The University of Sydney
School of Mathematics and Statistics

Tutorial Week 9

MATH1905: Statistics (Advanced) — Semester 2, 2017

Web Page: `http://sydney.edu.au/science/maths/MATH1905`
Lecturer: Michael Stewart

Due to the public holiday, some material relevant to these exercises may not yet have been covered. Please take note of lecturer and tutor announcements relating to this.

1. A random sample of 80 observations from a population **known** to have a standard deviation of 12 gave a sample average of $\bar{x} = 6.55$.

(a) Taking the observed value as an *estimate* of the (unknown) population mean μ, compute a standard error to go with this estimate.

(b) Provide 3 (two-sided) confidence intervals for μ at confidence levels

(i) 90%

(ii) 95%

(iii) 99%

(c) State what assumption(s) and/or theoretical result(s) one needs for these confidence intervals to be

(i) valid;

(ii) *approximately* valid.

(d) Provide 3 lower confidence limits for μ at levels

(i) 90%

(ii) 95%

(iii) 99%

(e) Compute a value for the z-statistic for testing the hypothesis $H_0 \colon \mu = 5$.

(f) Compute a p-value for a test of H_0 against the (one-sided) alternative $H_1 \colon \mu > 5$.

(g) Write a sentence giving an interpretation of the p-value.

(h) Is this significant at the

(i) 10% level;

(ii) 5% level;

(iii) 1% level?

(i) How do your answers change if the alternative is two-sided?

(j) How do your hypothesis test answers relate to your confidence interval/limit answers?

2. A p-value of 0.98 indicates that the null hypothesis is true. Comment.

3. A random sample of 30 households was selected as part of a study on electricity usage, and the number of kilowatt-hours (kWh) was recorded for each household in the sample for the March quarter of 2013. The average usage was found to be 375kWh. In a very large study in the March quarter of the previous year it was found that the standard deviation of the usage was 81kWh. Assuming the standard deviation is unchanged and that the usage is normally distributed, which of the following options is an appropriate expression for the 99% confidence interval for the mean usage in the March quarter of 2013?

(a) $375 \pm 2.756 \times \frac{81}{\sqrt{30}}$

(b) $375 \pm 2.576 \times \frac{9}{\sqrt{30}}$

(c) $375 \pm 2.326 \times \frac{81}{\sqrt{30}}$

(d) $375 \pm 2.576 \times \frac{81}{\sqrt{30}}$

4. A p-value of 0.2 means:

(a) there is 20% chance H_0 true,

(b) there is 20% chance H_1 true,

(c) there is strong evidence against H_0,

(d) the data are consistent with H_0.

5. The performance in kilometres/litre (km/l) of a particular model of car tested by machine is 7 km/l and the distribution appears normal. Company engineers have redesigned the carburettor in an effort to improve the performance and have equipped a sample of 36 cars with this new carburettor. When tested the average performance of the sample was 7.6 km/l, with a sample standard deviation of 1.5 km/l. The sample has no "outliers". What are the null (H_0) and alternative (H_1) hypotheses to be tested using the sample?

(a) $H_0 : \mu = 7,\ H_1 : \mu \neq 7$,

(b) $H_0 : \mu < 7,\ H_1 : \mu = 7$,

(c) $H_0 : \mu > 7,\ H_1 : \mu \leq 7$,

(d) $H_0 : \mu = 7,\ H_1 : \mu > 7$.

6. In the preceding problem the appropriate p-value of the test is

(a) 0.6554,

(b) 0.0082,

(c) between 0.01 and 0.025,

(d) none of the these.

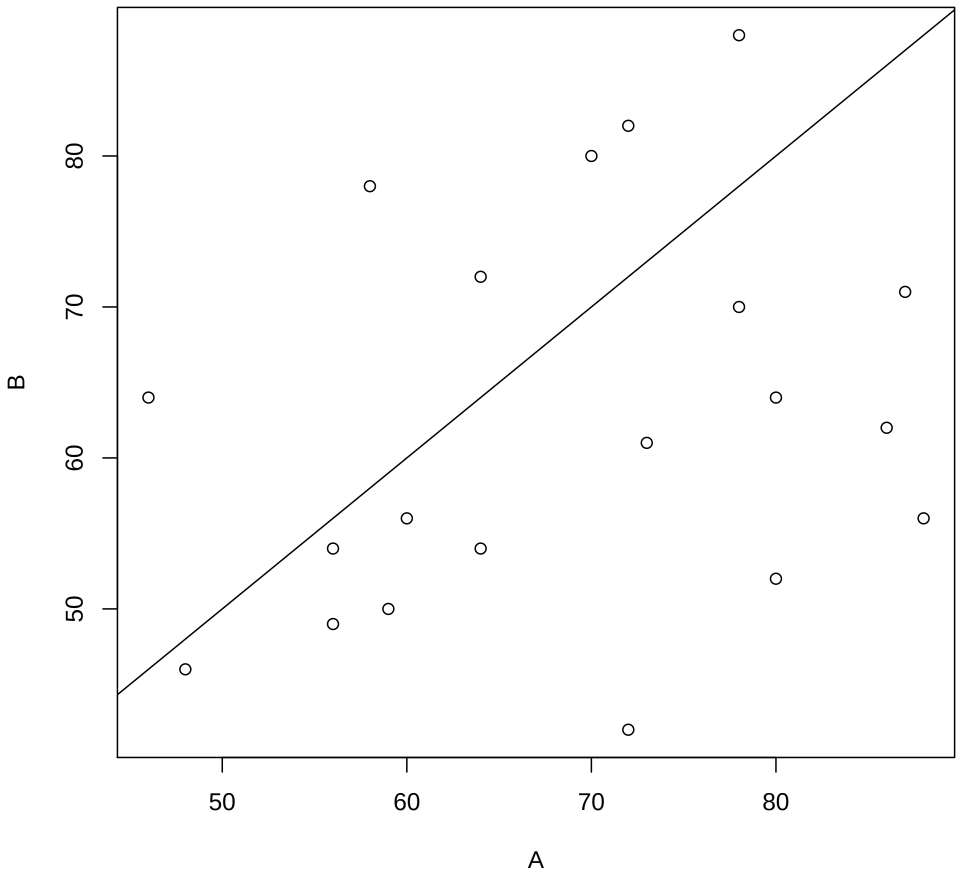
7. An insurance assessor has received estimates from two different repair garages (A and B) for minor repairs on 20 cars and wants to know if there strong evidence of a difference in estimates on average.

The estimates on each of the 20 cars, from each of the two garages are entered into the R objects `A` and `B` as below:

```
A=c(48,56,87,88,86,64,80,78,72,70,80,58,72,60,64,46,56,59,73,78)
B=c(46,49,71,56,62,54,52,88,82,80,64,78,42,56,72,64,54,50,61,70)
```

Based on the following output, answer the questions that follow:

```
plot(A,B)
abline(0,1)
```




```
mean(A)
```

```
[1] 68.75
```

```
mean(B)
```

```
[1] 62.55
```

```
sd(A)
```

```
[1] 12.71499
```

```
sd(B)
```

```
[1] 12.92275
```

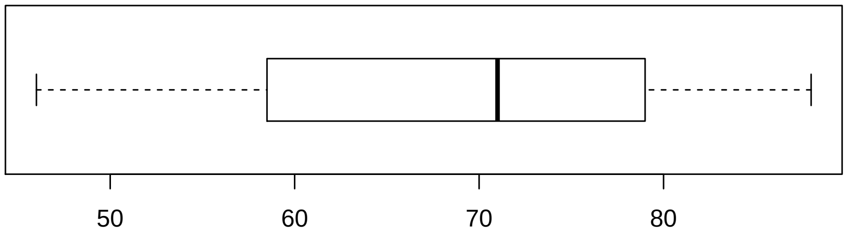
```
sd(A-B)
```

```
[1] 15.52112
```

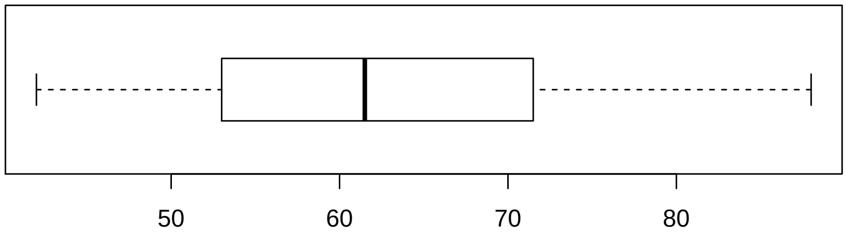
```
qt(c(0.9,0.95,0.975,0.99,0.995),df=19)
```

```
[1] 1.327728 1.729133 2.093024 2.539483 2.860935
```

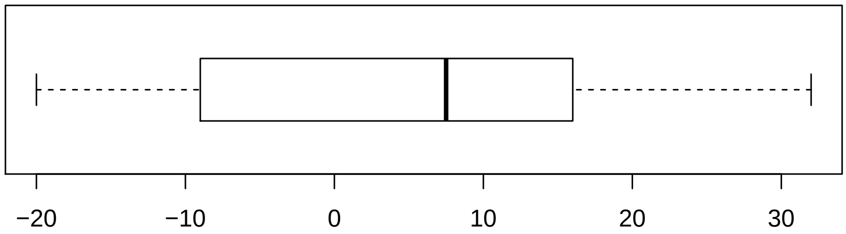
```
boxplot(A,horizontal=T)
```




```
boxplot(B,horizontal=T)
```




```
boxplot(A-B,horizontal=T)
```



(a) Let μ_d represent the true mean difference for minor repairs of this type. Explain why a paired test is appropriate.

(b) Making an appropriate normality assumption, set up the appropriate hypotheses and perform a paired t-test (i.e. compute observed value of test statistic and obtain p-value).

(c) Comment on the assumption of normality by referring to the appropriate boxplot(s) above.