Stats 506, F18, Problem Set 3

Chen Xie, chenxie@umich.edu

November 17, 2018

Question 1

Part a

The percent of homes have stucco construction as the major outside wall material within each division is shown in the table 1 and figure 1 below.

Table 1: *Proportion of homes with stucco construction within each census division in 2015. Estimates are based on the residential energy consumption survey.*

Census Division	% Stucco Homes (95% CI)
Mountain South	64.2% (55.4, 73.0)
Pacific	44.6% (41.3, 47.9)
Mountain North	16.6% (10.2, 23.0)
South Atlantic	10.6% (7.8, 13.4)
West North Central	4.9% (0.9, 8.8)
West South Central	3.0% (1.6, 4.3)
Middle Atlantic	2.1% (0.6, 3.5)
New England	1.2% (0.0, 2.8)
East North Central	0.7% (0.1, 1.2)
East South Central	0.4% (0.0, 1.2)

From this table, we can know that the division 'Mountain South' has the highest proportion, and 'East South Central' has the lowest proportion.

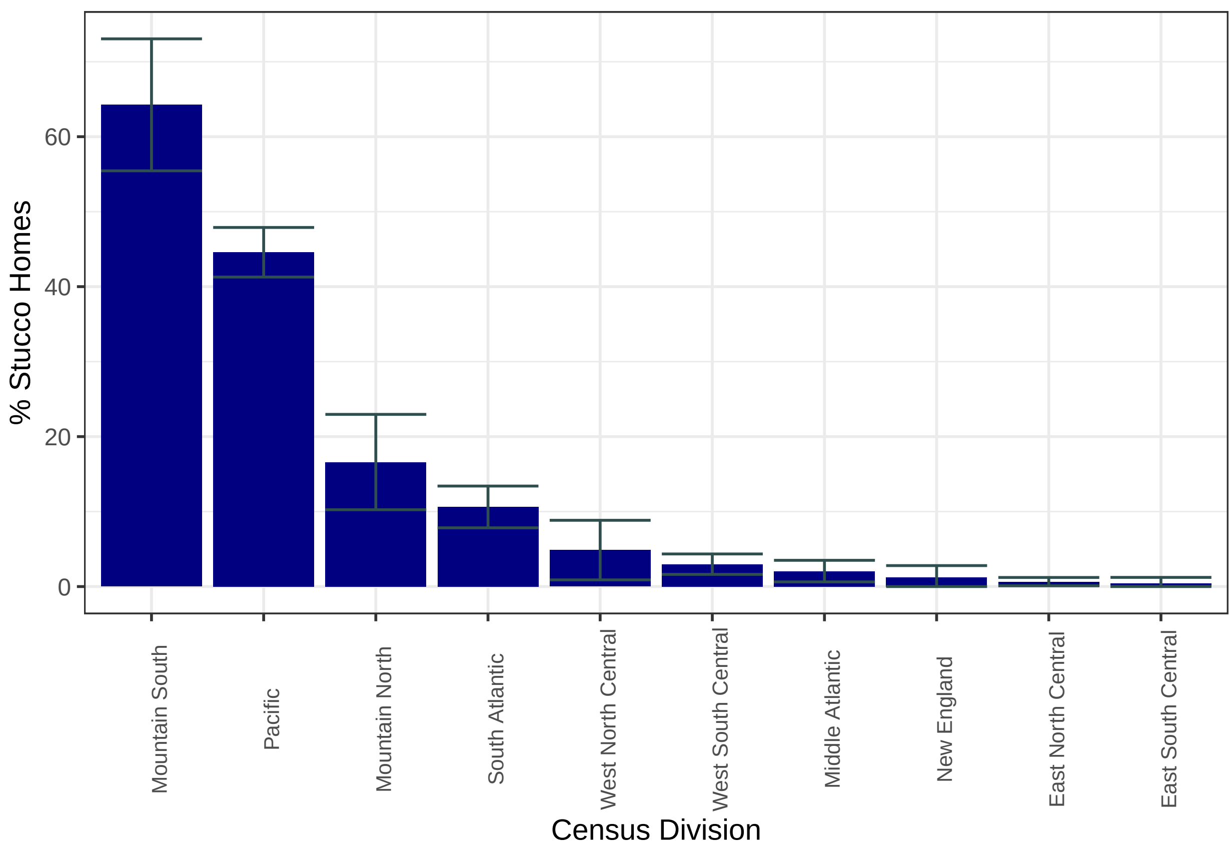

Figure 1: Estimated percent of homes within each census division with major wall type of stucco.

Part b

The average total electricity usage in kilowatt hours in each division is shown in the table 2 and figure 2 below.

Table 2: *Average annual electricity utilization by Census Division in kwh/home.*

Census Division	Average Electricity Usage, kwh/home (95% CI)
East South Central	14,536, (13,320 - 15,752)
West South Central	14,324, (13,495 - 15,153)
South Atlantic	13,447, (12,904 - 13,989)
West North Central	10,524, (9,635 - 11,413)
Mountain South	10,442, (7,950 - 12,934)
East North Central	9,129, (8,730 - 9,528)
Middle Atlantic	8,465, (8,071 - 8,860)
Mountain North	8,384, (7,121 - 9,648)
Pacific	8,100, (7,750 - 8,450)
New England	7,515, (6,472 - 8,557)

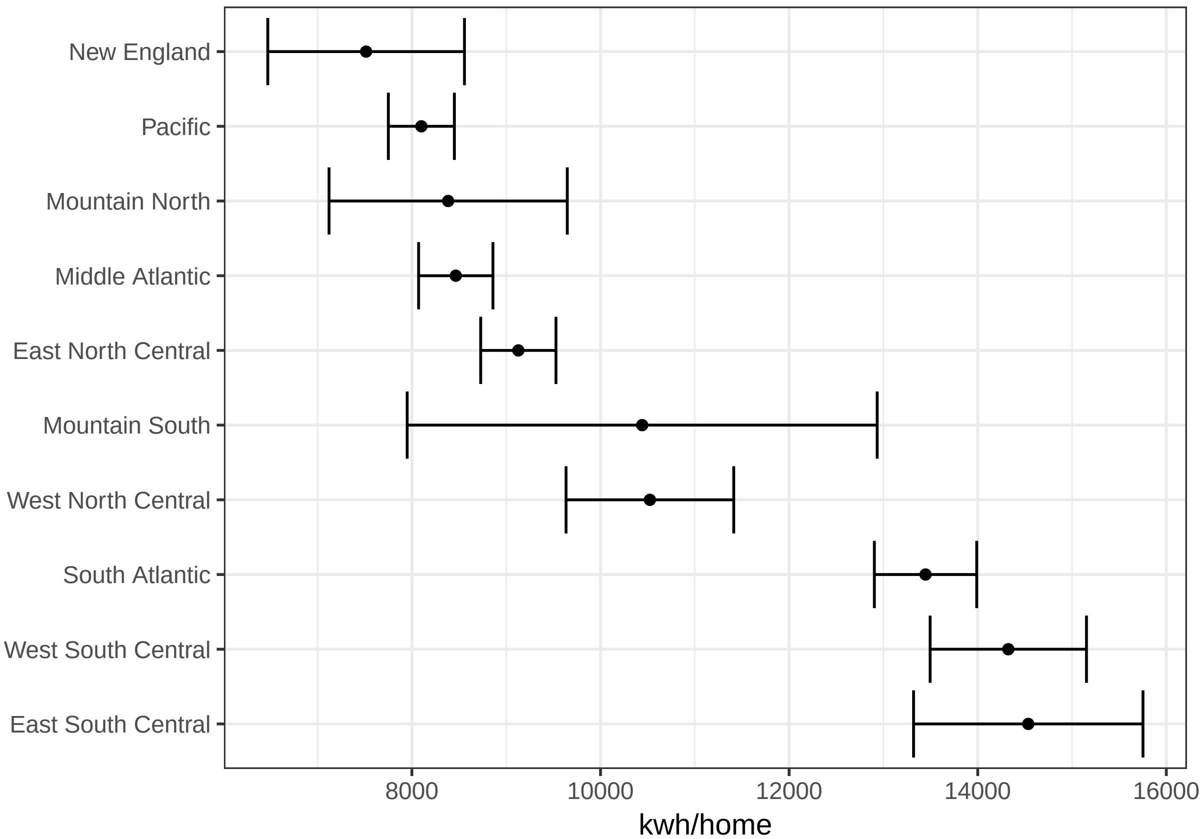

Figure 2: Estimated average annual electricity usage in khw/home for each of 10 census divisions.

The average total electricity usage in kilowatt hours in each division for urban and rural is shown in the table 3 and figure 3 below.

Table 3: *Average electricity utilization in kwh per home for urban and rural areas witihin each census division.*

Census Division	Rural, kwh/home (95% CI)	Urban, kwh/home (95% CI)
East South Central	16,333, (14,088 - 18,578)	13,747, (12,197 - 15,298)
West South Central	16,317, (14,067 - 18,567)	13,629, (12,852 - 14,405)
South Atlantic	15,942, (14,839 - 17,045)	12,725, (12,134 - 13,316)
West North Central	14,174, (12,608 - 15,740)	9,467, (8,722 - 10,211)
Pacific	14,115, (12,001 - 16,229)	7,349, (6,905 - 7,793)
East North Central	13,500, (12,022 - 14,978)	7,980, (7,552 - 8,408)
Middle Atlantic	12,223, (10,633 - 13,814)	7,987, (7,659 - 8,316)
Mountain North	9,356, (5,698 - 13,014)	8,099, (7,396 - 8,803)
New England	9,001, (6,766 - 11,236)	6,964, (5,918 - 8,010)
Mountain South	8,610, (6,536 - 10,685)	10,743, (8,178 - 13,308)

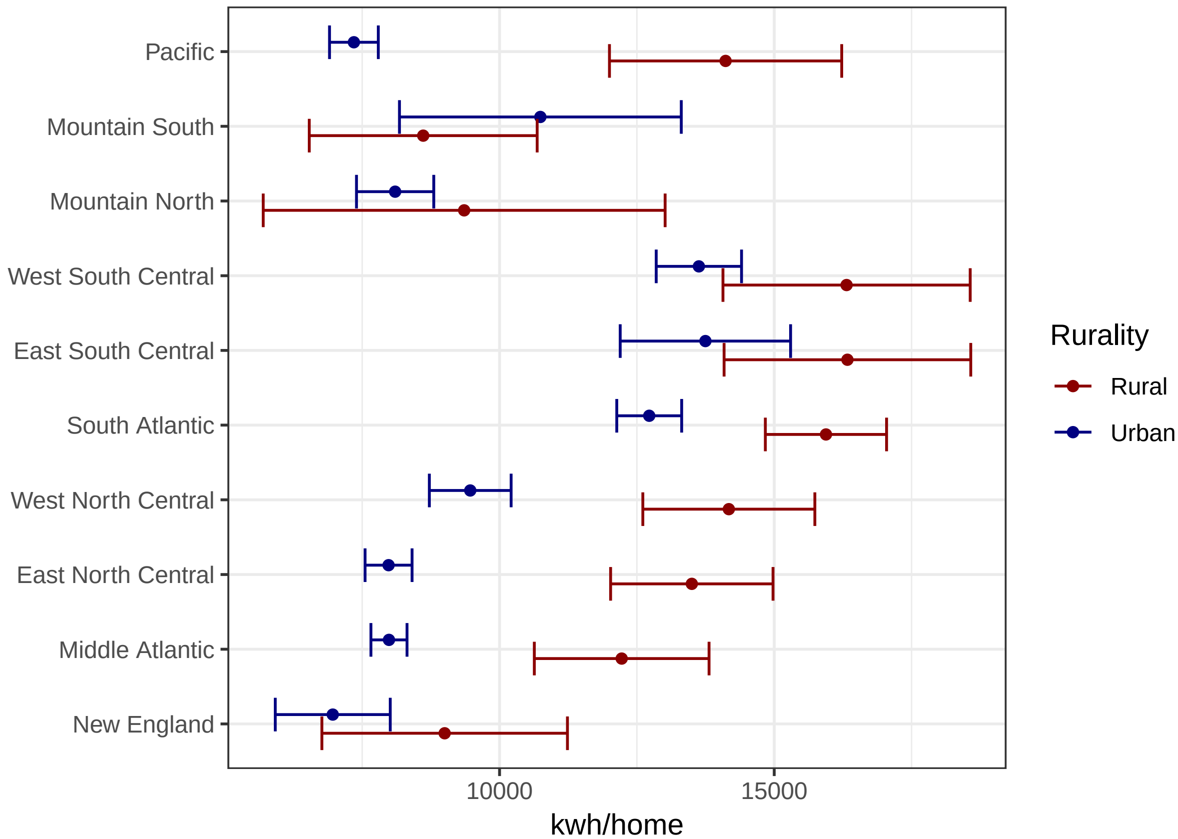

Figure 3: Estimated average annual electricity usage in khw/home for rural and urban areas in each of 10 census divisions.

Part c

The disparity between urban and rural in internet access within each division is shown in the table 4 and figure 4 below.

The division 'Mountain South' has the largest disparity between urban and rural areas in terms of the proportion of homes with internet access.

Table 4: *Urban and rural disparity in internet access for the ten US Census Division in 2015.*

Census Division	Urban Internet Access, % (95% CI)	Rural Internet Access, % (95% CI)	Difference, % (95% CI)
Mountain South	85.3% (81.3, 89.2)	66.7% (58.3, 75.2)	18.5% (7.2, 29.8)
East South Central	78.4% (70.5, 86.2)	69.0% (63.5, 74.6)	9.3% (-1.4, 20.1)
West North Central	88.0% (84.6, 91.4)	80.3% (71.5, 89.2)	7.7% (-2.5, 17.8)
Mountain North	87.4% (82.0, 92.9)	81.9% (73.8, 90.0)	5.5% (-6.2, 17.2)
West South Central	81.6% (76.4, 86.8)	76.5% (72.1, 80.9)	5.1% (-2.3, 12.5)
Pacific	88.7% (86.2, 91.2)	85.3% (77.4, 93.1)	3.4% (-4.5, 11.4)
South Atlantic	85.3% (82.6, 88.0)	82.0% (76.3, 87.8)	3.3% (-3.5, 10.1)
New England	87.6% (82.5, 92.6)	85.8% (82.4, 89.2)	1.8% (-2.5, 6.0)
East North Central	86.3% (83.8, 88.7)	86.2% (81.6, 90.8)	0.0% (-5.3, 5.4)
Middle Atlantic	89.3% (83.9, 94.8)	91.3% (85.3, 97.3)	-1.9% (-9.1, 5.2)

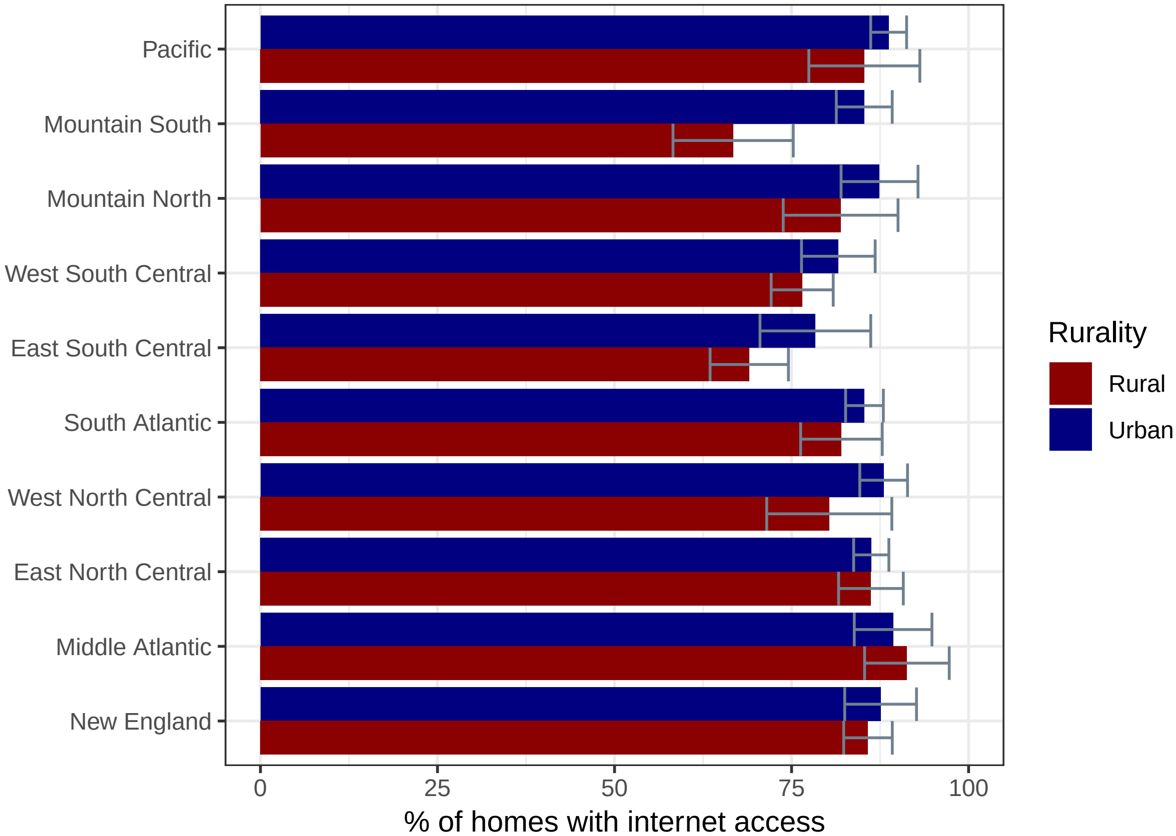

Figure 4: Urban and rural disparity in internet access for the ten US Census Division in 2015

Part d

The average number of bedrooms for different type of housing units is shown in the table 5 and figure 5. We can find that the type 'Single-family detached house' will have most bedrooms.

Table 5: *Average numbers of bedrooms for different type of housing units*

Type of housing units	Average numbers of bedrooms, (95% CI)
Mobile home	2.609, (2.59 - 2.627)
Single-family detached house	3.235, (3.235 - 3.235)
Single-family attached house	2.567, (2.558 - 2.575)
Apartment in a building with 2 to 4 units	1.755, (1.739 - 1.772)
Apartment in a building with 5 or more units	1.546, (1.536 - 1.556)

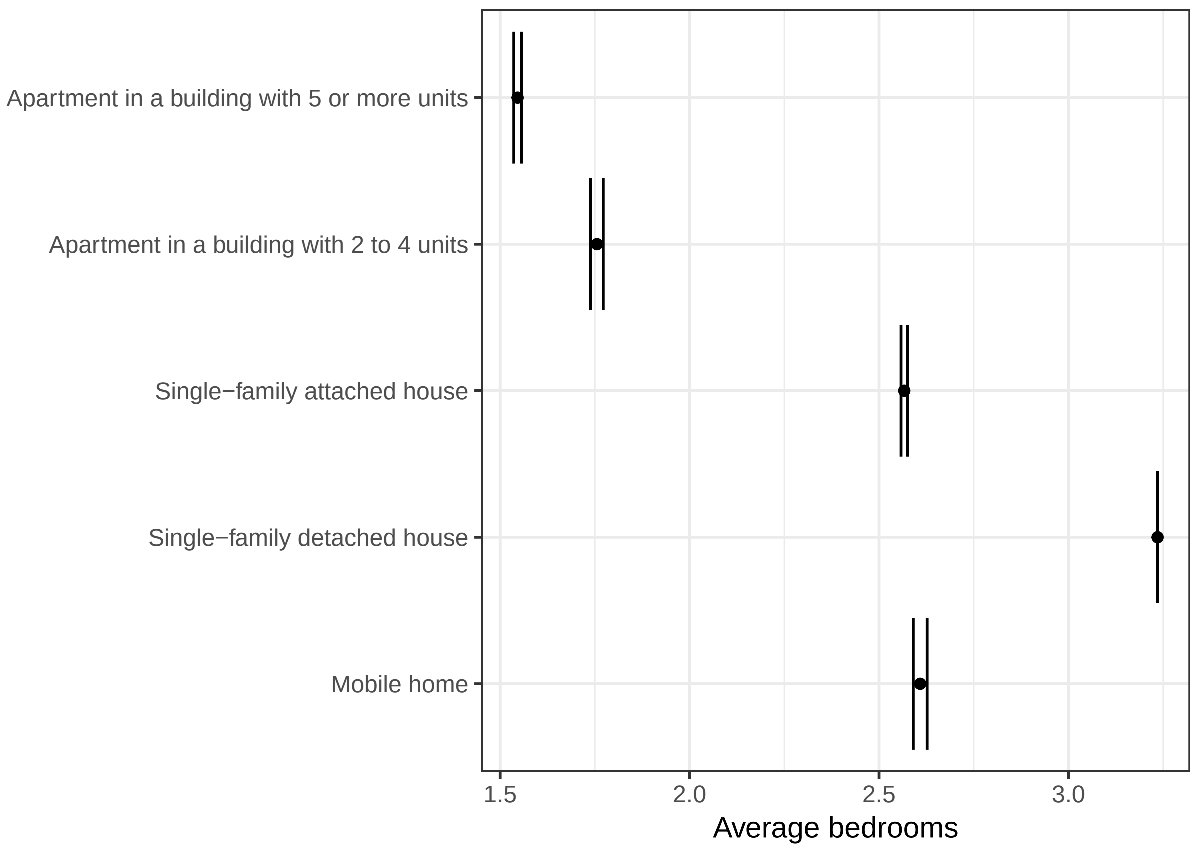

Figure 5: Average numbers of bedrooms for different type of housing units

Question 2

Part a & b

The function in Part a is called 'mc_pval'.

I choose that $\Sigma = I_{p \times p}$, Σ is a $p \times p$ Identity matrix, and $\sigma = 10$.

The estimates of parameters of between the function in Part a is the same as the result of R function lm().

The estimates of parameters of X1 to X6 is shown in the talbe 6, torelance = 1e-3,

Table 6: *Compare estimates of parameters of X1 to X6 between the function in Part a and the R function lm(), tolerance=1e-3*

	Pval of Part_a	Pval of lm()	Difference
X1	0.0042	0.0043	0
X2	0.0002	0.0002	0
X3	0.0926	0.0929	0
X4	0.0948	0.0952	0
X5	0.0203	0.0205	0
X6	0.1762	0.1765	0

Part c & d & e

The Monte Carlo estimates of the family wise error rate, the false discovery rate, the sensitivity, and the specificity of different types of adjusted p-values in multiple comparison are shown in the table 7 and figure 6.

When p-value is not adjusted, the family wise error rate is extremely large, and the false discovery rate is large also. But the adjusted p-values can fix this problem in some levels, and also improve the specificity.

Table 7: Monte Carlo estimates of the family wise error rate, the false discovery rate, the sensitivity, and the specificity of different types of adjusted p-values in multiple comparison

	Family wise error rate	False discovery rate	Sensitivity	Specificity
Unadjusted	0.989	0.3344	0.8465	0.9497
Bonferroni	0.001	0.0051	0.0212	1
Holm	0.001	0.0051	0.0212	1
BH(Benjamini-Hochberg)	0.203	0.0405	0.4805	0.9975
BY(Benjamini-Yekuteli)	0.007	0.0034	0.152	0.9999

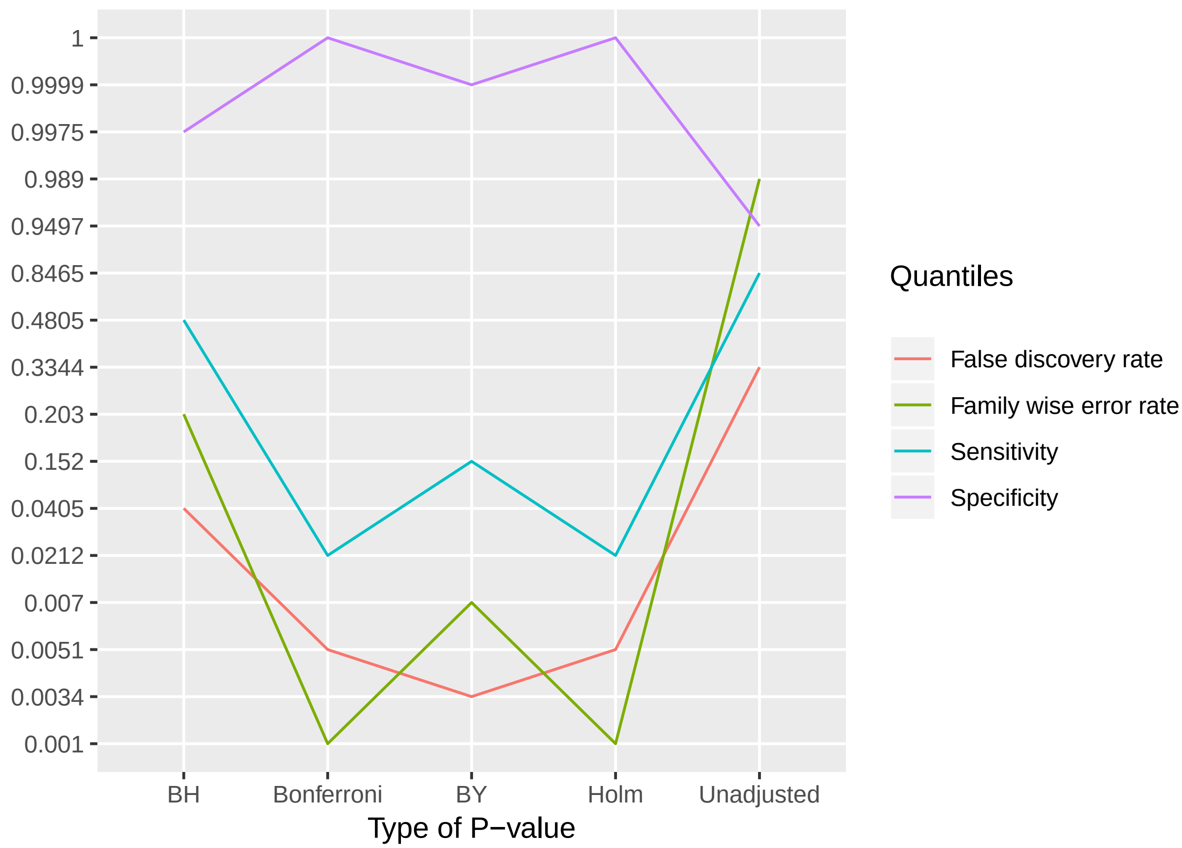

Figure 6: Monte Carlo estimates of the family wise error rate, the false discovery rate, the sensitivity, and the specificity of different types of adjusted p-values in multiple comparison

Question 3

Part a

Computed by data.table, the table of univariate regression coefficients with groups is shown below.

Table 8: *The univariate regression coefficients with groups computed by data.table*

Number of cylinders(Group)	Beta of Displacement	Beta of Gross horsepower	Beta of Weight
4	-0.1351	-0.1128	-5.6470
6	0.0036	-0.0076	-2.7801
8	-0.0196	-0.0142	-2.1924

Part b

beta_dt_fn is the function for an arbitrary dependent, independent, and grouping varaibles. The results computed by **beta_dt_fn** is shown below. It matches the Part a.

Table 9: *The univariate regression coefficients with groups computed by a function computing by data.table*

Number of cylinders(Group)	Beta of Displacement	Beta of Gross horsepower	Beta of Weight
4	-0.1351	-0.1128	-5.6470
6	0.0036	-0.0076	-2.7801
8	-0.0196	-0.0142	-2.1924

Part c

Computed by dplyr summarize_at, the table of univariate regression coefficients with groups is shown below.

Table 10: *The univariate regression coefficients with groups computed by dplyr*

Number of cylinders(Group)	Beta of Displacement	Beta of Gross horsepower	Beta of Weight
4	-0.1351	-0.1128	-5.6470
6	0.0036	-0.0076	-2.7801
8	-0.0196	-0.0142	-2.1924

Part d

beta__dplyr__fn is the function for an arbitrary dependent, independent, and grouping varaibles. The results computed by **beta__dplyr__fn** is shown below. It matches the Part c.

Table 11: *The univariate regression coefficients with groups computed by a function computing by data.table*

Number of cylinders(Group)	Beta of Displacement	Beta of Gross horsepower	Beta of Weight
4	-0.1351	-0.1128	-5.6470
6	0.0036	-0.0076	-2.7801
8	-0.0196	-0.0142	-2.1924