Statistics Exam October 2019

Problem 1

$$P(twin) = 0.10$$

$$P(left\ handed) = 0.034$$

$$P(left\ handed|twin) = 0.21$$

1.

$$P(twin\ and\ lefthanded) = P(twin) * P(lefthanded|twin) = 0.10 * 0.21 = 0.021$$

The probability that a person from Belgium is both left-handed and a twin **= 0.021**

2.

$$P(twin|left\ handed) = \frac{P(left\ handed|twin)}{P(left\ handed)} * P(twin) = \frac{0.21}{0.034} * 0.10 = 0.6176$$

Using Bayes Theorem, The probability that left-handed person is a twin = **0.6176**

Problem 2

μ = 674

σ = 42

In order to calculate the probability that the battery life on a new iPad lasts less than 10 hours (600), we would have to find out firstly how many standard deviations an observation of <600 would fall away from the mean:

$$z = \frac{X - \mu}{\sigma}$$

$$z = \frac{600 - 674}{42} = -1.761$$

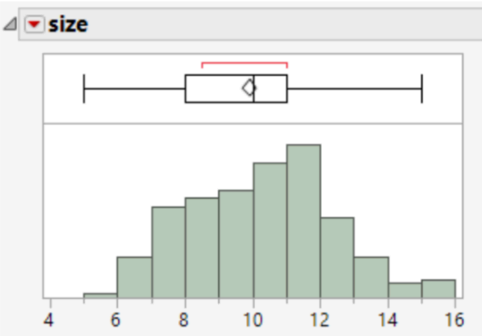
An observation with a battery life of less than 600 falls -1.761 standard deviations from the mean:

$$\mu - 1.761\sigma$$

A z-score of -1.761 corresponds to a cumulative probability of 0.0392 (Obtained from tables) under the standard normal curve, normally distributed, to the left of z.

This means, that the probability that the battery on a new iPad lasts less than 10 hours = **0.0392** $\approx$ **3,92%**

Problem 3



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

Minimum value: 5

First quantile Q1: 6

Median: 10

Third quantile Q3: 11

Maximum: 15

The mean of the distribution = 9.915 and the standard deviation = 2.063

The histogram shows a slightly left-skewed distribution. Additionally, it seems to be characterized by bimodality (one mode between 10 and 11 and one between 11 and 12)

Thus the middle 50% of the shoe sizes are between 6 and 11, with the median (10) not too far from the mean at 9.91. Both the maximum and minimum are equally far away from the middle 50%, which we can see from the distribution.

Problem 4

Large sample t-test comparing left-handed and right-handed proportions suffering from double-sided PLDM:

$H_0: p_1 = p_2$

$H_a: p_1 \neq p_2$

$p_1 = right\ handed = \frac{58}{84} = 0.69$

$p_2 = left\ handed = \frac{16}{15} = 0.93$

Estimate for difference = $(\widehat{p_1} - \widehat{p_2}) = -0.24701$

Pooled estimate = $\frac{n_1 * \widehat{p_1} + n_2 * \hat{p}_2}{n_1 + n_2} = 0.73$

Standard error of the difference = $\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}} = 0.07878$

$$Z = \frac{(\widehat{p_1} - \widehat{p_2}) - 0}{\sqrt{\hat{p}(1-\hat{p}) * (\frac{1}{n_1} + \frac{1}{n_2})}} = -2.040$$

Conclusion:

The test statistic of -2.040 is approximately standard normal distributed, with 0 degrees of freedom.

P- value of 0.04137 (obtained from JMP)

As the p-value is below the 5% significance level, there is strong evidence against the null-hypothesis of $p_1 = p_2$

We conclude that there is evidence, that the probability of suffering from double-sided PLDM depends on whether you are left-handed or right-handed. The probability of suffering from double-sided PLDM is much higher when you are a left-handed sufferer.

Problem 5

1.

Probability that a child perceives the bearded man to be stronger:

$$p = \frac{53}{135} = 0.3925$$

$$se = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.3925(1-0.3915)}{135}} = 0.04203$$

A 95% confidence interval is given by

$$\hat{p} \pm 1.96 \times \textit{the standard error}$$

$$\left]\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}};\right[$$

$$0.39259 \pm 1.96 \times 0.04203$$

$$\mathbf{]0.310; 0.4749[}$$

2.

$H_0: p_1 = p_2$

$H_a: p_1 \neq p_2$

McNemar's test statistic is:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{Standard\ error\ of\ difference}$$

$$z = \frac{b - c}{\sqrt{b + c}} = \frac{53 - 16}{\sqrt{53 + 16}} = 4.45428$$

P-value = <0.0001

which is normally distributed under the null hypothesis of no difference, giving us a **p-value <0.0001** (using JMP). This means that there is strong evidence against the null hypothesis of $p_1 = p_2$, and we can reject it on a 5% significance level.

Conclusion:

We conclude, that the test provides strong evidence, that more children perceive a bearded man as strong compared to a bearded man being child-like.

Problem 6:

Two sample t-test:

$$H_0: \bar{x}_1 = \bar{x}_2$$

$$H_a: \bar{x}_1 \neq \bar{x}_2$$

Assuming unequal variances:

Standard error of the difference = $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 239.94$

Degrees of freedom = $\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = 27.1562$

Mean difference = 342

The test statistic for the two-sample t-test is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{standard\ error\ of\ the\ difference} = \frac{342}{239.94} = 1.42535$$

Conclusion: Assuming unequal variances, the t-test provides a test statistic of 1.425. It is approximately t-distributed with 27.15 degrees of freedom.

This provides a P-value of **0.16381** (obtained from JMP)**,** which is above the 5% significance level.

The test does not provide strong evidence against the null-hypothesis of $\bar{x}_1 = \bar{x}_2$. From the test result we can therefore conclude that the annual USD contribution does not differ significantly between the two account types, TSA or 401K.

Problem 7:

1.

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*

Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
height	1	1	1315,6606	1459,394	<,0001*
sex	1	1	50,1202	55,5958	<,0001*
height*sex	1	1	0,2356	0,2614	0,6095

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
height	1	1	432,59639	479,8568	<,0001*
sex	1	1	49,84367	55,2890	<,0001*
height*sex	1	1	0,23563	0,2614	0,6095

The overall F-test for height's effect on shoe size = 497.408, with a P-value of <0.0001.

We see however, from the Sequential test, that there is an insignificant interaction between height and sex. The F-test statistic equals 0.2613 (F-distributed with 1 and 403 degrees of freedom), providing a p-value 0f 0.6095.

Conclusion: We can therefore not reject the null-hypothesis of no independence. Thus, there is an effect of height on shoe size (from what we see on the Effect Tests), but the 'height-effect' effect does not differ between the sexes (seen through the Sequential test).

Shoe size of a 70-inch-tall male:

From the prediction equation (attained in JMP output)

$$shoe\ size = \underline{\underline{\mathbf{10.90357}}}$$

The 95% prediction interval (attained from JMP)

$$]\mathbf{9.0322\ ;\ 12.774}[$$

For the 97.5%-quantile in the t-distribution, with n-1 (406) degrees of freedom, you would obtain a t-quantile of 1.960 (looking at tables, and assuming infinite degrees of freedom).

2.

H_0: That Y is statistically independent of all explanatory variables:

Taking sex into account we see from the JMP output:

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|
Intercept	-13,33308	1,106341	-12,05	<,0001*
height	0,3463461	0,015529	22,30	<,0001*
sex[F]	-0,979445	0,131239	-7,46	<,0001*

The test statistic for testing whether the effect of height when sex is taken into account equals 497.408. It is F-distributed with 1 and 404 degrees of freedom, and we get a p-value of <0.0001. We can therefore reject the null-hypothesis of no dependence.

To conclude, when we take sex into account, the effect of height on shoe size is significant.

95% confidence interval for effect of sex:

The slope parameter for sex = $-\ 0.979445$

Interpretation:

The parameter is a factor, which means that whenever the model is testing for the shoe size of a female, the parameter will become -0.979445. Whenever the model is testing for a male, the parameter will become 0. This means that being female has a more negative effect on the mean shoe size, compared to men; the mean shoe size of women is lower.

The standard error for the parameter = 0.131239

(obtained from JMP)

The slope parameter is t-distributed, with 374 degrees of freedom. At a 95% confidence interval, we obtain a t-quantile of 1.96 (obtained from tables, assuming infinite DFs).

$$b \pm t - quantile \times standard\ error$$

$$]-0.979445 \pm 1.96 \times 0.131239[$$

Our confidence interval for the slope parameter:

$$\mathbf{]-1.2366; -0.7222[}$$

As the confidence interval does not include the null value: we can reject the null hypothesis of no dependence. We conclude, that the impact of female gender on shoe size is significant, as it negatively impacts Y.

When the slope parameter = 0

$$b \pm t - quantile \times standard\ error$$

$$]0 \pm 1.96 \times 0.131239[$$

$$]\mathbf{-0.2573; 0.2573}[$$

When the confidence interval DOES include the null value, cannot reject the null-hypothesis; and there is no evidence for a significant impact of males on shoe size.