STAT 500 Homework 2

Vu Thi-Hong-Ha; NetID: 851924086

September 4, 2020

1 Question 1

(a)
(i) $P(Z \leq 0.66) = 0.745$
(ii)
$P(Z > 0.66) = 1 - P(Z \leq 0.66) = 0.255 = P(Z < -0.66)$ because standard normal distribution is symmetric.
$P(|Z| \leq 0.66) = P(-0.66 \leq Z \leq 0.66) = P(Z \leq 0.66) - P(Z < -0.66) = 0.745 - 0.255 = 0.49$
(iii) $P(Z = 0.66) = 0$
(b)
(i) $a = -0.84162$
(ii) $b = 1.28155$

2 Question 2

$Z \sim N(112, 81)$.
(a) $P(Z < 103) = 0.15866$. Then, 15.866% of the children in the population have hemoglobin level below 103 g/L.
(b) $P(100 < Z < 124) = 1 - P(Z < 100) - P(Z > 124) = 1 - 0.44111 - 0.44111 = 0.11778$
(c) Median = 112.
(d) $P(Z < q) = 0.2$, where q is the 20th percentile. Then $q = 43.82868$.
(e) $P(a < Z < b) = 0.8$. The interval (a, b) is the shortest when it is symmetric over the mean of the distribution. Then, $a = -b$ and $P(Z > b) = 0.1$, so $b = 215.80568$ and $a = -215.80568$.

3 Question 3

Hypothesis:
H_0: $\mu_1 = \mu_2$
H_A: $\mu_1 \neq \mu_2$

$Y_1 \sim N(\mu_1, \sigma_1^2)$
$Y_2 \sim N(\mu_2, \sigma_2^2)$
Y_1 and Y_2 are independent.
(a) Results:
$$\sum_{j=1}^{n_1} Y_{1j} \sim N(n_1\mu_1, n_1\sigma_1^2)$$
$$\sum_{j=1}^{n_2} Y_{2j} \sim N(n_2\mu_2, n_2\sigma_2^2)$$
$\bar{Y_1} \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$
$\bar{Y_2} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$
Hence, $\bar{Y_1} - \bar{Y_2} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$.
(b)
A test statistic for this hypothesis test is $T.S. = \dfrac{\bar{Y_1} - \bar{Y_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$.

(c)

The test statistic mentioned in (b): $T.S. = \dfrac{\bar{Y_1} - \bar{Y_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$ under the null hypothesis, because T.S. is a linear combination of normal distribution and $\bar{Y_1} - \bar{Y_2} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$.
With H_A: $\mu_1 \neq \mu_2$, $p - value = 2 \times P(|T.S.| > z)$, where z is the z-score based on the desired significance level.

4 Question 4

(a) The study is an experiment. It is an investigation where the subjects are randomly assigned to receive different treatments.
- Experimental units: hypertensive subjects.
- Response variable: systolic blood pressure level.
- Replicate: yes, 30 subjects per condition.
- Randomization: yes.
- Treatments: gold-standard drug and new drug.

5 Question 5

T-test for Difference in Blood Pressure Levels

The TTEST Procedure

Variable: y

trt	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1		15	130.5	11.9335	3.0812	110.0	153.0
2		15	139.3	12.7429	3.2902	119.0	167.0
Diff (1-2)	Pooled		-8.8667	12.3448	4.5077		
Diff (1-2)	Satterthwaite		-8.8667		4.5077		

trt	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		130.5	123.9	137.1	11.9335	8.7369	18.8204
2		139.3	132.3	146.4	12.7429	9.3294	20.0968
Diff (1-2)	Pooled	-8.8667	-18.1003	0.3669	12.3448	9.7966	16.6958
Diff (1-2)	Satterthwaite	-8.8667	-18.1021	0.3687			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	28	-1.97	0.0592
Satterthwaite	Unequal	27.88	-1.97	0.0592

(a)
$t = -1.97$; $d.f = 28$; two-sided p-value $= 0.0592$; one-sided p-value $= 0.0296$. At 5% significance level, we reject the null that the mean systolic blood pressure levels are the same for the two drugs, and conclude that the new drug treatment produces a smaller mean systolic blood pressure level in hypertensive patients than the gold standard treatment.
(b)
Let Y_1 be the group using the new drug, and Y_2 be the group using the gold standard drug. We have:

$\bar{Y_1} = 130.5$, $\bar{Y_2} = 139.3$; $\bar{Y_1} - \bar{Y_2} = -8.8667$
$n_1 = 15; n_2 = 15; d.f = 28$
$\alpha = 0.05$
$t_{n-1+n_2-2,1-\frac{\alpha}{2}} = t_{28,0.95} = 1.70113$
$S_p = 12.3448$

Then, the 95% confidence interval is $(\infty; (\bar{Y_1} - \bar{Y_2}) + t_{n-1+n_2-2,1-\alpha} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$: $(\infty; -8.245535)$.
The null mean 0 is not included in the 95% CI, so we reject the null that the mean systolic blood pressure levels are the same for the two drugs at the significance level of 5%. If we sample for a large number of times, $100(1-\alpha)\%$ of such intervals will contain the true value of the difference between two treatment groups.

6 Question 6

(a)

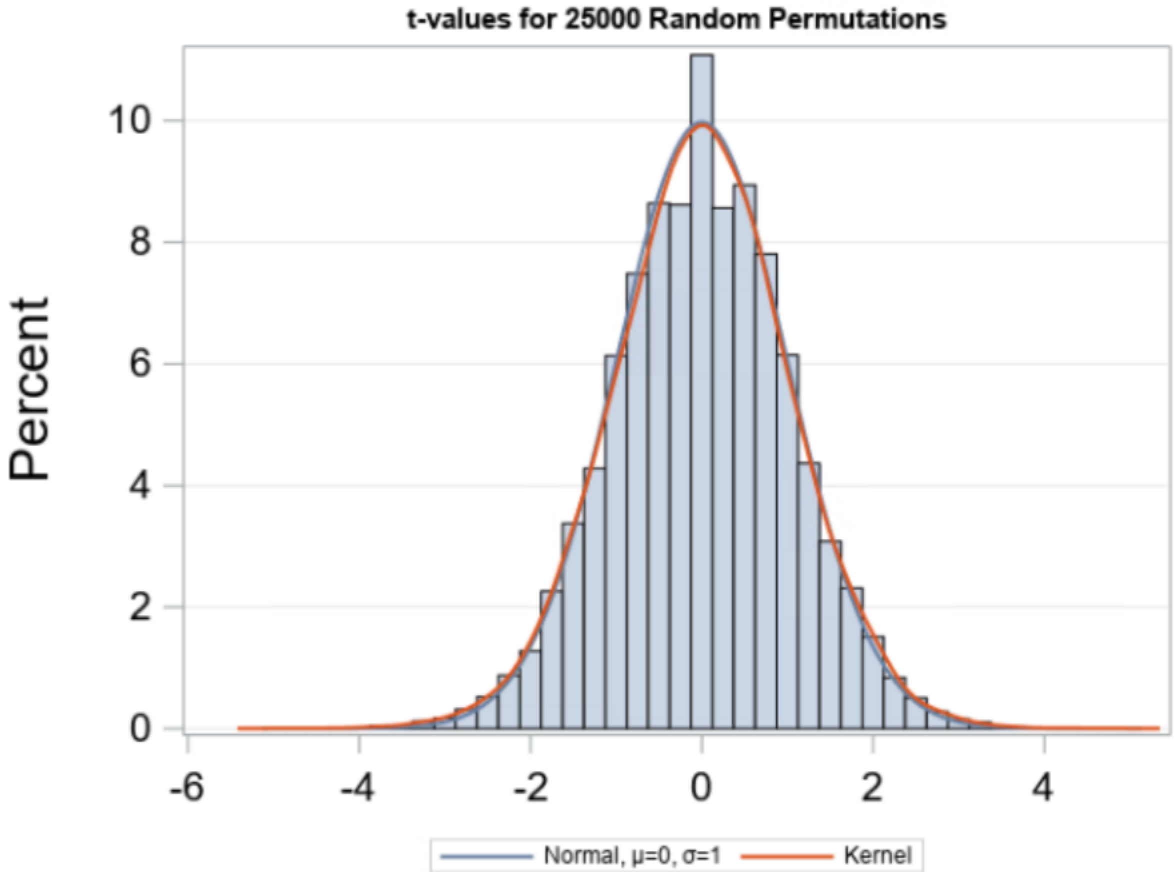
sample mean difference for 25000 Random Permutations

The FREQ Procedure

extreme_t	Frequency	Percent	Cumulative Frequency	Cumulative Percent
no	24250	97.00	24250	97.00
yes	750	3.00	25000	100.00

extreme_diff	Frequency	Percent	Cumulative Frequency	Cumulative Percent
no	24305	97.22	24305	97.22
yes	695	2.78	25000	100.00

The approximate p-value of t statistics by permutation test is 0.03. At the significance level of 5%, we reject the null that the mean systolic blood pressure levels are the same for the two drugs. This conclusion is consistent with that of Question 5a.



The histogram for the values of the t-statistics appear to be well approximated by a t-distribution with a bell shape.
(b)
Approximate p-value of difference statistics by permutation test is 0.0278. At the significance level of 5%, we reject the null that the mean systolic blood pressure levels are the same for the two drugs. This conclusion is consistent with that of Question 5a and 6a.