STATS 1000 / STATS 1004 / STATS 1504
Statistical Practice 1
Assignment 3
2020

DEADLINE:

- Wednesday 8^{th} April 2020 (Week 6) 5:00pm

CHECKLIST

☐: Have you shown all of your working, including probability notation where necessary?

☐: Have you given all numbers to **3 decimal** places.

☐: Have you included all R output and plots to support your answers where necessary.

☐: Have you made sure that all plots and tables each have a caption.

☐: If before the deadline, have you submitted your assignment via the online submission on MyUni?

☐: Is your submission a single word document or pdf file - correctly orientated, easy to read? If not, penalties apply.

☐: Penalties for more than one document - 10% of final mark for each extra document. Note that you may resubmit and your final version is marked, but the final document should be a single file.

☐: If after the deadline, but within 24 hours, have you contacted us via the enquiry page on MyUni and then submitted your assignment online via the online submission on MyUni?

☐: Penalties for late submission - within 24 hours 40% of final mark. After 24 hours, assignment is not marked and you get zero.

☐: Assignments emailed instead of submitted by the online submission on MyUni will not be marked and will recieve zero.

☐: Have you checked that the assignment submitted if the correct one, as we cannot accept other submissions after the due date.

☐: Do not write directly on the question sheet.

1. **Random variable**

   This question may be hand-written. Remember to attach it to the rest of the assignment.

   Notation recap: to write the probability that the random variable X has the value a, we write $P(X = a)$, not $P(X) = a$.

   A dishonest casino has a biased six-sided die. It still has the faces 1 to 6, but instead of each face being equally likely it has the following probabilities of each face:

Face	1	2	3	4	5	6
Probability of face	0.18	0.164	0.164	0.164	0.164	0.164

   Let X denote the random variable that is the face rolled. Using probability notation and showing your working, answer the following questions:

   (a) Verify that this is a valid discrete probability distribution. Remember we need to check two things.

   [2 marks]

   - All the probabilities lies between 0 and 1.
   - The sum of all the probabilities is

$$0.18 + 0.164 + \ldots + 0.164 = 1.$$

   Hence, this is a valid discrete probability distribution.

   (b) What is the probability of rolling a six or a five?

   [2 marks]

$$P(X = 6 \text{ or } X = 5) = P(X = 6) + P(X = 5) = 0.164 + 0.164 = 0.328.$$

   (c) What is the probability of rolling an even number?

   [2 marks]

$$\begin{aligned} P(X = 2 \text{ or } X = 4 \text{ or } X = 6) &= P(X = 2) + P(X = 4) + P(X = 6) \\ &= 0.164 + 0.164 + 0.164 \\ &= 0.492. \end{aligned}$$

(d) Calculate the mean value of X.

[2 marks]

$$\begin{aligned}\mu_X &= 1 \times P(X=1) + 2 \times P(X=2) + \cdots + 6 \times P(X=6) \\ &= 1 \times 0.18 + 2 \times 0.164 + \ldots + 6 \times 0.164 \\ &= 3.46.\end{aligned}$$

[Total: 8]

2. **Linear combinations**

This question may be hand-written. Remember to attach it to the rest of the assignment.

For each of the following games use the weighted die in Q1. Calculate your true mean winnings. Show your working for full marks. *Hint: Let your winnings be denoted by Y and write an equation relating Y to the number on the die X.*

(a) The casino pays you \$1.5 for each dot on the face of the dice corresponding to the number you rolled (e.g. rolling a 2 would get you \$3, rolling a six would get you \$9).

[2 marks]

$$Y = 1.5X.$$

So

$$\mu_Y = 1.5\mu_X = 1.5 \times 3.46 = \$5.19.$$

(b) The casino pays you \$1.5 for each spot of the number you roll, but you have to pay \$4 to play.

[2 marks]

$$Y = 1.5X - 4.$$

So

$$\mu_Y = 1.5\mu_X - 4 = 1.5 \times 3.46 - 4 = 5.19 - 4 = \$1.19.$$

[Total: 4]

3. **Random allocation of observations**

   This question must be typed up in Word. Remember to attach it to the rest of the assignment.

   In Practical 5, we randomly allocated 50 observations into two groups of size 25 each. We did the random allocation in R and then analysed the two groups (Group A) and (Group B). In your Word document, include the following:

   (a) A table of summary statistics for each group.

   The summary statistics needed are the sample mean, sample standard deviation, sample median, sample IQR.

   [1 mark]

   For my randomization, the summary statistics are given in Table **??**. Each of your tables should give different values.

   In Table **??** I have given the summary statistics for the wood dataset. If your summary statistics for your randomisation data matches this exactly then your randomization did not work. There is a 1 in 126410606437752 chance that your randomisation gave the same results as the wood dataset.

```
##         mean    sd median  IQR
## Low   11.484 7.560    8.6 11.1
## High  10.816 6.618    9.2 10.6
```

```
##         mean    sd median IQR
## Low   16.284 6.391   16.2 7.7
## High   6.016 2.330    5.6 3.4
```

   (b) A side-by-side box-plot of the observations for each group.

   [1 mark]

   Figure 1 gives the side-by-side box-plot for my randomised data.

   (c) A panel histogram of the observations for each group.

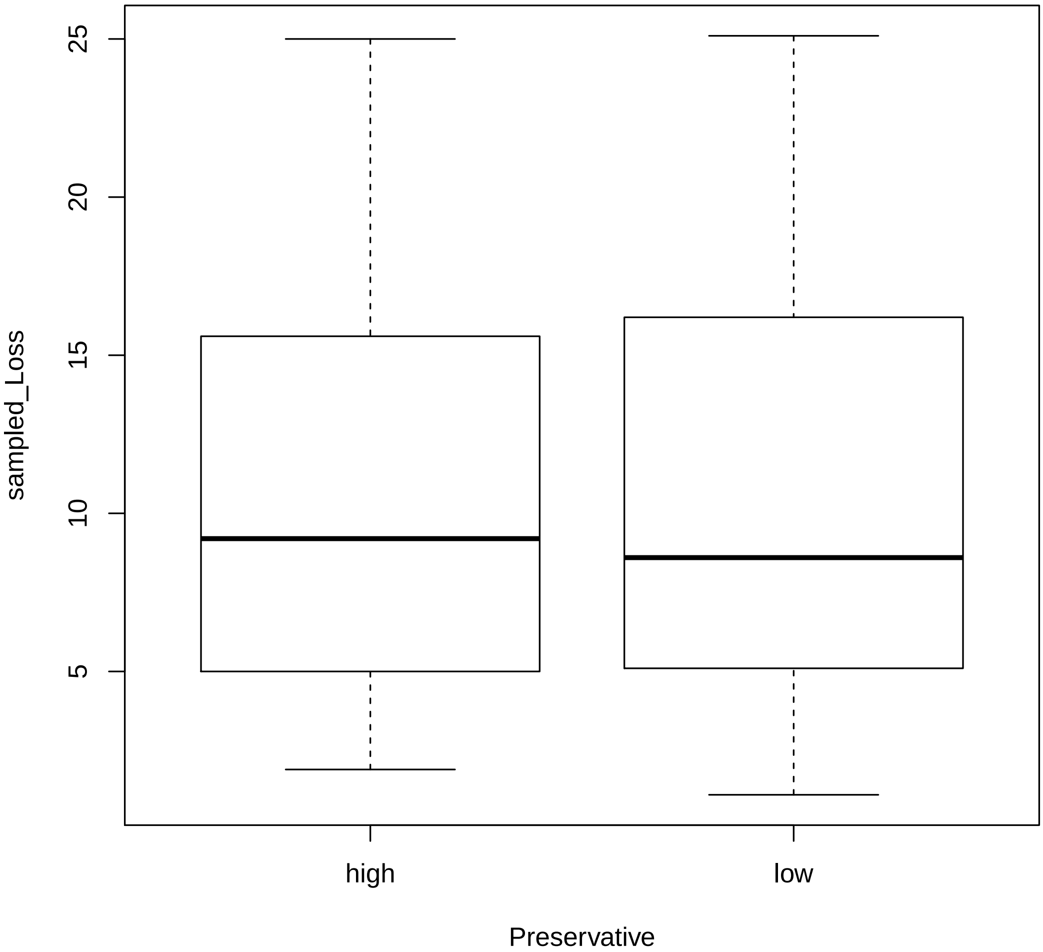

Figure 1: Side-by-side boxplot of observations for each group in the randomised data displaying larger spread in group A than in group B.

[1 mark]

Figure 2 is the panel histogram for my randomized data.

(d) Compare the distribution of the observations for each group.

Remember to consider shape, location, spread and outliers.

[4 marks]

shape: With my data, both appear unimodal. Both groups are right-skewed.

location: median and mean lower for group B compared to group A.

spread: IQR and standard deviation larger for group A compared to group B.

Outliers: No outliers in my randomised data.

(e) **Challenging question**

The observations in the spreadsheet given to you are in fact the percentage loss of timber given in the `wood.sav` (Practical 2). In that case we had two levels of preservative - high and low. I have removed this information from the data.

How do your summary statistics, box-plots, and histograms compare to those you obtained in Practical 2 (You do not need to include the woods dataset figures from the practical). Why do you think they are different? The main thing to focus on is the differences in the measures of location.

[3 marks]

In the wood data set there is a large difference in the mean loss for the two groups. This is not as obvious in the randomised dataset. When the data is randomised then the large losses seen with the low level of preservative are equally likely to be in both group A and group B and so we see less difference in the means.

[Total: 10]

4. **Binomial question**

For full marks, please show all working and use probability notation.

This question may be hand-written.

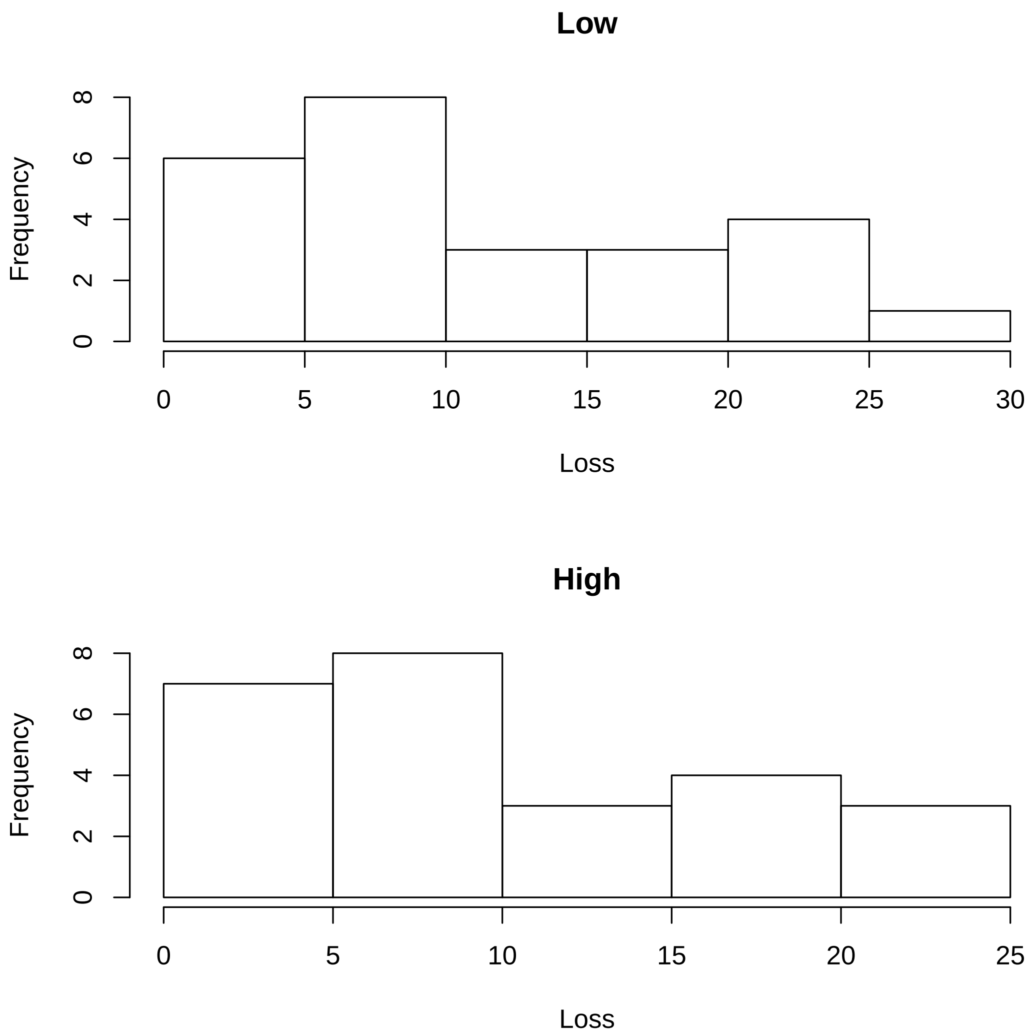

Figure 2: Panel histogram of observations for each group in the randomised data displaying unimodality for both groups.

A weighted coin is tossed 5 times. It is biased such that the probability of a head is 0.75.

Let X be the random variable that represents the total number of heads obtained.

(a) What is the distribution of X? Justify your answer.

[5 marks]

X has a binomial distribution.

- (B)inary - success - head, failure - tail.
- (I)ndependent - the result from one toss should not effect another.
- (N)umber - there is a fixed number of tosses, in this case $x = 5$.
- (S)uccess - the probability of success is constant, in this case we know that the probability of success p is 0.75.

(b) Using R, calculate the following.

i. The probability of obtaining no heads.

[2 marks]

$$\begin{aligned} P(X = 0) &= \text{dbinom(0, size = n, prob = p)} \\ &\approx 0.001. \end{aligned}$$

ii. The probability of at least one head.

[2 marks]

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \text{pbinom(0,size = n, prob = p)} \\ &= 1 - 9.765625 \times 10^{-4} \\ &\approx 0.999. \end{aligned}$$

iii. The probability of between 1 and 3 heads (inclusive).

[2 marks]

$$
\begin{aligned}
P(1 \leq X \leq 3) &= P(X \leq 3) - P(X = 0) \\
&= \text{pbinom(3,size} = \text{n, prob} = \text{p)}- \\
&\quad \text{pbinom(0,size} = \text{n, prob} = \text{p)} \\
&= 0.3671875 - 9.765625 \times 10^{-4} \\
&= 0.366.
\end{aligned}
$$

[Total: 11]

5. **Normal question**

   For full marks, please show all working and use probability notation.

   This question may be hand-written.

   The weight of male red kangaroos is normally distributed with a mean of 78 kg and a standard deviation of 6 kg.

   Let X be the random variable that represents the weight of a randomly selected male red kangaroo.

   Using R, calculate the following:

   (a) The probability that a randomly selected male red kangaroo weighs less than 60 kg?

   [2 marks]

$$
\begin{aligned}
P(X < 60) &= P(X \leq 60) \\
&= \text{pnorm(60,mu, sd)} \\
&= 0.001.
\end{aligned}
$$

   (b) The probability that a randomly selected male red kangaroo weighs more than 90 kg or less than 60 kg?

   [2 marks]

$$
\begin{aligned}
P(X < 60 \text{ or } X > 90) &= P(X \leq 60) + P(X > 90) \\
&= P(X \leq 60) + (1 - P(X \leq 90)) \\
&= \text{pnorm(60,mu, sd)} \\
&+ (1 - \text{pnorm(90, mu, sd)}) \\
&= 0.024.
\end{aligned}
$$

(c) The probability that a randomly selected male red kangaroo weighs between 60 and 90 kg?

[2 marks]

$$
\begin{aligned}
P(3 \leq X \leq 90) &= P(X \leq 90) - P(X \leq 60) \\
&= \text{pnorm(90,mu, sd)} \\
&- \text{pnorm(60, mu, sd)} \\
&= 0.976.
\end{aligned}
$$

(d) The weight c such that 10% of all male red kangaroo are less than this weight?

[2 marks]

$$
\begin{aligned}
P(X \leq c) &= 0.10 \\
\Rightarrow c &= \text{qnorm(0.1, mu, sd)} \\
&= 70.311\ kg.
\end{aligned}
$$

(e) Presentation marks (1 for presenting it in Word, 1 for appropriate figure captions, 1 for appropriate table captions).

[3 marks]

Question needs to be in Word, with figure captions, and table captions.

[Total: 11]

[[Assignment total: 44]]