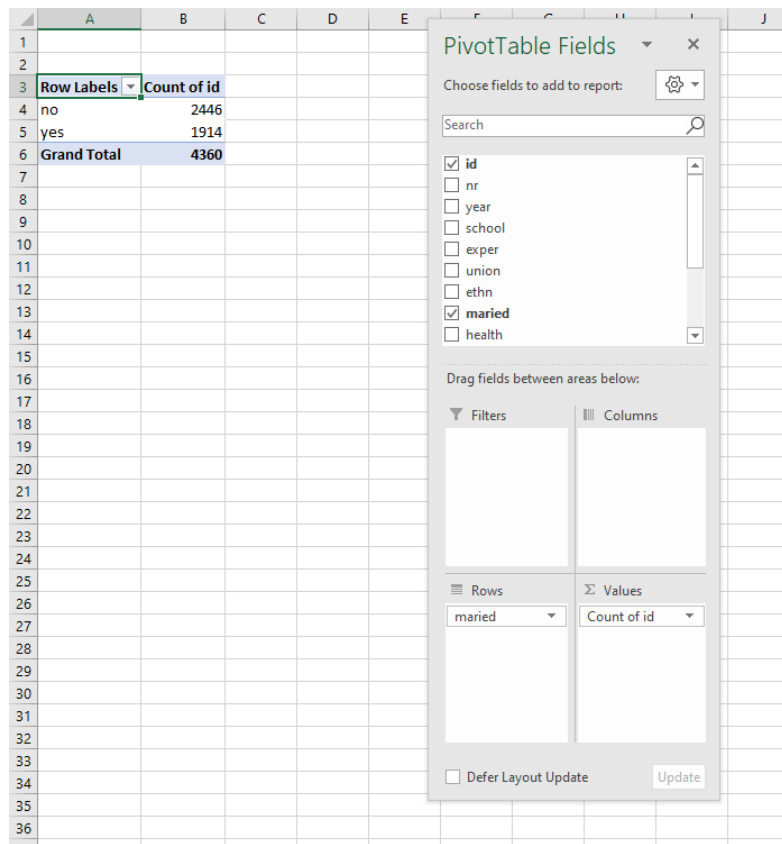


EXPLORATORY DATA ANALYSIS IN EXCEL– DEMO NOTES

Frequencies

Create a PivotTable from the source data.

1. Make a frequency table by selecting categories of interest in the Rows/Columns field, then place a Count of the ID field in the Values section.
 - a. To convert a field from a Sum to a Count, double-click on that variable header, and select Count in the “Summarize value field by” menu.



| Row Labels | Count of id |
|--------------------|-------------|
| no | 2446 |
| yes | 1914 |
| Grand Total | 4360 |

PivotTable Fields

Choose fields to add to report:

Search

- ☒ id
- ☐ nr
- ☐ year
- ☐ school
- ☐ exper
- ☐ union
- ☐ ethn
- ☒ married
- ☐ health

Drag fields between areas below:

Filters

Columns

Rows

married

Values

Count of id

☐ Defer Layout Update Update

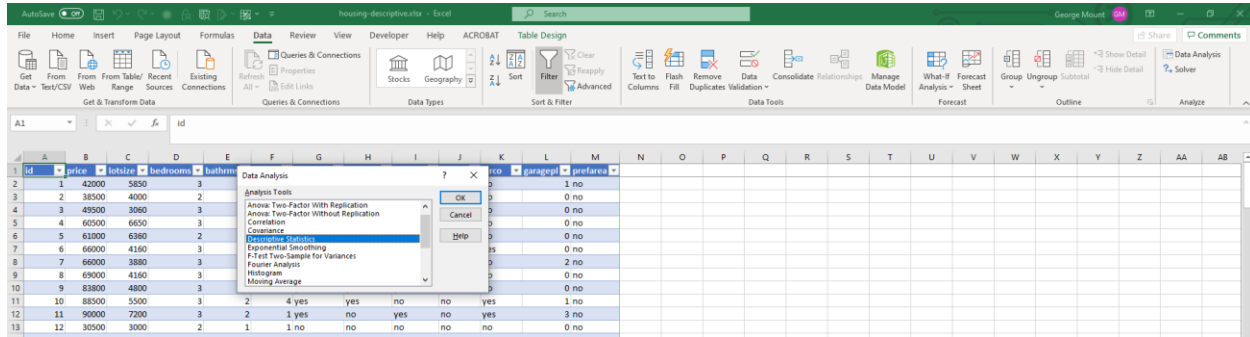
Downloading the Analysis ToolPak

[See instructions from Microsoft here](#). Note the process is different for Windows and Mac.



Descriptive Statistics

1. Go to the Data tab on the home ribbon.
2. Select Data Analysis from the Analyze group (far right of the menu).
3. Select Descriptive Statistics from the menu.



4. Select your Input Range. This will be Column J, year. If your selection includes a header row, make sure to check on the “Labels in First Row” option.
5. By default, the output will be placed in a new worksheet. If you want it elsewhere, click inside “Output Range.” Make sure to double-click inside the dialog box before selecting a new range, otherwise the input range will be re-written.
6. Check on “Summary Statistics.”

| 1 | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X |
|------|------|-------|---------|------|--------|-------|--------|---------|--------|-------------|----------------------------------|--------------------------------------|------------|---|---|---|---|---|---|---|---|---|---|---|
| | id | price | lotsize | year | school | union | ethnic | married | health | wage | industry | occupation | residence | | | | | | | | | | | |
| 4324 | 4323 | 12451 | 1982 | 14 | 4 | no | other | yes | no | 1.696641043 | Professional_and_Related Service | Professional, Technical, and_kindred | south | | | | | | | | | | | |
| 4325 | 4324 | 12451 | 1983 | 14 | 5 | no | other | yes | no | 2.016516822 | Professional_and_Related Service | Professional, Technical, and_kindred | south | | | | | | | | | | | |
| 4326 | 4325 | 12451 | 1984 | 14 | 6 | no | other | yes | yes | 1.176785255 | Professional_and_Related Service | Professional, Technical, and_kindred | NA | | | | | | | | | | | |
| 4327 | 4326 | 12451 | 1985 | 14 | 7 | no | other | yes | no | 1.49183267 | Entertainment | Laborers_and_Farmers | south | | | | | | | | | | | |
| 4328 | 4327 | 12451 | 1986 | 14 | 8 | no | other | yes | no | 1.97198476 | Professional_and_Related Service | Professional, Technical, and_kindred | south | | | | | | | | | | | |
| 4329 | 4328 | 12451 | 1987 | 14 | 9 | no | other | yes | no | 1.876827282 | Professional_and_Related Service | Professional, Technical, and_kindred | south | | | | | | | | | | | |
| 4330 | 4329 | 12477 | 1980 | 12 | 4 | yes | other | no | no | 2.213467409 | Manufacturing | Operatives_and_kindred | north_east | | | | | | | | | | | |
| 4331 | 4330 | 12477 | 1981 | 12 | 5 | yes | other | no | no | 2.222542385 | Manufacturing | Craftsmen_Foremen_and_kindred | north_east | | | | | | | | | | | |
| 4332 | 4331 | 12477 | 1982 | 12 | 6 | yes | other | yes | no | 2.201329889 | Manufacturing | Operatives_and_kindred | north_east | | | | | | | | | | | |
| 4333 | 4332 | 12477 | 1983 | 12 | 7 | yes | other | yes | no | 2.171697191 | Manufacturing | Craftsmen_Foremen_and_kindred | north_east | | | | | | | | | | | |
| 4334 | 4333 | 12477 | 1984 | 12 | 8 | yes | other | yes | no | 2.1680542 | Manufacturing | Operatives_and_kindred | north_east | | | | | | | | | | | |
| 4335 | 4334 | 12477 | 1985 | 12 | 9 | yes | other | yes | no | 2.113686557 | Manufacturing | Craftsmen_Foremen_and_kindred | north_east | | | | | | | | | | | |
| 4336 | 4335 | 12477 | 1986 | 12 | 10 | yes | other | yes | no | 2.113618994 | Manufacturing | Operatives_and_kindred | north_east | | | | | | | | | | | |
| 4337 | 4336 | 12477 | 1987 | 12 | 11 | yes | other | yes | no | 2.202486397 | Manufacturing | Service_Workers | north_east | | | | | | | | | | | |
| 4338 | 4337 | 12500 | 1980 | 12 | 4 | no | other | yes | no | 0.972402635 | Trade | Managers_Officials_and_Proprietors | NA | | | | | | | | | | | |
| 4339 | 4338 | 12500 | 1981 | 12 | 5 | no | other | no | no | 1.324885733 | Trade | Service_Workers | NA | | | | | | | | | | | |
| 4340 | 4339 | 12500 | 1982 | 12 | 6 | no | other | no | no | 0.962707049 | Trade | Service_Workers | NA | | | | | | | | | | | |
| 4341 | 4340 | 12500 | 1983 | 12 | 7 | no | other | yes | no | 0.845251511 | Business_and_Repair_Service | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4342 | 4341 | 12500 | 1984 | 12 | 8 | no | other | yes | no | 1.929574308 | Business_and_Repair_Service | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4343 | 4342 | 12500 | 1985 | 12 | 9 | no | other | yes | no | 2.037502633 | Trade | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4344 | 4343 | 12500 | 1986 | 12 | 10 | no | other | yes | no | 1.148220688 | Business_and_Repair_Service | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4345 | 4344 | 12500 | 1987 | 12 | 11 | no | other | yes | no | 1.306740194 | Manufacturing | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4346 | 4345 | 12534 | 1980 | 11 | 2 | no | other | no | no | 1.84091654 | Construction | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4347 | 4346 | 12534 | 1981 | 11 | 3 | no | other | no | no | 2.174751722 | Construction | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4348 | 4347 | 12534 | 1982 | 11 | 4 | no | other | yes | no | 2.122766666 | Construction | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4349 | 4348 | 12534 | 1983 | 11 | 5 | no | other | yes | no | 2.130875197 | Construction | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4350 | 4349 | 12534 | 1984 | 11 | 6 | no | other | yes | no | 2.1680542 | Construction | Laborers_and_Farmers | NA | | | | | | | | | | | |
| 4351 | 4350 | 12534 | 1985 | 11 | 7 | no | other | yes | no | 2.207794529 | Construction | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4352 | 4351 | 12534 | 1986 | 11 | 8 | no | other | yes | no | 2.38188296 | Construction | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4353 | 4352 | 12534 | 1987 | 11 | 9 | no | other | yes | no | 2.342917012 | Construction | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4354 | 4353 | 12548 | 1980 | 9 | 5 | no | other | no | no | 1.130545336 | Construction | Laborers_and_Farmers | NA | | | | | | | | | | | |
| 4355 | 4354 | 12548 | 1981 | 9 | 6 | no | other | no | no | 1.111603469 | Construction | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4356 | 4355 | 12548 | 1982 | 9 | 7 | no | other | no | no | 0.832481541 | Construction | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4357 | 4356 | 12548 | 1983 | 9 | 8 | no | other | yes | no | 1.591878696 | Construction | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4358 | 4357 | 12548 | 1984 | 9 | 9 | yes | other | yes | no | 1.212542755 | Construction | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4359 | 4358 | 12548 | 1985 | 9 | 10 | no | other | yes | no | 1.765961776 | Construction | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4360 | 4359 | 12548 | 1986 | 9 | 11 | yes | other | yes | no | 1.745894213 | Professional_and_Related Service | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |
| 4361 | 4360 | 12548 | 1987 | 9 | 12 | yes | other | yes | no | 1.466542887 | Public_Administration | Craftsmen_Foremen_and_kindred | NA | | | | | | | | | | | |

Descriptive Statistics

Input Range: \$B\$1:\$G\$4361

Grouped By: Columns

☒ Labels in first row

Output options

☐ Output Range:

☒ New Worksheet Ply

☐ New Workbook

☒ Summary statistics

☐ Confidence Level for Means: 95 %

☐ Kth Largest:

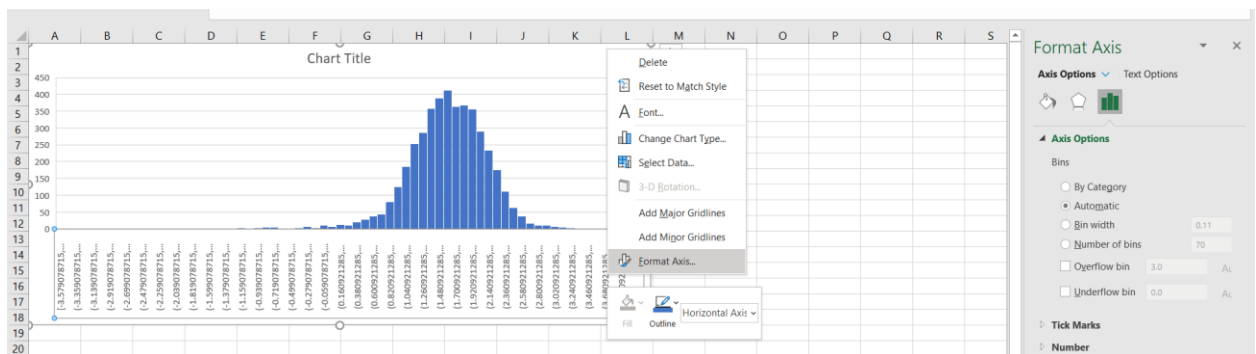
☐ Kth Smallest:

OK Cancel Help



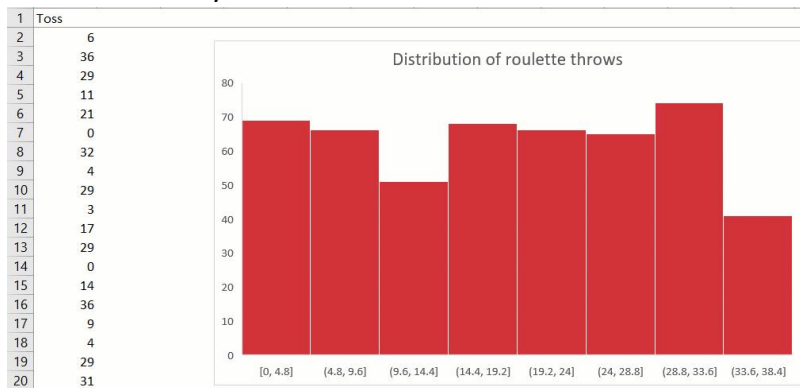
Histograms

1. Select your input range and go to Insert -> Charts. Histogram should be your third option. Select that. You can cut and paste the resulting histogram elsewhere in the workbook.
2. To change the number of bins in the histogram, right-click on the X-axis and select Format Axis. You can then customize the X-axis on the side menu. *Note: these features are not available on Excel for Mac.*



Central limit theorem

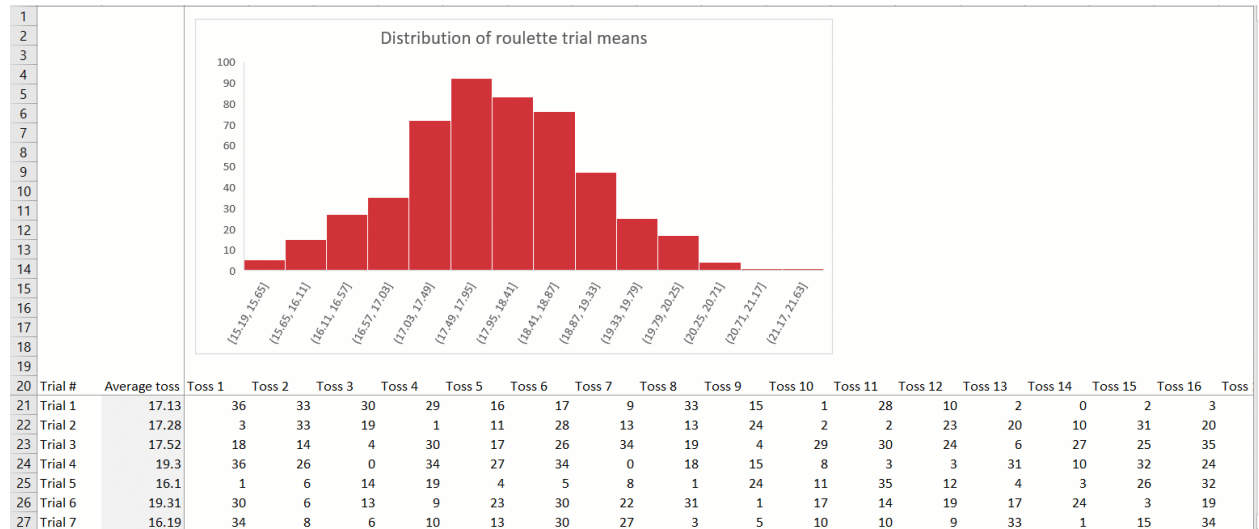
1. Simulate 500 rounds of a roulette spin using `RANDBETWEEN(0,36)`
2. Plot the resulting frequency distribution as a histogram.
3. Use F9 while in your workbook to refresh it.



4. This is a *uniform* distribution.



5. Now simulate a roulette spin 100 times and take the average spin. Do this 500 times and plot the resulting distribution of *sample* means.



6. This time we get a normal distribution, due to the central limit theorem.

Law of large numbers: large-numbers.xlsx

1. Simulate a roulette toss 500 times in Column B: `RANDBETWEEN(0,36)`
2. Take a running total in Column C: `SUM(B2:B2)`
3. Take a running total in Column D: `C2/A2`
4. Plot Column D as a line chart. Press F9 to recalculate.
 - a. The line converges to the expected mean due to the law of large numbers.

