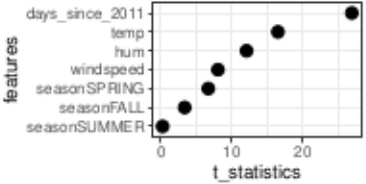
LINEAR REGRESSION - INTERPRETATION

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_p x_p + \epsilon = \mathbf{x}^\top \boldsymbol{\theta} + \epsilon$$

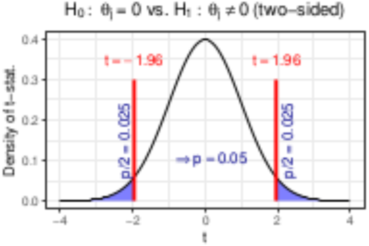
Feature importance:

- Absolute **t-statistic** value: $\hat{\theta}_j$ scaled with standard error ($SE(\hat{\theta}_j) \triangleq$ reliability of estimate)

$$|t_{\hat{\theta}_j}| = \left| \frac{\hat{\theta}_j}{SE(\hat{\theta}_j)} \right|$$



- High t-values $\Rightarrow$ important (significant) feat.
- **p-value**: probability of obtaining a more extreme test statistic assuming H_0 is correct (here: $\theta_j = 0$, i.e., feat. j not significant)
  $\rightsquigarrow$ High $|t| \Rightarrow$ small p-val. (speak against H_0)



EXAMPLE: LINEAR REGRESSION - MAIN EFFECTS

Bike data: predict no. of rented bikes using 4 numeric, 1 categorical feat. (season)

$$
\begin{aligned}
\hat{y} = & \hat{\theta}_0 + \hat{\theta}_1 \mathbb{1}_{x_{season}=SPRING} + \\
& \hat{\theta}_2 \mathbb{1}_{x_{season}=SUMMER} + \\
& \hat{\theta}_3 \mathbb{1}_{x_{season}=FALL} + \hat{\theta}_4 x_{temp} + \\
& \hat{\theta}_5 x_{hum} + \hat{\theta}_6 x_{windspeed} + \\
& \hat{\theta}_7 x_{days\_since\_2011}
\end{aligned}
$$

	Weights	SE	t-stat.	p-val.
(Intercept)	3229.3	220.6	14.6	0.00
seasonSPRING	862.0	129.0	6.7	0.00
seasonSUMMER	41.6	170.2	0.2	0.81
seasonFALL	390.1	116.6	3.3	0.00
temp	120.5	7.3	16.5	0.00
hum	-31.1	2.6	-12.1	0.00
windspeed	-56.9	7.1	-8.0	0.00
days_since_2011	4.9	0.2	26.9	0.00

Interpretation:

- **Intercept**: If all feature values are 0 (and season is WINTER reference cat.), the expected number of bike rentals is $\hat{\theta}_0 = 3229.3$

EXAMPLE: LINEAR REGRESSION - MAIN EFFECTS

Bike data: predict no. of rented bikes using 4 numeric, 1 categorical feat. (season)

$$\begin{aligned}\hat{y} = &\hat{\theta}_0 + \hat{\theta}_1 \mathbb{1}_{x_{season}=SPRING} + \\ &\hat{\theta}_2 \mathbb{1}_{x_{season}=SUMMER} + \\ &\hat{\theta}_3 \mathbb{1}_{x_{season}=FALL} + \hat{\theta}_4 x_{temp} + \\ &\hat{\theta}_5 x_{hum} + \hat{\theta}_6 x_{windspeed} + \\ &\hat{\theta}_7 x_{days\_since\_2011}\end{aligned}$$

	Weights	SE	t-stat.	p-val.
(Intercept)	3229.3	220.6	14.6	0.00
seasonSPRING	862.0	129.0	6.7	0.00
seasonSUMMER	41.6	170.2	0.2	0.81
seasonFALL	390.1	116.6	3.3	0.00
temp	120.5	7.3	16.5	0.00
hum	-31.1	2.6	-12.1	0.00
windspeed	-56.9	7.1	-8.0	0.00
days_since_2011	4.9	0.2	26.9	0.00

Interpretation:

- **Intercept**: If all feature values are 0 (and season is WINTER = reference cat.), the expected number of bike rentals is $\hat{\theta}_0 = 3229.3$
- **Categorical**: Rentals in SPRING are by $\hat{\theta}_1 = 862$ higher than in WINTER, c.p.

EXAMPLE: LINEAR REGRESSION - MAIN EFFECTS

Bike data: predict no. of rented bikes using 4 numeric, 1 categorical feat. (season)

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 \mathbb{1}_{x_{season}=SPRING} + \hat{\theta}_2 \mathbb{1}_{x_{season}=SUMMER} + \hat{\theta}_3 \mathbb{1}_{x_{season}=FALL} + \hat{\theta}_4 x_{temp} + \hat{\theta}_5 x_{hum} + \hat{\theta}_6 x_{windspeed} + \hat{\theta}_7 x_{days\_since\_2011}$$

	Weights	SE	t-stat.	p-val.
(Intercept)	3229.3	220.6	14.6	0.00
seasonSPRING	862.0	129.0	6.7	0.00
seasonSUMMER	41.6	170.2	0.2	0.81
seasonFALL	390.1	116.6	3.3	0.00
temp	120.5	7.3	16.5	0.00
hum	-31.1	2.6	-12.1	0.00
windspeed	-56.9	7.1	-8.0	0.00
days_since_2011	4.9	0.2	26.9	0.00

Interpretation:

- **Intercept**: If all feature values are 0 (and season is WINTER = reference cat.), the expected number of bike rentals is $\hat{\theta}_0 = 3229.3$
- **Categorical**: Rentals in SPRING are by $\hat{\theta}_1 = 862$ higher than in WINTER, c.p.
- **Numerical**: Rentals increase by $\hat{\theta}_4 = 120.5$ if temp increases by 1 °C, c.p.