**UnitedHealth Group Claims Analysis**

The company wants insight of the claims data containing the claims paid out for medical malpractice lawsuits this year. This study finds the differences in claims amounts between several groups under various conditions. It compares proportions between groups and runs an experiment and analysis on experiment data.

This report is divided into sections to answer each question. It provides assumptions, hypotheses, test results, test analyses and conclusions.

Assumptions (all tests):

* Simple random sampling
* Independent samples
* Normally distributed populations
* 95% confidence level is appropriate for confidence intervals
* 5.0% significance level

Assumptions (specific tests):

* z-test: 10 minimum counts of success, 10 minimum counts of failure
* t-test, Anova: Equal population variances
* Anova: Dependent variable should be continuous
* chi-square: Independent groups
* chi-square and two-way Anova: Equal sample size (min. 5) for each group

**Problems, Tests and Analysis**

**1. A. The proportion of mild or medium claims by patient gender**

**Hypothesis**

|  |  |
| --- | --- |
| H0: π1 - π2 = 0% | 𝑤ℎ𝑒re π1 and π2 are proportions of mild or medium by female and male, respectively |
| H1: π1 - π2 ≠ 0% | π1, π1 as above |

Test: z-test – see [Appendix 1.A.1](#Appendix1a1)

**Analysis**

102 of 130 (**78.46%**) females and 46 of 70 (**65.71%**) males have mild or medium claims.

z-critical values = -1.96 and +1.96 (two-tail)

z statistic 1.9603 > z-critical 1.96 and p-value 0.05 <= α 0.05 --> reject H0

Confidence Interval test to confirm the differences – see [Appendix 1.A.2](#Appendix1a2)

The Confidence interval are between 0% and 25.49%. 0 in the CI can be interpreted as negative or positive difference. But with p-value equal to 0.05 and other value is above 0, I can conclude that 0 is a real population different.

**Conclusion**

At 5% significance level, there is sufficient evidence to conclude there is difference in the proportion of mild or medium type severity claims by female patients compared to that of male patients

At 95% confidence, the true difference in proportion of Female and Male genders who claims mild or medium type of severity is from 0% (no difference) up to 25.49% difference

**1. B. Differences in the average claims amount when a private attorney is involved and when there is no private attorney.**

**Hypothesis**

|  |  |
| --- | --- |
| H0: μ1 - μ2 <= 0 | 𝑤ℎ𝑒re μ1 = mean claim amount when a private attorney is involved, and μ2 when there is no private attorney |
| H1: μ1 - μ2 > 0 | μ1, μ2 as above |

Test: One-tailed hypothesis t-test (upper) - see [Appendix 1.B.1](#Appendix1b1)

**Analysis**

Note that there are outliers in claims when a private attorney is involved, see [Appendix 1.B.3](#Appendix1b3)

S2/S1 = 30761.71/30414.88 = 1.01 which is < 2 -> equal variance

t-sample 4.8124 > t-critical 1.6526 (upper tail) and p-value 0 < 0.05 -> reject H0

I estimate the size of the difference using confident interval – see [Appendix 1.B.2](#Appendix1b2)

**Conclusion**

At 5% significance level, there is sufficient evidence to conclude that the average claim amount when a private attorney is involved is higher than when there is no private attorney involved.

I am 95% confident that the average claim amount when a private attorney is involved

is between 13,197.61 to 31,423.51 greater than the average claim amount when no private attorney is involved.

**1. C. Compare private attorney representation in severe and medium severity claims**

**Hypothesis**

|  |  |
| --- | --- |
| H0: π1 - π2 <= 0% | 𝑤ℎ𝑒re π1 = private attorney representation in severe claim, π2 = Private attorney representation in medium claim |
| H1: π1 - π2 > 0% | π1, π1 as above |

Test: z-test (upper) – see [Appendix 1.C](#Appendix1c)

**Analysis**

42 out of 52 (80.77%) severe claims represented by private attorneys

88 out of 123 (71.54%) medium claims represented by private attorneys

Upper tail, z-value 1.2760 < z-critical value 1.6449 and p-value 0.1010 > alpha 0.05 -> fail to reject H0

**Conclusion**

At 5% significance level, I cannot conclude that private attorney representation is higher for severe than for medium claims

**2. A. The comparation of the percentages of severe claims by Orthopaedic surgeons and by other specialists**

**Hypothesis**

|  |  |
| --- | --- |
| H0: π1 - π2 >= 0% | 𝑤ℎ𝑒re π1 = Severe claims by Orthopaedic, π2 = Severe claim by other specialists |
| H1: π1 - π2 < 0% | 𝑤ℎ𝑒re π1, π1 as above |

Test: z-test - see [Appendix 2.A](#Appendix2a1)

**Analysis**

11 of 48 (22.92%) Severe claims by Orthopaedic.

41 of 152 (26.97%) Severe claims by other specialists.

z-value-0.5586 > z-critical -1.6449 (It is lower tail), p-value = 0.2882 > α = 0.05 therefore we failed to reject H0

**Conclusion**

At 5% significance level, I cannot conclude that the percentage of severe claims with the involvement of an Orthopaedic surgeon is lower than that of all other specialists.

**2 B. Compare the average claims amount in severe claims when Orthopaedic involved and when other specialisations involved**

**Hypothesis**

|  |  |
| --- | --- |
| H0: μ1 - μ2 <= 0 | 𝑤ℎ𝑒re μ1, μ2 = *orthopaedic* and *other specialists* severe mean claim amount respectively |
| H1: μ1 - μ2 > 0 | 𝑤ℎ𝑒re μ1, μ2 as above |

Test: pooled t test (upper) - see [Appendix 2.B.1](#Appendix2b)

**Analysis**

s1/s2 = 36942.01 / 32968.42 = 1.12 -> equal variant

For upper tail, t-sample statistic 0.3413 > critical value 1.6759

And p-value 0.3672 which is higher than alpha 0.05, we fail to reject H0

Note that we found outliers in our sample – see [Appendix 2.B.2](#Appendix2b2)

**Conclusion**

At 5% significant level, there is no sufficient evidence to conclude that the average claim amount for severe claims when Orthopaedic surgeons are involved is greater than all other specialisations

**3. A. The average claim amount across the claimant's marital status.**

**Hypothesis**

|  |  |
| --- | --- |
| H0 : 𝜇1 = 𝜇2 = 𝜇3 = 𝜇3 = 𝜇4 | 𝑤ℎ𝑒re 1=Divorced,2= Married,3=Single,4=Widowed |
| H1 : 𝑁ot 𝑎ll 𝜇𝑗 𝑎re equal | 𝑤here 𝑗=1,2,3,4 |

Test: One-Way Anova – see [Appendix 3.A.1](#Appendix3A1)

**Analysis**

From the test result F-test > F-critical (6.705450468 > 2.65067651) and p-value 0.000249 < α 0.05 -> reject the H0

The outliers are given in [Appendix 3.A.3](#Appendix3A3)

**Conclusion**

With 5% significance level, there is sufficient evidence to conclude the average claim amount significantly differs across the claimant's marital status.

To find the differences, see Post Hoc Tukey Kramer - [Appendix 3.A.2](#Appendix3A2)

The Tukey Kramer result shows:

1. There is significant difference in average claims amounts between Divorced and Married, between Divorced and Single, between Divorced and Widowed. between Married and Widowed, between Single and Widowed groups.
2. There is no significant difference in average claim amounts between Married and Single groups.

**3. B. The average claim amount across surgeon specialities.**

**Hypothesis**

|  |  |
| --- | --- |
| H0 : 𝜇1 = 𝜇2 = 𝜇3 = 𝜇4 | 𝑤ℎ𝑒re 1= Anesthesiologists,2= Dermatologists,3= Obstetricians,4= Gynaecologists |
| H1 : 𝑁ot 𝑎ll 𝜇𝑗 𝑎re equal | 𝑤here 𝑗=1,2,3,4 |

Test: One-way Anova – see [Appendix 3.B.1](#Appendix3B)

**Analysis**

Variance: s hightest²/s lowest² = 1,195,468,219.88 / 645,888,118.77 = 1.85 which is < 4 -> Equal variance

Note that outliers found in the sample – see [Appendix 3.B.2](#Appendix3B2)

From the Anova result, the p-value is 0.990611 > α =0.05 and F-test 0.04 < F-critical 2.65 therefore It fails to reject the H0

**Conclusion**

At 5% significance level, there is no sufficient evidence to conclude that the average claims amount significantly differs across surgeon specialities

**3. C. Compare the proportion of claimants represented by a private attorney across the claimant's marital status**

**Hypothesis**

|  |  |
| --- | --- |
| H0: π1 = π2 = π3 = π4 | 𝑤ℎ𝑒re 1= Divorced,2= Married,3= Single,4= Widowed |
| H1: 𝑁ot 𝑎ll π 𝑗 𝑎re equal | 𝑤here 𝑗=1,2,3,4 |

Test: Chi-square - see [Appendix 3.C](#Appendix3C)

**Analysis**

Chi square is positively skewed, therefore if Chi square statistic 1.31 < Critical value 7.8147 and p-value 0.7260 > 0.05 -> fail to reject H0

**Conclusion**

At 5% significance level, there is no sufficient evidence that the proportion of claimants represented by a private attorney differs significantly across the claimant's marital status.

**4. The effect of private attorney representation and insuranceon claims amounts**

**Hypothesis**

The effect of **Private Attorney** on the mean of claims amounts

|  |  |
| --- | --- |
| H0: 𝜇1 = 𝜇2 | 𝑤ℎ𝑒re 1= Private attorney, 2= No Private Attorney |
| H1: 𝜇1 ≠ 𝜇2 | Where 1, 2 as above |

The effect of **Insurance type** on the mean of claims amounts

|  |  |
| --- | --- |
| H0 : 𝜇1 = 𝜇2 = 𝜇3 = 𝜇4 | 𝑤ℎ𝑒re 1= Medicare/Medicaid, 2= No Insurance,3= Private, 4=Workers Compensation |
| H1 : 𝑁ot 𝑎ll 𝜇𝑗 𝑎re equal | 𝑤here 𝑗=1,2,3,4 |

|  |  |
| --- | --- |
|  |  |

Interaction **private attorney and insurance**

H0: There is no interaction between Private Attorney and Insurance type

H1: There is interaction between Private Attorney and Insurance type

Test: Two-Way Anova – see [Appendix 4.1](#Appendix41)

To better understanding the main effect, a Marginal Mean table is presented in [Appendix 4.2](#Appendix42)

**Analysis**

Main effect of a private attorney representation:

The marginal means of 302,637 (Private Attorney), 229,278.37 (No Private Attorney). Looking at those means; I found the claim amounts are different.

With p-value = 0.022768732 (< 0.05), there is **significant evidence** that there is effect of private attorney to the claim amount.

Main effect of insurances:

The marginal means of 129,164.91 (Medicare/Medicate), 125,738.57 (No Insurance), 142,571.22 (Private) and 134,441.46 (Worker Compensations). No significant differences in claim amounts.

With p = 0.875840001 (> 0.05), there is **no significant** **evidence** there is effect of insurance to the claim amount

Interaction effect:

P = 0.63141386, there is **no significant evidence** that the interaction between private attorney and insurance affect the claim amount

To understand the interaction effect better, a cell means plot is drawn and it be seen in [Appendix 4.3](#Appendix43)

Analysis from the plot and Anova as the following:

* Overall, claims with **private attorney** representation have **higher average amounts** than without ones.
* **Insurance** types have **no impact** on the claim amounts except for Medicare/Medicaid patients.
* For Medicare/Medicaid patients, there is a large different amount between with a private attorney and without a private attorney representation. Without a private attorney, Medicare/Medicaid patients claimed **the least** $46,811.98 however, with a private attorney representation, they have **the highest** claim amount $ 82,352.94.

**5. The work plan as a data analyst at UnitedHealth.**

1. Making a list of tasks and deadlines.
2. Prioritising tasks based on deadlines and urgency.
3. Breaking down big tasks into smaller tasks.
4. Maintaining communication throughout the process

**Conclusion**

In the report, I compare claims amount between 2 groups and across different groups. I also compare proportions in 2 or more groups.

Using post hoc test, I calculate the differences between groups. I analyse my test results and conclude whether the alternative hypothesis is valid.

During the investigation, I found outliers and unequal variances in the sample, I suggest that we look into the population data to be able to conclude what cause those unusual data.

**Appendices**

**Appendix 1.A.1: Hypothesis test result for mild or medium by patient gender**



[Go back](#Appendix1a1_back)

**Appendix 1.A.2: Confidence Interval test result for mild or medium by patient gender**



[Go back](#Appendix1a2_back)

**Appendix 1.B.1: Pooled t-test – Claims amount means between private attorney representation and no private attorney**



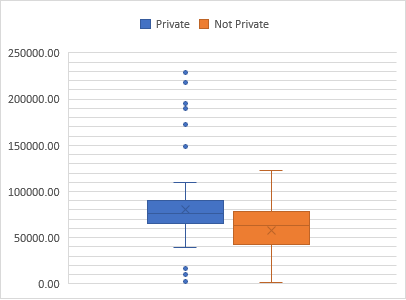
[Go back](#Appendix1b1_back)

**Appendix 1.B.2 – Confidence Interval - Claim amounts means between private attorney representation and no private attorney**



[Go back](#Appendix1b2_back)

**Appendix 1.B.3 – Checking outliers**



[Go back](#Appendix1b3_back)

**Appendix 1.C: Private attorney representation in severe and medium severity.**



[Go back](#Appendix1c_back)

**Appendix 2.A: Hypothesis z-test of the severe claims between Orthopaedic surgeon and other specialists claims**



[Go back](#Appendix2a1_back)

**Appendix 2.B.1: t-test of the average claims amount for severe claims between Orthopaedic surgeon and all other specialisations**



[Go back](#Appendix2b1_back)

**Appendix 2.B.2: Outliers**



[Go back](#Appendix2b2_back)

**Appendix 3.A.1: Anova - mean claim amount across the claimant's marital status**



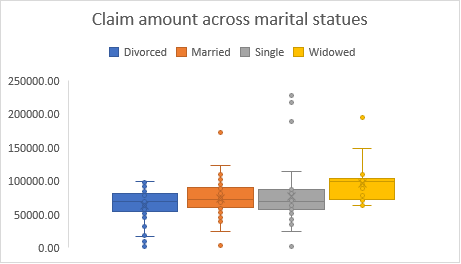
[Go back](#Appendix3A1_back)

**Appendix 3.A.2: Tukey Kramer**



[Go back](#Appendix3A1_back)

**Appendix 3.A.3: Outliers**



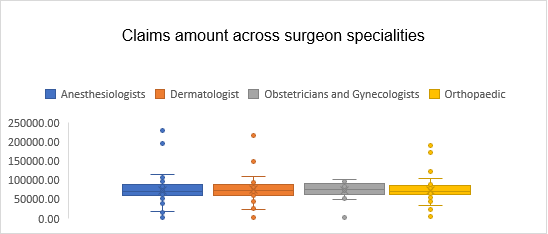
[Go back](#Appendix3A3_back)

**Appendix 3.B.1: Anova for the average claim amount across surgeon specialities.**



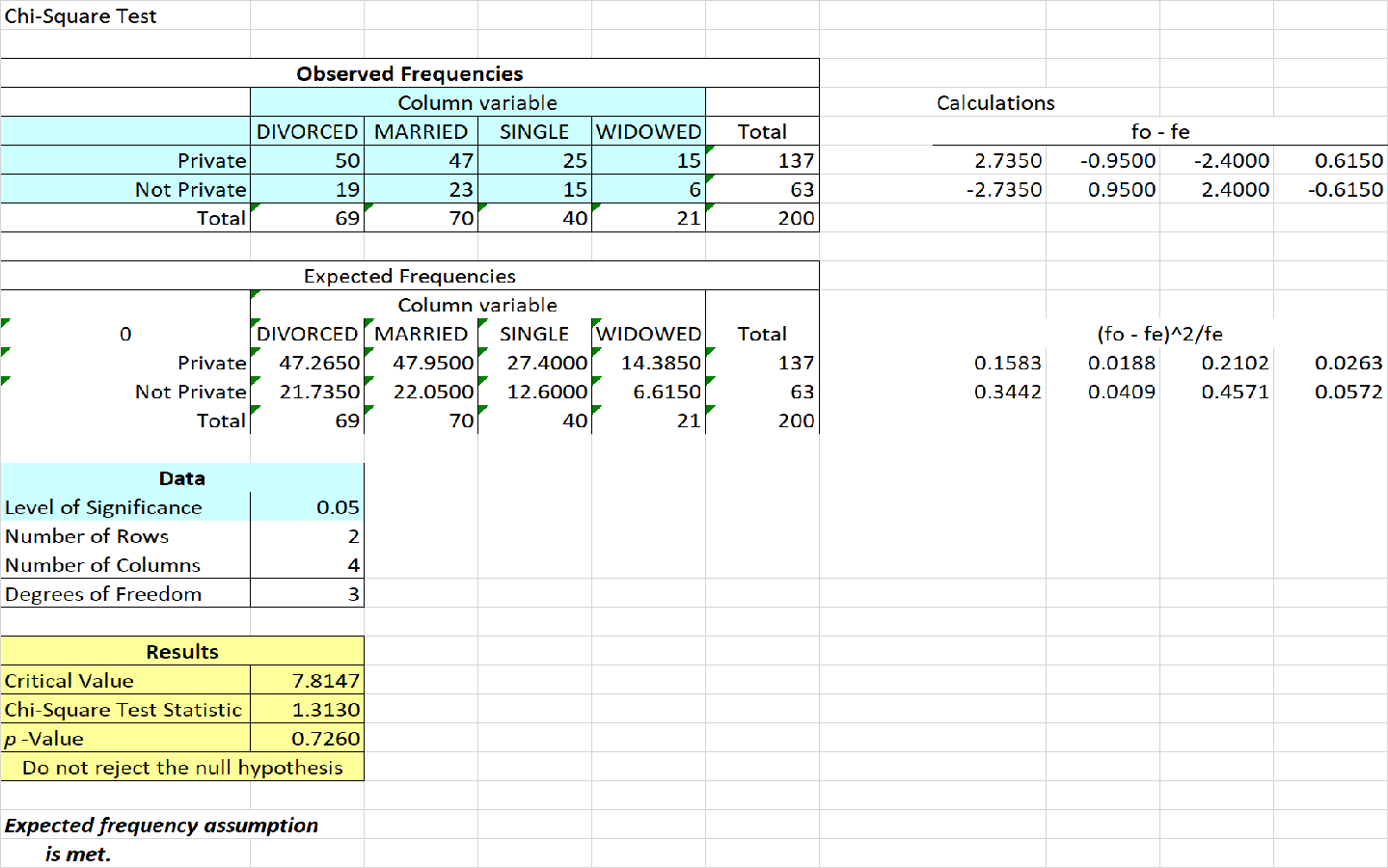
[Go back](#Appendix3B1_back)

**Appendix 3.B.2: Outliers**



[Go back](#Appendix3B2_back)

**Appendix 3.C: The chi-square test to calculate the differences of the proportion of claimants represented by a private attorney across the claimant's marital status**

****

[Go back](#Appendix3C_back)

**Appendix 4 .1: Two-Way Anova: The differences of private attorney representation and insuranceon the amount claimed**



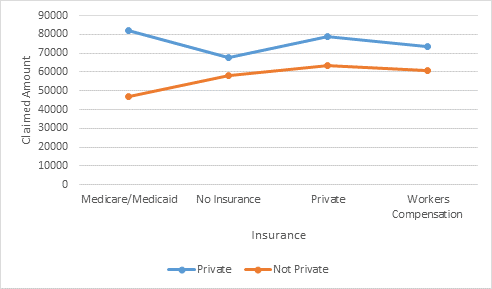
[Go back](#Appendix41_back)

**Appendix 4.2: Marginal Mean table**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Marginal mean table** |  |  |  |  |  |
|  | **Medicare/Medicaid** | **No Insurance** | **Private** | **Workers Compensation** | **Total** |
| **Private** | 82352.938 | 67788.72 | 78991.26 | 73504.874 | 302637.79 |
| **Not Private** | 46811.976 | 57949.854 | 63579.96 | 60936.586 | 229278.374 |
|  | 129164.914 | 125738.574 | 142571.2 | 134441.46 |  |

[Go back](#Appendix41_back)

**Appendix 4.3: Cell Mean Plot**



[Go back](#Appendix43_back)