Answers - Hypothesis tests: correlation and Normal mean

Year: Year 13

Answer to Question 1

a. Let μ be the mean mass of a bag of flour in kg. $H_0 : \mu = 1.5$, $H_1 : \mu$ (one-tailed).
b. Let ρ be the population product moment correlation coefficient. $H_0 : \rho = 0$, $H_1 : \rho > 0$ (one-tailed).
c. Let μ be the mean waiting time in minutes. $H_0 : \mu = 4$, $H_1 : \mu \neq 4$ (two-tailed, because the word "changed" does not give a direction).

Answer to Question 2

Method: state $H_0 : \rho = 0$ and H_1 ($\rho > 0$, ρ or $\rho \neq 0$) using the population correlation coefficient ρ. Compare $|r|$ with the critical value for the right number of tails, then write a conclusion in context that is not too definite (\"there is evidence that...\").

a. The test is for positive correlation, so it is one-tailed and we use 0.5494. $H_0 : \rho = 0$, $H_1 : \rho > 0$. Since $0.58 > 0.5494$ we reject H_0: there is sufficient evidence at the 5% level of positive correlation. So at the 5% level the data suggest that the two variables are positively correlated.
b. Now any correlation counts, so the test is two-tailed and we use 0.6319. $H_0 : \rho = 0$, $H_1 : \rho \neq 0$. Since 0.58 we do not reject H_0: there is insufficient evidence at the 5% level of any correlation. The same value of r gives a different conclusion because the 5% is now split between the two tails.

Answer to Question 3

Method: state $H_0 : \rho = 0$ and H_1 ($\rho > 0$, ρ or $\rho \neq 0$) using the population correlation coefficient ρ. Compare $|r|$ with the critical value for the right number of tails, then write a conclusion in context that is not too definite (\"there is evidence that...\").

a. The test is one-tailed in the negative direction, so we compare the size of r with 0.4973. $H_0 : \rho = 0$, $H_1 : \rho$. Compare $|r| = 0.55$ with 0.4973. Since $0.55 > 0.4973$, reject H_0: there is evidence of negative correlation. So at the 5% level there is evidence that the variables are negatively correlated.
b. The test is two-tailed, so the critical value is larger. $H_1 : \rho \neq 0$. Compare 0.55 with 0.5760. Since 0.55, do not reject H_0: insufficient evidence of correlation. So without a direction decided in advance, the evidence is not strong enough.

Answer to Question 4

Method: if $X \sim N(\mu, \sigma^2)$, the mean of a random sample of size n satisfies $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. The standard error is $\frac{\sigma}{\sqrt{n}}$; standardise with $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$.

a. The variance of the mean is $\frac{\sigma^2}{n} = \frac{144}{16} = 9$. $\bar{X} \sim N\left(50, \frac{12^2}{16}\right) = N(50, 3^2)$. So the sample mean has standard error 3, which is $\sqrt{16} = 4$ times smaller than σ.

b. Standardise: $z = \frac{53 - 50}{3} = 1$, then take the upper tail.
$P(\bar{X} > 53) = P(Z > 1) = 0.1587$. So about 15.9% of samples have a mean above 53.

c. Standardise: $z = \frac{46 - 50}{3} = -\frac{4}{3}$. $P(\bar{X}$. So about 9.1% of samples have a mean below 46.

Answer to Question 5

a. $\bar{X} \sim N\left(120, \frac{81}{36}\right) = N(120, 1.5^2)$.

b. $z = -2.3263$, so the critical value is $120 - 2.3263 \times 1.5 = 116.51$. The critical region is $\bar{x} \leq 116.51$.

c. $117.2 > 116.51$, so it is not in the critical region: do not reject H_0. There is insufficient evidence at the 1% level that the mean is below 120.

Answer to Question 6

Method: state the hypotheses about μ, find the distribution of $\bar{X}$ under H_0, calculate the test statistic or p-value, compare with the significance level, then conclude in context.

a. $\bar{X} \sim N\left(40, \frac{36}{20}\right)$, standard error $\sqrt{1.8} = 1.3416$. $z = \frac{42.5 - 40}{1.3416} = 1.863$.

b. $P(\bar{X} \geq 42.5) = 1 - 0.9688 = 0.0312$, so the two-tailed p-value is $2 \times 0.0312 = 0.0624$.

c. $0.0624 > 0.05$, so do not reject H_0. There is insufficient evidence at the 5% level that the mean differs from 40.

Note: with a one-tailed test ($H_1 : \mu > 40$) the p-value would be 0.0312 and H_0 would be rejected, which shows why the alternative hypothesis must be chosen before seeing the data.

Answer to Question 7

Method: state $H_0 : \rho = 0$ and H_1 ($\rho > 0$, ρ or $\rho \neq 0$) using the population correlation coefficient ρ. Compare $|r|$ with the critical value for the right number of tails, then write a conclusion in context that is not too definite (\"there is evidence that...\").

a. $r = 0.987$.
b. $H_0 : \rho = 0$, $H_1 : \rho > 0$. $0.987 > 0.6215$, so reject H_0: there is sufficient evidence of positive correlation between hours revised and mark.
c. Correlation does not prove causation. Other factors, such as prior ability or motivation, might affect both revision time and marks.

Answer to Question 8

a. $r = -0.996$.
b. $H_0 : \rho = 0$, $H_1 : \rho$. $|r| = 0.996 > 0.7155$, so reject H_0: there is sufficient evidence at the 1% level of negative correlation between temperature and hot chocolate sales.
c. $T = 40$ is outside the range of the data (extrapolation), so the linear relationship may not continue; sales cannot be negative.

Answer to Question 9

Method: state the hypotheses about μ, find the distribution of $\bar{X}$ under H_0, calculate the test statistic or p-value, compare with the significance level, then conclude in context.

a. The group suspects that bags are underweight, so the test is one-tailed. $H_0 : \mu = 200$, $H_1 : \mu$ where μ is the mean mass per bag.
b. Under H_0, $\bar{X} \sim N\left(200, \frac{16}{20}\right) = N(200, 0.8)$, so the standard error is $\sqrt{0.8} = 0.8944$. $z = \frac{198.6 - 200}{0.8944} = -1.565$, $p = P(Z$. So, if the true mean were 200 g, a sample mean as low as 198.6 g would happen about 5.9% of the time.
c. Compare the same p-value with each significance level. At 5%: $0.0588 > 0.05$, do not reject H_0 (insufficient evidence the mean is lower). At 10%: 0.0588, reject H_0 (sufficient evidence the mean is lower). So the conclusion depends on the significance level chosen.

Answer to Question 10

a. $H_0 : \mu = 42$, $H_1 : \mu > 42$. $\bar{X} \sim N(42, \frac{36}{30})$, standard error 1.0954. $z = \frac{2.1}{1.0954} = 1.917$, $p = 0.0276$. Reject H_0: there is sufficient evidence that the

mean journey time has increased.

b. The probability of a Type I error equals the significance level, 0.05.

c. The data is the mean of a sample of 30 journeys, so we need the distribution of a sample mean, which has a smaller standard deviation than a single journey.

Answer to Question 11

Method: state the hypotheses about μ, find the distribution of $\bar{X}$ under H_0, calculate the test statistic or p-value, compare with the significance level, then conclude in context.

a. $H_0 : \mu = 120$, $H_1 : \mu$. $\bar{X} \sim N(120, 3^2)$ since $15/\sqrt{25} = 3$. $z = \dfrac{114.2 - 120}{3} = -1.933$, $p = 0.0266$. Reject H_0: there is sufficient evidence that the mean battery life is less than 120 hours.

b. $0.0266 > 0.01$, so at the 1% level we do not reject H_0. The conclusion changes.

c. The battery life is normally distributed with standard deviation 15 hours, and the phones form a random sample.

Answer to Question 12

a. $H_0 : \mu = 230$, $H_1 : \mu \neq 230$. $\bar{X} \sim N(230, 5^2)$. $z = \dfrac{11}{5} = 2.2$. Two-tailed p-value $= 2 \times 0.0139 = 0.0278$. Reject H_0: there is sufficient evidence that the mean reading speed has changed.

b. At 2.5% in each tail: $230 \pm 1.96 \times 5 = 230 \pm 9.8$. The critical region is $\bar{x} \leq 220.2$ or $\bar{x} \geq 239.8$. Since $241 > 239.8$ the sample mean is in the critical region.

c. No. Without a control group, other factors might explain the change, such as practice with the test.

Answer to Question 13

a. $r = -0.973$.

b. $H_0 : \rho = 0$, $H_1 : \rho \neq 0$. $|r| = 0.973 > 0.6319$, reject H_0: there is significant evidence of correlation (older swimmers in this club are faster).

c. The data cover ages 10 to 19; beyond that range we have no evidence, and performance eventually levels off then declines, so the trend cannot be extended.

Answer to Question 14

a. $H_0 : \mu = 12$, $H_1 : \mu > 12$. $\bar{X} \sim N(12, 0.6^2)$. $z = \dfrac{1.1}{0.6} = 1.833$, $p = 0.0334$. Reject H_0 : sufficient evidence that growth has increased.

b. $0.0334 > 0.01$, do not reject H_0: insufficient evidence at the 1% level.
c. Use a larger sample and choose the plants at random; use a control group grown under the same conditions at the same time.

Answer to Question 15

a. $r = \dfrac{-190}{\sqrt{250 \times 480}} = \dfrac{-190}{346.4} = -0.548.$
b. One-tailed: $H_0 : \rho = 0$, $H_1 : \rho$. $|r| = 0.548 > 0.4973$, reject H_0 (negative correlation). Two-tailed: $H_1 : \rho \neq 0$. 0.548, do not reject H_0.
c. r is unchanged (-0.548) because linear coding with a positive multiplier does not change the correlation coefficient.
d. The two-tailed test: choosing a one-tailed test after seeing a negative r would be biased. The hypotheses must be set before the data is examined, and here the two tests give different conclusions.

Answer to Question 16

a. Under H_0, $\bar{X} \sim N(50, 2^2)$. $z = 1.6449$, critical value $50 + 1.6449 \times 2 = 53.29$. Critical region: $\bar{x} \geq 53.29$.
b. $54 > 53.29$, so reject H_0: sufficient evidence that the mean strength has increased.
c. It equals the significance level, 0.05 (a Type I error).
d. If $\mu = 55$, $\bar{X} \sim N(55, 2^2)$.

$$P(\bar{X} \geq 53.29) = P\left(Z \geq \frac{53.29 - 55}{2}\right) = P(Z \geq -0.855) = 0.804.$$

Answer to Question 17

Method: if $X \sim N(\mu, \sigma^2)$, the mean of a random sample of size n satisfies $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$. The standard error is $\dfrac{\sigma}{\sqrt{n}}$; standardise with $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$.

a. Standardise the sample mean and divide the inequality by the standard error $\dfrac{8}{\sqrt{n}}$. $\bar{X} \sim N\left(\mu, \dfrac{64}{n}\right)$ so $Z = \dfrac{\bar{X} - \mu}{8/\sqrt{n}}$. $|\bar{X} - \mu|$. This is the required result.
b. $P(|Z|$ when $k = 1.96$, since 2.5% is left in each tail. We need $\dfrac{\sqrt{n}}{8} \geq 1.96$, so $\sqrt{n} \geq 15.68$ and $n \geq 245.9$. The smallest n is **246**. So the engineer must measure at least 246 bolts.
c. Substitute $n = 100$ into the result of part a). $n = 100$: $\dfrac{\sqrt{100}}{8} = 1.25$. $P(|Z|$. So

with 100 bolts the probability is only about 79%, less than the required 95%.

Answer to Question 18

a. Calculator: $r = 0.792$.
b. $H_0 : \rho = 0$, $H_1 : \rho > 0$. $0.792 > 0.6215$, reject H_0.
c. $0.968 > 0.6694$, reject H_0 again.
d. The outlier reduced r from 0.968 to 0.792, because it does not lie on the line followed by the other points. Both tests give the same conclusion, but removing an outlier needs a reason, such as a recording error.

Answer to Question 19

a. A: 0.032, reject H_0. B: $0.081 > 0.05$, do not reject. C: 0.0049, reject.
b. A: $0.032 > 0.01$, do not reject. B: do not reject. C: 0.0049, reject.
c. If 0.081 is the tail probability, the two-tailed p-value is $2 \times 0.081 = 0.162$, which is far above 0.05, so do not reject. (If 0.081 is already two-tailed, the conclusion is the same.)
d. A hypothesis test never proves anything: it shows sufficient evidence at a stated significance level. Also there remains a 1% chance of a Type I error.

Answer to Question 20

a. $0.95^5 = 0.774$.
b. $1 - 0.95^5 = 1 - 0.7738 = 0.226$.
c. $X \sim B(5, 0.05)$. $P(X = 1) = 5 \times 0.05 \times 0.95^4 = 0.204$.
d. Each test carries a 5% chance of a Type I error, so the overall chance of at least one error grows with the number of tests (22.6% for 5 tests). A remedy is to use a lower significance level for each test, or to decide in advance which test is the main one.

Answer to Question 21

a. Reject if $z = \dfrac{52 - 50}{10/\sqrt{n}} > 1.6449$, i.e. $\dfrac{\sqrt{n}}{5} > 1.6449$, so $\sqrt{n} > 8.2245$ and $n > 67.64$. The smallest integer is 68.
b. A small difference (2 units, a fifth of a standard deviation) can be significant with a big enough sample, but statistical significance does not mean the difference is large or important.

Answer to Question 22

a. 5%: $0.45 > 0.4409$, reject H_0. 1%: 0.45, do not reject H_0.
b. 0.45, do not reject H_0.
c. A weak correlation can still be statistically significant for a big enough sample, and a non-significant result means only that there is insufficient evidence of correlation, not that the true correlation is zero.