Hypothesis Testing and the Normal Distribution

Prerequisites

You should be familiar with (1) hypothesis testing with the binomial distribution (2) the normal approximation to the binomial distribution (3) the central limit theorem.

Example (1)

A medium claims that she can predict the suit of a card drawn from a pack at random. The probability of predicting the suit correctly is 0.25. In an experiment to determine the validity of her claim 30 cards are drawn at random. The experimenters establish the following hypotheses

$H_0 : p = 0.25$

$H_1 : p > 0.25$

Letting X denote the number of times the medium successfully predicts the suit of the card chosen, the experimenters define the critical region to be $X \geq 13$

(*a*) Calculate the significance level of this procedure.

(*b*) Suppose the medium to have some telepathic powers, but rather than being able to predict the suit of the card with certainty that she has the power to increase the proportion of cards correctly predicted to 30%. Calculate the probability of drawing the correct conclusion if the value of p is actually 0.30.

Solution

(*a*) Under H_0: $X \sim B(30, 0.25)$

$$\begin{aligned} \alpha &= P(X \geq 13) \\ &= 1 - P(X \leq 12) \\ &= 1 - 0.9784 \\ &= 0.0216 \end{aligned}$$

(*b*) Suppose in fact $p = 0.30$. The critical region means that we reject $H_0 : p = 0.25$ if $X \geq 13$. Under $X \sim B(30, 0.30)$

$$P(X \geq 13) = 1 - P(X \leq 12)$$
$$= 1 - 0.9155$$
$$= 0.0845$$

Thus if in fact $p = 0.30$ the probability of reaching the conclusion that $p \neq 0.25$ is 0.0845 = 8.45%. If indeed the medium is a telepath this may not do justice to her claim.

The test in example (1) is not very sensitive to the case when the probability differs only slightly from the assumed value of $p = 0.25$, which poses a limitation on the use of this test.

Testing by means of the normal approximation to the binomial distribution

One way to reduce the problem of the lack of sensitivity when using a test based on the binomial distribution is to take a larger sample size. As tables tend to give values of binomial distributions up to a sample size of $n = 30$ this does require using a normal approximation to the binomial distribution and extending our concepts of hypothesis testing to cover the normal distribution.

Example (1) continued

In a second experiment with the medium 60 cards are drawn at random from the pack.

(*a*) Assuming that this is by chance and that the real proportion of her guesses is 25% find the probability that she should succeed in predicting 30% or more of these cards correctly.

(*b*) The experimenters set a significance level of 5% for this test and establish the hypotheses $H_0 : p = 0.25$ and $H_1 : p > 0.25$. Given that she succeeded in predicting 30% of the cards correctly, which of these two alternative hypotheses is accepted?

(*c*) The experimenters define the critical region to be $X \geq N$ for some integer *N*. Find the value of *N* that corresponds to a significance level of at least 5% and state the actual *p*-value of the critical region.

(*d*) Suppose the medium to have some telepathic powers and that she has the power to increase the proportion of cards correctly predicted to 30%. With the critical region as determined in part (*c*) calculate the probability of drawing the correct conclusion if the value of *p* is actually 0.30.

Solution

(a) $X \sim B(60, 0.25)$

$n = 60 \quad p = 0.25$

$\mu = np = 60 \times 0.25 = 15$

$\sigma^2 = npq = 15 \times 0.75 = 11.25$

$X \sim B(60,\ 0.25)$ is approximated by $X \sim N(15, 11.25)$

30% of 60 trials is a test value of $X_{\text{test}} = 0.30 \times 60 = 18$

We require the probability $P(X \geq 18)$

The value 18 is included in this critical region.

Using the continuity correction this corresponds to the region

$P(X > 17.5)$ under $X \sim N(15, 11.25)$.

$x_1 = 17.5 \qquad z_1 = \dfrac{x-\mu}{\sigma} = \dfrac{17.5 - 15}{\sqrt{11.25}} = 0.745\ \ (3\text{ s.f.}) \qquad \Phi(0.745) = 0.7719$

p-value $= P(X \geq 18) \approx P(Z > 0.745) = 1 - 0.7719 = 0.2281 = 0.228\ \ (3\text{ s.f.})$

(b) The significance level is $\alpha = 0.05$

The probability of the test result is

$P(X \geq 18) \approx P(Z > 0.745) = 1 - 0.7719 = 0.228\ \ (3\text{ s.f.})$

Since this probability is much greater than $\alpha = 0.05$ we accept H_0.

The true proportion is $p = 0.25$. We conclude that the medium's claim is false.

(c) The z-value corresponding to a significance level of $\alpha = 0.05$ is $z = 1.645$ for a one-tailed test. As before the sample statistic under $H_0 : p = 0.25$ is $X \sim B(60,\ 0.25)$ and is approximated by $X \sim N(15, 11.25)$. Then on substituting into $\dfrac{x-\mu}{\sigma} = z$, we obtain

$$\frac{x - 15}{\sqrt{11.25}} = 1.645 \quad \Rightarrow \qquad x = 20.517$$

For the critical value if we take the next integer up $(N = 21)$ we will, under the continuity correction, include $x = 20.5$ so the significance level will be greater than 0.05. Hence we must take $N = 22$ and the critical region is given by

Reject H_0 if $X \geq 22$.

To find the actual p-value of this critical region, we must use the continuity correction. 22 is included in the region, so the corresponding z-value is

$$z = \frac{21.5 - 15}{\sqrt{11.25}} = 1.938\ \ (3\text{ d.p.})$$

The p-value corresponding to this z-value is 0.9737

$P(z \geq 1.938) = 1 - 0.9737 = 0.0263$

and this is the actual significance level adopted by taking the critical region to be

Reject H_0 if $X \geq 22$.

(*d*) Suppose in fact $p = 0.30$

Then $X \sim B(60, 0.30)$ which is approximated by $X \sim N(18, 12.6)$

We suppose $X_{\text{test}} \geq 22$

Using the continuity correction $P(X \geq 22)$ under $X \sim B(60, 0.30)$

corresponds to the region $P(X > 21.5)$ under $X \sim N(18, 12.6)$.

$x = 21.5 \qquad z = \dfrac{x - \mu}{\sigma} = \dfrac{21.5 - 18}{\sqrt{12.6}} = 0.986 \quad (3 \text{ d.p.}) \qquad \Phi(0.986) = 0.8380$

$P(X \geq 22) \approx P(Z > 0.986) = 1 - 0.8380 = 0.162 \quad (3 \text{ s.f.})$

There is a 16.2% chance of getting the correct conclusion with this larger sample size.

An increase in the sample size (in this example from 30 to 60) improves the likelihood of arriving at the correct conclusion if in fact the null hypothesis is false. In this last example, with a sample size of 30, it was only 8.5% probable that we would conclude that the null hypothesis is false if indeed it the probability of a success is actually 0.30. Doubling the sample size to 60 has increased this to 16.2%.

Example (2)

The makers of scratch cards claim that 15% of the cards give a cash prize. The gambling regulators wish to check this claim. They believe that the percentage of cards actually giving a cash prize is less than 15%. They take a sample of 500 of these cards chosen at random and find that 57 of these cards give cash prizes.

(*a*) Use a distributional approximation to find the p-value of this result.

(*b*) Interpret your p-value in context using a 5% significance level.

Solution

(*a*) Let X denote the number of cards in the sample giving a cash prize.

$H_0 : p = 0.15 \qquad H_1 : p < 0.15$

Under H_0: $X \sim B(500, 0.15) \qquad n = 500 \quad p = 0.15$

$\mu = np = 500 \times 0.15 = 75 \qquad \sigma^2 = npq = 75 \times 0.85 = 63.75$

$X \sim B(500, 0.15)$ is approximated by $X \sim N(75, 63.75)$

$X_{\text{test}} = 57$

We require the probability $P(X \le 60)$.

The value 57 is included in this critical region.

Using the continuity correction this corresponds to the region

$P(X < 57.5)$ under $X \sim N(75, 63.75)$.

$x_1 = 60.5 \qquad z_1 = \dfrac{x-\mu}{\sigma} = \dfrac{57.5-75}{\sqrt{63.75}} = -2.192 \text{ (3 d.p.)} \qquad \Phi(2.192) = 0.9858$

$p\text{-value} = P(X < 57.5) \approx P(Z < -2.192) = 1 - 0.9858 = 0.0142$

(*b*) $p\text{-value} = 0.0142 < 0.05 = \alpha$

Reject H_0 Accept H_1

The true proportion is less than 15%

The claim made by the makers of the scratch cards is false.

Testing the mean of a normal distribution

In the examples we have considered so far the normal distribution arises as an approximation to the binomial distribution. We use it in order to test a hypothesis originally formulated on the assumption that the sample followed a binomial distribution. Naturally, problems also arise directly in the context of a normal distribution.

Imagine we have a factory process producing components. The machine is expected to produce these components to a specification. Suppose the mean width is intended to be μ and the variance σ^2. Every so often the company, in its drive for total quality, samples the components. The sample mean may vary from the expected mean owing to chance factors. However, the question is, by how much does the sample mean have to vary from the expected mean for the company to conclude that the sample is different from the expected population? If the sample mean does significantly differ from the expected mean for the population, then company will conclude that they have to reset their machines. In this context we are comparing the mean of a sample with mean of a population. Therefore, two parts of background theory that you should already have met are appropriate. (1) Finding the mean and variance of a sample; (2) The central limit theorem. Given this background theory, then in the light of what you have also learnt about hypothesis testing, you are already able to test a hypothesis concerning the deviation of a sample mean from the expected population mean.

Example (6)

Chocolate Company produces a *Tasty Bar* chocolate. The weight X of *Tasty Bars* is believed to be normally distributed with mean 80.3 g and standard deviation 1.3 g. Chocolate Company regularly take samples of size 24 as part of a quality control process to determine whether their machines are working properly.

(*a*) State the central limit theorem and use it to determine the distribution of the mean of the samples taken by Chocolate Company.

(*b*) What is the probability that the mean of a sample of 24 *Tasty Bars* selected at random is less than 80.0 g? Give your answer to 3 significant figures.

(*c*) One sample of 24 *Tasty Bars* is as follows.

78.1	80.4	79.6	81.2	78.6	80.3	78.7	77.3	80.8	78.1	79.2	80.6
79.4	79.2	79.8	80.9	78.5	80.5	79.6	79.7	80.9	79.3	80.3	78.4

(*i*) Find the sample mean and sample variance.

(*ii*) Calculate the p-value of these results.

(*d*) Let μ denote the mean of the sample given in part (*c*) and let $\mu_0 = 80.3$ denote the mean of the weight X of all *Tasty Bars*. The team in charge of quality control at Chocolate Company form the following hypotheses.

H_0 $\quad \mu = \mu_0 = 80.3$

H_1 $\quad \mu \neq \mu_0$

(*i*) Explain in context why this is a two-tailed as opposed to a one-tailed test, and why a two-tailed test is more appropriate here than a one-tailed test.

(*ii*) The Company set a significance level of 5%. In this test what is the size (probability) of the critical region for each tail?

(*iii*) State your conclusion in context.

Solution

(*a*) In this context the central limit theorem may be stated as

$$\text{If } X \sim N\left(\mu,\ \sigma^2\right) \text{ then } \bar{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right) \qquad \text{for samples of size } n.$$

(More generally the central limit theorem applies to any population whatsoever regardless of the background distribution and provided the sample size n is sufficiently large.)

We have $X \sim N\left(80.3, (1.3)^2\right)$. By the central limit theorem

$$\bar{X} \sim N\left(80.3, \frac{(1.3)^2}{24}\right) = N\left(80.3, (0.2653...)^2\right).$$

(*b*) This asks for $P\left(\bar{X} < 80.0\right)$.

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{80.0 - 80.3}{0.2653...} = -1.131 \text{ (3 d.p.)} \qquad \Phi(1.131) = 0.8710$$

$$P(X < 80.0) = P(Z < -1.131) = 1 - 0.8710 = 0.129 \text{ (3 s.f.)}$$

(*c*) (*i*) The sample mean is given by

$$\bar{x} = \frac{\text{Sum of values}}{n} \qquad \left[\bar{x} = \frac{\sum x}{n}\right]$$

Here $n = 24$ and the sum of the values is $\sum x = 1909.4$.

$$\bar{x} = \frac{1909.4}{24} == 79.558... = 79.6 \text{ g} \quad \text{(3 s.f.)}$$

The sample variance is given by

$$\sigma^2 = \frac{\text{Sum of squares}}{n} - (\bar{x})^2 \qquad \left[\sigma^2 = \frac{\sum x^2}{n} - (\bar{x})^2\right]$$

$$\sum x^2 = 151934$$

$$\sigma^2 = \frac{\sum x^2}{n} - (\bar{x})^2 = \frac{151934}{24} - (79.558...)^2 = 1.0632... = 1.06 \text{ (3 s.f.)}$$

(*ii*) The *p*-value is the probability $P(X < 79.558..)$ under the hypothesis $H_0 : \mu = \mu_0 = 80.3$ that the sample mean and variance have not changed.

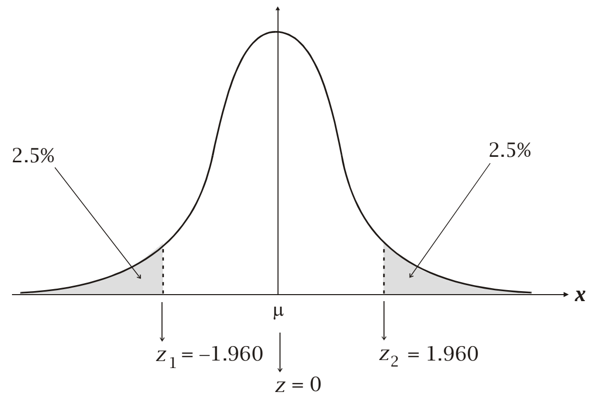
$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{79.558.. - 80.3}{0.2653...} = -2.796 \text{ (3 d.p.)} \qquad \Phi(2.796) = 0.9975$$

$$P(X < 79.558..) = P(Z < -2.796) < 1 - 0.9975 = 0.0025$$

p-value = 0.0025

(*d*) (*i*) In a two-tailed test we are testing the possibility that here the sample mean is either greater than or less than the expected population mean. That is $\mu > \mu_0$ or $\mu < \mu_0$. This is appropriate to this quality control test because in advance of the results it is possible that the output of the machine could vary in either direction – to create bars that were too heavy or too light.

(*ii*) The significance level is $\alpha = 0.05 = 5\%$. As this is a two-tailed test this means that the critical region in the test is divided between both tails of the distribution. Each tail contributes an area of $0.025 = 2.5\%$.



(*iii*) Here the p-value is

$$p\text{-value} = 0.0025 \ll 0.025 = \text{probability of the critical region for the left-hand tail}$$

Therefore we conclude

Reject H_0, accept H_1

The machines producing *Tasty Bars* do need recalibrating.

Example (7)

It is thought that a distribution has mean 2 and variance 0.4. A sample of size 8 has sample mean 2.45. Test at the 5% level whether the mean is 2.

Solution

H_0 $\quad \mu = 2$
H_1 $\quad \mu \neq 2$
Significance level = α=0.05

As this is a two-tailed test the significance level for each tail is $\frac{0.05}{2} = 0.025$

$X \sim N(2,\ 0.4)$ then $\overline{X} \sim N\left(2, \frac{0.4}{8}\right)$

$\overline{X} = 2.45$
$\alpha = 0.05$

$z_{test} = \frac{2.45 - 2}{\sqrt{0.05}} = 2.012$ (3 s.f.) $\qquad \Phi(2.012) = 0.9779$

p-value $= P\left(\overline{X} > 2.45\right) = P(Z > 2.012) = 1 - 0.9779 = 0.0221$

p-value $= 0.0221 < 0.025 =$ significance level for each tail

Reject H_0, accept H_1