

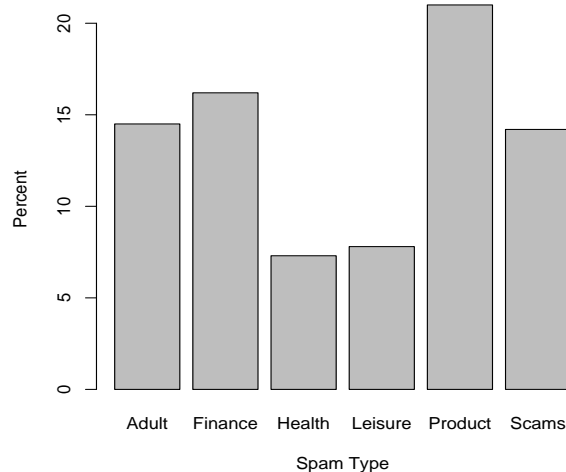
PROJECT 3 SOLUTIONS

Statistics 401: Fall 2006

Due: Friday, September 22

1. (1 pt) A distribution is the range of values that a variable can take on and the frequency with which it takes on these values
2. (1 pt) Figure 1 contains a bar plot of the spam data set. See Appendix 1 for R code.

Figure 1: Bar Chart of Spam Type



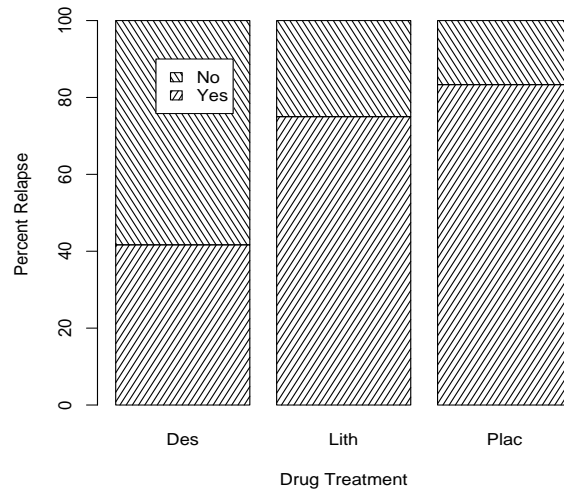
3. (a) (1 pt) The stem-plot showing the distribution of the number of hikers observed at Bear Trap Canyon trailhead is

```
3 | 245
4 | 1228
5 | 78
6 | 4
```
 - (b) (2 pts) The median number of hikers is 42 while the mean number of hikers is 45.3. The mean is larger than the median due to the presence of the right skew in the stem-plot.
 - (c) (2 pts) The IQR of the data is 18.25 hikers, the standard deviation is 11.086 hikers, and the variance is 122.9 hikers².
4. (a) (2 pts) Among 500,000 registrants at authentic happiness.com, the 577 who had completed “baseline happiness-level questionnaires” were in the sample. Since this is not a random sample, we can not make general conclusions from this data to the larger population of all human beings happiness.
 - (b) (2 pts) Seligman is referring to the fact that the 577 subjects were randomly assigned to one of 6 treatment groups (five “experimental interventions” and one control) and the control group performed a “placebo” intervention by simply writing something about some childhood memory. This is the “best” type of experiment if all possible extraneous variables

were controlled through direct control. If gender or income levels are suspected of affecting happiness, a better experimental design could block on these.

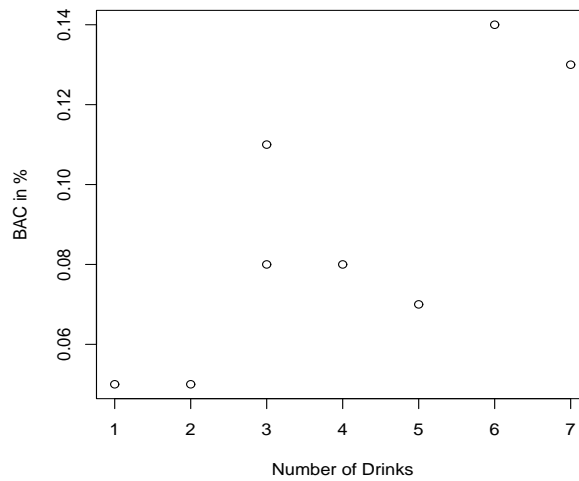
- (c) (2 pts) The “placebo” intervention required a subject to write something about some childhood memory, which is purportedly unrelated to the happiness of the subject. The purpose of the placebo is to control for the psychological positive effect that subjects may experience while taking part in an experiment which might increase their happiness.
 - (d) (1 pt) Measurement or response bias can result when subjects self-report their response.
5. (a) (1pt) This problem is an example of a completely randomized design or CRD.
- (b) (2 pts) The segmented bar chart is given in Figure 2.

Figure 2: Segmented Bar Chart of Drug Treatment and Cocaine Relapse



- (c) (2 pts) It appears that desipramine is the best drug treatment with about half as many relapses as the other two treatments, while lithium is only a little better than the placebo.
6. (a) (1pt) A scatterplot with number of drinks as an explanatory variable and BAC as the response is given in Figure 3.

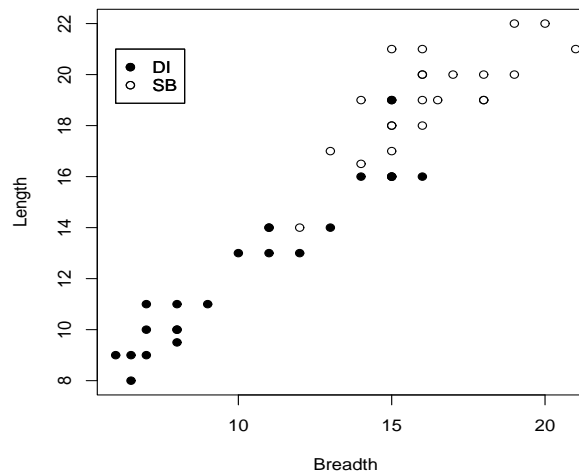
Figure 3: Scatterplot of BAC versus number of drinks



(b) (2 pts) The form of the relationship appears to be linear. The direction is positive since increased number of drinks correspond to an increase in BAC. The strength of the relationship between drinks and BAC appears to be moderate since the points do not appear too spread out from the line.

7. (a) (2 pts) Figure 4 shows a scatterplot of Length versus Breadth for Dangar Island and Salamander Bay.

Figure 4: Scatterplot of Length versus Breadth for Dangar Island and Salamander Bay.



(b) (1 pt) According to the scatterplot, the breadth and length measurements of the jellyfish at Salamander Bay appear to be larger on average than the measurements of jellyfish at Dangar Island.

(c) (2 pts) The 5 number summary of the lengths at Dangar Bay is

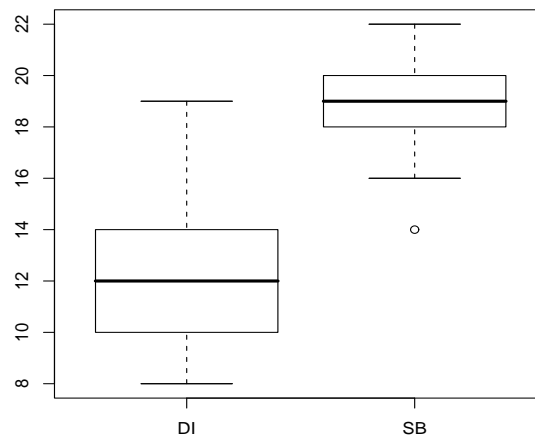
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
8.00	10.00	12.00	12.34	14.00	19.00

The 5 number summary of the lengths at Salamander Bay is

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
14.00	18.00	19.00	18.98	20.00	22.00

(d) (2 pts) The comparative boxplots of length for each of the two locations is given in Figure 5.

Figure 5: Side by side boxplots of Lengths for Dangar Island and Salamander Bay.



(e) (1 pt) The distribution of the lengths of jellyfish from Dangar Island is below the distribution of the lengths of jellyfish from Salamander Bay, and there is very little overlap. In fact, the minimum jellyfish length from Salamander Bay is larger than 75% of the jellyfish observed from Dangar Island (that is, minimum jellyfish length from Salamander Bay is larger than the third quartile from Dangar Bay). This supports the conclusion from 7b.

Appendix 1

```
> # Problem 2
> prop=c(.145,.162,.073,.078,.21,.142)
> percent=prop*100
> barplot(percent,names=c("Adult","Finance","Health","Leisure","Product","Scams"),
          xlab="Spam Type",ylab="Percent")
```

```
> # Problem 3
> hike<-c(64, 48, 42, 41, 57, 32, 34, 35, 42, 58)
> stem(hike)
```

The decimal point is 1 digit(s) to the right of the |

```
3 | 245
4 | 1228
5 | 78
6 | 4
```

```
> median(hike)
[1] 42
> mean(hike)
[1] 45.3
> IQR(mpg)
[1] 4.75
> sd(mpg)
[1] 2.884826
>
>
> # Problem 5
> freq=matrix(c(10,18,20,14,6,4),nrow=2,byrow=T)
> freq
      [,1] [,2] [,3]
[1,]  10  18  20
[2,]  14   6   4
> rownames(freq)=c("Yes","No")
> freq
      [,1] [,2] [,3]
Yes   10  18  20
No    14   6   4
> colnames(freq)=c("Des","Lith","Plac")
> freq
      Des Lith Plac
Yes   10  18  20
No    14   6   4
> percent=prop.table(freq,2)*100
> ang=c(60,120)
> index=c(2,1)
```

