ØAMET2200 · Fall 2019
Worksheet 9

Instructor: Fenella Carpena

November 1, 2019

Exercise 1 Suppose we have regression models to predict the salary of employees. Let *salary* be an employee's annual salary in kroner, *bachelor* if the employee has a bachelor's degree, and *exper* be the employee's years of work experience.

(a) Consider the sample regression

$$\widehat{salary} = b_0 + b_1 bachelor$$

(i) What is the sample regression for employees with a bachelor degree?

(ii) What is the sample regression for employees without a bachelor degree?

(iii) Interpet b_0.

(iv) Interpret $b_0 + b_1$.

(v) Interpret b_1.

(b) Consider the sample regression

$$\widehat{salary} = b_0 + b_1 bachelor + b_2 exper$$

(i) What is the sample regression for employees with a bachelor degree?

(ii) What is the sample regression for employees without a bachelor degree?

(iii) Interpet b_0.

(iv) Interpret $b_0 + b_1$.

(v) Interpret b_1.

(vi) Interpret b_2.

(c) Consider the sample regression

$$\widehat{salary} = b_0 + b_1 bachelor + b_2 exper + b_3 bachelor * exper$$

(i) What is the sample regression for employees with a bachelor degree?

(ii) What is the sample regression for employees without a bachelor degree?

(iii) Interpet b_0.

(iv) Interpret $b_0 + b_1$.

(v) Interpret b_1.

(vi) Interpret b_2.

(vii) Interpret $b_2 + b_3$.

(viii) Interpret b_3.

Exercise 2 Does choosing the right airline help you get into business class? How about planning ahead? For this exercise, we have data from a random sample of 202 recent round-trip business class tickets for travel between London Heathrow Airport (LHR) and New York's John F. Kennedy Airport (JFK). We have two airlines in our dataset, American Airlines and Delta Airlines. The data contain the following three variables:

- `rtrip_fare`: the price (in dollars) of a round-trip ticket in business class
- `delta`: a dummy variable for Delta Airlines
- `advance`: the number of days in advance (before the flight date) that the ticket was purchased

We are interested in finding out how we can purchase business class tickets at an affordable price. We have two ideas that we want to explore. First, it might be the case that one of the two airlines we're considering (American and Delta) offers cheaper business class seats. Second, by purchasing the business class ticket in advance, we might save some money, and how much we save by purchasing in advance may differ by airline.

To determine the effects of airline and advanced ticket purchases on business class ticket costs, we estimate the following multiple regression model.

```
. regress rtrip_fare i.delta##c.advance ;
```

Source	SS	df	MS
Model	168142808	3	56047602.7
Residual	2.4423e+09	198	12334998
Total	2.6105e+09	201	12987424.9

Number of obs	202
F(3, 198)	4.54
Prob > F	0.0042
R-squared	0.0644
Adj R-squared	0.0502
Root MSE	3512.1

rtrip_fare	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.delta	-2262.889	1693.512	-1.34	0.183	-5602.524	1076.746
advance	-507.367	155.2434	-3.27	0.001	-813.5097	-201.2242
delta#c.advance						
1	454.8766	173.3933	2.62	0.009	112.9421	796.8112
_cons	10622.75	396.2652	26.81	0.000	9841.312	11404.2

(a) Does the model have explanatory power? What statistic do you look at in this regression output to find out?

(b) Does the slope estimate for the interaction suggest that buying tickets in advance is actually more expensive on Delta than buying tickets closer to the date of the flight? Explain.

(c) What are the sample regression equations for Delta and American? Interpret the coefficients in each regression.

(d) Draw the two regression lines that you wrote in part (c). Make sure to label the x-axis and y-axis.

(e) If it is within three days of your flight, which airline offers you the best deal? If it is 10 days before your flight, which airline offers you the best deal?

(f) Would you exclude any of the variables included in the above regression? Why or why not? Explain.

Exercise 3 (Stata) Scandinavian Airlines (SAS) offer credit cards that reward customers who use the card with frequent-flyer miles. The more the customer uses the card, the more miles earned. Do these cards work? Do customers who get such a card fly more on that airline? To find out, SAS compared the number of miles flown by a sample of 250 members in its SAS Eurobonus frequent-flyer program. Some of these Eurobonus members have the SAS credit card. `HasCard` in the data is a dummy variable, coded 1 for those who have the card and 0 for those who do not.

For this exercise, parts (a) to (c) are conceptual questions that do not require Stata. For the remaining parts, the dataset is `freq_flier.xlsx`.

Motivation

(a) How would the results of this comparison affect the use of this promotion by the airline?

Method

(b) Explain why the airline should be concerned about the effects of possible confounding variables in this analysis.

(c) One possible confounding variable is the number of miles flown by the customer in the year prior to getting the airline credit card. How can the airline use regression to mitigate the problems introduced by this lurking variable?

Mechanics

(d) Use a two-sample t-test to compare the current mileage of those who have a card to those who do not.

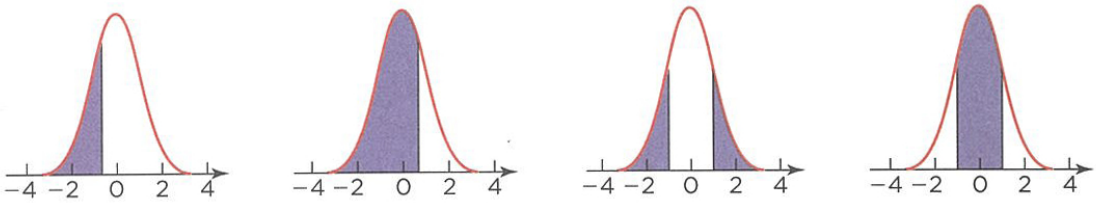
(e) Assume that all MRM conditions hold. Fit a regression model with current mileage as the y variable, and has card, miles flown last year, and their interaction as the x variables.

  (i) Interpret the coefficient on the interaction variable.

  (ii) Should we drop the interaction variable from the regression? Why or why not? Assume a significance level of 5%.

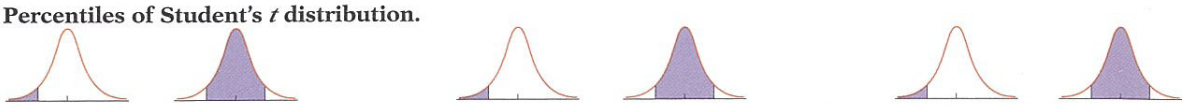
Message

(f) Present the results of this analysis to SAS. Be sure to explain why the regression in part (e) gives a different answer from the two-sample t-test in part (d).

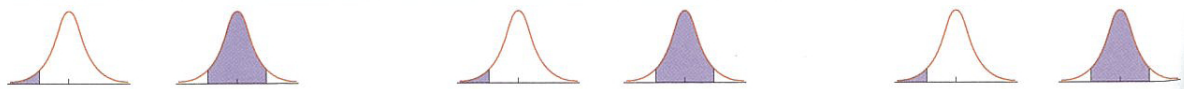


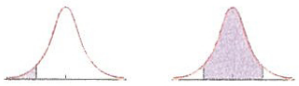
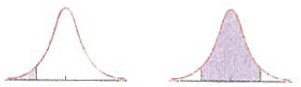
z	$P(Z \leq -z)$	$P(Z \leq z)$	$P(\|Z\| > z)$	$P(\|Z\| \leq z)$
0	0.50	0.50	1	0
0.0502	0.48	0.52	0.96	0.04
0.1004	0.46	0.54	0.92	0.08
0.1510	0.44	0.56	0.88	0.12
0.2019	0.42	0.58	0.84	0.16
0.2533	0.40	0.60	0.80	0.20
0.3055	0.38	0.62	0.76	0.24
0.3585	0.36	0.64	0.72	0.28
0.4125	0.34	0.66	0.68	0.32
0.4677	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.



$df = 1$			$df = 2$			$df = 3$		
t	$P(T_1 \le -t)$	$P(-t \le T_1 \le t)$	t	$P(T_2 \le -t)$	$P(-t \le T_2 \le t)$	t	$P(T_3 \le -t)$	$P(-t \le T_3 \le t)$
3.078	0.1	0.8	1.886	0.1	0.8	1.638	0.1	0.8
6.314	0.05	0.9	2.920	0.05	0.9	2.353	0.05	0.9
12.71	0.025	0.95	4.303	0.025	0.95	3.182	0.025	0.95
31.82	0.01	0.98	6.965	0.01	0.98	4.541	0.01	0.98
63.66	0.005	0.99	9.925	0.005	0.99	5.841	0.005	0.99
318.3	0.001	0.998	22.33	0.001	0.998	10.21	0.001	0.998
636.6	0.0005	0.999	31.60	0.0005	0.999	12.92	0.0005	0.999
6366	0.00005	0.9999	99.99	0.00005	0.9999	28.00	0.00005	0.9999
$df = 4$			$df = 5$			$df = 6$		
t	$P(T_4 \le -t)$	$P(-t \le T_4 \le t)$	t	$P(T_5 \le -t)$	$P(-t \le T_5 \le t)$	t	$P(T_6 \le -t)$	$P(-t \le T_6 \le t)$
1.533	0.1	0.8	1.476	0.1	0.8	1.440	0.1	0.8
2.132	0.05	0.9	2.015	0.05	0.9	1.943	0.05	0.9
2.776	0.025	0.95	2.571	0.025	0.95	2.447	0.025	0.95
3.747	0.01	0.98	3.365	0.01	0.98	3.143	0.01	0.98
4.604	0.005	0.99	4.032	0.005	0.99	3.707	0.005	0.99
7.173	0.001	0.998	5.893	0.001	0.998	5.208	0.001	0.998
8.610	0.0005	0.999	6.869	0.0005	0.999	5.959	0.0005	0.999
15.54	0.00005	0.9999	11.18	0.00005	0.9999	9.082	0.00005	0.9999
$df = 7$			$df = 8$			$df = 9$		
t	$P(T_7 \le -t)$	$P(-t \le T_7 \le t)$	t	$P(T_8 \le -t)$	$P(-t \le T_8 \le t)$	t	$P(T_9 \le -t)$	$P(-t \le T_9 \le t)$
1.415	0.1	0.8	1.397	0.1	0.8	1.383	0.1	0.8
1.895	0.05	0.9	1.860	0.05	0.9	1.833	0.05	0.9
2.365	0.025	0.95	2.306	0.025	0.95	2.262	0.025	0.95
2.998	0.01	0.98	2.896	0.01	0.98	2.821	0.01	0.98
3.499	0.005	0.99	3.355	0.005	0.99	3.250	0.005	0.99
4.785	0.001	0.998	4.501	0.001	0.998	4.297	0.001	0.998
5.408	0.0005	0.999	5.041	0.0005	0.999	4.781	0.0005	0.999
7.885	0.00005	0.9999	7.120	0.00005	0.9999	6.594	0.00005	0.9999
$df = 10$			$df = 11$			$df = 12$		
t	$P(T_{10} \le -t)$	$P(-t \le T_{10} \le t)$	t	$P(T_{11} \le -t)$	$P(-t \le T_{11} \le t)$	t	$P(T_{12} \le -t)$	$P(-t \le T_{12} \le t)$
1.415	0.1	0.8	1.397	0.1	0.8	1.383	0.1	0.8
1.895	0.05	0.9	1.860	0.05	0.9	1.833	0.05	0.9
2.365	0.025	0.95	2.306	0.025	0.95	2.262	0.025	0.95
2.998	0.01	0.98	2.896	0.01	0.98	2.821	0.01	0.98
3.499	0.005	0.99	3.355	0.005	0.99	3.250	0.005	0.99
4.785	0.001	0.998	4.501	0.001	0.998	4.297	0.001	0.998
5.408	0.0005	0.999	5.041	0.0005	0.999	4.781	0.0005	0.999
7.885	0.00005	0.9999	7.120	0.00005	0.9999	6.594	0.00005	0.9999
$df = 13$			$df = 14$			$df = 15$		
t	$P(T_{13} \le -t)$	$P(-t \le T_{13} \le t)$	t	$P(T_{14} \le -t)$	$P(-t \le T_{14} \le t)$	t	$P(T_{15} \le -t)$	$P(-t \le T_{15} \le t)$
1.350	0.1	0.8	1.345	0.1	0.8	1.341	0.1	0.8
1.771	0.05	0.9	1.761	0.05	0.9	1.753	0.05	0.9
2.160	0.025	0.95	2.145	0.025	0.95	2.131	0.025	0.95
2.650	0.01	0.98	2.624	0.01	0.98	2.602	0.01	0.98
3.012	0.005	0.99	2.977	0.005	0.99	2.947	0.005	0.99
3.852	0.001	0.998	3.787	0.001	0.998	3.733	0.001	0.998
4.221	0.0005	0.999	4.140	0.0005	0.999	4.073	0.0005	0.999
5.513	0.00005	0.9999	5.363	0.00005	0.9999	5.239	0.00005	0.9999
$df = 16$			$df = 17$			$df = 18$		
t	$P(T_{16} \le -t)$	$P(-t \le T_{16} \le t)$	t	$P(T_{17} \le -t)$	$P(-t \le T_{17} \le t)$	t	$P(T_{18} \le -t)$	$P(-t \le T_{18} \le t)$
1.337	0.1	0.8	1.333	0.1	0.8	1.33	0.1	0.8
1.746	0.05	0.9	1.740	0.05	0.9	1.734	0.05	0.9
2.120	0.025	0.95	2.110	0.025	0.95	2.101	0.025	0.95
2.583	0.01	0.98	2.567	0.01	0.98	2.552	0.01	0.98
2.921	0.005	0.99	2.898	0.005	0.99	2.878	0.005	0.99
3.686	0.001	0.998	3.646	0.001	0.998	3.610	0.001	0.998
4.015	0.0005	0.999	3.965	0.0005	0.999	3.922	0.0005	0.999
5.134	0.00005	0.9999	5.044	0.00005	0.9999	4.966	0.00005	0.9999



$df = 19$			$df = 20$			$df = 22$		
t	$P(T_{19} \le -t)$	$P(-t \le T_{19} \le t)$	t	$P(T_{20} \le -t)$	$P(-t \le T_{20} \le t)$	t	$P(T_{22} \le -t)$	$P(-t \le T_{22} \le t)$
1.328	0.1	0.8	1.325	0.1	0.8	1.321	0.1	0.8
1.729	0.05	0.9	1.725	0.05	0.9	1.717	0.05	0.9
2.093	0.025	0.95	2.086	0.025	0.95	2.074	0.025	0.95
2.539	0.01	0.98	2.528	0.01	0.98	2.508	0.01	0.98
2.861	0.005	0.99	2.845	0.005	0.99	2.819	0.005	0.99
3.579	0.001	0.998	3.552	0.001	0.998	3.505	0.001	0.998
3.883	0.0005	0.999	3.850	0.0005	0.999	3.792	0.0005	0.999
4.897	0.00005	0.9999	4.837	0.00005	0.9999	4.736	0.00005	0.9999

$df = 24$			$df = 26$			$df = 28$		
t	$P(T_{24} \le -t)$	$P(-t \le T_{24} \le t)$	t	$P(T_{26} \le -t)$	$P(-t \le T_{26} \le t)$	t	$P(T_{28} \le -t)$	$P(-t \le T_{28} \le t)$
1.318	0.1	0.8	1.315	0.1	0.8	1.313	0.1	0.8
1.711	0.05	0.9	1.706	0.05	0.9	1.701	0.05	0.9
2.064	0.025	0.95	2.056	0.025	0.95	2.048	0.025	0.95
2.492	0.01	0.98	2.479	0.01	0.98	2.467	0.01	0.98
2.797	0.005	0.99	2.779	0.005	0.99	2.763	0.005	0.99
3.467	0.001	0.998	3.435	0.001	0.998	3.408	0.001	0.998
3.745	0.0005	0.999	3.707	0.0005	0.999	3.674	0.0005	0.999
4.654	0.00005	0.9999	4.587	0.00005	0.9999	4.530	0.00005	0.9999

$df = 30$			$df = 32$			$df = 34$		
t	$P(T_{30} \le -t)$	$P(-t \le T_{30} \le t)$	t	$P(T_{32} \le -t)$	$P(-t \le T_{32} \le t)$	t	$P(T_{34} \le -t)$	$P(-t \le T_{34} \le t)$
1.31	0.1	0.8	1.309	0.1	0.8	1.307	0.1	0.8
1.697	0.05	0.9	1.694	0.05	0.9	1.691	0.05	0.9
2.042	0.025	0.95	2.037	0.025	0.95	2.032	0.025	0.95
2.457	0.01	0.98	2.449	0.01	0.98	2.441	0.01	0.98
2.75	0.005	0.99	2.738	0.005	0.99	2.728	0.005	0.99
3.385	0.001	0.998	3.365	0.001	0.998	3.348	0.001	0.998
3.646	0.0005	0.999	3.622	0.0005	0.999	3.601	0.0005	0.999
4.482	0.00005	0.9999	4.441	0.00005	0.9999	4.405	0.00005	0.9999

$df = 36$			$df = 40$			$df = 50$		
t	$P(T_{36} \le -t)$	$P(-t \le T_{36} \le t)$	t	$P(T_{40} \le -t)$	$P(-t \le T_{40} \le t)$	t	$P(T_{50} \le -t)$	$P(-t \le T_{50} \le t)$
1.306	0.1	0.8	1.303	0.1	0.8	1.299	0.1	0.8
1.688	0.05	0.9	1.684	0.05	0.9	1.676	0.05	0.9
2.028	0.025	0.95	2.021	0.025	0.95	2.009	0.025	0.95
2.434	0.01	0.98	2.423	0.01	0.98	2.403	0.01	0.98
2.719	0.005	0.99	2.704	0.005	0.99	2.678	0.005	0.99
3.333	0.001	0.998	3.307	0.001	0.998	3.261	0.001	0.998
3.582	0.0005	0.999	3.551	0.0005	0.999	3.496	0.0005	0.999
4.374	0.00005	0.9999	4.321	0.00005	0.9999	4.228	0.00005	0.9999

$df = 60$			$df = 75$			$df = 100$		
t	$P(T_{60} \le -t)$	$P(-t \le T_{60} \le t)$	t	$P(T_{75} \le -t)$	$P(-t \le T_{75} \le t)$	t	$P(T_{100} \le -t)$	$P(-t \le T_{100} \le t)$
1.296	0.1	0.8	1.293	0.1	0.8	1.290	0.1	0.8
1.671	0.05	0.9	1.665	0.05	0.9	1.660	0.05	0.9
2.000	0.025	0.95	1.992	0.025	0.95	1.984	0.025	0.95
2.390	0.01	0.98	2.377	0.01	0.98	2.364	0.01	0.98
2.660	0.005	0.99	2.643	0.005	0.99	2.626	0.005	0.99
3.232	0.001	0.998	3.202	0.001	0.998	3.174	0.001	0.998
3.460	0.0005	0.999	3.425	0.0005	0.999	3.390	0.0005	0.999
4.169	0.00005	0.9999	4.110	0.00005	0.9999	4.053	0.00005	0.9999

$df = 125$			$df = 150$			$df = \infty$		
t	$P(T_{125} \le -t)$	$P(-t \le T_{125} \le t)$	t	$P(T_{150} \le -t)$	$P(-t \le T_{150} \le t)$	t	$P(Z \le -t)$	$P(-t \le Z \le t)$
1.288	0.1	0.8	1.287	0.1	0.8	1.282	0.1	0.8
1.657	0.05	0.9	1.655	0.05	0.9	1.645	0.05	0.9
1.979	0.025	0.95	1.976	0.025	0.95	1.960	0.025	0.95
2.357	0.01	0.98	2.351	0.01	0.98	2.326	0.01	0.98
2.616	0.005	0.99	2.609	0.005	0.99	2.576	0.005	0.99
3.157	0.001	0.998	3.145	0.001	0.998	3.090	0.001	0.998
3.370	0.0005	0.999	3.357	0.0005	0.999	3.291	0.0005	0.999
4.020	0.00005	0.9999	3.998	0.00005	0.9999	3.891	0.00005	0.9999