Analysing Datasets

**Analysis of Brandprefs.xlsx**

Step 1. Hypotheses:

* H₀: Brand preference is independent of area (no difference between areas).
* H₁: Brand preference depends on area (distribution differs between areas).

Step 2. Data summary (contingency table):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of Brand** | **Column Labels** |  |  |  |
| **Row Labels** | **A** | **B** | **Other** | **Grand Total** |
| 1 | 11 | 17 | 42 | 70 |
| 2 | 19 | 30 | 41 | 90 |
| **Grand Total** | **30** | **47** | **83** | **160** |

Step 3. Test used:

* To find the perform a chi square test, we use the formula:
  + =CHISQ.TEST(actual\_range, expected\_range)
* To find the expected range:
  + The expected range is the table of counts you would expect if brand preference and area were independent.
  + For example:
    - Expected count for Area 1, Brand A = (Row total for Area 1 × Column total for Brand A) ÷ Grand total = (70 × 30) ÷ 160 = 13.125

|  |  |  |  |
| --- | --- | --- | --- |
| **Expected range:** |  |  |  |
| **Area** | **Brand A** | **Brand B** | **Other** |
| 1 | 13,125 | 20,5625 | 36,3125 |
| 2 | 16,875 | 26,4375 | 46,6875 |

Step 4. Results:

* =CHISQ.TEST(actual\_range, expected\_range)
  + This fives us the result p = 0.193

Step 5. Decision:

* At α = 0.05, p > 0.05, we fail to reject the null hypothesis (H₀)
* Interpretation:
  + There is no significant evidence that brand preferences differ between Area 1 and Area 2.
  + In both areas, Other brands are most popular.
    - Brand B is more popular than A in both areas.
    - While the percentages shift slightly (Area 2 shows higher proportions for A and B), the differences are not statistically significant.

Conclusion: Brand preferences are broadly similar across the two demographic areas.

**Analysis of Designs.xlsx**

Designs.xlsx (Data Set F) — Paired t-test of container designs

Step 1. Hypotheses (paired t-test):

* Two-tailed test (general difference):
  + H₀: μ₁ = μ₂
  + H₁: μ₁ ≠ μ₂

* One-tailed test (expect Con1 > Con2):
  + H₀: μ₁ ≤ μ₂
  + H₁: μ₁ > μ₂

Step 2. Descriptive statistics:

* Mean Con1 = 172.6
* Mean Con2 = 159.4
* Difference = 13.2 (Con1 higher)

Step 3. Paired samples t-test:

* Test used: Paired samples t-test (same stores, two designs).
  + To perform tests, use:
    - =T.TEST(range\_con1, range\_con2,2,1) # two-tailed paired
    - =T.TEST(range\_con1, range\_con2,1,1) # one-tailed paired

Step 4. Results:

* t statistic = 2.875 (df = 9)
* p (two-tailed) = 0.018
* p (one-tailed, Con1 > Con2) = 0.009

Step 4. Decision:

* Two-tailed: reject H₀ at 5% level.
* One-tailed: reject H₀ at 1% level → strong evidence Con1 > Con2.

Interpretation:

There is significant evidence that container design affects sales. Design 1 (Con1) sold more items on average (≈13 units higher per store). Therefore, Design 1 should be preferred.

**Analysis of Diets.xlsx**

Diets.xlsx (Data Set B) — Independent samples t-test of weight loss

Hypotheses (two-tailed test):

* H₀: μ₁ = μ₂ (mean weight loss is the same for Diet A and Diet B)
* H₁: μ₁ ≠ μ₂ (mean weight loss differs between Diet A and Diet B)

Descriptive statistics (from Excel):

* Diet A:  
  • n = 50  
  • Mean = 5.34 kg  
  • Standard deviation = 2.54 kg
* Diet B:  
  • n = 50  
  • Mean = 3.71 kg  
  • Standard deviation = 2.77 kg

Step 1. Test for equal variances (F-test):

* F = 0.839
* p (two-tailed) = 0.540 (> 0.05)
* → Variances are not significantly different → assume equal variances.

Step 2. Independent samples t-test (equal variances):

* t = 3.072
* df = 98
* p (two-tailed) = 0.0028
* p (one-tailed) = 0.0014

Decision:

* At α = 0.05: reject H₀.
* At α = 0.01: also reject H₀.

Interpretation:

There is strong evidence that Diet A produces greater weight loss than Diet B. On average, Diet A participants lost about 1.63 kg more than those on Diet B.

**Analysis of Heather.xlsx**

Step 1. Observed frequencies (your actual table)

From your sheet:

|  |  |  |
| --- | --- | --- |
| Prevalence | Location A | Location B |
| Absent | 8 | 20 |
| Sparse | 22 | 14 |
| Abundant | 26 | 10 |

Put this into cells B3:C5.

Step 2. Row and column totals

Calculate the row total, column total, and grand total in the actual range table.

Step 3. Expected frequencies

Find the expected frequencies (range) in the following way:

* (Row total for Absent × Column total for Location A) ÷ Grand total = 15.68
  + Do the same for the other rows/columns

|  |  |  |
| --- | --- | --- |
| Expected range |  |  |
|  | Location A | Location B |
| Absent | 15,68 | 12,32 |
| Sparse | 20,16 | 15,84 |
| Abundant | 20,16 | 15,84 |

Step 4. Results:

* To find the perform a chi square test, we use the formula:
  + =CHISQ.TEST(actual\_range, expected\_range)
    - This gives us the result: p-value = 0,00168

Step 5. Interpretation:

* This is the p-value of the chi-square test.
* Since p ≈ 0.0017 < 0.05, you reject H₀.
* That means: prevalence of heather is significantly associated with location.

Step 6. Conclusion

Since p < 0.05 → we reject the null hypothesis (H₀).

The prevalence of heather differs significantly between Location A and Location B.

**Analysis of Superplus.xlsx**

Superplus.xlsx (Data Set C) — Independent samples t-test of male vs female incomes

Hypotheses (one-tailed test):

* H₀: μₘ ≤ μᶠ (male mean income ≤ female mean income)
* H₁: μₘ > μᶠ (male mean income > female mean income)

Descriptive statistics (Excel):

* Males  
  • n = 60  
  • Mean = 52.91  
  • Std dev = 15.27

* Females  
  • n = 60  
  • Mean = 44.23  
  • Std dev = 13.79
* Difference in means = 8.68

Step 1. F-test for equality of variances:

* F = 1.226
* p (two-tailed) = 0.436 (> 0.05)
* → Equal variances assumed

Step 2. Independent samples t-test (equal variances):

* t = 3.268
* df = 118 (n₁ + n₂ − 2)
* p (two-tailed) = 0.0014
* p (one-tailed) = 0.00071

Step 3: Interpretation:

* At α = 0.05 (and even at α = 0.01), we reject the null hypothesis (H₀)
* Conclusion:
  + There is strong evidence that the population mean income of male cardholders is greater than that of female cardholders, with a sample difference of about £8,680 (since the values are in £’000).