Hypothesis testing with the Binomial distribution

Starter

1. *(Review of last lesson)* Given that $X \sim \text{B}\left(60, \frac{1}{6}\right)$, use the BinomialCD function on your calculator to find:

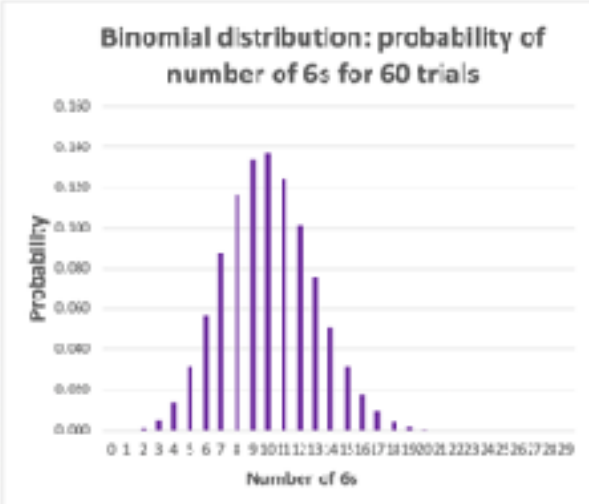
(a)	$P(X = 0)$	(b)	$P(X \le 1)$	(c)	$P(X \le 2)$	(d)	$P(X \le 3)$
(e)	$P(X \le 4)$	(f)	$P(X \le 5)$	(g)	$P(X \ge 15)$	(h)	$P(X \ge 16)$

Notes

Important note: the working out shown in the first few examples is not good enough for examinations. Hopefully though it will help you understand the method and the formal written mathematics (given after E.g. 2) can be added afterwards.

Hypothesis test for a biased dice

If a six-sided dice was rolled 60 times, we would expect ten 6s but we wouldn't be surprised if nine or eleven 6s appeared. We probably wouldn't be surprised if eight or twelve 6s appeared. At what point though would we start to believe that the dice was biased?



The graph shows the probability of rolling the number of 6s indicated on the horizontal axis. For example, the probability of rolling ten 6s is a little less than 0.0140.

In fact, for $X \sim \text{B}\left(60, \frac{1}{6}\right)$:

$$P(X = 10) = {}^{60}C_{10} \times \left(\frac{1}{6}\right)^{10} \times \left(\frac{5}{6}\right)^{50} = 0.137$$

As can be seen from the graph, the probability of getting 21 or more 6s is negligible. Similarly, for zero or one 6.

Biased against rolling a 6

Imagine five 6s were rolled i.e. half as many as expected. Assuming a $5\ \%$ significance level, would this show the dice is biased?

$P(X = 5) = {}^{60}C_5 \times \left(\frac{1}{6}\right)^{5} \times \left(\frac{5}{6}\right)^{55} = 0.0310 < 0.05$ which suggests the result is significant and the dice is biased.

However, the hypothesis test does not say the dice must return precisely five 6s. It asks whether the dice is biased against a 6 i.e. if $P(6) < \frac{1}{6}$. Therefore, we consider values which are ***at least as extreme as the observed value*** i.e. $P(X \le 5)$.

From the starter, $P(X \le 5) = 0.0512 \not< 0.05$. The result is not significant and so there is ***insufficient evidence*** to say the dice is biased against a 6.

Biased towards rolling a 6

Imagine sixteen 6s were rolled. Would this indicate the dice is biased towards getting a 6 at the $5\ \%$ level? The hypothesis test asks whether $P(6) > \frac{1}{6}$.

Again, we consider values which are ***at least as extreme as the observed value*** i.e. $P(X \geq 16)$. From the starter $P(X \geq 16) = 1 - P(X \leq 15) = 1 - 0.966 = 0.034 < 0.05$ so this time the result is significant i.e. there is ***sufficient evidence*** to suggest the dice is biased towards a 6.

However, if we carried out a hypothesis test at the $2\ \%$ level, the result would not be significant.

Other scenarios

A hypothesis test using the Binomial distribution is not just restricted to biased dice or coins.

E.g. 1 At the start of last season, Erik carried out a football team supporter's survey and his results suggested that $30\ \%$ supported Team A. After a poor season Erik now believes their support will have gone down and carries out another survey. From a sample of 40 people, six people said they support Team A. Is Erik right (use a $4\ \%$ significance level)?

Working:

Last season, $P(\text{support Team A}) = 0.3$.
If this is still true with the sample of 40 people, then $X \sim \text{B}(40, 0.3)$.
Erik now believes $P(\text{support Team A}) < 0.3$
Consider values that are ***at least as extreme as the observed value.***
$P(X \leq 6) = 0.0238 < 0.04$
The result is significant.
There is ***sufficient evidence*** to suggest that support for Team A has gone down.

One-tailed tests vs. two-tailed tests

All the tests above consider whether the probability, or proportion, has gone up or gone down — these are called ***one-tailed tests***. Hypothesis tests can be carried which simply consider whether the probability has changed — these are called ***two-tailed tests***.

The difference is that with ***two-tailed tests***, we ***halve the significance level*** when deciding whether the result is significant or not.

E.g. 2 Results from previous years indicate that $91\ \%$ of students pass the Grade 8 practical music exam. Following coronavirus, the music organisation that runs the exam wants to know whether this has changed. Recent results indicate the out of 318 students who took the Grade 8 practical music exam, 300 passed. Is there evidence, at the $5\ \%$ level, that the pass rate has changed? Try to give an explanation for your conclusion.

Working:

Previously $P(\text{pass}) = 0.91$
If it is still true with the sample of 3184 students, then $X \sim \text{B}(318, 0.91)$.
Consider values that are ***at least as extreme as the observed value.***
Since $\frac{300}{318} > 0.91$, calculate $P(X \geq 300)$:
$P(X \geq 300) = 1 - P(X \leq 299) = 1 - 0.981 = 0.019$
Since only a change is considered, ***halve the significance level*** when comparing it with the calculated probability.
$P(X \geq 300) = 0.019 < 0.025$.
So the result is significant. There is ***sufficient evidence*** to suggest that the pass rate for the Grade 8 practical music exam ***has changed***.
This may be because students had more time to practice their musical instrument during lockdown.

Formal way of writing a hypothesis test

The problem with our working in the out in the above examples is that it does not follow the formal way that hypothesis tests should be written and hence would receive few marks in an exam.

It cannot be stressed enough that the ***correct notation and language must be followed***.

Success criteria — Hypothesis testing with the binomial distribution

1. Write down the null hypothesis, H_0, ***clearly stating what p refers to***:
   $H_0 : p = \ldots$ where p is the proportion of…
2. State the alternative hypothesis, H_1.
   $H_1 : p < \ldots$ or $H_1 : p > \ldots$ (one-tailed test)
   $H_1 : p \neq \ldots$ (two-tailed test)
3. State the distribution under H_0: $X \sim \text{B}(n, p)$.
4. State the significance level: $\alpha =$
5. Use the Binomial CD function on your calculator to find the probability of results ***at least as extreme as the observed value*** i.e. $P(X \leq \ldots) =$ or $P(X \geq \ldots) =$
6. To reject or not to reject:
   if $P(X \leq \ldots)$ or $P(X \geq \ldots) < \alpha \Rightarrow$ reject H_0 (result is significant).
   if $P(X \leq \ldots)$ or $P(X \geq \ldots) > \alpha \Rightarrow$ do not reject H_0 (result is not significant)
   N.B. Remember to ***halve the significant level*** for ***two-tailed tests***.
7. Write a conclusion based on the practical situation beginning…
   "There is ***sufficient evidence*** at the α % level to suggest that the proportion…" or
   "There is ***insufficient evidence*** at the α % level to suggest that the proportion…"

Do these questions using the full and formal working.

E.g. 3 A manufacturer claims to have a $90\ \%$ germination rate for their seeds. A gardener believes it is less and to test the theory plants 100 seeds from which 84 germinate. Test the manufacturer's claim at the $5\ \%$ level.

Working: "believes it is less" $\Rightarrow$ one-tailed test
$H_0 : p = 0.9$ where p is the proportion of seeds that germinate.
$H_1 : p < 0.9$
Under H_0, $X \sim \text{B}(100, 0.9)$.
$\alpha = 0.05$
$P(X \leq 84) = 0.0399 < 0.05 = \alpha$
The result is significant so reject H_0.
There is sufficient evidence to suggest, at the $5\ \%$ level, that the proportion of seeds that germinate as stated by the manufacturer is too high.

E.g. 4 A political party's support is stagnant at 3 out of 10 people. After their party conference, they believe their support has increased. Would you accept this assertion if a further survey revealed that 36 people in a random sample of 95 people supported the party? Test at the $3\ \%$ level.

E.g. 5 A driving instructor changes their car. With the previous car, $73\ \%$ of their students passed the practical side of the driving test at the first attempt. After a period of time with the new car, 16 out of 18 passed first time. Determine, at the $10\ \%$ significance level whether the new car has made a difference in the pass rate.

Working: "made a difference in the pass rate" $\Rightarrow$ two-tailed test
$H_0 : p = 0.73$ where p is the proportion of students who pass the driving test first time
$H_1 : p \neq 0.73$
Under H_0, $X \sim$ B $(18, 0.73)$.
Although $\alpha = 0.1$***, it is a two-tailed test so we use*** 0.05***.***
Since $\frac{16}{18} > 0.73$ ***find*** $P(X \geq 16)$***.***
$P(X \geq 16) = 1 - P(X \leq 15) = 1 - 0.901 = 0.099 \not< 0.05$
The result is not significant so do not reject H_0.
There is insufficient evidence to suggest, at the $10\ \%$ level, that the pass rate for the practical side of the driving test has changed.

E.g. 6 A lottery company claims that $15\ \%$ of all tickets sold win a prize. Believing the lottery to be fraudulent, the Office for Trading Standards buys 250 tickets and wins 27 prizes. Carry out a test at the (a) $2\ \%$ and (b) $5\ \%$ to decide whether this indicates the lottery is fraudulent.

Different conclusions can be reached ***depending on the significance level*** of the hypothesis test. In addition, **another sample** may also lead to a ***different outcome***.

Video: **Hypothesis test with the Binomial distribution**
Video: **Lower tail test**
Video: **Upper tail test**
Video: **Two-tailed tests**

Solutions to Starter and E.g.s

Exercise
p400 18B Qu 1i, 2i, 3-11, (12-13 red)

Summary
Success criteria — Hypothesis testing with the binomial distribution:

1. Write down the null hypothesis, H_0, ***clearly stating what*** p ***refers to***:
   $H_0 : p = \ldots$ where p is the proportion of…
2. State the alternative hypothesis, H_1.
   $H_1 : p < \ldots$ or $H_1 : p > \ldots$ (one-tailed test)
   $H_1 : p \neq \ldots$ (two-tailed test)
3. State the distribution under H_0: $X \sim$ B (n, p).
4. State the significance level: $\alpha =$
5. State the test statistic, X
6. Use the Binomial CD function on your calculator to find the probability of results ***at least as extreme as the observed value*** i.e. $P(X \leq \ldots) =$ or $P(X \geq \ldots) =$
7. To reject or not to reject:
   if $P(X \leq \ldots)$ or $P(X \geq \ldots) < \alpha \Rightarrow$ reject H_0 (result is significant).
   if $P(X \leq \ldots)$ or $P(X \geq \ldots) > \alpha \Rightarrow$ do not reject H_0 (result is not significant)
   N.B. Remember to ***halve the significant level*** for ***two-tailed tests***.
8. Write a conclusion based on the practical situation beginning…
   "There is ***sufficient evidence*** at the $\alpha\ \%$ level to suggest that the proportion…" or
   "There is ***insufficient evidence*** at the $\alpha\ \%$ level to suggest that the proportion…"