16. The success probability of a new strategy is 0.8. Use tables to find the probability of more than 6 successes in 12 independent repetitions of applying the new strategy.

(a) 0.0039 (b) 0.9961 (c) 0.0194 (d) 0.9806 (e) none of the above

Solution

Let X = number of successes in 12 independent repetitions $\sim B(n = 12, p = 0.8)$.
$P(X > 6) = 1 - P(X \leq 6) = 1 - 0.0194 = 0.9806$ (using Binomial tables), answer (d).

Note: To answer this with tables with values $p \leq 0.5$, we swap it around as follows.
Let X = number of failures in 12 independent repetitions $\sim B(n = 12, p = 0.2)$.
More than 6 successes = 5 or less failures (7 successes = 5 failures, 8 successes = 4 failures etc)
$P(X \leq 5) = 0.9806$ (using Binomial tables), answer (d).

Check:

```
> 1-pbinom(6,12,0.8)
[1] 0.9805947
> pbinom(5,12,0.2)
[1] 0.9805947
```

17. Two doping tests are available, test A and test B. If it is known that a certain blood sample was taken from a doped athlete then test A detects it with probability 0.5 and test B detects it with probability 0.45 whereas the probability that both tests simultaneously detect it is 0.4. What is the probability that neither of the tests detects that the blood sample is suspicious?

(a) 0.55 (b) 0.4 (c) 0.45 (d) 0.6 (e) 0.7

Solution

$P(A \text{ detects}) = 0.5, P(B \text{ detects}) = 0.45, P(A \cap B) = 0.4$.
$P(\text{neither detects}) = 1 - P(\text{at least 1 detects}) = 1 - P(A \cup B) = 1 - (P(A) + P(B) - P(A \cap B)) = 1 - (0.5 + 0.45 - 0.4) = 0.45$ (c).
Alternatively: draw a Venn diagram to visualise.

18. The random variable X is described by the following probabilities.

i	0	1	3	4
$P(X = i) = p_i$	0.5	0.4	0.05	0.05

What is the expected value of X?

(a) 0.55 (b) 0.4 (c) 0.45 (d) 0.75 (e) 0.7

Solution

$$E(X) = \sum_i xP(X = x) = 0 \times 0.5 + 1 \times 0.4 + 3 \times 0.05 + 4 \times 0.05 = 0.75 \text{ (d)}.$$

19. For the random variable in the previous question, what is the expected value of X^3?

(a) 0.81 (b) 6.90 (c) 2.25 (d) 0.42 (e) 4.95

Solution

$$E(X^3) = \sum_i x^3 P(X = x) = 0^3 \times 0.5 + 1^3 \times 0.4 + 3^3 \times 0.05 + 4^3 \times 0.05 = 4.95 \text{ (e)}.$$

20. Suppose that X is the number of Heads obtained in 12 independent tosses of a coin, for which a Head is twice likely as a Tail. The expected value of X is

(a) 9 (b) 7.5 (c) 9.6 (d) 8 (e) 7

Solution

Let $p = P(H)$.
Given: $P(H) = 2 \times P(T) = 2 \times (1 - P(H))$, hence $3P(H) = 2$, giving $P(H) = \frac{3}{3}$.
So we have $X \sim B(n = 12, p = \frac{2}{3})$, giving $E(X) = np = 12 \times \frac{2}{3} = 8$ (d).

21. For X in the previous question, $SD(X)$ is closest to

(a) 2.67 (b) 1.5 (c) 1.63 (d) 1.39 (e) 7.2

Solution

$X \sim B(n = 12, p = \frac{2}{3})$, giving $SD(X) = \sqrt{Var(X)} = \sqrt{np(1-p)} = \sqrt{12 \times \frac{2}{3} \times \frac{1}{3}} \approx 1.63$ (c).

22. If $X \sim B(12, 0.8)$, then using Binomial tables, $P(X = 6)$ is

(a) 0.0194 (b) 0.9845 (c) 0.0155 (d) 0.0039 (e) 0.0010

Solution

$P(X = 6) = P(X \geq 6) - P(X \leq 5) = 0.0194 - 0.0039 = 0.0155$ (c).

23. IfX has mean 5 and variance 25, then if $Y = 3X - 4$, what is $Var(Y) + 2E(Y)$?

(a) 269 (b) 247 (c) 93 (d) 154 (e) none of the above

Solution

Given: $E(X) = 5$ and $Var(X) = 25$.
$E(Y) = E(3X - 4) = 3E(X) - 4 = 3 \times 5 - 4 = 11$.
$Var(Y) = Var(3X - 4) = 3^2E(X) = 9 \times 25 = 225$.
Hence $Var(Y) + 2E(Y) = 225 + 2 \times 11 = 247$ (b).

24. If $X \sim B(10, 0.3)$, then $P(X \geq 5)$ is closest to

(a) 0.0493 (b) 0.3669 (c) 0.8497 (d) 0.9527 (e) 0.1503

Solution

$P(X \geq 5) = 1 - P(X \leq 4) = 0.1503$ (using Binomial tables), giving answer (e).

25. If $Y \sim B(11, 0.3)$, then $P(|Y - 5| \geq 3)$ is closest to

(a) 0.3170 (b) 0.0059 (c) 0.0086 (d) 0.0043 (e) 0.3127

Solution

Given: $\{|x \geq a|\} = \{x \geq a\} \cup \{x \leq -a\}$.
$P(|Y - 5| \geq 3) = P(Y - 5 \geq 3) + P(Y - 5 \leq -3) = P(Y \geq 8) + P(Y \leq 2) = 1 - P(X \leq 7) + P(Y \leq 2) = 1 - 0.9957 + 0.3127 = 0.317$ (using Binomial table twice), giving answer (a).

26. If $Z \sim N(0, 1)$, then $P(Z > 2)$ is closest to

(a) 0.9772 (b) 0.1587 (c) 0.5793 (d) 0.9861 (e) 0.0228

Solution

$P(Z > 2) = 1 - \Phi(2) = 0.0228$ (using Normal table), giving (e).

27. If $X \sim N(5, 16)$, then $P(X \geq 10)$ is closest to

(a) 0.1057 (b) 0.8944 (c) 0.6227 (d) 0.3773 (e) 0.9772

Solution

Using standardising and the Normal tables:
$P(X \geq 10) = P(\frac{X-5}{4} \geq \frac{10-5}{4}) = P(Z \geq 1.25) = 1 - \Phi(1.25) = 1 - 0.8943 = 0.1057$ (a).

28. If $Y \sim N(5, 9)$, then $P(|Y - 5| \geq 6)$ is closest to

(a) 0.3694 (b) 0.0456 (c) 0.2611 (d) 0.0885 (e) 0.9115

Solution

Using standardising and the Normal tables:
$P(|Y-5| \geq 6) = P(\frac{|Y-5|}{3} \geq \frac{6}{3}) = P(|\frac{Y-5}{3}| \geq 2) = P(|Z| \geq 2) = 2 \times P(Z \geq 2) = 2 \times (1 - \Phi(2)) = 2 \times (1 - 0.9772) = 0.0456$ (b).

29. Suppose that $\bar{X}$ is the average of a random sample of size 16 from a $N(12, 5^2)$ population. $P(\bar{X} \geq 15)$ is closest to

(a) 0.0082 (b) 0.7258 (c) 0.9918 (d) 0.2742 (e) 0.0164

Solution

Sample Mean: $\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) = N(12, \frac{5^2}{16}) = N(12, (\frac{5}{4})^2)$.
$P(\bar{X} \geq 15) = P(\frac{\bar{X}-12}{5/4} \geq \frac{15-12}{5/4}) = P(Z \geq 12/5) = 1 - \Phi(2.4) = 1 - 0.9918 = 0.0082$ (a).

30. When rolling a fair 6-sided die, the number showing is a random number X with $E(X) = 3.5$ and $Var(X) = 35/12$. The sum of three independent rolls od such a die is then a random variable who expected value and standard deviation are, respectively are:

(a) $21/2$ and $\sqrt{35/12}$ (b) 10.5 and $\sqrt{35}/2$ (c) 9.5 and $\sqrt{3 \times 35/12}$

(d) 3×3.5 and $3 \times 35/12$ (e) 9.5 and $3\sqrt{35/12}$

Solution

Given: 3 independent variables X_1, X_2, X_3, where $X_i \sim (3.5, 35/12)$, for $i = 1, 2, 3$.

Sample Total: $\sum_{i=1}^{3} X_i \sim (3 \times 3.5, 3 \times 35/12)$.

Hence the expected value and standard deviation are 3×3.5 and $\sqrt{3 \times 35/12}$ which is 3×3.5 and $\sqrt{35}/2$ (b).

31. If S denotes the sample total of a random sample of size 25 from a population with expectation 10 and variance 9 then the expectation and variance of S are, respectively,

(a) 10 and 9 (b) 10 and 9/25 (c) 250 and 225
(d) 250 and 75 (e) 10 and 3/5

Solution

Given: 25 independent variables $X_1, X_2, \dots X_{25}$, where $X_i \sim (10, 9)$, for $i = 1, 2, \dots, 25$.

Sample Total: $\sum_{i=1}^{25} X_i \sim (25 \times 10, 25 \times 9) = (250, 225)$ (c).

32. Suppose that weights (in kg) of people in a certain population have mean 70 and standard deviation 10kg. Suppose also that a passenger vehicle that can seat 40 such people can safely take 3000kg. According to the Central Limit Theorem, when carrying 40 passengers what is the probability, to 4 decimal places, that the total weight exceeds 3000kg?

(a) 0.0014 (b) 0.0083 (c) 0 (d) 0.0100 (e) 0.0209

Solution

Given: 40 independent people with weights $X_1, X_2, \dots X_{40}$, where $X_i \sim (70, 10^2)$, for $i = 1, 2, \dots, 40$.

Total Weight: $\sum_{i=1}^{40} X_i \sim (40 \times 70, 40 \times 10^2) = (2800, 4000)$.

According to the CLT, $\sum_{i=1}^{40} X_i \sim N(2800, 4000)$ (approximately).

Hence

$$P(\sum_{i=1}^{40} X_i \geq 3000) = P(\frac{\sum_{i=1}^{40} X_i - 2800}{\sqrt{4000}} \geq \frac{3000 - 2800}{\sqrt{4000}}) = P(Z \geq 3.162278) = 1 - \Phi(3.162278) \approx 0.0014$$

which is answer (a).

Note: use the closest value on the Normal table.

33. If X is an integer-valued random variable that has $E(X) = 10.5$, $Var(X) = 35/12$ then a normal approximation with continuity correction to $P(X > 12)$ is closest to

(a) 0.3859 (b) 0.2776 (c) 0.1894 (d) 0.1210 (e) 0.0721

Solution

Given: $X \sim (10.5, 35/12)$.
Using the CLT, we have an apprximating Normal, $Y \sim N(10.5, 35/12)$.
Hence, using continuity correction,

$$P(X > 12) \approx P(Y \geq 12.5) = P(\frac{Y - 10.5}{\sqrt{35/12}} \geq \frac{12.5 - 10.5}{\sqrt{35/12}}) = P(Z \geq 1.17108) = 1 - \Phi(1.17108) = 1 - 0.8790 = 0.121$$

giving answer (d).

Proportion Test

34. In testing the hypothesis that $X \sim B(12, .5)$ using large values of X as evidence, an observed value of $x = 10$ gives a p-value of

(a) 0.9807 (b) 0.0386 (c) 0.0193 (d) 0.0730 (e) 0.9270

Solution

This is equivalent to a one-sided hypothesis test. When the null hypothesis is true, $X \sim B(12, 0.5)$. Thus the p-value is

$$P(X \geq 10) = 1 - P(X \leq 9) \approx 1 - 0.9807 = 0.0193 \quad \text{(using tables)} \quad (c)$$

35. In testing the hypothesis that $X \sim B(12, .5)$ using large values of $|X - 6|$ as evidence, an observed value of $x = 10$ gives a p-value of

(a) 0.9807 (b) 0.0386 (c) 0.0193 (d) 0.0730 (e) 0.9270

Solution

Note that 'using large values of $|X - 6|$' is equivalent to a two-sided hypothesis test: ie both large and small values of $|X - 6|$ will give evidence against H_0. Note also that $|X - 6|$ is the test statistic required for a non-symmetric Binomial (where $E(X) = np = 6$). However, commonly $p = 0.5$ (which gives a symmetric Binomial) and hence we need only consider the test statistic as X.

When the null hypothesis is true, $X \sim B(12, 0.5)$. Thus the p-value is

$$2 \times P(X \geq 10) = 2 \times P(X \leq 2) = 2 \times 0.0193 = 0.0386 \quad \text{(using tables)} \quad (b)$$

36. Suppose that we need to test the hypotheses $H_0 : p = 0.8$ against $H_1 : p > 0.8$, where p is the probability of success in the binomial model. Based on $n = 10$, the number of 'successes' is 8. This indicates that:

(a) p-value is 0.32 and we have strong evidence against H_0.

(b) p-value is 0.68 and the data is consistent with H_0.

(c) p-value is 0.32 and the data is consistent with H_0.

(d) p-value is 0.68 and we have strong evidence against H_0.

Solution

$\boxed{\text{H}}$ Define $X \sim B(10, p)$ where we want to test $H_0 : p = 0.8$ versus $H_1 : p > 0.8$.
$\boxed{\text{T}}$ Under H_0, $X \sim B(10, 0.8)$.
$\boxed{\text{P}}$ Since the observation is $x = 8$, the p-value is

$$P(X \geq 8) = 1 - P(X \leq 7) \approx 1 - 0.3222 = 0.6778 \quad \text{(using tables)}$$

Given the p-value is large, we say 'the data are consistent with H_0 (b).

37. A random variable $X \sim B(12, p)$ is observed to take the value 8. If this is used to test $H_0 \colon p = 0.7$ against $H_1 \colon p > 0.7$, the p-value is closest to

(a) 0.25 (b) 0.72 (c) 0.81 (d) 0.49 (e) 0.09

Solution

$\boxed{\text{T}}$ Let $X \sim B(12, p)$. Under H_1, X tends to take larger values than under H_0. Thus we use larger values of X as more evidence against H_0.
$\boxed{\text{P}}$ Hence the p-value is the probability, under H_0, of at least as much evidence (against H_0) as the observation $x = 8$.

$$p\text{-value} = P(X \geq 8) = 1 - P(X \leq 7) \approx 1 - 0.2763 = 0.7237 \quad \text{(using tables)} \quad (b)$$

38. The p-value for testing the hypothesis that $X \sim B(10, 0.5)$ for a 2 sided alternative hypothesis, with an observation of 8, is

(a) 0.0547 (b) 0.1094 (c) 0.9893 (d) 0.0107 (e) 0.0214

Solution

$\boxed{\text{H}}$ $H_0 : p = 0.5$ vs $H_1 : p \neq 0.5$.

$\boxed{\text{T}}$ $X \sim B(10, 0.5)$, under H_0.

$\boxed{\text{P}}$

$$p\text{-value} = 2 \times P(X \geq 8) = 2(1 - P(X \leq 7)) = 2(1 - 0.9453) = 0.1094$$

which is (b).

39. The standard medication for a certain ailment is known to give relief to 60% of sufferers. A new drug has been developed which is claimed to be better than the standard medication. A random sample of 50 patients is given the new drug and 35 of them obtain relief. A p-value for the test of H_0:"the new drug is the same as the standard" versus H_1:"the new drug is better than the standard" is, using a normal approximation with continuity correction, closest to

(a) 0.0329 (b) 0.0968 (c) 0.0262 (d) 0.1241 (e) 0.0745

Solution

$\boxed{\text{H}}$ $H_0 : p = 0.6$ vs $H_1 : p > 0.6$.

$\boxed{\text{T}}$ $X \sim B(50, 0.6)$, under H_0.

$\boxed{\text{P}}$ The p-value is $P(X \geq 35)$, which we can't look up on Binomial table.

Hence, the approximating normal is $Y \sim N(np, np(1-p)) = N(50 \times 0.6, 50 \times 0.6 \times 0.4) = N(30, 12)$. Hence, using continuity correction, the p-value is $P(X \geq 35) \approx P(Y \geq 34.5)$.
Now standardising,

$$P(\frac{Y-30}{\sqrt{12}} \geq \frac{34.5-30}{\sqrt{12}}) = P(Z \geq 1.299038) = 1 - \Phi(1.299038) = 1 - 0.9032 = 0.0968$$

which is (b).

40. The number of heads in 12 flips of a coin is modelled as a random variable $X \sim B(12, p)$ for some $0 < p < 1$. If X takes the value 2, then the p-value for testing $H_0 : p = 0.5$ versus $H_1 : p < 0.5$ is closest to

(a) 0.0032 (b) 0.0064 (c) 0.0193 (d) 0.0386 (e) 0.1938

Solution

$\boxed{\text{H}}$ $H_0 : p = 0.5$ vs $H_1 : p < 0.5$.

$\boxed{\text{T}}$ $X \sim B(12, 0.5)$, under H_0, with $x_{obs} = 2$.

$\boxed{\text{P}}$ The p-value is $P(X \leq 2) = 0.0193$ (c).

$\boxed{\text{Sign Test}}$

41. 14 students each taste-tested two different brands of drink (brand X and brand Y) but the brands were hidden from them. The object of the exercise was to see if students preferred one brand over the other, but there was no indication of which this might be before the test. In all, 8 subjects preferred brand X, 4 preferred brand Y and 2 had no preference either way. Using a sign test (removing ties), the p-value is closest to

(a) 0.25 (b) 0.39 (c) 0.02 (d) 0.19 (e) 0.01

Solution

$\boxed{\text{Preparation}}$ We have 14 students, but we need to discard the 2 'ties' (zeroes), so $n = 12$. Let T = number preferring brand X, where $T \sim B(12, p)$

$\boxed{\text{H}}$ $H_0\colon p = 0.5$ (no preference) versus $H_1\colon p \neq 0.5$

$\boxed{\text{T}}$ Under H_0, $T \sim B(12, 0.5)$. The observed value of T is $t_{obs} = 8$.

$\boxed{\text{P}}$ Thus the p-value is

$$2 \times P(T \geq 8) = 2P(T \leq 4) \text{ (by symmetry)} = 2 \times 0.1938 \text{ (tables)} = 0.3876 \quad (b)$$

42. Fifteen test subjects are asked if they prefer brand A, brand B, or have no preference. Nine prefer brand A, 3 prefer brand B and 3 have no preference. The P-value for a sign-test of no difference between brands is closest to

(a) 0.0386 (b) 0.1460 (c) 0.0730 (d) 0.0193 (e) 0.1758

Solution

$\boxed{\text{Preparation}}$ We have 15 subjects, but we need to discard the 3 'ties' (no preference, zeroes), so $n = 12$. Let T = number preferring brand X, where $T \sim B(12, p)$

$\boxed{\text{H}}$ $H_0\colon p = 0.5$ (no preference) versus $H_1\colon p \neq 0.5$

$\boxed{\text{T}}$ Under H_0, $T \sim B(12, 0.5)$. The observed value of T is $t_{obs} = 9$.

$\boxed{\text{P}}$ Thus the p-value is

$$2 \times P(T \geq 9) = 2P(T \leq 3) \text{ (by symmetry)} = 2 \times 0.0730 \text{ (tables)} = 0.146 \quad (b).$$

$\boxed{\text{I Sample Z Test}}$

43. A random sample of size 25 from a $N(\mu, 9)$ population yields a sample mean of 21. To test $H_0\colon \mu = 20$, $H_1\colon \mu > 20$ an expression for the p-value is

(a) $\Phi(5/2)$ (b) $\Phi(-0.1)$ (c) $P(t_9 > 0.9)$ (d) $1 - \Phi(5/3)$ (e) none of these

Solution

$\boxed{\text{Preparation}}$ We have $\bar{X} \sim N(\mu, 9/25) = N(\mu, (3/5)^2)$.

$\boxed{\text{H}}$ $H_0 : \mu = 20$ (no preference) vs $H_1 : \mu > 20$

$\boxed{\text{T}}$ Under H_0, $\bar{X} \sim N(20, (3/5)^2)$. The observed value of $\bar{X}$ is $\bar{x} = 21$.

$\boxed{\text{P}}$ Thus the p-value is

$$P(\bar{X} \geq 21) = P(\frac{\bar{X} - 20}{3/5} \geq \frac{21 - 20}{3/5}) = P(Z \geq 5/3) = 1 - \Phi(5/3) \ \ (d)$$

OR Using the standardised version of test statistic:

$\boxed{\text{H}}$ $H_0 : \mu = 20$ vs $H_1 : \mu > 20$

$\boxed{\text{T}}$ Under H_0, $\bar{X} \sim N(20, (3/5)^2)$, or standardising $\dfrac{\bar{X} - 20}{3/5} \sim N(0, 1) = Z$. As $\bar{x} = 21$, the observed value of Z is $z_{obs} = \dfrac{21 - 20}{3/5} = 5/3$.

$\boxed{\text{P}}$ Thus the p-value is

$$P(Z \geq 5/3) = 1 - \Phi(5/3)$$

44. A random sample of size 64 from a $N(\mu, 16)$ population yields a sample mean of 19.1. To test $H_0\colon \mu = 20$ against $H_1\colon \mu < 20$ an expression for the p-value is

(a) $1 - \Phi(1.8)$ (b) $\Phi(-0.1)$ (c) $2[1 - \Phi(0.9)]$ (d) $\Phi(9/5)$ (e) $2\Phi(-1.8)$

Solution

$\boxed{\text{Preparation}}$ We have $\bar{X} \sim N(\mu, 16/64) = N(\mu, (.5)^2)$.

$\boxed{\text{H}}$ $H_0 : \mu = 20$ (no preference) vs $H_1 : \mu < 20$

$\boxed{\text{T}}$ Under H_0, $\bar{X} \sim N(20, (.5)^2)$. The observed value of $\bar{X}$ is $\bar{x} = 19.1$.

$\boxed{\text{P}}$ Thus the p-value is

$$P(\bar{X} \leq 19.1) = P(\frac{\bar{X} - 20}{0.5} \leq \frac{19.1 - 20}{0.5}) = P(Z \leq -1.8) = 1 - \Phi(1.8) = 1 - 0.9641 = 0.0359 \ \ (a)$$

OR Using the standardised version of test statistic:

$\boxed{\text{H}}$ $H_0 : \mu = 20$ vs $H_1 : \mu > 20$

$\boxed{\text{T}}$ Under H_0, $\bar{X} \sim N(20, (.5)^2)$, or standardising $\dfrac{\bar{X} - 20}{.5} \sim N(0, 1) = Z$. As $\bar{x} = 19.1$, the observed value of Z is $z_{obs} = \dfrac{19.1 - 20}{0.5} = -1.8$.

$\boxed{\text{P}}$ Thus the p-value is

$$P(Z \leq 1.8) = 0.0359$$

45. Suppose that we need to test the hypotheses $H_0 : \mu = 8$ against $H_1 : \mu > 8$. Based on n $(n > 25)$ observations and using $z = \dfrac{\bar{x} - 8}{\sigma/\sqrt{n}}$ (σ is known), the corresponding value of z is 1.78 . This indicates that the

(a) p-value is 0.9625 and we have strong evidence against H_0 .

(b) p-value is 0.9625 and the data are consistent with H_0 .

(c) p-value is 0.0375 and we have evidence against H_0 .

(d) p-value is 0.0375 and the data is consistent with H_0

Solution

$\boxed{\text{T}}$ Since the alternative hypothesis is $\mu > 8$, larger values of the Z-statistic indicate more evidence (against $H_0\colon \mu = 8$).

$\boxed{\text{P}}$ Hence the p-value is

$$P(Z \geq 1.78) = 1 - \Phi(1.78) = 1 - 0.9625 = 0.0375$$

$\boxed{\text{C}}$ As the p-value is small (cf $\alpha = 0.05$), we would conclude that there is evidence against H_0. (c)

46. Measurement errors on an electric balance are normally distributed with a known standard deviaiton of 0.005 grams and an unknown mean μ. If $\mu = 0$ then the balance is working properly, otherwise it needs to be repaired. A standard object whose weight is known exactly is measured 25 times. The measurement errors average out to -0.00175. The p-value for testing H_0: 'balance is working properly' versus H_1 : 'balance needs repairs' is closest to

(a) 0.04 (b) 0.08 (c) 0.36 (d) 0.72

(e) 0.64

Solution

$\boxed{\text{H}}$ H_0: $\mu = 0$ vs H_1: $\mu \neq 0$.

$\boxed{\text{T}}$ The test statistic is $\bar{X} \sim N(\mu, \dfrac{0.005^2}{25}) = N(\mu, 0.001^2)$, with observed value $\bar{x} = -0.00175$.

$\boxed{\text{P}}$ The p-value is

$$2P(\bar{X} \leq -0.00175) = 2P(\frac{\bar{X} - 0}{0.001} \leq \frac{0.00175 - 0}{0.001}) = 2P(Z \leq 1.75) = 2(1 - \Phi(1.75)) = 2(1 - 0.9599) = 0.0802 \ \ (b)$$

$\boxed{\text{1 Sample t Test}}$

47. A random sample of size 25 from a $N(\mu, \sigma^2)$ population with unknown variance yields a sample mean of 21 and sample variance 9. To test $H_0\colon \mu = 20$, $H_1\colon \mu > 20$ an expression for the p-value is

(a) $P(t_{25} > 5/3)$ (b) $1 - \Phi(5/3)$ (c) $2P(t_{24} > 5/3)$ (d) $P(t_{24} > 5/3)$

Solution

$\boxed{\text{T}}$ The test statistic is $T = \dfrac{\bar{X} - 20}{\sqrt{9}/\sqrt{25}} \sim t_{24}$, when H_0 is true.

The observed value of the test statistic is $t_{obs} = \dfrac{21 - 20}{3/5} = 5/3$.

$\boxed{\text{P}}$ Larger values of T are more evidence against H_0, hence the p-value is

$$P(T \geq 5/3) = P(t_{24} > 5/3) \ \ (d)$$

$\boxed{\text{2 Sample t Test}}$

48. When performing a two-sample t-test, if one sample has size 13 and sample sd 4.1, while the other has size 7 and sample sd 5.7, the pooled estimate of the unknown common population sd is closest to

(a) 10.06 (b) 4.69 (c) 13.64 (d) 3.15 (e) 22.04

Solution

Given: $n_x = 13, s_x = 4.1, n_y = 7, s_y = 5.7$.

Pooled Standard Deviation: $s_p = \sqrt{\dfrac{12 \times 4.1^2 + 6 \times 5.7^2}{18}} = 4.69$ (b).

49. Two samples have been taken from two independent normal populations with equal variances. From these samples ($n_x = 12, n_y = 15$) we calculate $\bar{x} = 119.4$, $\bar{y} = 112.7$, $s_x = 9.2$, $s_y = 11.1$. For the test $H_0 : \mu_x = \mu_y$ against $H_1 : \mu_x \neq \mu_y$, the p-value is included in

(a) (0.05,0.1) (b) (0.1,0.2) (c) (0.25,0.5) (d) (0.001,0.005) (e) none of these

Solution

Given: $n_x = 12, \bar{x} = 119.4, s_x = 9.2, n_y = 15, \bar{y} = 112.7, s_y = 11.1$.

Pooled Standard Deviation: $s_p = \sqrt{\dfrac{11 \times 9.2^2 + 14 \times 11.1^2}{25}} = 10.30724$.

$\boxed{\text{T}}$ Observed value of test statistic: $\tau_{obs} = \dfrac{119.4 - 112.7 - 0}{s_p\sqrt{1/12 + 1/15}} = 1.678366$.

$\boxed{\text{P}}$ P-value: $2P(t_{25} \geq \tau_{obs}) = 2P(t_{25} \geq 1.678366) \; \epsilon \; (0.1, 0.2)$ (b).

Check:

```
> 2*(1-pt(1.68,18))
[1] 0.1102295
```

Goodness of Fit Test

50. The following table gives the observed frequencies of genotypes A, B, and C of 100 plants:

Genotype	A	B	C	Total
Observed frequency, O_i	18	55	27	100

Under the null hypothesis that A, B, and C are in the ratio of 1:2:1, the expected frequencies, E_i's (respectively) are:

(a) 25, 25, 75 (b) 25, 25, 50 (c) 50, 25, 25 (d) 25, 50, 25 (e) none of these.

Solution

$O_1 = O_3 = 1/(1 + 2 + 1) \times 100 = 25$; $O_2 = 2/(1 + 2 + 1) \times 100 = 50$. (d)

51. The observed value τ_{obs} of the statistic $\tau = \sum \dfrac{(O_i - E_i)^2}{E_i} = \sum \dfrac{O_i^2}{E_i} - n$ for the data in Q50 is:

(a) $\tau_{\text{obs}} = 2.62$ (b) $\tau_{\text{obs}} = 102.62$ (c) $\tau_{\text{obs}} = 100$ (d) $\tau_{\text{obs}} = 0.262$ (e) $\tau_{\text{obs}} = 1.62$

Solution

$\boxed{\text{T}}$ $\tau_{obs} = \dfrac{18^2}{25} + \dfrac{55^2}{50} + \dfrac{27^2}{25} = 2.62$. (a)

52. 100 observations are made on a random variable only taking values 0, 1, 2 and 3. The frequencies are shown below:

Value	0	1	2	3
Frequency	22	38	32	8

A goodness of fit test is applied to see if these frequencies are well-described by $\mathcal{B}(3, 0.5)$ probabilities. The χ^2 goodness-of-fit statistic is 9.653. The corresponding p-value is somewhere in the interval

(a) (0.1,0.9) (b) (0.9,0.95) (c) (0.01,0.025) (d) (0.05,0.1) (e) (0.025,0.05)

Solution

Value	0	1	2	3
Observed Frequency O_i	22	38	32	8
Expected Frequency E_i	12.5	37.5	37.5	12.5

where

$E_0 = P(X = 0) = \binom{3}{0}(0.5)^0(0.5)^3 \times 100 = 12.5 = E_3$ (by symmetry).

$E_1 = P(X = 1) = \binom{3}{1}(0.5)^1(0.5)^2 \times 100 = 37.5 = E_2$ (by symmetry).

$\boxed{\text{T}}$ $\tau_{obs} = \frac{22^2}{12.5} + \frac{38^2}{37.5} + \frac{32^2}{37.5} + \frac{8^2}{12.5} = 9.653.$

$\boxed{\text{P}}$ $P - value = P(\chi^2_3 \geq 9.653) \;\; \epsilon \; (0.01, 0.025)$ (c).