Semester 1	**Tutorial Week 9**	2013

If $X_i \sim N(\mu, \sigma^2)$, a $(1-\alpha)\%$ CI for the true mean μ is $\bar{x} \mp t_{n-1,\alpha/2}\,\frac{s}{\sqrt{n}}$

If $X_i \sim Bern(p)$, a $(1-\alpha)\%$ CI for the true prop. p is $\hat{p} \mp z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ when n is large and $\hat{p} = \frac{X}{n}$.

A *Z-test* can be conducted to test the *population proportion* with $H_0 : p = p_0$ based on a random sample of binary variables of size n where n is large and the sample total $X \sim B(n, p)$. The test statistic is $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0, 1)$ where the sample proportion $\hat{p} = \frac{X}{n}$.

A *paired t-test* can be conducted to *compare population means* with $H_0 : \mu_d = 0$ based on a random sample of differences $d_i = x_i - y_i$ from n pairs of measurements drawn from normal populations. The test statistic is $t_{\text{obs}} = \frac{\bar{d} - 0}{s_d/\sqrt{n}} \sim t_{n-1}$ where $\bar{d}$ and s_d are the sample mean and sd of d_i respectively.

Tutorial discussion: Q3-7 and Q9-11 marked with *

1. **Multiple choice** To compute a 90% confidence interval (CI) for μ based on a random sample of size 25 from $N(\mu, \sigma^2)$ (μ and σ^2 are unknown), the value of t from the t-table is (approx):

   (a) 2.064 (b) 1.708 (c) 1.316 (d) 1.318 (e) 1.711

2. **Multiple choice** A random sample of size 25 from a normal population with mean μ and (unknown) variance gave the sample mean, $\bar{x} = 42.7$ and sd, $s = 5$. A 90% CI for μ is:

   (a) $42.7 \mp 1.711 \times \frac{5}{\sqrt{25}}$ (b) $42.7 \mp 1.711 \times \frac{25}{\sqrt{25}}$ (c) $42.7 \mp 1.645 \times \frac{5}{\sqrt{24}}$

   (d) $42.7 \mp 1.708 \times \frac{25}{25}$ (e) $42.7 \mp 1.708 \times \frac{5}{\sqrt{24}}$

3. **Multiple choice** *A random sample of $n = 200$ observations shows that there are $X = 36$ successes. The objective is to test at the 1% significance level that the true proportion of successes in the population is less than 24% ($p_0 = 0.24$). The test statistic should be:

   (a) $Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$ (b) $T_{n-1} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$ (c) $X \sim B(n, p_0)$

   (d) $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0, 1)$ (e) $T_{n-1} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim t_{n-1}$

4. **Multiple choice** *The value of the test statistic in Q3 is

   (a) -1.99 (b) -0.99 (c) 0 (d) 0.99 (e) 1.99

5. **Multiple choice** *The information in Q3 and Q4 tell us that:

   (a) the P-value< 0.01 and we have evidence against H_0.

   (b) the P-value< 0.01 and the data are consistent with H_0.

(c) the P-value> 0.01 and we have strong evidence against H_0.

(d) the P-value> 0.01 and the data are consistent with H_0.

(e) none of the above.

6. *Construct a 99% confidence interval (CI) for the population proportion in Q3

7. *(Table 8.22, p345, Rosner) An important hypothesis in hypertension research is that sodium restriction may lower blood pressure. The data on urinary sodium were obtained on 8 individuals enrolled in a sodium restricted group. Data were collected at baseline and after 1 week dietary counselling.

Person	1	2	3	4	5	6	7	8	Mean	sd
Week 0	7.85	12.03	21.84	13.94	16.68	41.78	14.97	12.07	17.65,	10.56
Week 1	9.59	34.50	4.55	20.78	11.69	32.51	5.46	12.95	16.50,	11.63
Diff	-1.74	-22.47	17.29	-6.84	4.99	9.27	9.51	-0.88	1.14,	12.22

Assuming that the differences are normally distributed, perform a test of the hypothesis that the dietary counseling made no difference in reducing sodium intake against the alternative hypothesis that there is a difference.

Use R to answer Q8 to Q11

8. Find the exact P-value in Q5 and Q7 using the test statistic and R.

Hint for Q8: Without a set of data, we will use the `pnorm` and `pt` functions and the calculated test statistics. Make sure to specify the correct `df` in `pt`, and don't forget to subtract from 1 and/or multiply by 2.

9. *Consider the differenced data in Q7. Create a vector from the data in the row named 'Diff' by typing `diff=scan()`. Then copy the row of 'Diff' from the pdf file, paste it into the R session window and hit `enter` twice. Use a suitable one-sample t-test, `t.test(diff)`, for these differences to test $H_0 : \mu_d = 0$ against $H_0 : \mu_d \neq 0$.

10. *R can be used directly for paired data. Create two vectors, say called `x` and `y`, from the data "Week 0" and "Week 1" in Q7 in a similar way as the vector `diff`. Now execute `t.test(x, y, paired = T)` to test $H_0 : \mu_d = 0$ against $H_0 : \mu_d \neq 0$.

Hint for Q9-10: With a set of data, we will use the `t.test` function. Note that the hypothesized value in H_0 is 0 which is set as default value and so you don't have to specify it. For the alternative, it can be `alt="greater"`, or `alt="less"`. The default for `alt` is two-sided and so you don't have to specify it too. The P-value is reported in the output, together with the test statistic, df and CI.

Hint for Q10: If you forget to specify the option `paired = T`, R will treat the data as two independent samples (default setting is `paired = F`) and the result will be different.

11. *The data in `lead.txt` contain binary variables for 124 people. The variable is 1 if the blood lead levels are above 40 micrograms/100ml measured in both 1972 and 1973 and 0 otherwise. Read the data using

```
lead=scan(file=url("http://www.maths.usyd.edu.au/math1015/r/lead.txt"))
```

(a) Count `x` the number of people having blood lead levels above 40 micrograms/100ml measured in both 1972 and 1973 and hence calculate such sample proportion.

(b) Test if the proportion of people having blood lead levels above 40 micrograms/100ml measured in both 1972 and 1973 is 0.15. Calculate the test statistic and the P-value and draw conclusion.

Hint: There is no specific R module to perform the Z-test. One should set suitable equations to calculate the test statistic and P-value.

1. A manufacturer claims that the breaking strength of its main product (steel wires) is normally distributed with mean 1800N. A new technique in the manufacturing process states that the breaking strength can be increased. To test this claim a sample of 50 steel wires is tested and it is found that the mean breaking strength is 1850N with sd 100N. Can you claim that the new technique is better? Justify your decision using a suitable statistical argument.

2. There are three politicians who are attempting to win the Democratic nominatin for senator. In a survey of 1000 Democrats, candidate A is favored by 550 people, candidate B is supported by 300 people, and the remaining 150 respondents favor candidate C. Do these results provide enough statistical evidence at the 5% significance level to indicate that candidate A will receive at least 50% of the vote?

3. A psychologist has performed the following experiment. For each of 10 sets of identical twins who were born 30 years ago, he recorded their annual incomes according to which twin was born first. The results (in $1000s) are shown below. Can he infer at 5% significance level that there is a difference in income between the twins?

Twin	1	2	3	4	5	6	7	8	9	10
First	32	36	21	30	49	27	39	38	56	44
Second	44	43	28	39	51	25	32	42	64	44