

Problem 1

1. $0.84 \cdot 0.2 + 0.22 \cdot 0.16 = 0.2032$

The probability that a Norwegian bicyclist sustains a head injury in a bicycle accident is 20.32%.

2. The probability that she sustained it in a collision is approximately 17.32%.

3. The expected length of rehabilitation after a bicycle accident is 0.8832 months.

Problem 2

1. $1978/3120 = 0.633974359$

$$Se = \sqrt{\frac{0.633974359(1-0.633974359)}{3120}} = 0.00862411$$

$$0.633974359 \pm 0.00862411 * 1.96 = 0.617071103; 0.650877615$$

The confidence interval for the probability of agreeing in the survey goes from 0.61707 to 0.65088, (using the formula in the book and the table standard normal cumulative probabilities; 97.5% quantile(normal distribution))

From this interval we can say with 95% certainty that the danes that agree in the survey is between 61.707% and 65.0878%. We can therefore with 95% certainty say that the majority of danes agree with this statement.

Problem 3

1. The liar index shows that the probability of lying in 1996 is 0.14 and 0.1852941 in 2001. However by testing these numbers, using r-console (normal distribution, exact value of the 97.5% quantile), we get a confidence interval for the difference in probabilities that include 0, (-0.1022793; 0.0116911), therefore we cannot conclude that there has been a significant difference in the index. Using r-console; we also get a $X^2 = 2.3841$ (df=1, chi-squared-distribution), with a non-significant p-value 0.1226

meaning that we cannot conclude that there has been a significant difference in the index.

Problem 4

1.

$$3.9 - 1.8 \pm 0.86188 * 1.969734 = 2.1 \pm 1.69767899 = 0.402321; 3.79767$$

Using R-console to get the t^* , with $df=244$ and t-distribution.

Using the formula in the book; the mean-difference 95% confidence interval is approximately 0,402321;3,79767. Therefore we can with 95% confidence say that the experimental group has improved between 0,402321 and 3,79767 units more than the control group.

Problem 5

1.

Chi-Squared Tests

	Value	df	p
X^2	15.384	3	0.002
N	401		

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

Using Jasp we get a X^2 value of 15.384 (chisquare-distribution, $df=3$) with a significant p-value, 0.002, therefore we can reject null (independence), the values indicates that there is in fact a difference between Graduates/undergraduates interest in getting AI-guidance.

From the table below we can see that graduates tend to be more interested than Undergraduates.

Contingency Tables

interest		Undergraduate/graduate		Total
		Graduate	Undergraduate	
very interested	Count	30.000	58.000	88.000
	Standardized residuals	2.385	-2.385	
interested	Count	19.000	57.000	76.000
	Standardized residuals	0.126	-0.126	
somewhat interested	Count	35.000	86.000	121.000
	Standardized residuals	1.374	-1.374	
Not at all interested	Count	14.000	102.000	116.000
	Standardized residuals	-3.678	3.678	
Total	Count	98.000	303.000	401.000

Problem 6

1.

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

Using jasp I get a F-value= 40.066 (F-distribution, df=2 , 728) with a significant p-value<0.001, we can therefore reject null and assume that there is in fact an effect of weather on the number of rented bikes.

Letter-Based Grouping - weather ▼

weather	Letter
clear	c
cloudy	b
rainy	a

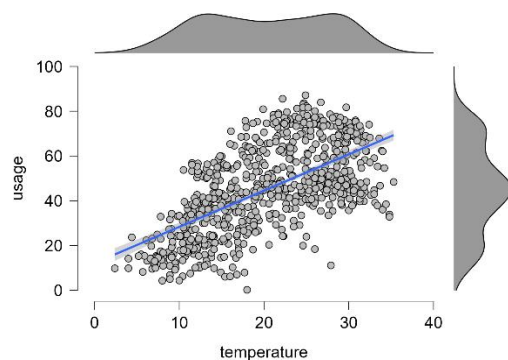
Using letter-based grouping table we can see that all weather types' effects differ significantly from one another.

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

From Post-Hoc comparisons- weather, we can see that clear weather gives the most rentals, and cloudy weather generates more rentals than rainy weather.

2.

temperature - usage

The higher temperature, the more rentals.

Using linear regression in Jasp we get a regression equation: $\text{Rentals} = 1.62 \cdot \text{temperature} + 12.149$.

Using jasp (normal distribution, exact number of the 97.5% quantile); The 95% confidence interval for the slope is 1.473;1.766.

3.

The effect of weather is still significant when taking the effect of temperature into account.

ANCOVA - usage

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

The effect of weather, $F=40.712$ and a significant p-value, indicates that the effect of weather is still significant.

The “order” remains the same; clear weather attracts most rentals, and cloudy attracts more than rainy weather

By comparing the marginal means we can see that when accounting for temperature, all of the marginal means (clear, rainy, cloudy) increases → higher temperature increases rentals in every weather condition.

The effect of weather*temperature is not significant, meaning that they are not affecting each other.

ANCOVA - usage

Cases	Sum of Squares	df	Mean Square	F	p
temperature	11414.971	1	11414.971	55.549	< .001
weather	2565.501	2	1282.751	6.242	0.002
weather * temperature	377.230	2	188.615	0.918	0.400
Residuals	148983.179	725	205.494		

Note. Type III Sum of Squares