AP Statistics
Inference for Comparisons
Take Home Test Answer Sheet

Name: Ima Knowitall
Date: ____________
Period: ________

Show all this work on this answer sheet! Do not attach additional pages.

Question 1

a. Let p_1 = proportion of dementia cases with Prempro.
Let p_2 = proportion of dementia cases with placebo

$H_0: p_1 = p_2$
$H_a: p_1 < p_2$
$\alpha = .05$

- Patients were randomly to treatment groups.
- Independence is a safe assumption that more than 45320 women, age 65 and older, were available for this study. Randomization assures independence between groups.
- $n_1 \hat{p}_c = n_2 \hat{p}_c = 30.5$
  $n_1(1-\hat{p}_c) = n_2(1-\hat{p}_c) = 2235.5$
  We have at least 10 women in each group who would develop dementia and at least 10 who would not.
  $\therefore (\hat{p}_1 - \hat{p}_2)$ follows a Normal model.

$n_1 = n_2 = 2266$
$x_1 = 40;\ x_2 = 21$
$\hat{p}_c = \frac{40+21}{4532} = .0135$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = 2.449\ ;\quad p\text{-value} = P(z < 2.449) = .9928$$

Because p-value > α, we fail to reject Ho. We have sufficient evidence that Prempro is not effective for reducing cases of dementia.

E/I

b. $(.00168, .01509)$

E/P/I

c. If we were to perform this experiment in the same way many times, 95% of the sample differences would create intervals that capture the true difference between the proportion of dementia cases of patients on Prempro and those on a placebo.

Question 2

a. Let μ_1 = avg. stopping distance for site 1
Let μ_2 = avg. stopping distance for site 2

$H_0: \mu_1 = \mu_2$
$H_a: \mu_1 \neq \mu_2$
$\alpha = .05$

$\bar{x}_1 = 157.77$
$s_1 = 1.68$
$n_1 = 10$

$\bar{x}_2 = 164.14$
$s_2 = 2.77$
$n_2 = 12$

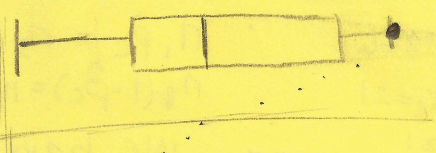
$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \ df = 18.441$$

$t = -6.636$

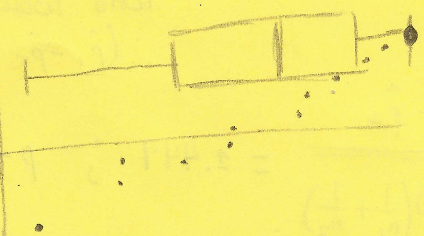
$p\text{-value} = P(|t| > 6.636)$
$= .00000278$

Because p-value < α, we reject H_0. We have sufficient evidence to conclude that the braking distances are different.

- No info provided regarding randomization (Proceed with caution)
- Braking distances at each site are independent because different drivers were used.
- It is reasonable to assume that more than 100 and 120 drivers were available for Sites 1 and 2 respectively.
- Sample 1:



Sample 2:



Both samples show approx. linear normal prob. plots. Both have one outlier at the highest value. Sample 1 shows right skewness. Sample 2 shows slight left skewness. Normality assumption is safe for each population, and thus $(\bar{x}_1 - \bar{x}_2)$ follows a Normal model.

b. $$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \ df = 18.441$$

$(-8.382, -4.356)$

E|P|I

If the braking distances were the same (i.e. no difference), we would expect to see 0 as a possible difference contained in the C.I. However, all values in the interval are less than 0, which indicates not only that the braking distances are different, but that site 1 has an avg. braking distance that is shorter than that of site 2.

Question 3

a.

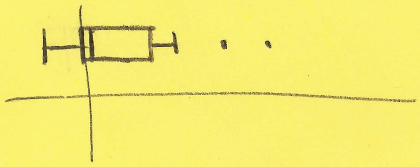
→ There is no mention of randomization.

→ Differences from 1968 to 1972 should be independent as information was gathered from different cities.

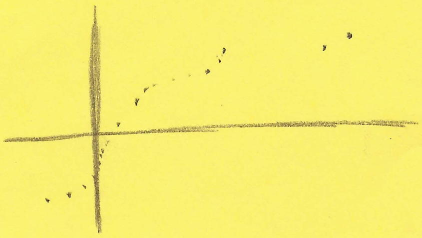
→ Normality of causes concern

→ Population of differences is not known to be Normal

→ $n < 30$; CLT will not assure Normality



Boxplot is skewed right with 2 outliers



Normal prob. plot has a cury, non-linear pattern.

The t-test is not appropriate.

b.

Let μ_D = avg difference in percentage of women in the labor force (1972−1968)

$H_0: \mu_D = 0$

$H_a: \mu_D \neq 0$ $\alpha = .05$

$\bar{x}_D = 3.368$

$s_D = 5.974$

$n = 19$

$$t = \frac{\bar{x}_D}{s_D/\sqrt{n}} = 2.458$$

df = 18

p-value = $P(|t| > 2.458) = .0246$

Because p-value < α, we reject H_0. We have sufficient evidence that the average percentage of women in the labor force changed between 1968 and 1972.