

Problem 1

1.

Let $P(S)$ denote a solo accident, let $P(H)$ the event of a head injury and let $P(C)$ the event that the accident is a collision:

$$P(S)=0.84$$

$$P(H|S)=0.2$$

$$P(H|C)=0.22$$

$$P(H) = P(H \text{ and } S) + P(H \text{ and not } S)$$

Which can be written as the following:

$$P(H) = P(S) * P(H|S) + P(\text{not } S) * P(H|\text{not } S)$$

$$P(H) = 0.84 * 0.2 + (1 - 0.84) * 0.22 = 0.2032$$

2.

I will find $P(C|H)$:

$$P(C|H) = \frac{P(C \text{ and } H)}{P(H)}$$

$$P(C|H) = \frac{P(C) * P(H|C)}{P(H)} = \frac{(1 - 0.84) * 0.22}{0.2032} = 0.1732$$

3.

$$0 * 0.7968 + 4 * 0.1680 + 6 * 0.0352 = 0.8832$$

Problem 2

1.

$$\frac{1978}{3120} = 0.63397$$

95%-confidence interval with standard error of 0.008624:

$$\hat{p} \pm z * se =]0.6171, 0.6509[$$

Based on the confidence interval, we can with 95% confidence say that there is between 0.6171 and 0.6509 probability that a Dane will agree with the fine.

Since 0.5 is not in the interval, and the two endpoints in the confidence interval are both above 0.5, we can with 95% confidence determine that a majority of Danes agree with the fine.

Problem 3

1.

Jude Werra's Liar Index for 1996:

$$\frac{42}{300} = 0.14$$

Jude Werra's Liar Index for 2001:

$$\frac{63}{340} = 0.1853$$

The pooled proportion is 0.1641, giving us the following:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p} * (1 - \hat{p}) * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = -1.544$$

The test is z-distributed with a z-statistic of -1.544, giving us the p-value of 0.1225. Thus, we do not reject the null hypotheses and conclude that Jude Werra's Liar Index for 1996 (14%) does not differ significantly from Jude Werra's Liar Index for 2001 (18.53%).

Problem 4

1.

Assuming equal standard deviation, the standard error is 0.8619 and df=485. This gives us:

$$(\bar{x}_1 - \bar{x}_2) \pm t * (pooled\ se(\bar{x}_1 - \bar{x}_2)) =]0.4064, 3.7936[$$

Based on the confidence interval, we can with 95% confidence say that the experimental group will have a mean improvement between 0.4064 and 3.7936 larger than the control group.

Problem 5

1.

Output from JASP:

Contingency Tables*Contingency Tables*

Interest		Education		Total
		Undergraduate	Graduate	
very interested	Count	58.000	30.000	88.000
	Expected count	66.494	21.506	88.000
	% within column	19.142 %	30.612 %	21.945 %
interested	Count	57.000	19.000	76.000
	Expected count	57.426	18.574	76.000
	% within column	18.812 %	19.388 %	18.953 %
somewhat interested	Count	86.000	35.000	121.000
	Expected count	91.429	29.571	121.000
	% within column	28.383 %	35.714 %	30.175 %
not at all interested	Count	102.000	14.000	116.000
	Expected count	87.651	28.349	116.000
	% within column	33.663 %	14.286 %	28.928 %
Total	Count	303.000	98.000	401.000
	Expected count	303.000	98.000	401.000
	% within column	100.000 %	100.000 %	100.000 %

Chi-Squared Tests

	Value	df	p
X ²	15.384	3	0.002
N	401		

Note. Continuity correction is available only for 2x2 tables.

Pearson's χ^2 -test statistic equals 15.384. Under the null, it is approximately χ^2 - distributed with 3 degrees of freedom. This gives us a p-value of 0.002, so we reject the null hypothesis and conclude that interest in AI tools depends on whether the student is an undergraduate or graduate.

Looking at the column proportions, we see that the graduates are generally more interested in AI tools than undergraduates. 19.142% of undergraduates are very interested in AI tools, which is lower than the 30.612% of graduates that are very interested in AI tools.

Furthermore, 33.663% of undergraduates are not interested at all in AI tools, which is higher than the 14.286% of graduates that are not interested at all in AI tools.

Problem 6

1.

Output from JASP:

ANOVA

ANOVA - usage

Cases	Sum of Squares	df	Mean Square	F	p
weather	27164.457	2	13582.229	40.066	< .001
Residuals	246789.082	728	338.996		

Note. Type III Sum of Squares

Descriptives

Descriptives - usage

weather	N	Mean	SD	SE	Coefficient of variation
clear	463	48.768	18.795	0.873	0.385
cloudy	247	40.359	18.091	1.151	0.448
rainy	21	18.033	12.403	2.707	0.688

Post Hoc Tests**Standard (HSD)**

Post Hoc Comparisons - weather

		Mean Difference	SE	df	t	P _{Tukey}
clear	cloudy	8.409	1.451	728	5.797	< .001
	rainy	30.735	4.108	728	7.482	< .001
cloudy	rainy	22.326	4.185	728	5.335	< .001

Note. P-value adjusted for comparing a family of 3 estimates.

Letter-Based Grouping - weather

weather	Letter
clear	c
cloudy	b
rainy	a

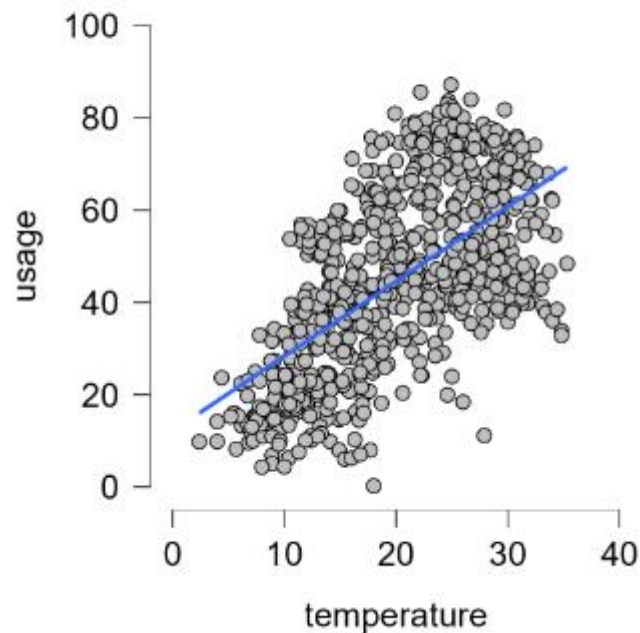
Note. If two or more means share the same grouping symbol, then we cannot show them to be different, but we also did not show them to be the same.

The one-way Anova test statistic for testing whether the means differ equals 40.066. Under the null hypothesis it is F-distributed with 2 and 728 degrees of freedom, giving a p-value of <0.001. Thus, we reject the null hypothesis of no difference, and we may conclude that at least one of the weather types differ from the other two.

Tukey's procedure shows that all three weather types differ, as none of them "share a letter": The expected bikes rented is the smallest for rainy weather and largest for clear weather.

2.

Output from JASP:

Scatter Plots**temperature - usage***Model Summary - usage*

Model	R	R ²	Adjusted R ²	RMSE
M ₀	0.000	0.000	0.000	19.372
M ₁	0.627	0.394	0.393	15.094

Note. M₁ includes temperature

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
M ₁	Regression	107864.603	1	107864.603	473.441	< .001
	Residual	166088.936	729	227.831		
	Total	273953.539	730			

Note. M₁ includes temperature

Note. The intercept model is omitted, as no meaningful information can be shown.

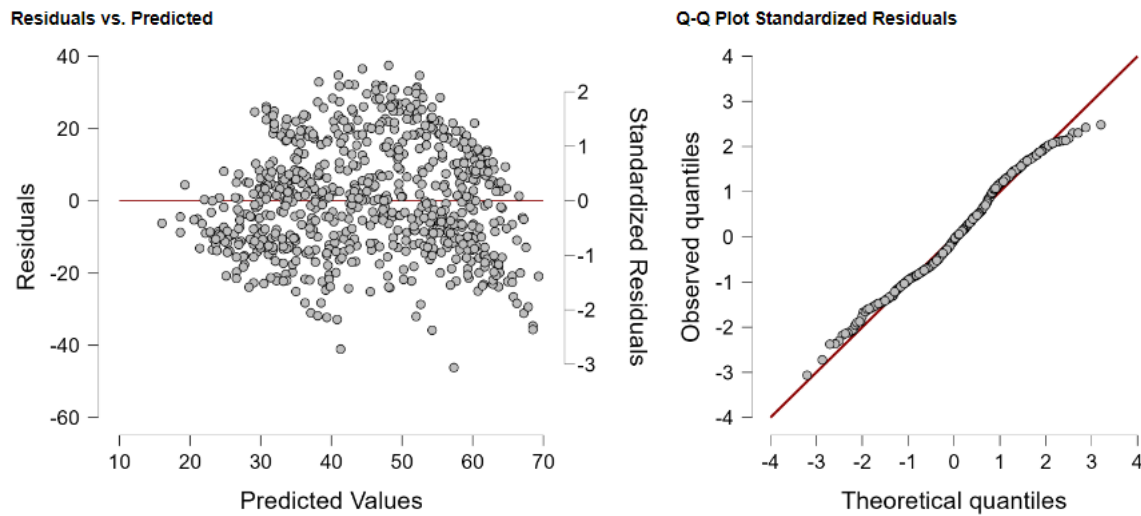
Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p	95% CI	
							Lower	Upper
M ₀	(Intercept)	45.043	0.717		62.866	< .001	43.637	46.450
M ₁	(Intercept)	12.149	1.612		7.539	< .001	8.985	15.313
	temperature	1.620	0.074	0.627	21.759	< .001	1.473	1.766

The scatterplot indicates a relatively strong positive linear relationship between number of bikes rented and temperature, with scatter of points having almost similar variability at different x-values. One might argue that there is a slight trend of increasing variation when the temperature increases.

The linear relationship shows that the higher the temperature, the greater number of bikes are rented.

Output from JASP:



For the linear regression, the QQ-plot looks very nice, with the points being sufficiently on the line.

The residual plot fluctuates almost randomly about the horizontal line, but with a slight trend of change in variation, as the temperature increases, which we also saw in the scatterplot.

Regression equation:

$$\widehat{bikes\ rented} = 12.149 + 1.62 * temperature$$

95%-confidence interval for the slope:

$$]1.473, 1.766[$$

3.

Output from JASP:

ANCOVA - usage

Cases	Sum of Squares	df	Mean Square	F	p
weather	16728.527	2	8364.264	40.712	< .001
temperature	97428.673	1	97428.673	474.226	< .001
Residuals	149360.409	727	205.448		

Note. Type III Sum of Squares

The test on whether the number of bikes rented depends on the weather when the effect of the temperature is taken into account is f-distributed with degrees of freedom equal to 2 and 727 and a f-statistic of 40.712. This gives us a p-value of <0.001. Thus, we reject the null hypothesis of no difference, and we may conclude that at least one of the weather types differ from the other two, when the effect of temperature is taken into account.

In question 1, we reject the null hypothesis of no difference and conclude that at least one of the weather types differ from the other two. In question 3, we also reject the null hypothesis of no difference and conclude that at least one of the weather types differ from the other two, when the effect of temperature is taken into account. Thus, the inclusion of temperature does not qualitatively change the conclusion from question 1.