Data Analysis Worksheet #12: T-Test

Part 1: Hypotheses

T-tests are primarily for comparing a

⬜ numeric ⬜ categorical DV across ⬜two ⬜ three ⬜ four groups

To determine whether a store's average daily sales are above $10k or not, one needs to run a(n)

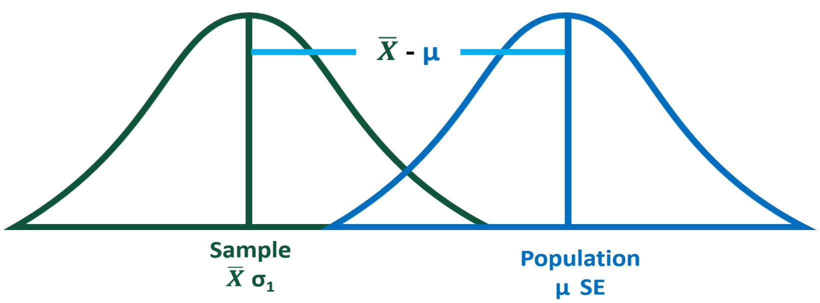
⬜ one sample ⬜ independent samples ⬜ paired samples T Test

To answer this question, we can rephrase it as a pair of hypotheses:

H0: average daily sales are no greater than $10k

H1: average daily sales are greater than $10k

Mark what you know in the given scenario on the diagram below. (Note that, obviously, not all values are given to you in this scenario, and so some will remain as variables.)



The t statistic should be computed as = $\frac{\bar{X}-\mu}{SE}$

To compare average daily sales between stores run by male managers versus stores run by female managers, one needs to run a(n)

⬜ one sample ⬜ independent samples ⬜ paired samples T Test

To answer this question, we can rephrase it as a pair of hypotheses:

H_0: average daily sales at stores run by male managers are no different from average daily sales at stores run by female managers

H_1: average daily sales at stores run by male managers are different from average daily sales at stores run by female managers

To compare stores daily sales in July versus August, one needs to run a(n)

⬜ one sample ⬜ independent samples ⬜ paired samples T Test

To answer this question, we can rephrase it as a pair of hypotheses:

H_0: average daily sales at stores in July are no different from average daily sales in August

H_1: average daily sales at stores in July are different from average daily sales in August

To determine if average property price in Williamsburg is over $1 million or not, one needs to run a(n)

⬜ one sample ⬜ independent samples ⬜ paired samples T Test

To answer this question, we can rephrase it as a pair of hypotheses:

H0: average property price in Williamsburg is no greater than $1 million

H1: average property price in Williamsburg is greater than $1 million

Part 2: Group Comparisons

To compare property prices between 23185 and 23188, one needs to run a(n)

⬜ one sample ⬜ independent samples ⬜ paired samples T Test

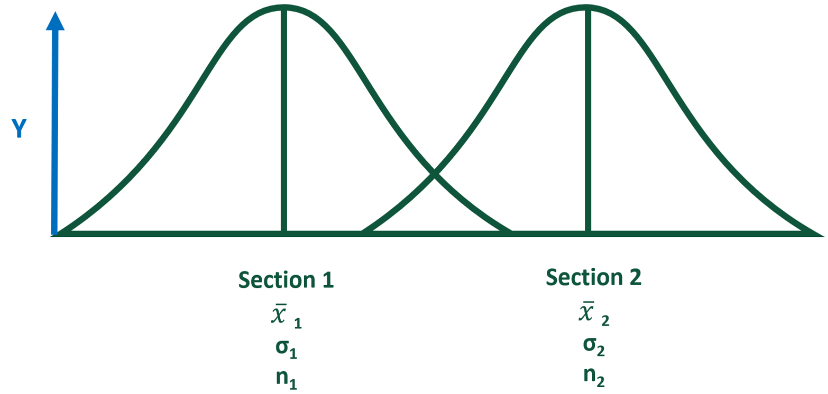
To answer this question, we can rephrase it as a pair of hypotheses:

H0: average property price of 23185 is no different from that of 23188

H1: average property price of 23185 is different from that of 23188

Answer the following questions using the appropriate T Test with the FROZEN class dataset.

Mark values from JASP results on the diagram below.



INDEPENDENT T-TEST

The parametric independent t-test, also known as Student's t-test, is used to determine if there is a statistical difference between the means of two independent groups. The test requires a continuous dependent variable (i.e. body mass) and an independent variable comprising 2 groups (i.e. males and females).

This test produces a t-score which is a ration of the differences between the two groups and the differences within the two groups:

$$t = \frac{\textit{mean group } 1 - \textit{mean group } 2}{\textit{standard error of the mean differences}}$$

$$t = \frac{(X_1 - X_2)}{\sqrt{\frac{(S_1)^2}{n_1} + \frac{(S_2)^2}{n_2}}}$$

X = mean

S = standard deviation

n = number of data points

A large t-score indicates that there is a greater difference between groups. The smaller the t-score, the more similarity there is between groups. A t-score of 5 means that the groups are five times as different from each other as they are within each other.

The null hypothesis (H_o) tested is that the population means from the two unrelated groups are equal

The T-Test statistic t is computed as = $\frac{\textit{Mean group } 1 - \textit{Mean group } 2}{\textit{Pooled Standard Error}}$(Enter values from JASP.)

Why do we need to use "pooled" standard error, and not just standard error, as we did previously?

Because in an **independent-samples t-test** we are comparing **two separate groups.** Each group has its own variance and sample size, so we cannot rely on a single group's standard error. Instead, we use the **pooled standard error**, which combines information from both groups' variances and sample sizes to create a fair estimate of the variability of the difference between means.

What does it mean if the T-Test statistic t were negative?

It would just indicate **direction**: a **negative t** means the **mean price in 23185 is lower than 23188**. (For a two-tailed test, the p-value is based on |t|, so the conclusion about "different" vs "not different" wouldn't change—only the direction does.)

Write up these results in APA style:

An independent-samples t-test compared average property prices in ZIP codes 23185 and 23188. Prices were **higher in 23185** (M = \$1,018,000, SD = \$678,555, *n* = 66) than in 23188 (M = \$662,373, SD = \$274,586, *n* = 28), **t**(92) = 2.67, **p** = .009.

`One can also run a(n)` ⬜ ANOVA ⬜ `Repeated Measures ANOVA`

Answer the following questions using the appropriate ANOVA test with the FROZEN class dataset.

ANOVA - Price

Cases	Sum of Squares	df	Mean Square	F	p
Zipcode	$2.482 \times 10^{+12}$	1	$2.482 \times 10^{+12}$	7.144	.009
Residuals	$3.196 \times 10^{+13}$	92	$3.474 \times 10^{+11}$		

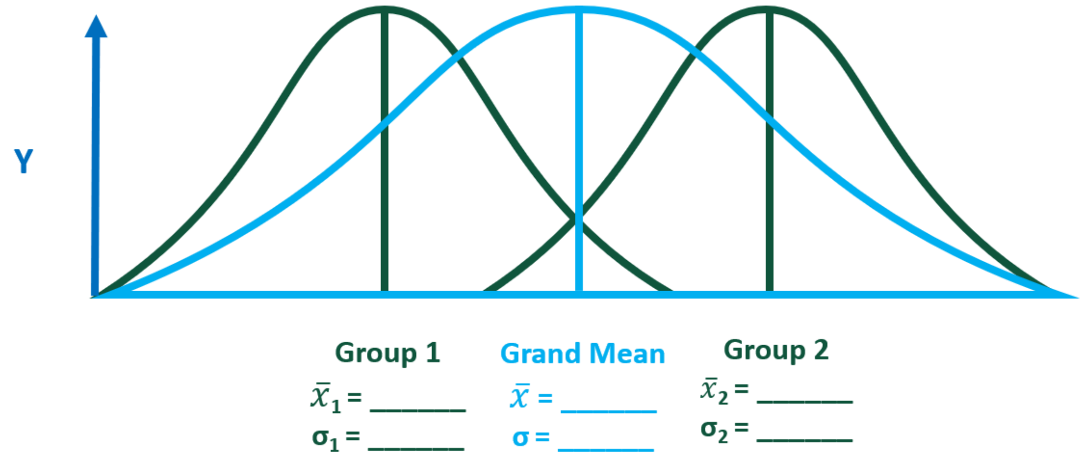
Note. Type III Sum of Squares

Descriptives

Descriptives - Price

Zipcode	N	Mean	SD	SE	Coefficient of variation
23,185	66	$1.018 \times 10^{+6}$	678,555	83,524	0.667
23,188	28	662,373	274,586	51,892	0.415

Mark values from JASP results on the diagram below.



Group1(Zip 23185): Mean:1,018,000 | SD: 678,555

Group2 (Zip 23188): Mean:662,373 | SD: 274,586

Grand Mean= (66*1018000+28*662373)/94=912,069

Grand SD: SQRT(Variance)=sqrt($3.4442*10^{13}$/93)=608,559

The F ratio for this ANOVA is computed as = $\frac{MSb}{MSw}$ (Enter values form JASP.) 7.144

Is it possible for the F ratio to be negative? ⬜ yes ⬜ no

Write up these results in APA style:

A one-way ANOVA compared mean property prices across ZIP codes 23185 and 23188. The results showed a significant difference in prices, $F(1, 92) = 7.144$, $p = .009$. On average, homes in 23185 (M = \$1,018,000, SD = \$678,555, n = 66) were priced higher than those in 23188 (M = \$662,373, SD = \$274,586, n = 28).

Part #3: Self Comparisons

To determine if Zestimate is a different Price, one needs to run a(n)

⬜ one sample ⬜ independent samples ⬜ paired samples T Test

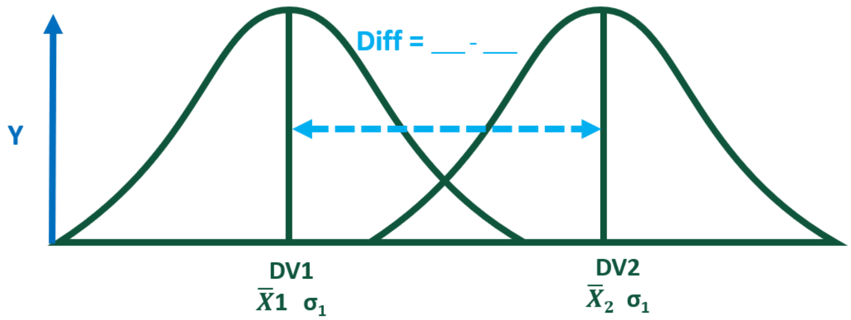
To answer this question, we can rephrase it as a pair of hypotheses:

H0: The average Zestimate is equal to the average actual Price.

H1: The average Zestimate is different from the average actual Price.

Answer the following questions using the appropriate T Test with the FROZEN class dataset.

Mark values from JASP results on the diagram below.



Paired Samples T-Test

Paired Samples T-Test

Measure 1		Measure 2	t	df	p
Zestimate	-	Price	−3.411	93	< .001

Note. Student's t-test.

Descriptives

Descriptives

	N	Mean	SD	SE	Coefficient of variation
Zestimate	94	897,629	587,391	60,585	0.654
Price	94	911,861	608,597	62,772	0.667

The T-Test statistic t is computed as = $\frac{mean\ difference}{SE\ of\ the\ differences}$(Enter values from JASP.)

-3.411

$$SD_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}}$$

= (897629-911861)/4172

=-14232/4172

=-3.411

What does it mean if the T-Test statistic t were negative?

We have defined the difference as Zestimate − Price, a negative t means Zestimate is lower than Price on average (i.e., the mean difference is negative).

For this question, one can also run a(n) ⬜ ANOVA ⬜ Repeated Measures ANOVA.

Write up these results in APA style:

A paired-samples t-test compared Zestimate and actual Price for the same homes. Zestimate (M = \$897,629, SD = \$587,391) was lower than Price (M = \$911,861, SD = \$608,597), $t(93) = -3.41$, $p < .001$. On average, Zestimate was \$14,232 below the actual sale price.