ØAMET2200 · Fall 2019
Worksheet 3

Instructor: Fenella Carpena

September 6, 2019

Exercise 1 (Adapted from Stine and Foster, Chapter 15, Question 37) Consider the situations described below. Identify the population and sample, explain what the parameter p or μ represents, and tell whether we can use the given information to create a confidence interval. If so, find the 90% confidence interval.

(a) A survey of customers at a Rema 1000 asks whether they found shopping at Rema 1000 to be a better experience than at Kiwi. Of the 2,500 forms distributed to consumers, 325 were completed, and 250 of these said that the experience was at Rema 1000 was better.

(b) A sample of 1,000 customers of DNB who were given loans during the last two years contains 2 customers who have defaulted.

Exercise 2 (Adapted from Stine and Foster, Chapter 15, Question 53) Click fraud has become a major concern as more and more companies advertise on the internet. Click fraud is an internet crime where a person or computer program is used to register clicks on pay-per-click advertising.

For example, suppose that OsloMet advertises with Google. Then, Google places an ad for OsloMet's website at the top of the search results page, and OsloMet pays a fee to Google anytime someone clicks on the link. That's fine when the person who clicks the link is someone who is interested in the OsloMet website, but it is a crime when it's a computer program pretending to be a customer.

Suppose that an analysis of 1,200 clicks coming into OsloMet's website last week identified 175 of these clicks as fraudulent clicks.

(a) Under what conditions does it makes sense to treat these 1,200 clicks as a sample? What would be the population?

(b) Show the 95% confidence interval for the population proportion of fraudulent clicks. Round your final answer to two decimal points.

(c) If a company pays Google \$4.50 for each click, give a confidence interval (again, rounding your final answer to two decimal points) for the mean cost due to fraud *per click*.

Exercise 3 (Stine and Foster, Chapter 15, Question 29) A summary of sales made in the quarterly report of a department store says that the average retail purchase was \$125 with a margin of error equal to \$15. What does the margin of error mean in this context?

Exercise 4 (Stine and Foster, Chapter 15, Question 31) To prepare a report on the economy, analysts need to estimate the percentage of businesses that plan to hire additional employees in the next 60 days.

(a) How many randomly selected employers must you contact in order to guarantee a margin of error of no more than 4%?

(b) If analysts believe that at most 20% of firms are likely to be hiring, how many must they survey to obtain a margin of error of 0.04?

Exercise 5 (Orkla Foods) Orkla Foods is planning a line of frozen organic dinners, and the company plans to conduct a survey to measure the calorie intake of Norwegians. The company nutritionist wants to know the average calorie intake to within ± 50 calories with 95% confidence. What sample size do they need for the survey? A pilot study of 25 people shows a standard deviation of 430 calories.

Exercise 6 (Video Games) A video game developer wants to know how intensely customers play its most recently released game. It is planning to run a small survey of its customers.

(a) If the developer wants to know the average number of minutes spent daily playing the game to within ± 10 minutes, how small does the standard error of the mean of a sample need to be?

(b) Would a sample of n = 200 customers be enough to meet the goal stated in part (a), or is there missing information?

(c) The developer would like to know the proportion of customers who have played its game in the last week to within 5% (i.e., ± 0.05). How many customers does it need to survey?

Exercise 7 (Credit Cards) Mastercard is considering offering an OsloMet alumni credit card. The profitability of offering such a card depends on: (1) the proportion of all alumni who will sign up for the credit card, p, and (2) the average monthly balance (in kroner) that those who sign up for the card will carry, μ. Before deciding whether to launch the OsloMet alumni card, Mastercard conducts a pilot study. Specifically, pre-approved credit card offers are sent by mail to a simple random sample of 1,000 OsloMet alumni. Summary statistics for the pilot sample are shown below.

Summary Statistics for Pilot Sample

Number of offers	1,000
Number accepted	140
Proportion who accepted	$\widehat{p} = 0.14$
Average of monthly balance	$\overline{x} = 1,990.50$ kroner
SD of monthly balance	$s = 2,833.33$ kroner
Kurtosis K_4 of monthly balance	2.975

(a) What is the 95% CI for p, the proportion of all OsloMet alumni who will accept the card offer?

(b) Given the results in part (a), if MasterCard launches the OsloMet credit card, would 20% of all OsloMet alumni accept the offer?

(c) What is the 80% CI for p?

(d) How would the answer in part (c) change if we had a larger sample size?

(e) What is the 95% CI for μ, the average monthly balance that those who sign up for the credit card will carry?

(f) What is the 80% CI for μ?

(g) Let's look at the profit that Mastercard earns from the credit cards. Let Y_i be the profit per credit card offer. The summary statistics for Y_i are shown below.

Summary Statistics for Y_i

$\overline{y}$	12.867 kroner
s	117.674 kroner
n	1,000
$s/\sqrt{n}$	3.721 kroner
Kurtosis	44.061

Do the data satisfy the sample size condition?

(h) What is the 95% CI for the average profit per offer?

(i) Suppose there are currently 100,000 OsloMet alumni. If Mastercard makes credit card offers to these 100,000 alumni, what is the 95% CI for Mastercard's total profit?

(j) On the basis of the results in part (i), would you recommend that Mastercard launch the OsloMet credit card?

Exercise 8 (Stocking Supermarket Shelves) A company that stocks shelves in supermarkets is considering expanding the supply that it delivers. Items that are not sold must be discarded at the end of the day, so it only wants to schedule additional deliveries if stores regularly sell out. A break-even analysis indicates that an additional delivery cycle will be profitable if items are selling out in more than 60% of markets. A survey during the last week in 45 markets found the shelves bare in 35. Should the company add an additional delivery cycle?

(a) State the null and alternative hypothesis.

(b) Is the hypothesis test in part (a) one-sided or two-sided?

(c) Describe the Type I and Type II errors.

(d) What is the test statistic?

Exercise 9 (Stata) For this exercise, use Stata's built-in `nslw88` dataset, which contains data from a simple random sample of 1,878 women. You can access this dataset by typing `sysuse nlsw88` in the command line.

(a) Tabulate the variable `union` (Hint: use the `tabulate` command). What do you think the values of the variable `union` mean?

(b) What is the 90% confidence interval for the proportion of the female population that are union members? (Hint: use the `ci` command)

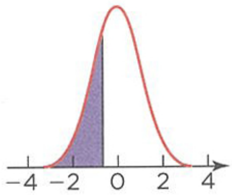
(c) Does the sample size condition hold for the CI you constructed in part (b)? (Hint: use the results from part (a))

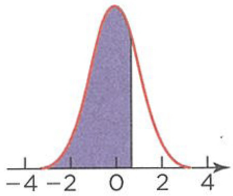
(d) What are the sample mean and SD of the variable `union`? (Hint: use the `summarize` command)

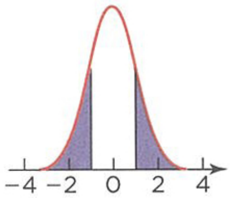
(e) Use the results from part (d) to calculate **by hand (i.e., without using Stata)** the 90% confidence interval for the proportion of the female population that are union members. Do you get the same confidence interval as in part (b)? (You should have, with some rounding error).

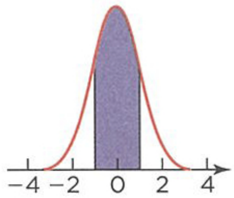
(f) Repeat parts (b) and (e) for the variable `wage`; this variable corresponds to the individual's hourly wage (in dollars).

(g) Does the sample size condition hold for the CI you constructed in part (f)? (Hint: use the command `summarize wage, detail` to find the kurtosis)



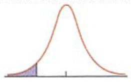
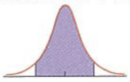
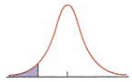
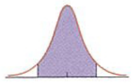
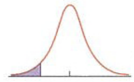
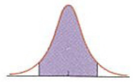






z	$P(Z \leq -z)$	$P(Z \leq z)$	$P(\|Z\| > z)$	$P(\|Z\| \leq z)$
0	0.50	0.50	1	0
0.0502	0.48	0.52	0.96	0.04
0.1004	0.46	0.54	0.92	0.08
0.1510	0.44	0.56	0.88	0.12
0.2019	0.42	0.58	0.84	0.16
0.2533	0.40	0.60	0.80	0.20
0.3055	0.38	0.62	0.76	0.24
0.3585	0.36	0.64	0.72	0.28
0.4125	0.34	0.66	0.68	0.32
0.4677	0.32	0.68	0.64	0.36
0.4959	0.31	0.69	0.62	0.38
0.5244	0.30	0.70	0.60	0.40
0.5828	0.28	0.72	0.56	0.44
0.6433	0.26	0.74	0.52	0.48
0.6745	0.25	0.75	0.50	0.50
0.7063	0.24	0.76	0.48	0.52
0.7388	0.23	0.77	0.46	0.54
0.7722	0.22	0.78	0.44	0.56
0.8064	0.21	0.79	0.42	0.58
0.8416	0.20	0.80	0.40	0.60
0.8779	0.19	0.81	0.38	0.62
0.9154	0.18	0.82	0.36	0.64
0.9542	0.17	0.83	0.34	0.66
0.9945	0.16	0.84	0.32	0.68
1.0364	0.15	0.85	0.30	0.70
1.0803	0.14	0.86	0.28	0.72
1.1264	0.13	0.87	0.26	0.74
1.1750	0.12	0.88	0.24	0.76
1.2265	0.11	0.89	0.22	0.78
1.2816	0.10	0.90	0.20	0.80
1.3408	0.09	0.91	0.18	0.82
1.4051	0.08	0.92	0.16	0.84
1.4758	0.07	0.93	0.14	0.86
1.5548	0.06	0.94	0.12	0.88
1.6449	0.05	0.95	0.10	0.90
1.7507	0.04	0.96	0.08	0.92
1.8808	0.03	0.97	0.06	0.94
1.9600	0.025	0.975	0.05	0.95
2.0537	0.02	0.98	0.04	0.96
2.3263	0.01	0.99	0.02	0.98
2.5758	0.005	0.995	0.01	0.99
2.8070	0.0025	0.9975	0.005	0.995
3.0902	0.001	0.999	0.002	0.998
3.2905	0.0005	0.9995	0.001	0.999
3.7190	0.0001	0.9999	0.0002	0.9998
3.8906	0.00005	0.99995	0.0001	0.9999
4.2649	0.00001	0.99999	0.00002	0.99998
4.4172	0.000005	0.999995	0.00001	0.99999

T–TABLE **Percentiles of Student's *t* distribution.**

$df = 1$

t	$P(T_1 \le -t)$	$P(-t \le T_1 \le t)$
3.078	0.1	0.8
6.314	0.05	0.9
12.71	0.025	0.95
31.82	0.01	0.98
63.66	0.005	0.99
318.3	0.001	0.998
636.6	0.0005	0.999
6366	0.00005	0.9999

$df = 2$

t	$P(T_2 \le -t)$	$P(-t \le T_2 \le t)$
1.886	0.1	0.8
2.920	0.05	0.9
4.303	0.025	0.95
6.965	0.01	0.98
9.925	0.005	0.99
22.33	0.001	0.998
31.60	0.0005	0.999
99.99	0.00005	0.9999

$df = 3$

t	$P(T_3 \le -t)$	$P(-t \le T_3 \le t)$
1.638	0.1	0.8
2.353	0.05	0.9
3.182	0.025	0.95
4.541	0.01	0.98
5.841	0.005	0.99
10.21	0.001	0.998
12.92	0.0005	0.999
28.00	0.00005	0.9999

$df = 4$

t	$P(T_4 \le -t)$	$P(-t \le T_4 \le t)$
1.533	0.1	0.8
2.132	0.05	0.9
2.776	0.025	0.95
3.747	0.01	0.98
4.604	0.005	0.99
7.173	0.001	0.998
8.610	0.0005	0.999
15.54	0.00005	0.9999

$df = 5$

t	$P(T_5 \le -t)$	$P(-t \le T_5 \le t)$
1.476	0.1	0.8
2.015	0.05	0.9
2.571	0.025	0.95
3.365	0.01	0.98
4.032	0.005	0.99
5.893	0.001	0.998
6.869	0.0005	0.999
11.18	0.00005	0.9999

$df = 6$

t	$P(T_6 \le -t)$	$P(-t \le T_6 \le t)$
1.440	0.1	0.8
1.943	0.05	0.9
2.447	0.025	0.95
3.143	0.01	0.98
3.707	0.005	0.99
5.208	0.001	0.998
5.959	0.0005	0.999
9.082	0.00005	0.9999

$df = 7$

t	$P(T_7 \le -t)$	$P(-t \le T_7 \le t)$
1.415	0.1	0.8
1.895	0.05	0.9
2.365	0.025	0.95
2.998	0.01	0.98
3.499	0.005	0.99
4.785	0.001	0.998
5.408	0.0005	0.999
7.885	0.00005	0.9999

$df = 8$

t	$P(T_8 \le -t)$	$P(-t \le T_8 \le t)$
1.397	0.1	0.8
1.860	0.05	0.9
2.306	0.025	0.95
2.896	0.01	0.98
3.355	0.005	0.99
4.501	0.001	0.998
5.041	0.0005	0.999
7.120	0.00005	0.9999

$df = 9$

t	$P(T_9 \le -t)$	$P(-t \le T_9 \le t)$
1.383	0.1	0.8
1.833	0.05	0.9
2.262	0.025	0.95
2.821	0.01	0.98
3.250	0.005	0.99
4.297	0.001	0.998
4.781	0.0005	0.999
6.594	0.00005	0.9999

$df = 10$

t	$P(T_{10} \le -t)$	$P(-t \le T_{10} \le t)$
1.415	0.1	0.8
1.895	0.05	0.9
2.365	0.025	0.95
2.998	0.01	0.98
3.499	0.005	0.99
4.785	0.001	0.998
5.408	0.0005	0.999
7.885	0.00005	0.9999

$df = 11$

t	$P(T_{11} \le -t)$	$P(-t \le T_{11} \le t)$
1.397	0.1	0.8
1.860	0.05	0.9
2.306	0.025	0.95
2.896	0.01	0.98
3.355	0.005	0.99
4.501	0.001	0.998
5.041	0.0005	0.999
7.120	0.00005	0.9999

$df = 12$

t	$P(T_{12} \le -t)$	$P(-t \le T_{12} \le t)$
1.383	0.1	0.8
1.833	0.05	0.9
2.262	0.025	0.95
2.821	0.01	0.98
3.250	0.005	0.99
4.297	0.001	0.998
4.781	0.0005	0.999
6.594	0.00005	0.9999

$df = 13$

t	$P(T_{13} \le -t)$	$P(-t \le T_{13} \le t)$
1.350	0.1	0.8
1.771	0.05	0.9
2.160	0.025	0.95
2.650	0.01	0.98
3.012	0.005	0.99
3.852	0.001	0.998
4.221	0.0005	0.999
5.513	0.00005	0.9999

$df = 14$

t	$P(T_{14} \le -t)$	$P(-t \le T_{14} \le t)$
1.345	0.1	0.8
1.761	0.05	0.9
2.145	0.025	0.95
2.624	0.01	0.98
2.977	0.005	0.99
3.787	0.001	0.998
4.140	0.0005	0.999
5.363	0.00005	0.9999

$df = 15$

t	$P(T_{15} \le -t)$	$P(-t \le T_{15} \le t)$
1.341	0.1	0.8
1.753	0.05	0.9
2.131	0.025	0.95
2.602	0.01	0.98
2.947	0.005	0.99
3.733	0.001	0.998
4.073	0.0005	0.999
5.239	0.00005	0.9999

$df = 16$

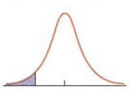
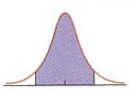
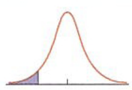
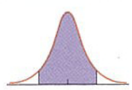
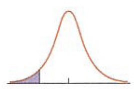
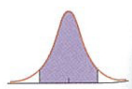
t	$P(T_{16} \le -t)$	$P(-t \le T_{16} \le t)$
1.337	0.1	0.8
1.746	0.05	0.9
2.120	0.025	0.95
2.583	0.01	0.98
2.921	0.005	0.99
3.686	0.001	0.998
4.015	0.0005	0.999
5.134	0.00005	0.9999

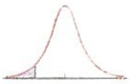
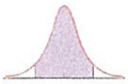
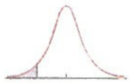
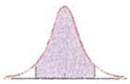
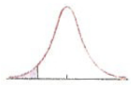
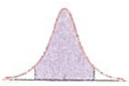
$df = 17$

t	$P(T_{17} \le -t)$	$P(-t \le T_{17} \le t)$
1.333	0.1	0.8
1.740	0.05	0.9
2.110	0.025	0.95
2.567	0.01	0.98
2.898	0.005	0.99
3.646	0.001	0.998
3.965	0.0005	0.999
5.044	0.00005	0.9999

$df = 18$

t	$P(T_{18} \le -t)$	$P(-t \le T_{18} \le t)$
1.33	0.1	0.8
1.734	0.05	0.9
2.101	0.025	0.95
2.552	0.01	0.98
2.878	0.005	0.99
3.610	0.001	0.998
3.922	0.0005	0.999
4.966	0.00005	0.9999

$df = 19$

t	$P(T_{19} \leq -t)$	$P(-t \leq T_{19} \leq t)$
1.328	0.1	0.8
1.729	0.05	0.9
2.093	0.025	0.95
2.539	0.01	0.98
2.861	0.005	0.99
3.579	0.001	0.998
3.883	0.0005	0.999
4.897	0.00005	0.9999

$df = 20$

t	$P(T_{20} \leq -t)$	$P(-t \leq T_{20} \leq t)$
1.325	0.1	0.8
1.725	0.05	0.9
2.086	0.025	0.95
2.528	0.01	0.98
2.845	0.005	0.99
3.552	0.001	0.998
3.850	0.0005	0.999
4.837	0.00005	0.9999

$df = 22$

t	$P(T_{22} \leq -t)$	$P(-t \leq T_{22} \leq t)$
1.321	0.1	0.8
1.717	0.05	0.9
2.074	0.025	0.95
2.508	0.01	0.98
2.819	0.005	0.99
3.505	0.001	0.998
3.792	0.0005	0.999
4.736	0.00005	0.9999

$df = 24$

t	$P(T_{24} \leq -t)$	$P(-t \leq T_{24} \leq t)$
1.318	0.1	0.8
1.711	0.05	0.9
2.064	0.025	0.95
2.492	0.01	0.98
2.797	0.005	0.99
3.467	0.001	0.998
3.745	0.0005	0.999
4.654	0.00005	0.9999

$df = 26$

t	$P(T_{26} \leq -t)$	$P(-t \leq T_{26} \leq t)$
1.315	0.1	0.8
1.706	0.05	0.9
2.056	0.025	0.95
2.479	0.01	0.98
2.779	0.005	0.99
3.435	0.001	0.998
3.707	0.0005	0.999
4.587	0.00005	0.9999

$df = 28$

t	$P(T_{28} \leq -t)$	$P(-t \leq T_{28} \leq t)$
1.313	0.1	0.8
1.701	0.05	0.9
2.048	0.025	0.95
2.467	0.01	0.98
2.763	0.005	0.99
3.408	0.001	0.998
3.674	0.0005	0.999
4.530	0.00005	0.9999

$df = 30$

t	$P(T_{30} \leq -t)$	$P(-t \leq T_{30} \leq t)$
1.31	0.1	0.8
1.697	0.05	0.9
2.042	0.025	0.95
2.457	0.01	0.98
2.75	0.005	0.99
3.385	0.001	0.998
3.646	0.0005	0.999
4.482	0.00005	0.9999

$df = 32$

t	$P(T_{32} \leq -t)$	$P(-t \leq T_{32} \leq t)$
1.309	0.1	0.8
1.694	0.05	0.9
2.037	0.025	0.95
2.449	0.01	0.98
2.738	0.005	0.99
3.365	0.001	0.998
3.622	0.0005	0.999
4.441	0.00005	0.9999

$df = 34$

t	$P(T_{34} \leq -t)$	$P(-t \leq T_{34} \leq t)$
1.307	0.1	0.8
1.691	0.05	0.9
2.032	0.025	0.95
2.441	0.01	0.98
2.728	0.005	0.99
3.348	0.001	0.998
3.601	0.0005	0.999
4.405	0.00005	0.9999

$df = 36$

t	$P(T_{36} \leq -t)$	$P(-t \leq T_{36} \leq t)$
1.306	0.1	0.8
1.688	0.05	0.9
2.028	0.025	0.95
2.434	0.01	0.98
2.719	0.005	0.99
3.333	0.001	0.998
3.582	0.0005	0.999
4.374	0.00005	0.9999

$df = 40$

t	$P(T_{40} \leq -t)$	$P(-t \leq T_{40} \leq t)$
1.303	0.1	0.8
1.684	0.05	0.9
2.021	0.025	0.95
2.423	0.01	0.98
2.704	0.005	0.99
3.307	0.001	0.998
3.551	0.0005	0.999
4.321	0.00005	0.9999

$df = 50$

t	$P(T_{50} \leq -t)$	$P(-t \leq T_{50} \leq t)$
1.299	0.1	0.8
1.676	0.05	0.9
2.009	0.025	0.95
2.403	0.01	0.98
2.678	0.005	0.99
3.261	0.001	0.998
3.496	0.0005	0.999
4.228	0.00005	0.9999

$df = 60$

t	$P(T_{60} \leq -t)$	$P(-t \leq T_{60} \leq t)$
1.296	0.1	0.8
1.671	0.05	0.9
2.000	0.025	0.95
2.390	0.01	0.98
2.660	0.005	0.99
3.232	0.001	0.998
3.460	0.0005	0.999
4.169	0.00005	0.9999

$df = 75$

t	$P(T_{75} \leq -t)$	$P(-t \leq T_{75} \leq t)$
1.293	0.1	0.8
1.665	0.05	0.9
1.992	0.025	0.95
2.377	0.01	0.98
2.643	0.005	0.99
3.202	0.001	0.998
3.425	0.0005	0.999
4.110	0.00005	0.9999

$df = 100$

t	$P(T_{100} \leq -t)$	$P(-t \leq T_{100} \leq t)$
1.290	0.1	0.8
1.660	0.05	0.9
1.984	0.025	0.95
2.364	0.01	0.98
2.626	0.005	0.99
3.174	0.001	0.998
3.390	0.0005	0.999
4.053	0.00005	0.9999

$df = 125$

t	$P(T_{125} \leq -t)$	$P(-t \leq T_{125} \leq t)$
1.288	0.1	0.8
1.657	0.05	0.9
1.979	0.025	0.95
2.357	0.01	0.98
2.616	0.005	0.99
3.157	0.001	0.998
3.370	0.0005	0.999
4.020	0.00005	0.9999

$df = 150$

t	$P(T_{150} \leq -t)$	$P(-t \leq T_{150} \leq t)$
1.287	0.1	0.8
1.655	0.05	0.9
1.976	0.025	0.95
2.351	0.01	0.98
2.609	0.005	0.99
3.145	0.001	0.998
3.357	0.0005	0.999
3.998	0.00005	0.9999

$df = \infty$

t	$P(Z \leq -t)$	$P(-t \leq Z \leq t)$
1.282	0.1	0.8
1.645	0.05	0.9
1.960	0.025	0.95
2.326	0.01	0.98
2.576	0.005	0.99
3.090	0.001	0.998
3.291	0.0005	0.999
3.891	0.00005	0.9999