Statistics Exam

15th October, 2019

Problem 1

In Belgium, 21% of all twins are left-handed. Overall, 10% are twins and 3.4% are lefthanded.

1. **Calculate the probability that a person from Belgium is both left-handed and a twin.**

Let a denote the probability that a person is a twin

Let b denote the probability that a person is lefthanded

P(A and B)= P(A) * P(B|A) = 0,1*0,21= 0,021

The probability that a person from Belgium is both left-handed and a twin us 0,021

2. Imagine that you have just met a left-handed person from Belgium. What is the probability that this person is a twin?

P(A|B) = $\frac{P(B|A)*P(A)}{P(B)} = \frac{0{,}21*0{,}1}{0{,}034} = 0{,}6176$

The probability is 0,6176

CHECK UP ON THIS AS EXCEL SAYS DIFFERENT:

Problem 2

The battery life for a new iPad is normally distributed with a mean of 674 minutes and a standard deviation of 42 minutes.

1. What is the probability that the battery on a new iPad lasts less than 10 hours (600 minutes)?

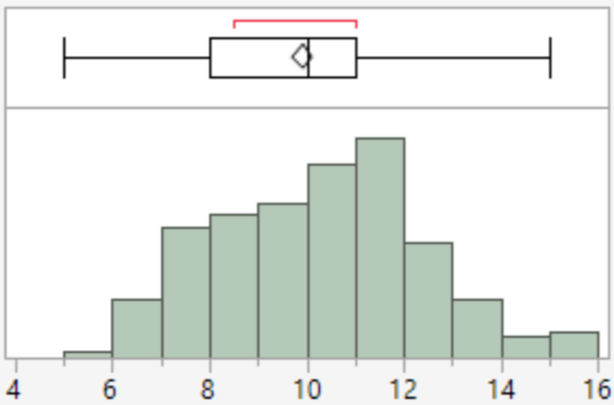
Through calculator, we find that the probability is 0,039

Problem 3

The data file contains shoe sizes (size) for 407 American college students.

1. **Make a histogram for the shoe sizes, comment briefly on its shape and report a fivenumber summary for the shoe sizes.**

size



Quantiles

100.0%	maximum	15
99.5%		15
97.5%		14
90.0%		12,5
75.0%	quartile	11
50.0%	median	10
25.0%	quartile	8
10.0%		7
2.5%		6
0.5%		5,52
0.0%	minimum	5

Through JMP, we get the histogram as shown above that suggests that the distribution shape is slightly left skewed. We also get the five number summary from JMP, where

Minimum value: 5
First quartile: 8
Median: 10
Third quartile: 11
Maximum value: 15

Problem 4

People with periodic limb movement disorder (PLMD) experience repetitive movements, typically in the lower limbs, that occur about every 20-40 seconds, while they sleep. PLMD can be double-sided (movement of both right and left limbs) or one-sided. In a study performed at the University of Toledo, 58 of 84 right-handed test persons, all suffering from PLDM, had doubled-sided PLDM, whereas the rest had one-sided PLDM. For the 16 lefthanded test persons, 15 had doubled-sided PLDM and only 1 had one-sided PLDM disorder.

1. **Test whether the probability of suffering from double-sided PLDM depends on whether you are left-handed or right-handed? If it does, briefly describe how.**

Proportion of right-handed that has double-sided PLDM: $\frac{58}{84} = 0{,}690476$
Proportion of left-handed that has double-sided PLDM: $\frac{15}{16} = 0{,}9375$

Difference in proportions: 0,690476-0,9375 = -0,247024

The test statistic is -2,0398. Under the null hypothesis, the test statistic is approximately normally distributed. Using a calculator, we get that the p-value is 0,04137. We therefore reject the null hypothesis and may conclude that the probability of suffering from double-sided PLDM does depend on whether you are left-handed or right-handed where the probability of left-handed suffering from double-sided PLDM is higher than for right-handed which e.g. can be seen in the difference in proportions which is negative.

Problem 5

One hundred and thirty five (135) children where presented with two pictures of a man, one in which he is bearded and one where he is clean-shaven. The children were asked which of the two looked stronger and which looked most like a dad. The results are given in the table below.

		Who looks stronger?	
		clean-shaven	bearded
Who looks most like a dad?	clean-shaven	26	53
	bearded	16	40

1. **Estimate the probability that a child perceives the bearded man to be stronger than the clean shaven man, and calculate a 95% confidence interval for this probability.**

The probability that a child perceives the bearded man to be stronger than the clean shaven man:

$\frac{53+40}{53+40+16+26} = 0{,}6889$

With a standard error of 0,039844 (calculated by Excel)

The 95% confidence is then [0,610796; 0,766982].

2. **Test whether the probability that a child perceives a bearded man as stronger differs from the probability that a child perceives a bearded man as more dad-like. If the probabilities differ, briefly describe how.**

The probability that a child perceives the bearded man to be stronger than the clean shaven man:

$\frac{53+40}{53+40+16+26} = 0{,}6889$

The probability that a child perceives a bearded man as more dad-like than the clean shaven man:

$\frac{16+40}{16+40+53+26} = 0{,}4148$

Difference in probabilities: 0,6889-0,4148=0,2741

The test statistic is 4,5279. Under the null hypothesis, the test statistic is approximately normally distributed. Using a calculator, we get that the p-value is less than 0,0001. We therefore reject the null hypothesis and may conclude that the probability that a child perceives a bearded man as stronger does indeed differ from the probability that a child perceives a bearded man as more dad-like.
As the difference in probabilities is positive, the probability of that a child perceives a bearded man as stronger is bigger than the probability that a child perceives a bearded man as more dad-like.

Problem 6

The typical American retirement accounts are either tax-sheltered annuities (“TSA”) or 401(k)s. Below are some summary statistics for annual contributions (USD) to such retirement accounts for a sample of adults from North Dakota.

	Mean	StdDev	n
TSA	2119.70	709.70	15
401(k)	1777.70	593.90	15

1. **Test whether the annual contribution differs between the two account types. If it does, briefly explain how.**

The difference in means is 342. We assume same standard deviations in the two groups (as the ratio of the larger standard deviation to the smaller is less than 2 ($\frac{709,7}{593,9}$ = 1,195)).

The standard error is 238,9408992

With 28 degrees of freedom

With the help of a calculator, we then that get the test statistic 1,4313 which is t-distrubuted under the null and that the p-value is 0,1634. Thus, we do not reject the null hypothesis as we not do not have enough evidence to prove that the contributions differ. We then conclude that contributions very well might not differ.

Problem 7

The data file contains shoe sizes (size), heights in inches (height) and information about sex (sex) for 407 American college students.

1. **Fit a model describing how shoe size depends on height and sex, and test in this model whether the effect of height differs between sexes; if it does, briefly describe how. Still in this model, predict the shoe size of a 70 inch tall male and provide a 95% prediction interval.**

To test whether the effect of height on shoe size differs between the sexes, I include an interaction term in our model. The output is shown below:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	1366,0165	455,339	505,0838
Error	403	363,3091	0,902	Prob > F
C. Total	406	1729,3256		<,0001*

Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-13,81311	1,451839	-9,51	<,0001*
sex[F]	-0,990937	0,133268	-7,44	<,0001*
height	0,3530954	0,020393	17,31	<,0001*
sex[F]*(height-68,4238)	-0,016106	0,031503	-0,51	0,6095

The test statistic for the hypothesis of no interaction is -0,51, it is t-distributed under the null hypothesis with 403 degrees of freedom. The p-value is then 0,6095. Hence, we do not reject and conclude that there is no indication that the effect of height on shoe size differs between sexes.

	size	sex	height	Pred Formula size	Lower 95% Indiv size	Upper 95% Indiv size
408	•	M	70	10,903570208	9,0322400857	12,77490033

Using JMP, we find that the predicted shoe size of a 70 inch tall male is 10,9 (roughly 11), with a 95% prediction interval of [9,0322; 12,7749].

2. **Fit a model describing how shoe size depends on height and sex assuming no interaction between the two explanatory variables. Test whether shoe size depends on height when sex is taken into account. Find a 95%-confidence interval for the effect of sex in this model. Explain what this confidence interval tells you as well as exactly what this parameter measures (how it can be interpreted).**

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	1365,7808	682,890	758,8825
Error	404	363,5447	0,900	Prob > F
C. Total	406	1729,3256		<,0001*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
height	1	1	447,59913	497,4080	<,0001*
sex	1	1	50,12024	55,6976	<,0001*

Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	-13,33308	1,106341	-12,05	<,0001*	-15,50798	-11,15817
height	0,3463461	0,015529	22,30	<,0001*	0,3158176	0,3768745
sex[F]	-0,979445	0,131239	-7,46	<,0001*	-1,237441	-0,721449

We get the output as shown above through JMP. We see that

The test statistic for testing the effect of height when sex is taken into account is 497,4080 and is F-distributed with 1 and 404 degrees of freedom, which gives us a p-value of less than 0,001. Thus we reject the null hypothesis, and can conclude that the shoe size of men and women at different heights differ.

The 95%-confidence interval for the effect of sex is [-1,2374; -0,7214], which tells me that with a 95% confidence, the shoe size for a woman will be between -1,2374 to -0,7214 times smaller than for a man when height is the same.

In general it can then be interpreted as that the predicted shoe size is smaller for women than for men if both has the same height.