Copenhagen Business School

Exam in Statistics

HA-IB 1st year
BSc International Shipping and Trade 1st year

Tuesday 19 October 2021 9:00-13:00

Solutions

Problem 1

1. The probability of being Rhesus positive is

$$\begin{aligned} &P(\text{Rhesus positive } \textbf{and} \text{ antigens}) + P(\text{Rhesus positive } \textbf{and} \text{ no antigens}) \\ =&0.49 + 0.35 = 0.84 \end{aligned}$$

   or 84%.

2. The probability of having antigens if you are Rhesus positive is

$$\begin{aligned} P(\text{antigens}|\text{Rhesus positive}) &= \frac{P(\text{antigens } \textbf{and} \text{ Rhesus positive})}{P(\text{Rhesus positive})} \\ &= \frac{0.49}{0.84} = 0.583333 \end{aligned}$$

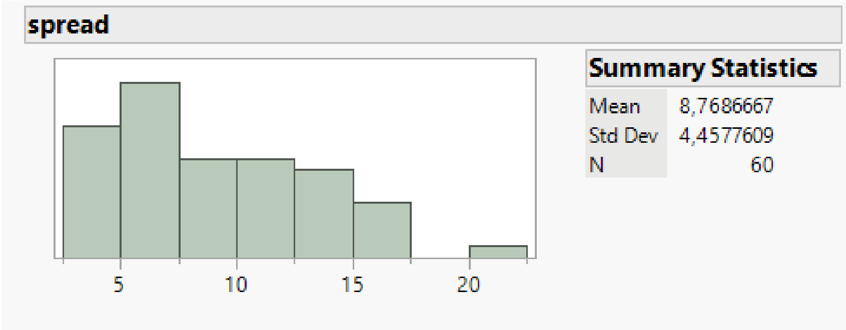
   or 58.3%.

3. The number, X, of Rhesus positive with antigens of the five Danes is binomially distributed with size parameter 5 and success probability 0.49. Thus

$$\begin{aligned} P(X \geq 3) &= P(X = 3) + P(X = 4) + P(X = 5) \\ &= \binom{5}{3} 0.49^3 (1 - 0.49)^2 + \binom{5}{4} 0.49^4 (1 - 0.49) + \binom{5}{5} 0.49^5 \\ &= 0.481255 \end{aligned}$$

   or 48.5%.

Problem 2



The histogram (above, left) indicates that the distribution of the speads is right-skewed. It seems that there may be a single outlier (20.13), but hardly an extreme one.

The mean and standard deviation of the measurements are 8.768667 and 4.457761 respectively. Hence the empirical rule would suggest that

- 68% of the meaurements are between 4.310906 and 13.226428 inches
- 95% of the measurements are between -0.1468551 and 17.6841884 inches
- almost all of the measurements are between -4.604616 and 22.141949 inches

Obviously, none of the spread measurements *can* be negative.

Problem 3

The test statistic for testing the hypothesis $H_0 : p = 0.53$ equals

$$\frac{58000/135000 - 0.53}{\sqrt{0.53(1-0.53)/135000}} = -73.88998$$

Under the null, this is approximately standard normally distributed giving us a p-value indistinguishable from 0. Hence, we reject the null hypothesis and conclude that the probability that student, who defaults on their student loan, is a woman, is smaller ($58/135 = 0.43$) than 0.53.

Using JMP instead, the test statistic equals 5459.729. Under the null, it is approximately χ^2-distributed with 1 degree of freedom. The p-value and the conclusion is as above.

Test Probabilities

Level	Estim Prob	Hypoth Prob
no	0,57037	0,47
yes	0,42963	0,53

Test	ChiSquare	DF	Prob>Chisq
Likelihood Ratio	5452,592	1	<,0001*
Pearson	5459,729	1	<,0001*

Method: Fix hypothesized values, rescale omitted

Problem 4

A 95% confidence interval is given by

$$\frac{79}{302} - \frac{104}{639} \pm 1.96\sqrt{\frac{79}{302} \cdot \left(1 - \frac{79}{302}\right)/302 + \frac{104}{639} \cdot \left(1 - \frac{104}{639}\right)/639} =]0.0415959; 0.15607421[$$

(1.96 is the 97.5%-quantile in the standard normal distribution). Thus the probability that a female professor is hired is with 95% confidence between 4.16 and 15.61 percentage points higher than the corresponding probability for a male professor. In particular, it is significantly higher (at a 5%-level) for women.

The conlusion is not that more women than men are hired as professors; clearly more men were hired. It does not even imply that a women will have larger probability of being hired as a full professor if a man also applies.

Problem 5

Tests

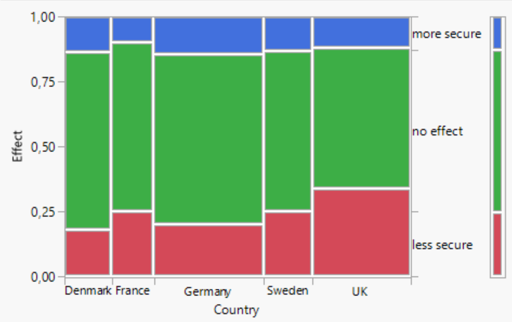
N	DF	-LogLike	RSquare (U)
4978	8	50,547362	0,0112

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	101,095	<,0001*
Pearson	102,373	<,0001*

The Pearson χ^2-test statistic for testing the hypothesis of same distribution of job security in the five European countries equals 102.373. Under the null it is χ^2-distributed with 8 degrees of freedom, which leads to a p-value below 0.0001. Hence, I reject the null hypothesis and conclude that the corona virus outbreak has had different effects on job security in the five countries.

From the mosaic plot and or the "row percentages" below it seems that compared to the other countries job security has been worse in the UK and better in Germany and Denmark.

Mosaic Plot



Contingency Table

Effect

Country: Count / Row %	less secure	no effect	more secure	Total
Denmark	124	464	93	681
	18,21	68,14	13,66	
France	151	393	60	604
	25,00	65,07	9,93	
Germany	316	1038	226	1580
	20,00	65,70	14,30	
Sweden	175	432	93	700
	25,00	61,71	13,29	
UK	478	769	166	1413
	33,83	54,42	11,75	
Total	1244	3096	638	4978

Problem 6

1. Assuming that the two standard deviations are equal (4.26≈4.69) the 90%-confidence interval for the difference between the mean spreads is

$$0.680 \pm 1.67155 \cdot 1.1574 =]-1.2506; 2.6186[.$$

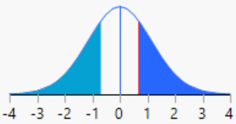
using the 95% quantile from the t-distribution with 58 degrees of freedom. As the confidence interval contains 0, there is no evidence to support that the mean spread differs between the types of cartridges (at a 10%-significance level, and hence neither at lower levels).

Pooled t Test

b-a

Assuming equal variances

Difference	0,6840	t Ratio	0,590985
Std Err Dif	1,1574	DF	58
Upper CL Dif	3,0008	Prob > |t|	0,5568
Lower CL Dif	-1,6328	Prob > t	0,2784
Confidence	0,95	Prob < t	0,7216



2. The regression equation is

$$\widehat{\text{spread}} = -0.07808 + 0.29489 \cdot \text{distance}$$

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0,078083	0,451343	-0,17	0,8633
distance	0,2948917	0,013609	21,67	<,0001*

The test statistic for testing the hypothesis of proportionality –$H_0 : \alpha = 0$– equals -0.173. Under the null it is t-distributed with 58 degrees of freedom giving us a p-value of 86.3%. Thus we cannot reject the null hypothesis, and it is therefore not unreasonable to assume, that the expected spread is indeed proportional to the distance to the target.

3. The test statistic for testing the hypothesis that the effect of distance on the spread does not depend on the type of cartridge equals 0.3894. Under the null it is F-distributed with 1 and 56 degrees of freedom, gving us a p-value of 53.51%. The null is therefore not rejected and it seems that the effect of distance on the spread of the pellets does not depend on the type of cartridge.

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
type	1	1	7,0178	3,2470	0,0769
distance	1	1	1043,5331	482,8233	<,0001*
type*distance	1	1	0,8417	0,3894	0,5351