**FRAGE 1 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 01**

Which of the following is an example of numeric data?

**Wählen Sie eine Antwort:**

Marital status Gender Color

*Age*

**FRAGE 2 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 01**

Which of the following is a measure of center?

**Wählen Sie eine Antwort:**

*Mode* Range

Standard deviation

First quartile

**FRAGE 3 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 01**

Which of the following is an example of a probability sample?

**Wählen Sie eine Antwort:**

Snowball sample Purposeful sample *Simple random sample* Convenience sample

**FRAGE 4 VON 78**

**MMET02-01\_E\_MC\_mittel/Unit 01**

What is the classic probability of rolling a 3 with a die?

**Wählen Sie eine Antwort:**

3 out of 3

1. *out of 6*
2. out of 3

6 out of 6

**FRAGE 5 VON 78**

**MMET02-01\_E\_MC\_mittel/Unit 01**

What is the classic probability of choosing an ace in a deck of cards?

**Wählen Sie eine Antwort:**

*4 out of 52*

3 out of 26

5 out of 52

4 out of 4

**FRAGE 6 VON 78**

**MMET02-01\_E\_MC\_schwer/Unit 01**

With a mean of 2 and a standard deviation of 3 which is the most likely value of an

observation sampled from a normal distribution?

**Wählen Sie eine Antwort:**

*2*

6

1

3

**FRAGE 7 VON 78**

**MMET02-01\_E\_MC\_schwer/Unit 01**

If every observation in a population has a chance of being selected which is greater

than zero than this is an example of what?

**Wählen Sie eine Antwort:**

Convenience sample *Probability sample* Simple random sample Purposeful sample

**FRAGE 8 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 02**

A 2 x 2 crosstabulation contains what number of variables?

**Wählen Sie eine Antwort:**

Three Four *Two* Eight

**FRAGE 9 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 02**

What is another name for crosstabulation table of crosstab?

**Wählen Sie eine Antwort:**

Summary table Alternate table Pivot table *Contingency table*

**FRAGE 10 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 02**

Which type of test would typically be used when the data is normally distributed?

**Wählen Sie eine Antwort:**

Non-parametric test Live test *Parametric test* Ordinary test

**FRAGE 11 VON 78**

**MMET02-01\_E\_MC\_mittel/Unit 02**

The percentage of variation from knowing x from a y variable in a Pearson Product

Moment correlation of .4 equals …

**Wählen Sie eine Antwort:**

25 %.

4 %.

90 %.

*16 %.*

**FRAGE 12 VON 78**

**MMET02-01\_E\_MC\_mittel/Unit 02**

A perfect Pearson Product Moment correlation would be represented by which number?

**Wählen Sie eine Antwort:**

100

0.5

0

*1*

**FRAGE 13 VON 78**

**MMET02-01\_E\_MC\_schwer/Unit 02**

The distribution of one variable given the levels of the other variable is described

through which of the following in a crosstabulation table?

**Wählen Sie eine Antwort:**

Standard deviation

Linear probability distributions Mode

*Conditional probability distributions*

**FRAGE 14 VON 78**

**MMET02-01\_E\_MC\_schwer/Unit 02**

Which non-parametric test compares the median of the two samples to see if the

difference falls outside of what one would expect from the error of drawing a sample?

**Wählen Sie eine Antwort:**

T-test ANOVA

F-Test

*Wilcoxon signed rank test*

**FRAGE 15 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 03**

... can be derived by recording data from experiments.

**Wählen Sie eine Antwort:**

Everything

New experiments *Random variables* The real truth

**FRAGE 16 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 03**

What is another name for a discrete probability distribution?

**Wählen Sie eine Antwort:**

*Probability mass function* Ordinary function Inverted function Random function

**FRAGE 17 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 03**

What is another name for a normal distribution?

**Wählen Sie eine Antwort:**

Beta distribution *Gaussian distribution* Uniform distribution Gamma distribution

**FRAGE 18 VON 78**

**MMET02-01\_E\_MC\_mittel/Unit 03**

What is one of the unique characteristics of the t-distribution?

**Wählen Sie eine Antwort:**

Uniform distribution *Heavier tails*

Does not resemble a curve

Same as a normal distribution

**FRAGE 19 VON 78**

**MMET02-01\_E\_MC\_mittel/Unit 03**

What is the most commonly used continuous probability distribution?

**Wählen Sie eine Antwort:**

Dirichlet distribution Poisson distribution *Normal distribution* Gamma distribution

**FRAGE 20 VON 78**

**MMET02-01\_E\_MC\_schwer/Unit 03**

What is the pseudonym for the inventor of the t-test?

**Wählen Sie eine Antwort:**

Professor Assistant John Doe *Student*

**FRAGE 21 VON 78**

**MMET02-01\_E\_MC\_schwer/Unit 03**

Which distribution can also approximate the Poisson distribution as well as the binomial

and hypergeometric distribution?

**Wählen Sie eine Antwort:**

General distribution Dirichlet distribution *Normal distribution* Beta distribution

**FRAGE 22 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 04**

A mean value of 5 from a sample represents what type of estimate?

**Wählen Sie eine Antwort:**

Hypothetical *Point* Negative Interval

**FRAGE 23 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 04**

A proportion of .5 derived from a sample is an example of ... estimate.

**Wählen Sie eine Antwort:**

*a point*

an interval a statistical

a percentage

**FRAGE 24 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 04**

When a researcher desires a range of values to estimate a population parameter, they

will utilize what kind of estimate?

**Wählen Sie eine Antwort:**

*Interval* General Best Point

**FRAGE 25 VON 78**

**MMET02-01\_E\_MC\_mittel/Unit 04**

Both point estimates and interval estimates are considered to be ... statistics.

**Wählen Sie eine Antwort:**

multivariate government population *inferential*

**FRAGE 26 VON 78**

**MMET02-01\_E\_MC\_mittel/Unit 04**

What is the goal of obtaining a point estimate?

**Wählen Sie eine Antwort:**

To find a biased estimate of the population

*To achieve the most likely estimate of the population parameter* To prepare for taking a sample

To obtain a range of values around the population parameter

**FRAGE 27 VON 78**

**MMET02-01\_E\_MC\_schwer/Unit 04**

Which estimator better reflects the difference in sample sizes?

**Wählen Sie eine Antwort:**

Normal estimator General estimator *Interval estimator* Point estimator

**FRAGE 28 VON 78**

**MMET02-01\_E\_MC\_schwer/Unit 04**

The mean wages for a firm are somewhere between 45K and 80K Euros is an example

of …

**Wählen Sie eine Antwort:**

a truth interval. a point estimate. a time interval.

*an interval estimate*.

**FRAGE 29 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 05**

Which of the following is an example where one would support the alternate

hypothesis?

**Wählen Sie eine Antwort:**

There is no significant difference between means. The mean cannot be calculated.

The population mean was found.

*There is a significant difference between means*.

**FRAGE 30 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 05**

Which of the following is an assumption of the null hypothesis?

**Wählen Sie eine Antwort:**

The mean was calculated incorrectly.

*There is no significant difference between means.* There is a significant difference between means. There is no correlation present.

**FRAGE 31 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 05**

Which t-test is used to compare the means of two separate samples?

**Wählen Sie eine Antwort:**

*Independent-samples t-test* Paired t-test

F-test

R-squared

**FRAGE 32 VON 78**

**MMET02-01\_E\_MC\_mittel/Unit 05**

Which t-test is used to compare the means of two dependent samples?

**Wählen Sie eine Antwort:**

R-square *Paired t-test* F-test

Independent-samples t-test

**FRAGE 33 VON 78**

**MMET02-01\_E\_MC\_mittel/Unit 05**

What is the probability distribution which is used with a known population standard

deviation?

**Wählen Sie eine Antwort:**

T-distribution F-distribution

Gamma distribution

*Normal distribution*

**FRAGE 34 VON 78**

**MMET02-01\_E\_MC\_schwer/Unit 05**

... are an interval estimate version of hypothesis testing.

**Wählen Sie eine Antwort:**

*Confidence intervals* Regression analyses Clinical trials Analyses of variance

**FRAGE 35 VON 78**

**MMET02-01\_E\_MC\_schwer/Unit 05**

We ... the null hypothesis when the p-value is ≥ alpha.

**Wählen Sie eine Antwort:**

always accept *fail to reject* restate

do nothing when

**FRAGE 36 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 06**

Which statistic is typically used to evaluate the performance of a regression equation?

**Wählen Sie eine Antwort:**

Correlation coefficient Y-value

Multiple comparison tests

*R-Squared*

**FRAGE 37 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 06**

Most p values in a regression equation should be less than what value?

**Wählen Sie eine Antwort:**

0.95

0.9

*0.05*

0.1

**FRAGE 38 VON 78**

**MMET02-01\_E\_MC\_leicht/Unit 06**

A simple linear regression has the following characteristic?

**Wählen Sie eine Antwort:**

Multiple dependent variables *Single independent variable* Categorical dependent variable Multiple independent variables

**FRAGE 39 VON 78**

**MMET02-01\_E\_MC\_mittel/Unit 06**

Which term in a regression analysis represents the value of y when x = 0?

**Wählen Sie eine Antwort:**

Parameter Beta weight Error *Intercept*

**FRAGE 40 VON 78**

**MMET02-01\_E\_MC\_mittel/Unit 06**

... is the attempt to use a regression equation to predict values outside of the original

range when the equation was developed.

**Wählen Sie eine Antwort:**

Summation

Non-linear prediction *Extrapolation* Outside estimation

**FRAGE 41 VON 78**

**MMET02-01\_E\_MC\_schwer/Unit 06**

Which type of variable may be present whenever the relationship between two variables

is significantly affected by the presence of a third variable which has **not** been included in the modeling effort?

**Wählen Sie eine Antwort:**

Supporting variable *Lurking variable* Independent variable Dependent variable

**FRAGE 42 VON 78**

**MMET02-01\_E\_MC\_schwer/Unit 06**

Which type of analysis will help us determine above how far away the predicted data

points are from the actual data points?

**Wählen Sie eine Antwort:**

Post hoc analysis *Residual analysis* Regression analysis Sensitivity analysis

**FRAGE 43 VON 78**

**MMET02-01\_E\_Offen\_leicht/Unit 01**

Explain the three main measures of center.

The three main measures of center are the mean, median, and mode. (2 points) The mean represents the arithmetic average. (2 points) The median represents the middle value when all values are ranked, and the mode represents the most frequent value in a distribution. (2 points)

**FRAGE 44 VON 78**

**MMET02-01\_E\_Offen\_leicht/Unit 01**

Explain why a convenience sample is considered to be a non-probability sample.

A convenience sample is not chosen at random (3 points) and therefore it is considered to be a non-probability sample. (3 points)

**FRAGE 45 VON 78**

**MMET02-01\_E\_Offen\_mittel/Unit 01**

Describe the three different types of probability.

Classical probability is the theoretical probability of process. (2 points) Empirical probability is the actual probability calculated as the result of an experiment (2 points) and judgmental probability is a probability created from a human judgment. (2 points)

**FRAGE 46 VON 78**

**MMET02-01\_E\_Offen\_mittel/Unit 01**

Explain why classical probability is different than empirical probability on the same process.

Empirical probability can have random variation. (3 points) However, on repeated attempts, the empirical probability of events should approach the classical probability. (3 points)

**FRAGE 47 VON 78**

**MMET02-01\_E\_Offen\_schwer\_F1/Unit 01**

Explain what effect has a random, normal distribution on the three measurements of center.

Explain which of them would be the best choice to describe random data from a normal distribution and why.

If the data are at chosen at random from a normal distribution, the mean, median, and mode would be approximately equal. (3 points) Therefore, one could choose all three measures of center (mean, media, and mode) (3 points) if the data was entirely at random and from a normal distribution (3 points). However, the best choice in this case would usually be the mean (3 points) because it uses all of the values in the data in its calculation and is therefore considered to be the most stable measure of central tendency. (3 points) It's also the most useful for future analyses, as many more advanced statistical techniques incorporate the mean. (3 points)

**FRAGE 48 VON 78**

**MMET02-01\_E\_Offen\_schwer\_F1/Unit 01**

Describe two advantages and challenges each with using a convenience sample. Why is it used so often by researchers?

A convenience has the advantage of being "convenient" (3 points) and easy to use in a variety of settings (3 points). In fact, it is often used by researchers as a quick way to gain insights on a particular topic. (3 points) However, a convenience sample is not a probability sample where all elements of a population have a probability of being selected which is greater than 0 (3 points). Therefore, convenience samples can contain biases (3 points) which prevent researchers who use it from generalizing about the entire population. (3 points)

**FRAGE 49 VON 78**

**MMET02-01\_E\_Offen\_leicht/Unit 02**

Explain what could it mean if there was a very high p-value when testing the difference between two means.

What effect does that have on the H0-Hypothesis?

The high p-value means that there is no statistically significant difference between the two means. (3 points) Therefore, one would not reject the null hypothesis. (3 points)

**FRAGE 50 VON 78**

**MMET02-01\_E\_Offen\_leicht/Unit 02**

Explain the conditions when would one choose to use a non-parametric statistical test.

When the data is not normally distributed (3 points) and when the assumptions for a parametric statistical test will not hold. (3 points)

**FRAGE 51 VON 78**

**MMET02-01\_E\_Offen\_mittel/Unit 02**

Explain the joint probability distribution in a crosstab and how to calculate it.

The joint probability distribution in a crosstab describes the proportion of the subjects jointly classified by a category of X and a category of Y. (3 points) Dividing the cells of the crosstab by the total provides the joint distribution. (3 points)

**FRAGE 52 VON 78**

**MMET02-01\_E\_Offen\_mittel/Unit 02**

Explain why researchers generally prefer to use the correlation coefficient over the covariance.

The correlation coefficient is very closely related to covariance but has some useful additional features, such as remaining unaffected by the change in scale, dimensions, and location, (3 points) and can also be used to compare two pairs of variables across different domains (3 points). Also possible (3 points): The correlation coefficient provides knowledge about the strength of the correlation.

**FRAGE 53 VON 78**

**MMET02-01\_E\_Offen\_schwer\_F1/Unit 02**

Interpret whether a statistically significant Pearson Product Moment correlation of 0.7 is considered to be good or **not**. Explain the coefficient of determination for this correlation.

The coefficient of determination (3 points) for a 0.7 correlation is 49 (3 points) meaning that 49 percent (3 points) of the variation in the x variable can be explained by knowing the y variable. (3 points) Therefore, this would be considered an "acceptable" (3 points) rather than a "good" correlation. (3 points)

**FRAGE 54 VON 78**

**MMET02-01\_E\_Offen\_schwer\_F1/Unit 02**

Explain why would a researcher choose a parametric statistical test over a non- parametric statistical test.

Which assumptions do we have to make in order to use a parametric test?

Non-parametric tests are used when the researcher is not sure about the data distribution (3 points) or is concerned that the data will not meet the assumptions of parametric statistical tests. (3 points) Some of the main assumptions include that the data is randomly sampled from a normal (or Gaussian) distribution. (3 points) Also, parametric statistical techniques assume that data

exhibit similar degrees of homoscedasticity, (3 points) or homogeneity of variance. (3 points) When the assumptions are met, parametric statistical tests are generally more powerful at being able to detect statistical significance in a hypothesis test. (3 points)

**FRAGE 55 VON 78**

**MMET02-01\_E\_Offen\_leicht/Unit 03**

Explain what is a key differentiator between a continuous and a discrete probability distribution.

The probability that a continuous random variable will assume any particular value (3 points) in a distribution is zero is the key differentiator between a continuous and a discrete probability distribution. (3 points)

**FRAGE 56 VON 78**

**MMET02-01\_E\_Offen\_leicht/Unit 03**

Explain why the normal distribution is often referred to as the bell-curve.

The normal distribution is often referred to as the bell-curve because many continuous random variables in business and life have central tendency and (3 points) display this bell-shaped curve (with many of the observations clustering toward the mean) when compiled and graphed. (3 points)

**FRAGE 57 VON 78**

**MMET02-01\_E\_Offen\_mittel/Unit 03**

Explain the two classifications of random variables and how they can be determined.

Discrete random variables can be determined through the process of counting. (3 points) Continuous random variables, on the other hand, can be determined only through some type of measurement. (3 points)

**FRAGE 58 VON 78**

**MMET02-01\_E\_Offen\_mittel/Unit 03**

List three of the most popular discrete probability distributions.

Some of the most popular discrete probability include the binomial probability distribution, hypergeometric probability distribution, multinomial probability distribution, negative binomial distribution, and the Poisson probability distribution. (2 points per distribution)

**FRAGE 59 VON 78**

**MMET02-01\_E\_Offen\_schwer\_F1/Unit 03**

List the six conditions that must be satisfied in order for a probability distribution to be valid.

The following conditions must be satisfied in order for a probability distribution to be valid. Let x be a discrete random variable with possible outcomes x1, x2, …, xn. (3 points) Let x also be a continuous random variable where because there are infinite values that x can assume (3 points) then the probability of x taking on any one specific value is 0. (3 points) Each value of the random variable has a probability between 0 and 1, (3 points) inclusive. (3 points) Moreover, the sum of all the probabilities is 1. (3 points)

**FRAGE 60 VON 78**

**MMET02-01\_E\_Offen\_schwer\_F1/Unit 03**

Explain and give five reasons why researchers typically prefer to utilize discrete probability distributions over continuous probability distributions?

Give an example of how a discrete probability distribution can be used in practice.

Discrete probability distributions are easier to work with since they utilize discrete mathematics and closed form solutions. (3 points) Many continuous probability distributions are not closed-form solutions (3 points) and therefore will require integral calculus to perform calculations with them (3 points). With discrete distributions, you can calculate the probability that X is precisely equal to some value. (3 points) The same is not possible with a continuous probability distribution. (3 points) For example, you can use the discrete Poisson distribution to describe the number of customers calling into the help desk within a week. (3 points)

**FRAGE 61 VON 78**

**MMET02-01\_E\_Offen\_leicht/Unit 04**

Explain the meaning of point estimators.

Which statistical values can be used for point estimates?

Point estimators are metrics or functions that are utilized in inferential statistics to ascertain an approximate value of a population parameter from a random sample selected from a population. (2 points) Point estimators extract the data from a sample of the population to calculate a point estimate or a statistic that serves as the most likely estimate of an unknown population parameter. (2 points) Point estimators can be a variety of statistics including means, proportions, medians, and standard deviation. (2 points)

**FRAGE 62 VON 78**

**MMET02-01\_E\_Offen\_leicht/Unit 04**

Explain the meaning of interval estimators. Which measures typically incorporate them?

Interval estimators contain a range of values, of which the actual population parameter has an estimated probability of falling within. (2 points) A confidence interval is one of the most popular interval estimation techniques. (2 points) Confidence intervals can be calculated for a variety of statistics including means and proportions. (2 points)

**FRAGE 63 VON 78**

**MMET02-01\_E\_Offen\_mittel/Unit 04**

Give two examples of when an interval estimate as opposed to point estimate is more useful.

An interval estimate, as opposed to point estimate, can be useful if a researcher decides that a point estimate is not representative as a population inference due to a skewed distribution (3 points) or from being uncomfortable from a risk perspective of reporting a point estimate. (3 points)

**FRAGE 64 VON 78**

**MMET02-01\_E\_Offen\_mittel/Unit 04**

Explain why would it be risky for an individual or a business to rely on a single point estimate.

Any point estimate calculated from a sample of the population will always vary based upon random sampling error. (3 points) Therefore, if the estimation application will rely on an accurate inference then this could introduce risk into the decision. (3 points)

**FRAGE 65 VON 78**

**MMET02-01\_E\_Offen\_schwer\_F2/Lektion 04**

Describe the decision process in choosing a point estimate versus an interval estimate.

Depending on the application being measured, (3 points) one would choose a point estimate only if they were comfortable that it was an unbiased estimator of the population. (3 points) Even if the former was the case, some researchers prefer to give a range of values (3 points) due to anticipated error (3 points) and the consequences of having made an error in estimation, (3 points) which is still not within statistical significance. (3 points)

**FRAGE 66 VON 78**

**MMET02-01\_E\_Offen\_schwer\_F2/Lektion 04**

Identify when a point estimate is to be considered an unbiased estimate of the population?

An unbiased estimator closely matches the population parameter. (3 points) The probability of assuming any particular value is zero (3 points) in a continuous probability distribution and therefore researchers are only trying to generate a value that is as close (3 points) as possible to the population parameter. (3 points) If this objective is met, (3 points) then the point estimator is deemed an unbiased estimator. (3 points)

**FRAGE 67 VON 78**

**MMET02-01\_E\_Offen\_leicht/Unit 05**

Describe the five steps to conducting a hypothesis test.

1. State the null and alternate hypothesis. H0 = 0/H0 ≠ 0 (2 points)

0

1. Choose the appropriate distribution to use. (1 point)
2. Ascertain the rejection and non-rejection regions. (1 point)
3. Compute the value of the test statistic. (1 point)
4. Make a decision based on the data. (1 point)

**FRAGE 68 VON 78**

**MMET02-01\_E\_Offen\_leicht/Unit 05**

Explain the difference in hypothesis testing when the population standard deviation is unknown.

When the population standard deviation is known, one can use the normal distribution and deploy the z-score. (3 points) When the population standard deviation is unknown, the t-test used. (3 points)

**FRAGE 69 VON 78**

**MMET02-01\_E\_Offen\_mittel/Unit 05**

Explain who the name "student" is in reference to in the Student's t-test? Why was this name assigned?

The t-test founder William Sealy Gosset (3 points) used "student" as a pseudonym to prevent getting in trouble with his employer the Guinness Brewery. (3 points)

**FRAGE 70 VON 78**

**MMET02-01\_E\_Offen\_mittel/Unit 05**

Explain what the reject region and non-reject region mean in hypothesis testing.

The rejection region represents the set of values for the test statistic that leads to rejection of H0. (3 points) The non-rejection region represents the set of values not in the rejection region that leads to non-rejection of H0. (3 points)

**FRAGE 71 VON 78**

**MMET02-01\_E\_Offen\_schwer\_F2/Lektion 05**

Explain the meaning of a p-value in hypothesis testing. Why is it important?

The p-value represents another quantitative measure (3 points) for reporting the result of a hypothesis test. (3 points) When the p-value is low, there is a greater likelihood of obtaining the same result. (3 points) Therefore, a low p-value provides statistical evidence (3 points) that the results of the test are not due to random sampling error alone. (3 points) A p-value is equal to the chance of obtaining a test statistic equal to or more extreme value than the observed value of H0. (3 points)

**FRAGE 72 VON 78**

**MMET02-01\_E\_Offen\_schwer\_F2/Lektion 05**

Frame the interpretation of the dependent-samples t-test as a particular case of the one-sample t-test after the preparation process is completed.

In addition, state the null and alternative hypotheses.

The preparation process in a paired t-test specifically involves subtracting the two scores for each participant to create a single difference score. (3 points) With this first step completed, the dependent-samples t-test becomes a one-sample t-test on the difference scores. (3 points) The hypothetical population mean (µ0) of interest is 0 (3 points) because this represents what the mean difference score would be if there were no difference on average between the two observations or two conditions of the same population. (3 points) The null hypothesis can now be stated that the mean difference score in the population is 0 (µ0 = 0) (3 points) and the alternative hypothesis is stated as being that the mean difference score in the population is not 0 (µ0 ≠ 0). (3 points)

**FRAGE 73 VON 78**

**MMET02-01\_E\_Offen\_leicht/Unit 06**

Describe the equation for a simple linear regression. What does each component represent?

The equation for simple linear regression follows the form Y =*a* + *b*X , where X represents the explanatory variable and Y represents the dependent variable. (3 points) The slope of the line is represented by *b*, and the intercept is represented by*a*.(3 points)

**FRAGE 74 VON 78**

**MMET02-01\_E\_Offen\_leicht/Unit 06**

Explain how the optimization process of least squares was discovered. What was the application it was originally used for?

Carl Friedrich Gauss and Adrien-Marie Legendre were scientists who both independently discovered an essential feature of regression analysis which is the method of least squares. (3 points) Least squares is a statistical optimization technique where the sum of the squared errors are minimized. Both of these scientists used the method to understand the orbits of celestial bodies (3 points).

**FRAGE 75 VON 78**

**MMET02-01\_E\_Offen\_mittel/Unit 06**

Write a simple regression equation that would be common for all forms of regression.

Dependent Variable (y) (3 points) = Constant (b) + Slope (m)\*Independent Variable (x) + Error (e) (3 points)

**FRAGE 76 VON 78**

**MMET02-01\_E\_Offen\_mittel/Unit 06**

Describe what it could mean and what to do if your p-value is greater than or less than

.05 with your regression equation.

If your regression equation p-value is less than 0.05, then you are generally confident in your analysis. (2 points) If you discover that significance F is greater than 0.05, you probably need to search for a new independent variable and rerun your regression analysis. (2 points) This should be repeated until significance F drops below 0.05. (2 points)

**FRAGE 77 VON 78**

**MMET02-01\_E\_Offen\_schwer\_F2/Lektion 06**

Identify and describe four key assumptions when running a linear regression analysis.

1. Homogeneity of variance (1 point): With homoscedasticity we assume the size of the error in our prediction doesn’t vary significantly across the values of the independent variable (3 points).
2. Independence of observations (1 point): With independent observations we are assuming there are no hidden relationships among the observations and that all of the observations in the dataset were gathered using statistically valid sampling methods (3 points).
3. Normality (1 point): Normality assumes the data are following a pattern that closely approximates the normal distribution (3 points).
4. The relationship between the dependent and independent variable is linear (3 points): The line of best fit through the data points is a straight line (rather than a curve which would then make the relationship nonlinear, or curvilinear). (3 points)

**FRAGE 78 VON 78**

**MMET02-01\_E\_Offen\_schwer\_F2/Lektion 06**

Explain the algebraic origins of least-squares regression.

Under what conditions can we be confident that they have the best fit line?

In its simplest form, appropriately referred to as simple linear regression, our goal is to find the best fit line (from which we can then predict with new observations) through a set of data points: (x1, y1), (x2, y2), … (xn, yn). (3 points)

We take our equation for a straight line from algebra where y is out independent variable (3 points) and x is our independent variable. (3 points) X is the slope of the line, and b (c in the UK) is the constant and also referred to as the y-intercept. (3 points) Remember the constant is y when x = 0. (3 points) Now, if we can find a slope and an intercept for a single line that passes through all the possible data points, then that it would certainly be considered the best fit line. (3 points)