

## Exercise Sheet Lecture 8 — Significance tests: The single-sample Z test

1. Assume that IQ scores for a certain population are approximately  $N(\mu, 100)$ . To test  $H_0 : \mu = 110$  against the one-sided alternative hypothesis  $H_1 : \mu > 110$ , we take a random sample of size  $n = 16$  from this population and observe  $\bar{x} = 113.5$ . Do we accept or reject  $H_0$  at the
  - (a) 5% level?
  - (b) 10% level?
  - (c) What is the p-value of this test?
2. At the end of the eighteenth century, the French mathematician Pierre-Simon de Laplace observed that out of 2,009 births in the community of Carcelle-le-Grignon there had been born 983 boys and 1026 girls; Laplace considered this to be a deviation from the expected mean of  $\frac{1}{2}$ , due to chance. A special investigation, initiated by him, counted babies born in numerous districts of France in the following years 1800, 1801, and 1802. It yielded a count of 110,312 boys and 105,287 girls.

Is the difference from  $1/2$  found in the sex ratio of newborns in the first study statistically significant? In the second study?

Compute a 95% confidence interval for the probability of a birth being female. What assumptions do you need to make?
3. The John Radcliffe Hospital in Oxford has about 8300 births a year. Royal Berkshire Hospital in Redding has about 5700 births a year. (Data from <http://www.birthchoiceuk.com/Access.htm>. Which is likelier to have 52% or more male births in a given year? Or is it equally likely? Explain. (There is about a 51% chance for a live-born infant to be male.)
4. One year, hypothetically, there were 252 trading days on the London Stock Exchange, and the composite index went up on 131 of them:  $131/252 \approx 52\%$ . A statistician attaches a standard error to this percentage as follows:

$$\text{SE for number} = \sqrt{252} \times \sqrt{0.52 \times 0.48} \approx 8$$

$$\text{SE for percent} = \frac{8}{252} \times 100\% \approx 3\%.$$

Is this the right SE? Answer yes or no, and explain.

5. Suppose a large research laboratory is conducting an intensive research effort to disprove a certain hypothesis  $H$ , that unfortunately happens to be true. They conduct 30 different experiments. Each time, they perform a hypothesis test at the 0.05 significance level, in which  $H$  is the null hypothesis. What is the probability that at least one of the tests will produce a result that allows them to reject the null hypothesis, and thus publish the (erroneous) claim that they have proved  $H$  to be false?
6. **The trial of the Pyx:** (*This is more or less from [FPP98].*) S. Stigler [Sti77] has described the “trial of the Pyx[...], an ancient ceremony of the Royal Mint of Great Britain. The avowed purpose of this ceremony is to ascertain that coinage issued by the Royal Mint meets the Crown’s specifications.” This was a procedure of assaying the gold content of coins struck by the Royal Mint. The Master of the Mint was to be severely punished if the weight of the coins — which determined their value, since a coin was supposed to represent simply a fixed quantity of gold or silver — was below the specified quantity. Of course, perfection is impossible in sublunar affairs, and so the Master of the Mint was allowed a margin of error, called the “remedy”.

Suppose that in one year the procedure was as follows: 1000 new guineas were selected at random from a very large number of guineas, placed in the Pyx (a ceremonial box), and weighed. A guinea should have weighed 128 grains (5760 grains/pound). The total weight should then be 128000 grains, and the remedy was set at  $1/200$  of this, or 640 grains. If the actual weight of the coins in the Pyx differed from the standard by more than the remedy, the Master of the Mint was subject to severe penalties.

- (a) Suppose the Master of the Mint manufactures honest guineas, averaging 128 grains, with a standard error of  $1/200$  of the true weight. What is the chance that he will survive the trial of the Pyx?
- (b) What if he sets the machine to make guineas that weight 127.4 grains on average? (The random error stays the same.)
- (c) **Challenging – just for fun!** Suppose the Master of the Mint manufactures 100,000 guineas with average weight  $\mu < 128$  grains and standard deviation  $1/200$  grain. If he survives the trial of the Pyx, he pockets the difference. If he fails, he will be fined 100,000

guineas. What should he choose  $\mu$  to be to maximise his average (expected) ill-gotten gains?

## References

- [FPP98] David Freedman, Robert Pisani, and Roger Purves. *Statistics*. Norton, 3 edition, 1998.
- [Sti77] Stephen M. Stigler. Eight centuries of sampling inspection: The Trial of the Pyx. *Journal of the American Statistical Association*, 72(359):493–500, 1977.