Mathematical Statistics

Lab 8

1. Look at the table taken from an article and test the hypothesis

$$H0: E(\text{cortical BMD}) = 1110$$
$$H1: E(\text{cortical BMD}) \neq 1110$$

Use significance level 0.05.

Table 1

	ALSPAC (discovery)	
	n = 999	
Age, years	15.4	(0.22)
Men, no. (%)	466	(47%)
Height, cm	169.5	(8.2)
Weight, kg	61.0	(10.7)
Position of cortical section from distal end of tibia	50%	
cortical BA, mm^2	300.5	(48.3)
cortical BMC, mg	330.0	(50.5)
cortical BMD, mg/cm^3	1100.0	(38.1)
cortical Th, mm	5.40	(0.65)
cortical PC, mm	72.6	(6.0)
cortical EC, mm	38.7	(5.8)
Position of trabecular section from distal end of tibia	NA	
trabecular BMD, mg/cm^3	NA	
Total body BMD, g/cm^2	n = 4003	1.03
Femoral neck BMD, g/cm^2	n = 3328	0.98
Lumbar spine BMD, g/cm^2	NA	

Values are mean(SD), unless otherwise stated.
BA = bone area, BMC = bone mineral content, BMD = bone mineral density, Th = th
doi:10.1371/journal.pgen.1001217.t001

Hint:

Quantiles for t-distributions with $df > 100$ are practically equal to the quantiles for standard normal distribution.

2. Authors claim the sample mean was 10 and the estimated standard deviation was 5. The sample size was 100. Authors calculated the value of the t-statistic ($t = -4$) and p-value (p-value $= 0.00012$). What hypothesis exactly did they test?

3. Researchers wanted to test hypothesis about the expected lifetime for people born on the year 1930. They took a random sample of 500 individuals from birth records and searched official records for their death certificates. For most people they found the corresponding

death certificates and t-test was used to test hypothesis about the expected lifetime using the available data. The researchers wanted to publish their research, but the reviewers say the t-test was not appropriate test to use to analyse this kind of data. Why?

4. Researcher did four experiments and got four measurements (29; 41; 41; 33). The mean of the experiments was 36, but an interesting theory predicts the expected value of the experiment to be 48. Can the expected value of the measurements be 48?
   - Which statistical test one can use to test this hypothesis?
   - Find the value of the test statistic!
   - What you can conclude from the test statistic? Should you stay with H_0 or should you reject H_0?

5. There are two varieties of wheat available for farmers. A researcher from the Estonian University of Life Sciences decided to investigate the varieties in nine test fields. Each test field was split in two halves and in one (randomly selected) half variety A was grown and on the other half variety B was grown. In the autumn the yields were measured:

Field	Variety A	Variety B
1	86	71
2	69	72
3	96	84
4	68	60
5	77	70
6	98	87
7	56	48
8	68	61
9	69	71

   Test the hypothesis:

$$\begin{aligned} H_0 &: E(\text{yield of Variety } A) = E(\text{yield of Variety } B) \\ H_1 &: E(\text{yield of Variety } A) \neq E(\text{yield of Variety } B) \end{aligned}$$

6. Researcher measured the speed driver can complete a race track (lap time). Then a driver got to drink a small glass of vodka and had to complete the race track again. The time to complete the track was measured on both occasions. Later the data was analysed by using the paired t-test. It turned out, that on average the driver was speedier after drinking the small glass of vodka (the difference was statistically significant). Similar analysis was performed on the number of minor errors driver made during the race (weave in and out of lanes, for example). Again the mean number of errors was smaller after drinking the vodka (statistically significant difference).

   The researcher wanted to publish the findings (drinking vodka makes you a better driver), but a colleague intervened. The college suspected the results were misleading due to a confounder variable. Which variable might be the confounder in this example?

(1)

$H_0: E[BMD] = 1110$

$H_1: E[BMD] \neq 1110$

$n = 999$

$\alpha = 0{,}05$

Statistic

$$T(X) = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{1100 - 1110}{38.1/\sqrt{999}} = -8{,}29$$

$f = 998$

$$|t| \overset{?}{>} t_{\alpha/2}(f)$$

$|-8{,}29| > 1{,}96 \Rightarrow H_0$ rejected

$\boxed{2}$

$\bar{x} = 10$

$s = 5$

$n = 100$

$t = -4$

p-value = 0,00012

$H_0 : \mu = \mu_0$

$H_1 \; \mu \neq \mu_0$

$$T(X) = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \qquad \mu_0 = 12$$

$$\mu_0 = \bar{x} - T(x)\frac{s}{\sqrt{n}} = 10 - \overset{-2}{\cancel{(-4)}}\frac{\cancel{5}}{\underset{2}{\cancel{\sqrt{100}}}} = 12$$

3

- Censored data: some people may be still alive. data could be right censored. Lived at least an n number of years).
- Non normal distribution for lifetimes.

4

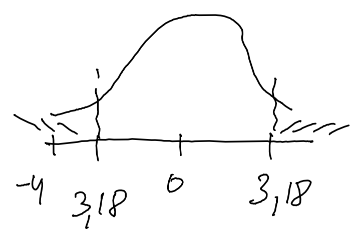
a) t - statistic

b)

$$T(x) = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{36 - 48}{6/\sqrt{4}} = \frac{-12}{3} = -4$$

$$S_2^2 = \frac{(29-36)^2 + 2(41-36)^2 + (33-36)^2}{3} = \frac{49 + 2\cdot 25 + 9}{3} = 36$$

c) Assume $\alpha = 0,05$



$$|t| > t_{\alpha/2, f}$$

$$|-4| > 3,18 \longrightarrow$$ reject H_0

5

H_0: E (yeald of Variety A) = E[B]

H_1: E[A] $\neq$ E[B]

Assume

$\alpha = 0{,}05$

$$\bar{x}_A = \frac{86+69+}{9} \approx 76{,}3$$

$$\bar{x}_B = \approx 69.3$$

$2{,}3 < 3{,}5$

$$S_A^2 = 14{,}19$$

$$S_B^2 = 12$$

$$T(X-Y) = \frac{\bar{x}_A - \bar{x}_B - (\mu_1 - \mu_2)}{\sqrt{\frac{S_A^2}{n} + \frac{S_B^2}{n}}} = \frac{76{,}3 - 69{,}3}{\sqrt{\frac{14{,}19}{9} + \frac{12}{9}}} \approx$$

$$\approx \frac{7}{\sqrt{2{,}99}} \approx 4{,}04$$

$$f = \left[\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2/n_2)^2}{n_2 - 1}} \right] = \frac{(2{,}99)^2}{\frac{4{,}52}{8}} = \frac{9}{0{,}56} \approx 18$$

$t_{\alpha/2, f} = 2{,}10$

(Another solution)

X - variety A

Y - variety B

$Z = X - Y$

$$\bar{x} = \frac{\overbrace{(86-71)}^{=15} + \overbrace{(69-72)}^{=-3} + \overbrace{(96-84)}^{=12} + \dots}{9} = 7$$

$S = 6$

$H_0: EZ = 0$

$H_1: EZ \neq 0$ $\qquad 1 - \alpha = 0{,}95$

$\alpha = 0{,}05$

$$T(Z) = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{7 - 0}{\cancel{6}^{2}/\cancel{3}} = 3{,}5$$

$t_{\frac{\alpha}{2}}(f) = 2{,}31$

$|T(Z)| > t_{\frac{\alpha}{2}}(f) \longrightarrow$ reject H_0

$\boxed{6}$ Confounder variable - an external factor that affects both the independent variable (was vodka drank) and dependent variable (lap time and number of errors) leading to potentially biased results. (Illusion of causation)

Confounder variable - drivers become more familiar with the track after the first lap.