STA258: Statistics with Applied Probability
Winter 2025

Review for Test 2

---

1. A botanist is interested in the heights of the red oak trees in a certain park in Ontario. Using sophisticated equipment she is able to accurately measure the heights of 125 trees and she calculates their sample mean to be 95.72 feet with a standard deviation of 15.1 feet.

   A colleague informed them the mean height of red oak was 90 feet when measurements were taken 10 years ago. The researcher was wondering whether the mean height of the red oak trees in the park has increased.

   (a) Conduct a hypothesis test to test the researcher's belief at the 1% level of significance.

   (b) In this test, did we assume the standard deviation of the population was known or unknown? How would the test differ if we had assumed the other (please only describe the changes)?

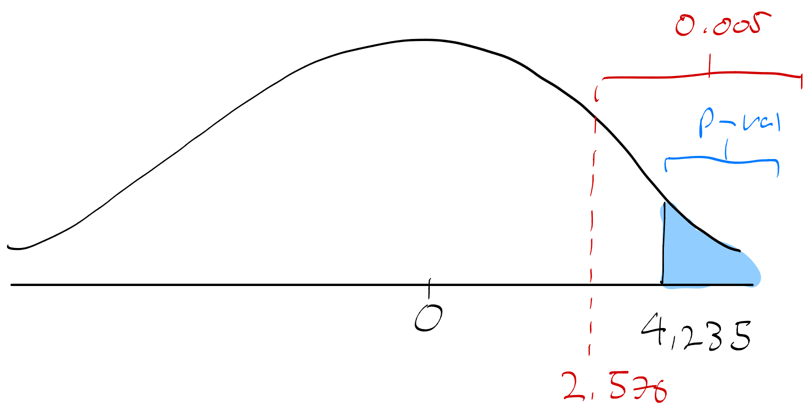
$n = 125 \quad \bar{x} = 95.72 \quad s = 15.1 \quad \alpha = 0.01$

(a) $H_0: \mu = 90 \qquad H_a: \mu > 90$

$$t^* = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{95.72 - 90}{15.1/\sqrt{125}} = 4.235$$

reference distribution: t-distribution at $n-1 = 124$ df

(not in t-table, use 100 df)



p-value $< 0.005 < 0.01$ (α)

(Strong)
Sufficient evidence against the null hypothesis at the 1% level of significance. Results from hypothesis test suggest the mean height of the trees is greater than 90 ft (trees have grown)

(b) In this test we assumed population standard deviation σ was not known (we used sample st. dev. s instead)

If we knew the population standard deviation we would calculate the test stat as

$$z^* = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

and reference distribution would be the standard normal one city in N. Ontario, the other in S. Ontario

2 cities in

2. A study was conducted on Seasonal affective disorder (SAD) in ~~Northern~~ Ontario which has very little sunshine over the winter. 10 people who are known to have experienced SAD (5 from each city) were randomly selected and the number of days they experienced SAD symptoms over the winter was recorded. The results are given in the table below:

Not paired.

2-sample inference

$n_1 = 5$

$n_2 = 5$

	Person in Northern city (1)	Person in Southern city (2)	Difference (North - South)
	80	74	6
	78	63	15
	84	67	17
	68	61	7
	75	65	10
Mean	77.0	66.0	11.00
St. Dev	6.0	5.0	4.85

Construct a 95% confidence interval for the difference in the mean number of days that people in the two cities experience SAD symptoms over the winter. Interpret the confidence interval and make an appropriate conclusion in the context of the study. You may assume that the population standard deviations are equal if necessary.

Not paired t-test. Use columns for North and South

since we assume equal population st devs ($\sigma_1 = \sigma_2$)

$$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(5-1)\cdot 6^2 + (5-1)\cdot 5^2}{5+5-2} = 30.5$$

$$S_p = 5.522$$

$$95\% \text{ CI} \longrightarrow 1-\alpha = 0.95 \qquad df = n_1+n_2-2$$

$$\alpha/2 = 0.025 \qquad = 8$$

$$t_{(n_1+n_2-2,\ \alpha/2)} = t_{(8,\ 0.025)} = 2.306$$

$$(\bar{X}_1 - \bar{X}_2) \pm t_{(n_1+n_2-2)} \cdot S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= (77-66) \pm (2.306)\cdot(5.522)\cdot\sqrt{\frac{1}{5}+\frac{1}{5}}$$

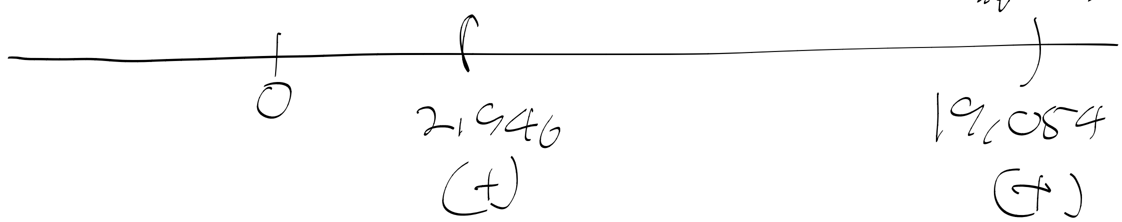
$$= (2.946,\ 19.054)$$

Interp:

We are 95% confident the difference in the mean number of days people diagnosed with SAD in the city in Norther Ontario and Southern Ontario which experience the disorder is between 2.946 days and 19.054 days

$\mu_N - \mu_S > 0$

$\mu_N > \mu_S$



The CI suggests the avg # days people experience SAD is larger in the city in Northern Ontario

3. A study was conducted to examine the effect of balance on the ability to complete tasks. A group of 7 people were randomly sampled and presented a puzzle to complete while stationary under lab conditions and the time taken to complete it was recorded. The same group was then asked to complete a different puzzle of similar difficulty while moving on a boat at moderate speed and the time taken to complete it was recorded. The results are given in the table below:

paired

μ_d

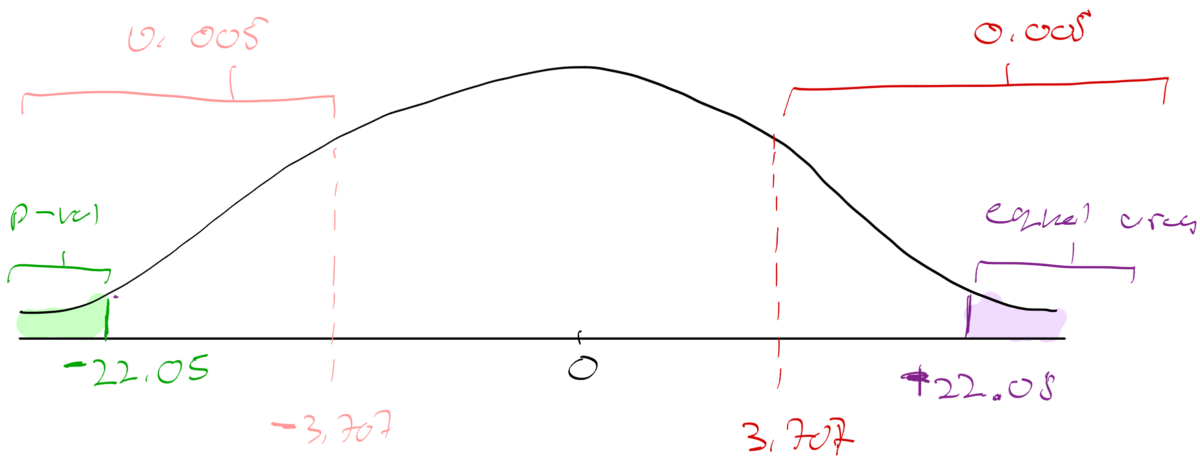
Person	Time taken on land	Time taken while in motion	Difference (stationary - Moving)
1	60	82	-22
2	55	73	-18
3	62	80	-18
4	48	68	-20
5	54	77	-23
6	51	68	-17
7	57	79	-22
Mean	~~55.3~~	~~75.3~~	-20.0
St. Dev	~~4.9~~	~~5.7~~	2.4

Conduct a hypothesis test to determine whether people are slower at tasks while in motion. Make a conclusion at the 5% level of significance. ~~You may assume equal standard deviations if necessary.~~

$H_0: \mu_d = 0$ $H_a: \mu_d < 0$

$$t^* = \frac{\bar{x}_d - \mu_0}{s_d/\sqrt{n}} = \frac{-20 - 0}{2.4/\sqrt{7}} = -22.05$$

reference distribution: t-distribution at $n-1=6$ df



p-value $< 0.005 < 0.05$ (α)

(strong) Sufficient evidence to reject the null hypothesis at 1% level of significance. The result from this hyp test suggests people in motion are slower at completing tasks

Inference on proportions

4. A farmer realized his orchard was infested with a certain type of insect. His options are to either use pesticide or to destroy the orchard through a controlled burn. He decides he will use pesticide if less than 30% of trees are infested. He hires an arborist to provide advise. The arborist takes a random sample of 100 trees and finds that 34 of them are infested. Condict a hypothesis test to determine whether the farmer should use pesticide or burn the orchard. Make a conclusion at the 5% level of significance.

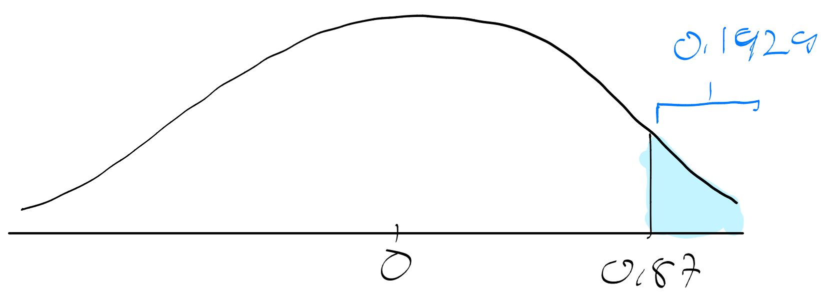
$n = 100 \qquad \hat{p} = \frac{34}{100} = 0.34$

$H_0: p = 0.30 \qquad H_a: p > 0.30$

makes most sense since sample prop is greater than hypothesized

$$z^* = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.34 - 0.30}{\sqrt{\frac{0.30(1-0.30)}{100}}} = 0.87$$

reference distribution: standard normal



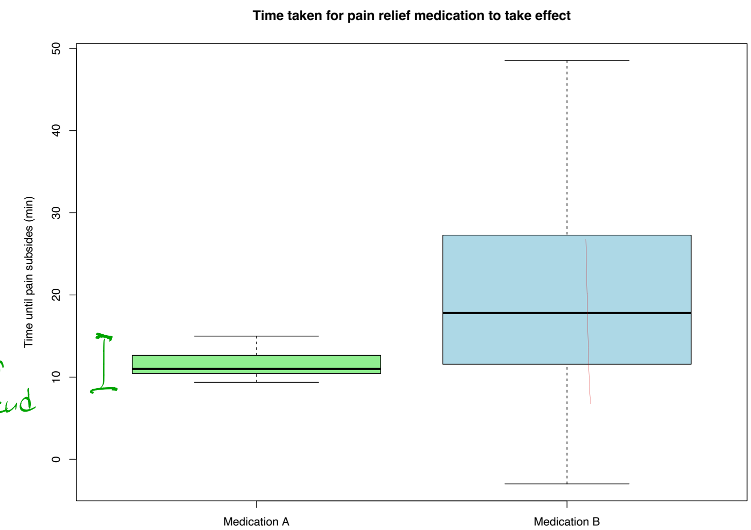
p-value = 0.1929 > 0.05 (α)

Insufficient evidence to reject the null at the 5% level of significance. The results from the test suggest the proportion of infested trees is 30% so the farmer should use pesticide

5. A test was conducted to determine the effectiveness of 2 types of pain relieving medication (A and B) on patients who experience headaches. The time taken for the headache to subside was recorded for the patients who participated in the study.

   Boxplots and R summary statistics for the study are given below:

Box-plots suggest unequal population variances



smaller spread

(much) larger spread

```
        Welch Two Sample t-test

data:  Med_A and Med_B
t = -2.2243, df = 17.974, p-value = 0.03918
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -14.4508655  -0.4115406
sample estimates:
mean of x mean of y
 11.57615  19.00735
```

assuming $\sigma_1^2 \neq \sigma_2^2$

$\mu_A - \mu_B < 0$

$\mu_A < \mu_B$

−14.45 (−) −0.41 (−) 0

```
        Two Sample t-test

data:  Med_A and Med_B
t = -1.5745, df = 25, p-value = 0.1279
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-17.151641   2.289235
sample estimates:
mean of x mean of y
11.57615  19.00735
```

Use the information given to determine which R output appears to be correct. Summarize the results of the correct R output. Make a conclusion at the 5% level of significance.

The R output using Welch's method appears correct since the box-plots suggest unequal variances due to box-plot for medication A having a much smaller spread than the boxplot for medication B.

$H_0: \mu_A - \mu_B = 0$ $H_a: \mu_A - \mu_B \neq 0$

test stat: $t = -2.2243$

reference dist: t-distribution at
17.974 df

p-value = 0.03918 < 0.05
(α)

Suff. evidence to reject H_0
that difference in mean time
of the medications is 0.

The test suggests the average time
taken by medication A is lower
(i.e. A works faster)